International Journal of
Molecular Sciences

Special Issue Reprint

Plant Genomics 2019 Volume II

Edited by
Frank M. You

www.mdpi.com/journal/ijms

MDPI

Plant Genomics 2019—Volume II

Plant Genomics 2019—Volume II

Editor

Frank M. You

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editor
Frank M. You
Ottawa Research and
Developmental Centre
Agriculture and Agri-Food Canada
Ottawa
Canada

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *International Journal of Molecular Sciences* (ISSN 1422-0067) (available at: https://www.mdpi.com/journal/ijms/special_issues/plant_genomics_2019).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

Volume II
ISBN 978-3-0365-8090-6 (Hbk)
ISBN 978-3-0365-8091-3 (PDF)

Volume I-II
ISBN 978-3-0365-8086-9 (Hbk)
ISBN 978-3-0365-8087-6 (PDF)

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.
The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editor

Frank M. You

Dr. Frank M. You is a highly accomplished senior research scientist in bioinformatics and genomics at the Agriculture and Agri-Food Canada (AAFC) Ottawa Research and Development Centre. He is also an Adjunct Professor in the Department of Plant Science at the University of Manitoba and a Guest Professor at Nanjing Agricultural University in China. Dr. You received his Ph.D. in plant genetics and breeding, with a specialization in statistical genetics, in 1989, and holds two Bachelor's degrees, one in agronomy from 1982 and another in computer science from 1999.

Dr. You is an expert in computational biology and bioinformatics, statistical genetics, and plant genetics and breeding. He has a wealth of experience in plant comparative and statistical genomics, quantitative genetics, genome assembly and annotation of complex genomes, gene expression and microarray data analysis, physical mapping and data analysis, high-throughput molecular marker design and development, and bioinformatics software development.

Dr. You's recent research projects focus on genome sequencing and annotation, QTL mapping, and the identification and characterization of genes associated with seed yield and disease resistance in flax and cereal crops. With his impressive expertise and extensive research accomplishments, Dr. You continues to contribute to the advancement of the field of plant genomics and genetics.

Preface to "Plant Genomics 2019—Volume II"

In recent years, researchers have uncovered the genes and genomic regions responsible for plants' growth, development, and stress responses. This reprint the 'Plant Genomics 2019' Special Issue comprises 57 papers exploring various aspects of plant genomics. These papers delve into gene discovery, genomic prediction, genome editing, plant chloroplast genome sequencing and comparative analysis, microRNA analysis, and comparative genomics.

The studies featured in this Special Issue employ a comprehensive research approach that combines bioinformatics and transcriptome analyses. With this approach, researchers have identified the genes associated with biotic and abiotic stress responses. Studies on the genome-wide identification of gene families, gene characteristics and distributions analysis, and gene expression profiles have shed light on various traits across multiple species.

In addition to gene discovery, the Special Issue also considers microRNAs (miRNAs) and their regulatory roles in gene expression. The roles of miRNAs in plant species have been explored, including the development of an artificial miRNA precursor system for gene silencing and the identification of miRNAs involved in seed development. These findings contribute to our understanding of how miRNA function in plant growth and development, offering potential avenues for crop improvement.

The Special Issue places a significant emphasis on the application of genomic tools for crop improvement. Papers on Chinese winter wheat and flax highlight the effectiveness of genomic prediction and marker-assisted selection in enhancing their yield, salt tolerance, and fruit ripening. Furthermore, the issue explores the revolutionary impact of the CRISPR/Cas9-mediated genome editing of plants, including targeted mutagenesis and gene replacement.

Comparative genomics is another key theme within this Special Issue, providing insights into the evolution of plant species. By comparing genomes, researchers can identify conserved gene families and regulatory elements, shedding light on plants' development and their adaptation to environmental stresses. Comparative genomics can also be used for phylogenetic analyses and resolving relationships between species.

Overall, the 57 papers featured in 'Plant Genomics 2019' exemplify the substantial progress made in understanding plant genetics and genomics. These studies offer valuable insights into the application of genomic tools for crop improvement, sustainable agriculture, and fundamental questions about the evolution and function of plant genes and genomes.

Frank M. You
Editor

International Journal of
Molecular Sciences

MDPI

Article

The Complete Chloroplast Genomes of *Punica granatum* and a Comparison with Other Species in Lythraceae

Ming Yan [1,2], Xueqing Zhao [1,2], Jianqing Zhou [1,2], Yan Huo [1,3], Yu Ding [1,2] and Zhaohe Yuan [1,2,*]

1 Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; punica24@njfu.edu.cn (M.Y.); zhaoxq402@gmail.com (X.Z.); wangying@njfu.edu.cn (J.Z.); xinhuizhang888@gmail.com (Y.H.); punica2421@gmail.com (Y.D.)
2 College of Forestry, Nanjing Forestry University, Nanjing 210037, China
3 College of Landscape Architecture, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: zhyuan88@gmail.com; Tel.: +86-25-85427056

Received: 7 April 2019; Accepted: 4 June 2019; Published: 13 June 2019

Abstract: Pomegranates (*Punica granatum* L.) are one of the most popular fruit trees cultivated in arid and semi-arid tropics and subtropics. In this study, we determined and characterized three complete chloroplast (cp) genomes of *P. granatum* cultivars with different phenotypes using the genome skimming approach. The complete cp genomes of three pomegranate cultivars displayed the typical quadripartite structure of angiosperms, and their length ranged from 156,638 to 156,639 bp. They encoded 113 unique genes and 17 are duplicated in the inverted regions. We analyzed the sequence diversity of pomegranate cp genomes coupled with two previous reports. The results showed that the sequence diversity is extremely low and no informative sites were detected, which suggests that cp genome sequences may be not be suitable for investigating the genetic diversity of pomegranate genotypes. Further, we analyzed the codon usage pattern and identified the potential RNA editing sites. A comparative cp genome analysis with other species within Lythraceae revealed that the gene content and organization are highly conserved. Based on a site-specific model, 11 genes with positively selected sites were detected, and most of them were photosynthesis-related genes and genetic system-related genes. Together with previously released cp genomes of the order Myrtales, we determined the taxonomic position of *P. granatum* based on the complete chloroplast genomes. Phylogenetic analysis suggested that *P. granatum* form a single clade with other species from Lythraceae with a high support value. The complete cp genomes provides valuable information for understanding the phylogenetic position of *P. gramatum* in the order Myrtales.

Keywords: pomegranate; chloroplast genome; sequence diversity; site-specific selection; phylogeny

1. Introduction

Pomegranates (*Punica granatum* L.) are an economically important fruit tree of the tropical and subtropical regions of the world. It is native to central Asia and has been highly praised in many human cultures since ancient times [1]. Pomegranates have showy edible fruit with a high content of anthocyanins and flavonoids [2,3]. It has been well demonstrated that pomegranates are valuable to human health due to high levels of flavonoids and anthocyanins, which are considered potent antioxidants offering protection against heart disease and cancer [4,5]. Also, the pomegranate tree is suitable for genetic analysis due to its short juvenile period, and the high number of progenies [6]. As important resources for basic research and crop improvement, the genome of *P. granatum* 'Taishanhong' has been determined [7]. This genome will shed new light on the understanding of some unique biological processes and pomegranate breeding. Compared to the nuclear genome, the

complete chloroplast genome is a low-cost and efficient way to get valuable genomic resources that can be used to understand evolution at multiple taxonomic levels [8,9] and analyze the population [10] because of its highly conserved structures and comparatively moderate substitution rates [11].

Chloroplasts (cp) are the photosynthetic organelles of the plant cells, which are derived from free-living cyanobacteria through endosymbiosis [12]. Apart from playing key roles in photosynthesis, chloroplasts are also responsible for other aspects of plant physiology and development [13]. A new study found that chloroplast retrograde signaling can regulate nuclear alternative splicing of a subset of *Arabidopsis thaliana* transcripts [14,15]. Interestingly, researchers have found that chloroplasts play diverse roles in plant defense, including contributing to the production of defense compounds [16]. Chloroplasts contain their own genome, the chloroplast DNA (cpDNA), which is highly conserved in genomic structure, gene content, and gene order. Cp genomes have been proved to be an effective biological tool for rapid and accurate species recognition as super-barcode [17,18]. With the advent of high-throughput sequencing technology, the increasing number of cp genomes of fruit crops has been published [19–21]. Before the development of next-generation sequencing technology, cp genome assembly was usually based on conventional primer walking strategies [22,23], which are laborious and costly. It is now convenient to obtain complete a cp genome by using genomic DNA extracted from the leaf tissue, because a large number of cp genomes are present in the sample [24]. Based on homology to cp from related species, these reads from cp can be assembled into circle genomes. Many bioinformatic tools have been developed to recover the cp genome sequence from total genomic DNA, such as NOVOPlasty [25], chloroExtractor [26], and GetOrganelle [27].

In the present study, we obtained cp genomes of pomegranates from three phenotypically different cultivars using the whole genome sequencing data. This study aimed to conduct a comprehensive analysis of the pomegranate cp genome, including gene content, genomic structure, codon usage, and potential RNA editing sites. In addition, combined with previously published cp genomes of Myrtales, phylogenetic analysis was performed to determine the taxonomic position of *P. granatum*. The results obtained here will provide valuable information for understanding the phylogenetic position of pomegranates and the evolutionary history of the order Myrtales.

2. Results and Discussion

2.1. General Features of Pomegranate Chloroplast Genomes

The complete cp genomes of 'Nana', 'Tunisia', and 'Taishanhong' were de novo assembled using whole genome sequencing data with GetOrganelle [27]. The cp genomes of 'Nana', 'Tunisia', and 'Taishanhong' were found to be 158,638, 158,639, and 158,638 bp in size, respectively. All of them exhibited a typical quadripartite structure, consisting of a pair of IRs separated by a large single copy region (LSC) and a small single copy region (SSC) (Figure 1). There are identical sets of 113 genes with the same gene order, including 79 protein-coding, 30 tRNA, and 4 rRNA genes. Six protein-coding genes (*rps7*, *rps12*, *rpl2*, *rpl23*, *ndhB*, *ycf2*), seven tRNA genes (*trnI-CAU*, *trnN-GUU*, *trnR-ACG*, *trnA-UGC*, *trnI-GAU*, *trnV-GAC*, *trnL-CAA*), and all rRNA genes (4.5S, 5S, 16S, 23S) are located at the IR regions. Eleven of the protein-coding genes and six of the tRNA genes contain introns, 14 of which contain a single intron, whereas three (*rps12*, *ycf3*, *clpP*) have two introns (Table 1). In particular, the *rps12* is a trans-spliced gene, whose first exon is located in the LSC, while the second and third exons reside in IRs. The *infA* gene was identified as a pseudogene because of the accumulation of the premature stop codons [28]. Another pseudogene *ycf1* existed because of the incomplete duplication of the normal copy of *ycf1* in the IRa and SSC junction, which is identical with previous reports [29,30]. There are some exceptions where non-ATG codons were identified as start codons, such as ACG for *psbL*, GTG for *rps19*, and ACG for *ndhD*. Alternate start codons have been found in other plant species [31]. Alternate start codons are still translated as Met when they are the start of a protein because a separate transfer RNA is used for initiation [32]. The overall GC content was 36.92%; this was consistent with previously

reported GC content of IRs (42.78%) being higher than that of the LSC (34.89%) and SSC (30.64%) [33]. The high GC percentage of IRs could be due to the presence of rRNA sequences in these regions [34].

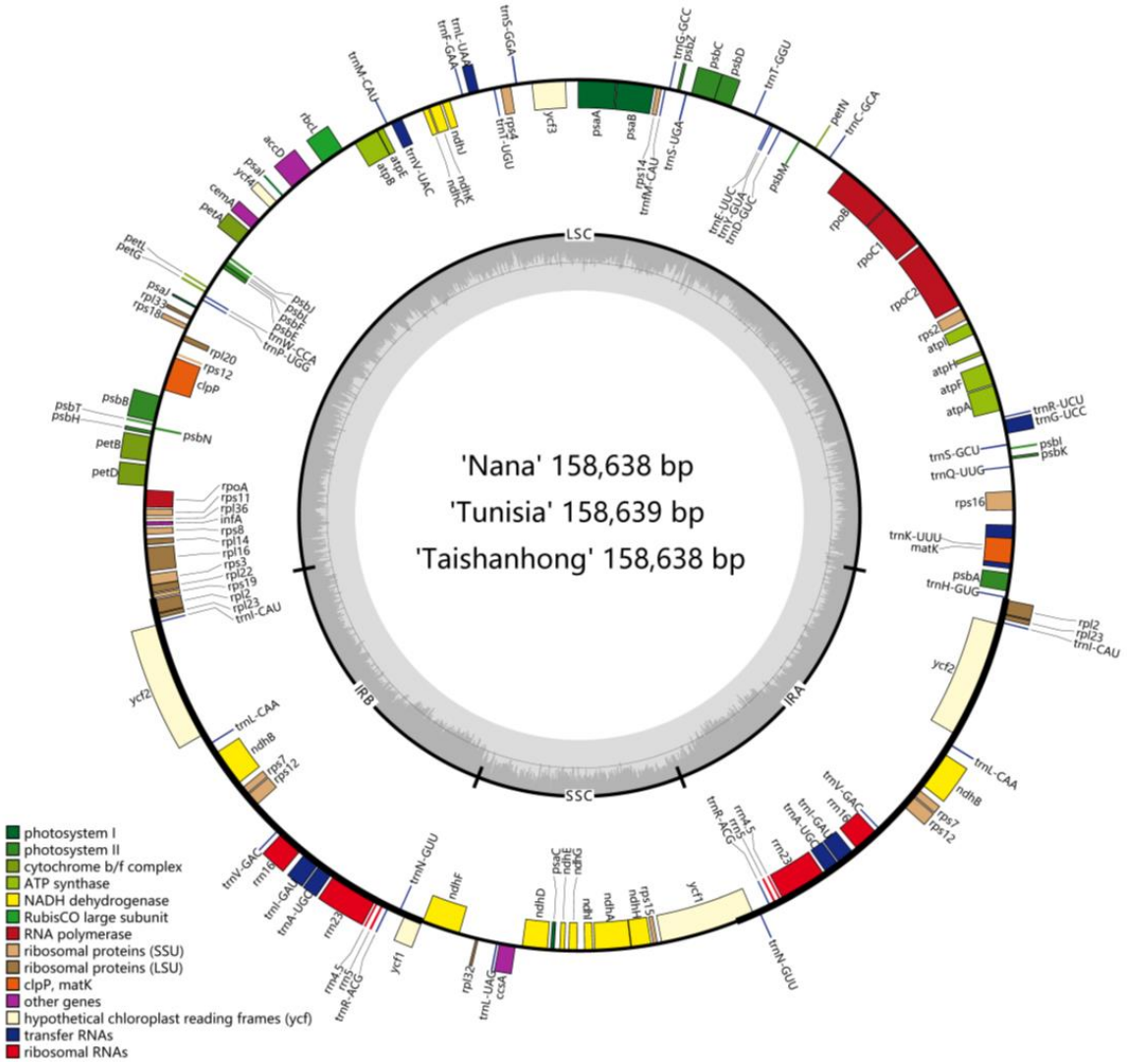


Figure 1. Chloroplast genome maps of *P. granatum*. Genes drawn outside the outer circle are transcribed clockwise, and those inside are transcribed counter-clockwise. Genes belonging to different functional groups are color-coded.

Chloroplast DNA has already been used in accessing the genetic diversity and phylogenetic structure at an intraspecies level. For instance, hypervariable regions of cp DNA such as *atpB-rbcL*, *trnL-trnF* and *rps16-trnQ* were used to assess the genetic diversity of Tunisian apricot accessions [35]. Also, chloroplast microsatellite loci were used to investigate the genetic diversity of Iranian pomegranate genotypes [36]. In our present study, the sequence diversity of pomegranate cp genomes was investigated combined with two previously reported cp genomes (NC_035240, MG878386). The results showed that the sequence diversity of pomegranates is extremely low (0.0008). Only 42 singleton variable sites were detected, and there were no parsimony variable sites in the alignment of the cp genomes of the five pomegranate accessions. Therefore, we propose that cp genome sequences might not be appropriate for investigating the genetic diversity of pomegranate genotypes.

Table 1. The groups of genes within the *P. granatum* chloroplast genome.

Group of Genes	Gene Names
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
Cytochrome b/f complex	*petA*, *petB*[a], *petD*[a], *petG*, *petL*, *petN*
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[a], *atpH*, *atpI*
NADP dehydrogenase	*ndhA*[a], *ndhB*[*a], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
RubisCO large subunit	*rbcL*
RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[a], *rpoC2*
Ribosomal proteins (SSU)	*rps2*, *rps3*, *rps4*, *rps7**, *rps8*, *rps11*, *rps12*[*b], *rps14*, *rps15*, *rps16*[a], *rps18*, *rps19*
Ribosomal proteins (LSU)	*rpl2*[*], *rpl14*, *rpl16*[a], *rpl20*, *rpl22*, *rpl23**, *rpl32*, *rpl33*, *rpl36*
Hypothetical chloroplast reading frames	*ycf1*, *ycf2**, *ycf3*[b], *ycf4*
Translation initiation factor IF-1	*infA*
Acetyl-CoA carboxylase	*accD*
Cytochrome c biogenesis Maturase	*matK*
ATP-dependent protease	*clpP*[b]
Inner membrane protein	*cemA*
Ribosomal RNAs	*rna4.5S**, *rna5S**, *rna16S*[*], *rna23S**
Transfer RNAs	*trnA-UGC*[*a], trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC*[a], *trnH-GUG, trnI-CAU*[*a], trnI-GAU*, trnK-UUU*[a], *trnL-CAA*, trnL-UAA*[a], *trnL-UAG, trnM-CAU, trnN-GUU*, trnP-UGG, trnQ-UUG, trnR-ACG*, *trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC*, trnV-UAC*[a], *trnW-CCA, trnY-GUA*

One asterisk indicates that the genes contained two copies. a and b indicate one- and two-intron containing genes, respectively.

2.2. Codon Usage Bias

As an essential evolutionary feature, the codon usage pattern has been widely investigated in many plant species [37–39]. In our study, we explored the codon usage pattern in the cp genomes of pomegranates. Protein-coding genes with more than 300 nucleotides were selected for further analysis. Firstly, the base composition on three different codon positions was determined, and the data are displayed in Figure 2A. The results indicated that the average GC content of the first (GC1), second (GC), and third codon positions (GC3) were 47.04, 39.79, and 28.34%, respectively. The base compositions of the three different positions were distributed unevenly. The average GC3 content was significantly lower than those of GC1 and GC2. The results of neutrality plots (Figure 2B) showed that no significant correlation (R2 = 0.0036) between GC12 and GC3 was observed, which suggests that selective pressure affects the codon usage bias in the pomegranate cp genomes [40,41]. The codon adaptation index (CAI) value (Figure 2C) ranged from 0.5 to 1 with a default *E. coli* reference gene set as the reference. According to their functions in the chloroplast, the protein-coding genes can be classified into three categories: photosynthesis related genes (photo-genes), genetic system related genes (genet-genes), and other genes. A recent study about codon usage bias of cp genomes in cultivated and wild Solanum species concluded that photo-genes always had higher CAI values than genet-genes because the

expression level of photo-genes is relatively higher than that of genet-genes [42]. The same result was also observed in the cp genome of pomegranates. The main reason is probably due to the fact that photo-genes may have a higher codon usage bias for the requirement of high gene expression than do the genet-genes in the plant cp genomes. Furthermore, the relationship between bas compositions and codon usage was investigated by ENC-plot (Figure 2D). Effective number of codons (ENC) values ranged from 35.73 to 61, suggesting that codon usage bias is relatively weak in the pomegranate cp genomes. The distribution of most genes was far away from the standard curve, which shows that there are other factors that affect the codon usage, other than base compositions [43–45]. As the cp genomes were highly AT-rich, it was not surprising that AT-ending codons would be predominant in the protein-coding genes. The results are also consistent with the mutational bias towards AT being the force driving the strong bias of codon usage of plan cp genomes [43]. Also, the Arg amino acid coded with the AGA codon was the most frequent codon with a relative synonymous codon usage (RSCU) value 2.02 (Table 2).

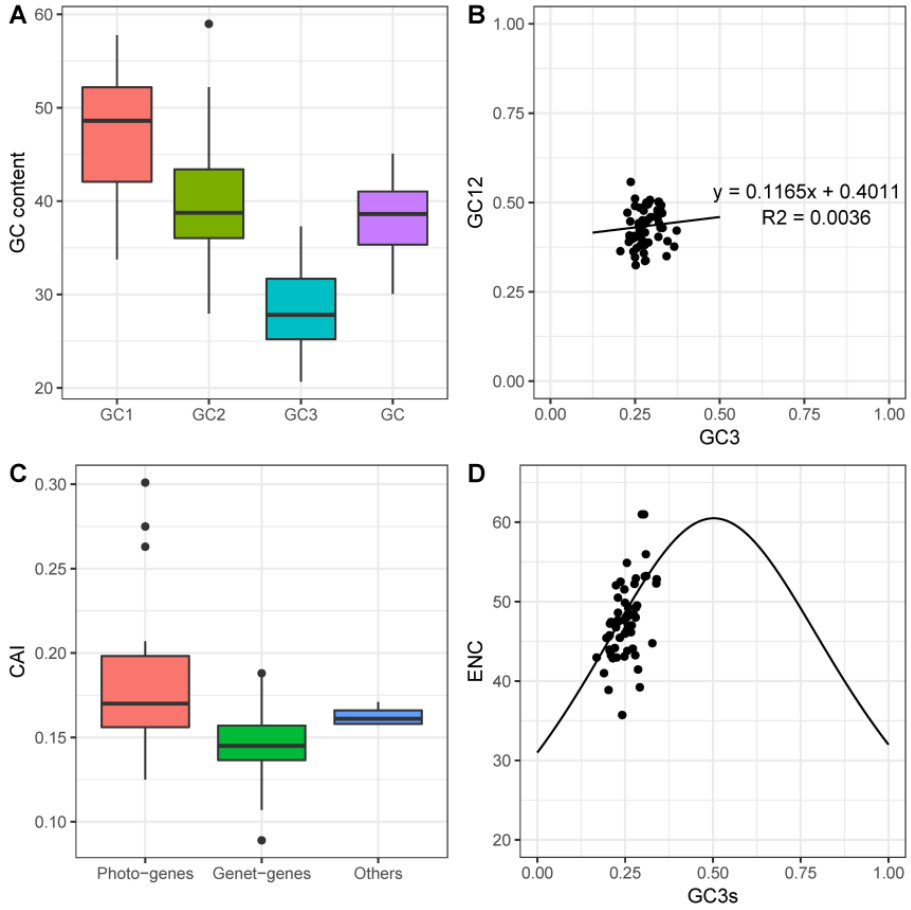


Figure 2. The codon usage pattern of the pomegranate chloroplast (cp) genome. (**A**) GC content on three different positions. (**B**) Neutrality plot (GC12 against GC3). (**C**) The codon adaptation index (CAI) value of gene sets with different functions. (**D**) Relationship between GC3 and effective number of codons (ENC) (ENC-plot). The expected ENC from GC3 is shown as a solid.

Table 2. Putative preferred codons in the *P. granatum* cp genome. RSCU = relative synonymous codon usage.

Amino Acid	Codon	Codon Frequency	RSCU	AA	Codon	Codon Frequency	RSCU
Phe	UUU*	4551	1.18	Ser	UCU*	2417	1.46
	UUC	3143	0.82		UCC	1577	0.96
Leu	UUA*	3112	1.41		UCA*	2278	1.38
	UUG*	2920	1.32		UCG	1268	0.77
	CUU*	2586	1.17	Pro	CCU*	1385	1.21
	CUC	1360	0.62		CCC	929	0.81
	CUA	1958	0.89		CCA*	1387	1.21
	CUG	1287	0.58		CCG	876	0.77
Ile	AUU*	4378	1.27	Thr	ACU*	1478	1.13
	AUC	2723	0.79		ACC	1095	0.84
	AUA	3246	0.94		ACA*	1759	1.35
Met	AUG	2760	1.00		ACG	886	0.68
Val	GUU*	2045	1.34	Ala	GCU*	1389	1.45
	GUC	1033	0.68		GCC	712	0.75
	GUA*	1891	1.24		GCA	1145	1.20
	GUG	1123	0.74		GCG	576	0.60
Tyr	UAU*	3606	1.37	Cys	UGU*	1410	1.17
	UAC	1665	0.63		UGC	993	0.83
TER	UAA*	2029	1.03	TER	UGA	2003	1.01
	UAG	1893	0.96	Trp	UGG	2392	1.00
His	CAU*	1908	1.38	Arg	CGU	888	0.7
	CAC	866	0.62		CGC	456	0.36
Gln	CAA*	2815	1.35		CGA*	1428	1.13
	CAG	1342	0.65		CGG	865	0.68
Asn	AAU*	3923	1.37	Ser	AGU	1441	0.87
	AAC	1800	0.63		AGC	923	0.56
Lys	AAA*	4768	1.31	Arg	AGA*	2560	2.02
	AAG	2538	0.69		AGG	1412	1.11
Asp	GAU*	2818	1.49	Gly	GGU	1642	1.01
	GAC	962	0.51		GGC	886	0.54
Glu	GAA*	3632	1.37		GGA*	2409	1.48
	GAG	1689	0.63		GGG	1569	0.96

Preferred codons (RSCU value > 1.0) are indicated with (*).

2.3. RNA Editing Sites

RNA editing is a posttranscriptional process, which has been experimentally identified in organellar transcriptomes from several species [46,47]. It mainly involves the conversion of cytidine to uridine, which generally results in amino acid changes. Therefore, knowing where sites of RNA editing exist in the organelle transcriptome could provide information for understanding the structure and function of the translated proteins [48]. The potential RNA editing sites in the pomegranate cp genome were predicted using the online program PREP. A total of 64 editing sites in 20 protein-coding genes were identified (Table 3). The *ndhB* gene had the highest number of potential editing sites (11), followed by the *ndhD* gene (9). In accordance with previous reports [49,50], we observed that most conversions at the codon positions changed from serine (S) to leucine (L) and most RNA editing sites led to amino acid changes from polar to apolar, which resulted in an increase in protein hydrophobicity.

Table 3. Predicted RNA editing sites in the cp genome of *P. granatum.*

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Score
matK	644	215	GCA (A) => GTA (V)	1
	1177	393	CGG (R) => TGG (W)	1
	1187	396	TCA (S) => TTA (L)	0.86
	1246	416	CAC (H) => TAC (Y)	1
atpA	791	264	CCC (P) => CTC (L)	1
atpF	92	31	CCA (P) => CTA (L)	0.86
atpI	23	8	ACC (T) => ATC (I)	1
rps2	422	141	TCG (S) => TTG (L)	1
rpoC2	3056	1019	GCA (A) => GTA (V)	0.86
	3998	1333	GCG (A) => GTG (V)	0.86
rpoC1	41	14	TCA (S) => TTA (L)	1
	1171	391	CCA (P) => TCA (S)	1
rpoB	338	113	TCT (S) => TTT (F)	1
	551	184	TCA (S) => TTA (L)	1
	566	189	TCG (S) => TTG (L)	1
	973	325	CTC (L) => TTC (F)	0.86
rps14	80	27	TCA (S) => TTA (L)	1
	149	50	TCA (S) => TTA (L)	1
atpB	1487	496	TCG (S) => TTG (L)	1
accD	794	265	TCG (S) => TTG (L)	0.8
	1403	468	CCT (P) => CTT (L)	1
psbL	2	1	ACG (T) => ATG (M)	1
psbF	77	26	TCT (S) => TTT (F)	1
clpP	559	187	CAT (H) => TAT (Y)	1
ndhB	28	10	CTC (L) => TTC (F)	1
	149	50	TCA (S) => TTA (L)	1
	467	156	CCA (P) => CTA (L)	0.8
	586	196	CAT (H) => TAT (Y)	1
	611	204	TCA (S) => TTA (L)	1
	737	246	CCA (P) => CTA (L)	1
	746	249	TCT (S) => TTT (F)	1
	830	277	TCA (S) => TTA (L)	1
	836	279	TCA (S) => TTA (L)	1
	1255	419	CAT (H) => TAT (Y)	1
	1481	494	CCA (P) => CTA (L)	1
ndhF	160	54	CTT (L) => TTT (F)	1
	586	196	CTT (L) => TTT (F)	0.8
ccsA	89	30	TCG (S) => TTG (L)	1
ndhD	2	1	ACG (T) => ATG (M)	1
	185	62	ACC (T) => ATC (I)	1
	313	105	CGG (R) => TGG (W)	0.8
	383	128	TCA (S) => TTA (L)	1
	674	225	TCG (S) => TTG (L)	1
	845	282	ACA (T) => ATA (I)	0.8
	878	239	TCA (S) => TTA (L)	1
	887	296	CCA (P) => CTA (L)	1
	1405	469	CTT (L) => TTT (F)	0.8
ndhG	155	52	CCA (P) => CTA (L)	1
	166	56	CAT (H) => TAT (Y)	0.8
	314	105	ACA (T) => ATA (I)	0.8
ndhA	341	114	TCA (S) => TTA (L)	1
	566	189	TCA (S) => TTA (L)	1
	1073	358	TCC (S) => TTC (F)	1

2.4. Sequence Diversity of the Chloroplast Genomes among Lythraceae Species

Four complete cp genomes within Lythraceae, available in GenBank with our newly assembled 'Taishanhong', were selected to analyze the sequence diversity. The mean P-distance value was designated to represent the level of divergence. The genetic distance of all 76 protein-coding genes (Figure 3A) ranged from 0.003053 (*psbN*) to 0.108932 (*rpl22*), with an average of 0.024379. The intergenic regions had a relatively higher genetic distance (Figure 3B) compared to the protein-coding regions, ranging from 0.005621 (*trnL-ndhB*) to 0.23463 (*trnL-ycf2*) with an average value of 0.069775. The results

agree with previous reports that intergenic regions showed greater divergence than coding regions (Figure 3D) [51]. The SSC region exhibited higher divergence levels than the LSC and IRs (Figure 3C). Three intergenic regions with genetic distance values over the 95th percentile were considered as highly divergent regions, including *trnH-psbA*, *trnL-ccsA*, and *trnL-ycf2*. These highly variable regions may be regarded as potential molecular markers for application in phylogenetic analyses in Lythraceae.

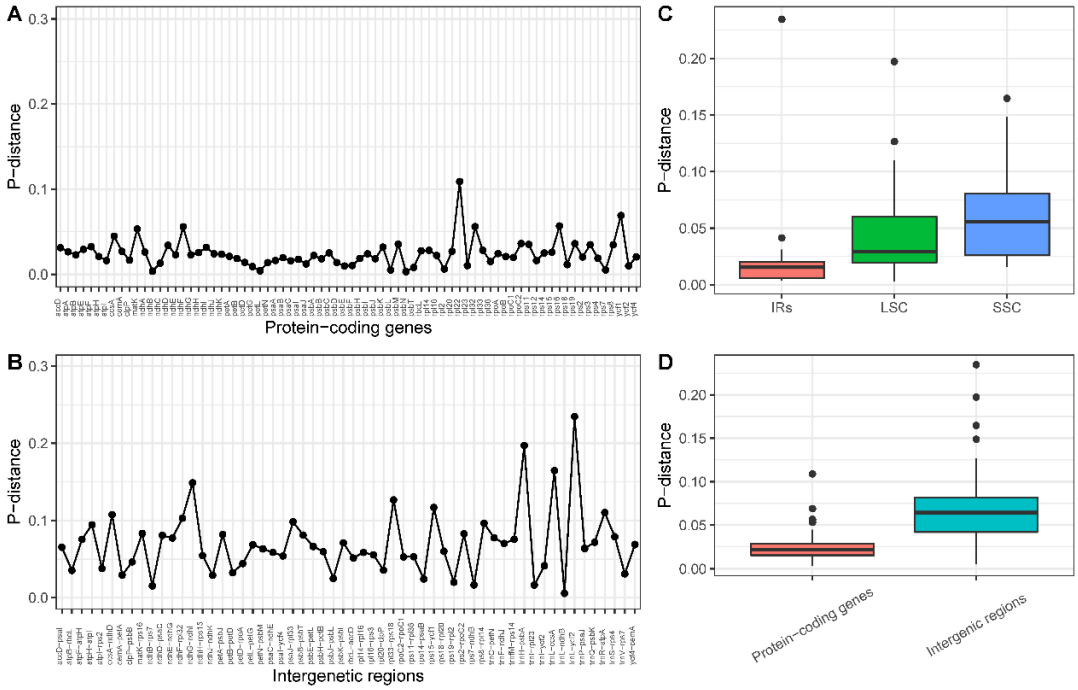


Figure 3. The genetic distance based on Kimura's two-parameter model. (**A**) The P-distance value of protein-coding genes. (**B**) The P-distance value of intergenic regions. (**C**) Boxplots of P-distance value difference among LSC, SSC, and IRs. (**D**) Boxplots of P-distance value differences between protein-coding genes and intergenic regions.

2.5. Structure Comparison among the Chloroplast Genomes of Lythraceae Species

Five complete cp genomes within Lythraceae were selected for comparison with each other. The genome sizes was ranged from 152,205 to 159,219 bp. The length of the LSC, SSC, and IRs varied in the range of 84,046–89,021 bp, 16,914–18,821 bp, and 23,902–25,914 bp, respectively. *Lagerstroemia indica* has the smallest genome, and this difference is mostly attributed to variation in the length of the LSC and IR regions (Table 4). A detailed comparison on four borders between the two IRs and the two single-copy regions showed that the border structures were highly similar with one another (Figure 4). However, a slight difference in junction positions was observed among these five cp genomes. For instance, the *ndhF* gene was located at the SSC region in *Sonneratia alba*, *Trapa maximowiczii*, *Punica granatum*, and *Heimia myrtifolia*, while it varied from 3 to 54 bp apart from the IRb/SSC junction. However, the *ndhF* gene crossed over the IRb/SSC region in *Lagerstroemia indica*. The *rps19* gene was located in the junction of the LSC/IRb in *Trapa maximowiczii*, *Lagerstroemia indica*, and *Punica granatum*, with 24–83 bp located in the IRb. However, in *Heimia myrtifolia* and *Sonneratia alba*, the *rps19* gene was fully located in the LSC region, and 4–16 bp apart from the LSC/IRb border. Overall, the IR boundary regions varied slightly in the Lythraceae cp genomes. IR expansion and contraction often results in genome size variations among various plant lineages, which can be used to study the phylogenetic classification and the genome evolution among plant lineages. Three reasons may explain the diversification of the IR boundary region sequences: the first is intramolecular recombination, the second is the presence of multiple repeat sequences, and the third is the indels, which caused a mismatch that resulted in the upstream sequence becoming a single copy [52].

Pairwise alignment of the *P. granatum* cp genome with the other Lythraceae species revealed a high degree of synteny and gene order conservation (Figure 5), suggesting an evolutionary conservation of the Lythraceae cp genomes at the genome-scale level.

Table 4. Summary of the complete chloroplast genome characteristics of five species in Lythraceae.

Species	*Punica granatum*	*Lagerstromeia indica*	*Sonneratia alba*	*Trapa maximowicizz*	*Heimia myrtifolia*
Genome size	158,638	152,025	153,061	155,577	159,219
LSC size	89,021	84,046	87,226	88,528	88,571
SSC size	18,684	16,914	18,032	18,272	18,821
IR size	25,467	25,623	23,902	24,389	25,914
Number of genes	113	113	107	110	112
Protein-coding genes	79 (6)	79 (7)	79 (6)	77 (5)	78 (7)
tRNA genes	30 (7)	30 (7)	24 (5)	29 (9)	30 (6)
rRNA genes	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)
Number of genes duplicated in IR	17	18	15	18	17
GC content	36.92	37.59	37.29	36.4	36.95
GenBank accession	MK603511	NC_030484	NC_039975	NC_037023	MG921615

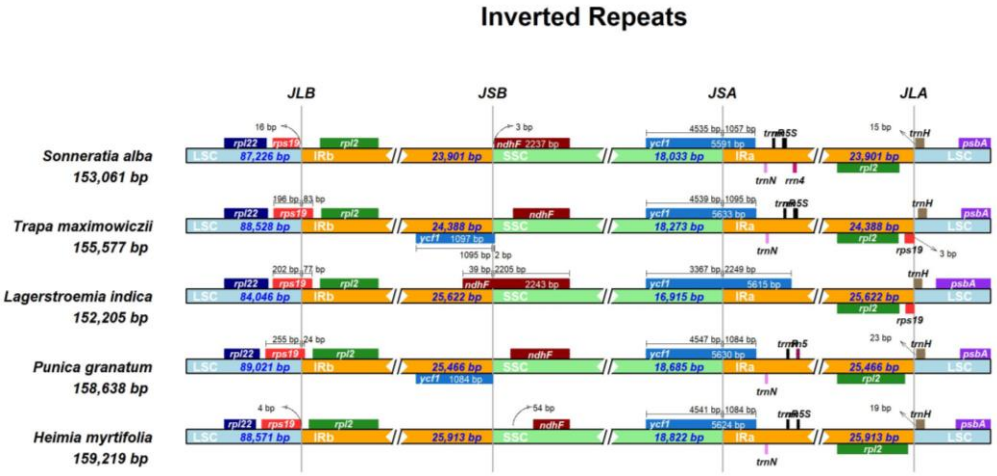


Figure 4. Comparison of the borders of the LSC, SSC, and IRs regions among five Lythraceae cp genomes.

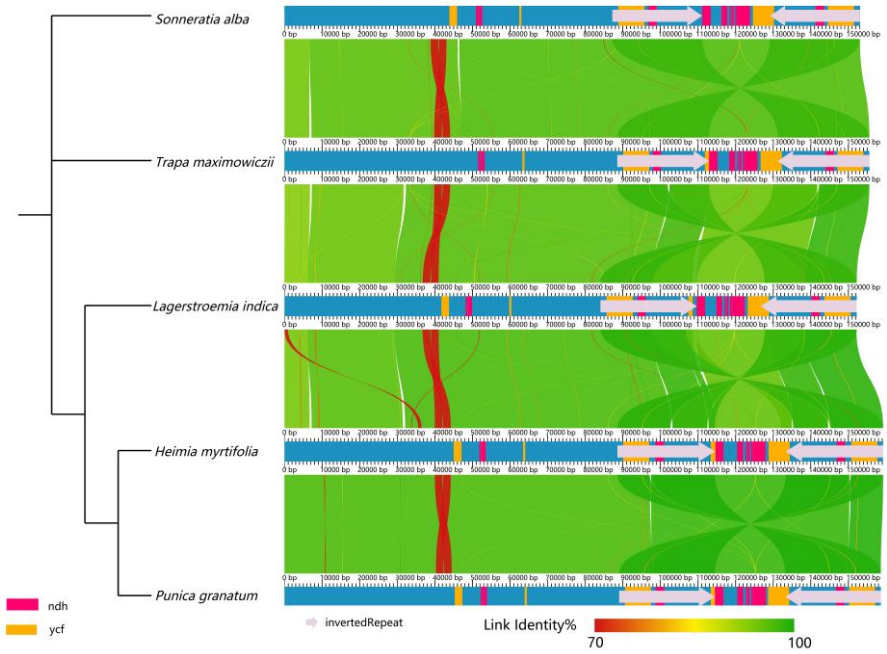


Figure 5. Co-linear analysis of five cp genomes within Lythraceae.

2.6. Positive Selection Analysis

The ratios of non-synonymous (dN) and synonymous (dS) substitutions for 75 protein-coding genes among five Lythraceae were calculated based on the site-specific model. Eleven genes with positively selected sites within the Lythraceae family were identified (Table 5). Those genes contained one subunit of acetyl-CoA carboxylase (*accD*), one photosystem I subunit gene (*psaI*), two NADH-dehydrogenase subunit genes (*ndhF*, *ndhJ*), one ribosome large subunit gene (*rpl22*), five ribosome small subunit genes (*rps2*, *rps4*, *rps7*, *rps8*, *rps12*), and the *ycf1* gene. According to the M8 model, the ycf1 gene possessed 10 positive sites, followed by *ndhF* (7) and *rpl22* (5). The other eight genes each had only one positive site. The Photo-genes included four genes (*accD*, *psaI*, *ndhF*, *ndhJ*). The Genet-genes included six genes (*rpl22*, *rps2*, *rps4*, *rps7*, *rps8*, *rps12*). The *ycf1* gene was considered as the other gene. Most positively selected genes were genetic system or photosynthesis related genes, which indicated that the chloroplast functional genes played vital roles during the plant evolution [53,54].

Table 5. Log-likelihood values of the site-specific models, with detected sites having non-synonymous/synonymous (dN/dS) values > 1.

Gene Name	Models (Number of Parameters)	lnL	Likelihood Ratio Test *p*-Value	Positively Selected Sites
accD	M8 (12) M7 (10)	−2534.400824 −2536.741156	0.0962956	125 G 0.955*
ndhF	M8(12) M7(10)	−4446.871610 −4466.787914	0.000000002	292 N 0.961 *; 486 R 0.999 **; 487 I 0.975 *; 490 K 0.985 *; 518 N 0.969 *; 648 S 0.983 *; 738 F 0.995 **
ndhJ	M8(12) M7(10)	−806.644367 −812.360744	0.003291615	121 R 0.970 *
psaI	M8(12) M7(10)	−139.985819 −143.4420	0.031547151	26 H 0.959 *
rpl22	M8(12) M7(10)	−942.833567 −950.440165	0.000497160	4 L 0.972 *; 5 Y 0.961 *; 73 P 0.962 *; 125 A 0.993 **; 126 R 0.994 **
rps12	M8(12) M7(10)	−529.317591 −535.294718	0.002536102	117 K 0.974 *
rps2	M8(12) M7(10)	−1171.788526 −1176.587212	0.008240568	173 E 0.982 *
rps4	M8(12) M7(10)	−994.666749 −999.616541	0.007084882	28 P 0.959 *
rps7	M8(12) M7(10)	−672.815064 −693.433775	0.000000001	84 T 1.000 **
rps8	M8(12) M7(10)	−718.655799 −724.953288	0.001840922	59 L 0.989 *
ycf1	M8(12) M7(10)	−11,993.590817 −12,006.745724	0.000001936	205 V 0.977 *; 206 F 0.975 *; 341 S 0.974 *; 495 S 0.952 *; 534 A 0.951 *; 1073 A 0.963 *; 1290 R 0.978 *; 1446 E 0.963 *; 1701 K 0.976 *; 1728 T 0.950 *

* $p < 0.05$; ** $p < 0.01$.

2.7. The Phylogenetic Position of *P. granatum*

Complete cp genomes are important as they can provide information for understanding the phylogenetic relationships at multiple taxonomic levels. For example, recent phylogenetic analyses based on protein-coding genes of the cp genome provided the broad phylogenetic framework for Viridiplantae, which has significant value both to evolutionary biologists and ecologists [55]. The genus *Punica* belongs to the order Myrtales and most likely originated from Saxifragales. However, the placement of genus *Punica* under different families such as Punicaceae, Lythraceae, and Myrtaceae has remained debatable [6]. Recent phylogenomic analysis based on 106 single-copy nuclear genes was performed and supported that *P. granatum* belongs to the Lythraceae family rather than the monogeneric Punicaeceae family [7].

To determine the position of *P. granatum* and to further analyze the relationships within Myrtales, we used the complete cp genome sequences to perform the phylogenetic analysis. Eighty-five species representing five families from the order Myrtales were selected. Two species from the order Vitales

(*Vitis acerifolia*, *Vitis vinifera*) were selected as outgroups. To avoid systematic errors produced by poor alignment, we removed poorly aligned sites using Gblocks. After the removal of the ambiguously aligned regions, the alignment contained 106,571 sites in total, including 20,088 parsimony-informative sites, 9178 singleton sites, and 77,305 constant sites. The method of data analysis (ML and BI) did not affect the results, and the treetopologies of ML and BI were found to be the same. The phylogenetic relationships among five families were fully resolved, and node support values for the ML type were higher than 90.

The tree showed (Figure 6) that all the accessions of the pomegranate were grouped into a single clade with other closely related species of the Lythraceae family. The monophyly of the family Lythraceae and the sister relation with family Onagraceae is highly supported (>90). Myrtaceae were strongly supported as monophyletic and formed a sister relationship with Melastomataceae. Five pomegranate accessions formed a clade with zero or nearly zero branches length, which suggests that the cp genome might be of limited use for cultivar identification and population genetic studies of *P. granatum*. Our full genomic data set resolved the phylogenomic placement of *Punica* and provided strong support for most relationships of Myrtales.

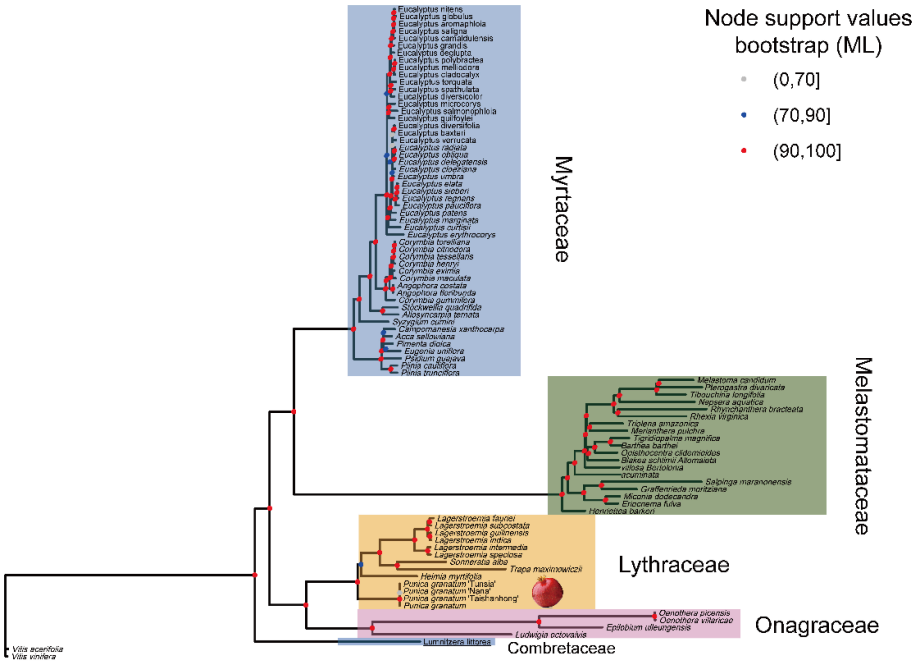


Figure 6. Phylogenetic tree was reconstructed using Maximum likelihood (ML) and Bayes inference (BI) methods based on complete cp genomes of the order Myrtales. Only the tree topology of the ML tree was presented.

3. Materials and Methods

3.1. Plant Material

Three *P. granatum* cultivars with distinct phenotypes were chosen to reconstruct the complete cp genome: 'Nana' is a dwarf pomegranate, which has a small and sour fruit with hard seeds. 'Tunisia' is a domesticated cultivar characterized as a normal-sized tree with sweet taste and soft seeds. 'Taishanhong' is a widely grown landrace in China, characterized as having a bright red peel with delicious taste and hard seeds. The materials used in this study were collected from the experimental orchard at Nanjing Forestry University. The voucher specimen was deposited in Nanjing Forestry University.

3.2. DNA Sequencing, Genome Assembly, and Annotation

Total genomic DNA was extracted from mature leaves using a modified CTAB protocol. Firstly, 1.0 μg Genomic DNA was sheared into an average fragment size of 350 bp by a Covaris S220 sonicator (Woburn, Massachusetts, MA, USA). Then, the size distribution and concentration of the libraries were determined using an Agilent 2100 Bioanalyzer and qualified by real-time PCR (2 nM), respectively. DNA libraries were sequenced on Illumina Hiseq X Ten (Nanjing, China) for at least 150 bp reads. The raw sequence data reported in this paper were deposited in the Genome Sequence Archive in Big Data Center [56], Beijing Institute of Genomics (BIG) [57], Chinese Academy of Sciences, under the BioProject with the accession number PRJCA001313. After the fragments were filtered and trimmed by the fastp program [58], clean reads were obtained. Subsequently, the high-quality paired-end reads were used to de novo assemble the complete cp genomes using the GetOrganelle program [27] with a combined k-mer of 95,105,125. Genome annotation was performed using the online program GeSeq [59] for the pomegranate cp genomes previously reported. The annotation results were inspected using Geneious [60] and adjusted manually as needed. The cp genome map was drawn using the online tool OGDRAW [61]. The complete cp genomes have been submitted to Genbank with accession number MK603511-MK603513.

3.3. Codon Usage

The complete cp genome of the pomegranate cultivar 'Taishanhong' was selected to analyze the codon usage pattern. The protein-coding genes with more than 300 nucleotides were extracted according to the annotation file. The GC content of GC1, GC2, and GC3 was calculated using an in-house python script. The codon usage indices were calculated by CodonW v1.4.4, including the relative synonymous codon usage (RSCU), codon adaptation index (CAI), and the effective number of codons (ENC). RSCU values were close to 1 indicating that all synonymous codons encoding the same amino acid were used equally. CAI is used to measure the extent of bias towards preferred codons in a gene. A higher CAI value means a stronger codon usage bias and a higher expression level. ENC is used to measure codon usage evenness. Its value ranges from 20 (extremely biased) to 61 (totally unbiased) [62].

3.4. RNA Editing Sites

Prediction of the possible RNA editing sites in *P. granatum* protein-coding genes were performed using the online program predictive RNA editor for plants (PREP) suite [63] with 35 genes as reference. Only those sites which had a cutoff value of 0.8 were kept.

3.5. Sequence Diversity

Four cp genomes from Lythraceae were downloaded from GenBank, including *Lagerstroemia indica* (NC_030484), *Heimia myrtifolia* (MG921615), *Sonneratia alba* (NC_039975), and *Trapa maximowiczii* (NC_037023). These four cp genomes together with that of our newly assembled 'Taishanhong' genome were used to detect the divergent hot spot. Intergenic and protein-coding regions from five Lythraceae cp genomes were extracted using an in-house python script. Multiple sequence alignment was performed using MAFFT [64] and the mean P-distances were calculated using R package 'ape' [65] with Kimura's two-parameter model.

3.6. Structure Comparison

IR expansion and contraction of cp genomes among the five Lythraceae species mentioned above were analyzed using IRscope (Helsinki, Finland) [66]. We also conducted a co-linear analysis. A pairwise alignment was achieved by the lastz program. The results were visualized using AliTV (Wurzburg, Germany) [67].

3.7. Positive Selection Analysis

In order to detect the protein-coding genes under selection in Lythraceae, the sequences for each gene were aligned separately using the Muscle (codon) implemented in MEGA [68], and the Maximum likelihood phylogenetic tree based on complete cp genome sequences was constructed using IQ-tree [69]. The site-specific model was performed to test for natural selection using the CODEML algorithm [70] implemented in EasyCodeML [71]. Six codon substitution models described as M0, M1a, M2a, M3, M7, and M8 were investigated. This model allowed the ω ratio to vary among sites with a fixed ω ratio in all branches in order to test for site-specific evolution in the gene phylogeny. Two likelihood ratio tests were performed to detect positively selected sites: M1a (neutral) vs. M2a (positive selection), M7 (β) vs. M8 (β and ω), and M0 (one-ratio) vs. M3 (discrete), which were compared using a site-specific model [72].

3.8. Phylogenetic Analysis

To determine the phylogenetic position of *P. granatum*, phylogenetic analysis was performed using the complete cp genomes in the Myrtales. The cp genomes previously published in the Myrtales and two species from Vitales were downloaded from NCBI using an in-house python script. Multiple sequences alignment was achieved by HomBlocks pipelines [73]. Some poorly aligned regions were removed with Gblocks [74]. Two methods were employed to construct a phylogenetic tree, including Maximum likelihood (ML), and Bayes inference (BI). The dataset was unpartitioned. ML was implemented in IQ-tree [69] under the best-fit model selected by using ModelFinder [75] according to Akaike information criterion (AIC). The ML tree was inferred with 1000 bootstrap replicates (the '-bb' options). BI was performed with MrBayes [76] under the best-fit model determined by Modeltest with the AIC. The Marjkov chain Monte Carlo (MCMC) analysis was run for 2 × 200,000 generations. Trees were sampled at every 1,000 generations with the first 25% discarded as burn-in. The stationarity was considered to be reached when the average standard deviation of split frequencies remained below 0.001. Phylogenetic trees with bootstrap values (BS) and posterior probabilities (PP) were visualized using R package 'ggtree' [77].

4. Conclusions

Next generation whole genome shotgun sequences of plant species often contain numerous reads that are derived from the cp genomes, which provides a unique opportunity to assemble complete cp genomes. This method of using low coverage of the whole genome sequencing data to recover highly repetitive genome regions such as organelle genomes is called the "whole genome skimming approach". In the present study, we determined and characterized the complete cp genomes of three *P. granatum* cultivars using the whole genome sequencing data. Sequence diversity analysis revealed that cp genome sequences may not be suitable for investigating the genetic diversity of pomegranate genotypes. The genome sequencing data of three different cultivars are valuable resources for pomegranate breeding programs. The complete cp genome sequences that were newly assembled in our study could provide valuable information for understanding the evolutionary relationships among the Myrtales.

Author Contributions: Data curation, M.Y., J.Z., and Y.D.; Formal analysis, M.Y., J.Z., and Y.H.; Funding acquisition, Z.Y.; Methodology, Z.Y.; Visualization, M.Y.; Writing—original draft, M.Y. and X.Z.; Writing—review and editing, X.Z., J.Z., and Z.Y.

Funding: This work was supported by the Initiative Project for Talents of Nanjing Forestry University (GXL2014070, GXL2018032), the Priority Academic Program Development of Jiangsu High Education Institutions (PAPD), and the Natural Science Foundation of Jiangsu Province (BK20180768).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Al-Khayri, J.M.; Jain, S.M.; Johnson, D.V. *Advances in Plant Breeding Strategies: Fruits*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 3.
2. Shahindokht, B.J. *Punica granatum* (Pomegranate) activity in health promotion and cancer prevention. *Oncol. Rev.* **2018**, *12*, 345.
3. Sreekumar, S.; Sithul, H.; Muraleedharan, P.; Azeez, J.-M.; Sreeharshan, S. Pomegranate Fruit as a Rich Source of Biologically Active Compounds. *BioMed Res. Int.* **2014**, *2014*, 686921. [CrossRef]
4. Malviya, S.; Arvind; Jha, A.; Hettiarachchy, N. Antioxidant and antibacterial potential of pomegranate peel extracts. *J. Food Sci. Technol.* **2014**, *51*, 4132–4137. [CrossRef] [PubMed]
5. Wang, J.-Y.; Zhu, C.; Qian, T.-W.; Guo, H.; Wang, D.-D.; Zhang, F.; Yin, X.-X. Extracts of black bean peel and pomegranate peel ameliorate oxidative stress-induced hyperglycemia in mice. *Exp. Ther. Med.* **2015**, *9*, 43–48. [CrossRef] [PubMed]
6. Silva, J.A.T.D.; Rana, T.S.; Narzary, D.; Verma, N.; Meshram, D.T.; Ranade, S.A. Pomegranate biology and biotechnology: A review. *Sci. Hortic.* **2013**, *160*, 85–107. [CrossRef]
7. Yuan, Z.-H.; Fang, Y.-M.; Zhang, T.-K.; Fei, Z.-J.; Han, F.-M.; Liu, C.-Y.; Liu, M.; Xiao, W.; Zhang, W.-J.; Wu, S. The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnol. J.* **2018**, *16*, 1363–1374. [CrossRef] [PubMed]
8. Sancho, R.; Cantalapiedra, C.-P.; Lopezalvarez, D.; Gordon, S.P.; Vogel, J.P.; Catalan, P.; Contrerasmoreira, B. Comparative plastome genomics and phylogenomics of Brachypodium: Flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* **2018**, *218*, 1631–1644. [CrossRef] [PubMed]
9. Ruhfel, B.R.; Gitzendanner, M.A.; Soltis, P.S.; Soltis, D.E.; Burleigh, J.G. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolut. Biol.* **2014**, *14*, 23. [CrossRef]
10. Yang, J.-B.; Tang, M.; Li, H.-T.; Zhang, Z.-R.; Li, D.-Z. Complete chloroplast genome of the genus Cymbidium: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolut. Biol.* **2013**, *13*, 84. [CrossRef]
11. Chen, N.; Sha, L.-N.; Dong, Z.-Z.; Tang, C.; Wang, Y.; Kang, H.-Y.; Zhang, H.-Q.; Yan, X.-B.; Zhou, Y.-H.; Fan, X. Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. *Gene* **2018**, *640*, 86–96. [CrossRef]
12. Mcfadden, G.I.; Van Dooren, G.G. Evolution: Red algal genome affirms a common origin of all plastids. *Curr. Biol.* **2004**, *14*, R514–R516. [CrossRef] [PubMed]
13. Daniell, H.; Lin, C.; Yu, M.; Chang, W.-J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genomo. Biol.* **2016**, *17*, 134. [CrossRef] [PubMed]
14. Petrillo, E.; Herz, M.A.G.; Fuchs, A.; Reifer, D.; Fuller, J.L.; Yanovsky, M.J.; Simpson, C.G.; Brown, J.W.S.; Barta, A.; Kalyna, M. A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* **2014**, *344*, 427–430. [CrossRef]
15. Godoy Herz, M.A.; Kubaczka, M.G.; Brzyzek, G.; Servi, L.; Krzyszton, M.; Simpson, C.; Brown, J.; Swiezewski, S.; Petrillo, E.; Kornblihtt, A.R. Light Regulates Plant Alternative Splicing through the Control of Transcriptional Elongation. *Mol. Cell* **2019**, *73*, 1066–1074. [CrossRef] [PubMed]
16. Alexia, T.; Cian, D.; Pooja, P.; Zachary, S.; María, E.S.; Lok, H.Y.; David, C.A.G.; Alexandre, Y.L.; Virendrasinh, K.; Andrew, D.W.; et al. Chloroplasts navigate towards the pathogen interface to counteract infection by the Irish potato famine pathogen. *bioRxiv* **2019**, 516443. [CrossRef]
17. Yang, J.; Feng, L.; Yue, M.; He, Y.L.; Zhao, G.F.; Li, Z.H. Species delimitation and interspecific relationships of the endangered herb genus Notopterygium inferred from multilocus variations. *Mol. Phylogenet. Evolut.* **2019**, *133*, 142–151. [CrossRef] [PubMed]
18. Chen, X.-L.; Zhou, J.-G.; Cui, Y.-X.; Wang, Y.; Duan, B.-Z.; Yao, H. Identification of *Ligularia* Herbs Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* **2018**, *9*, 695. [CrossRef] [PubMed]
19. Cheng, H.; Li, J.-F.; Zhang, H.; Cai, B.-H.; Gao, Z.-H.; Qiao, Y.-S.; Mi, L. The complete chloroplast genome sequence of strawberry (*Fragaria* × *ananassa* Duch.) and comparison with related species of Rosaceae. *PeerJ* **2017**, *5*, e3919. [CrossRef] [PubMed]

20. Redwan, R.-M.; Saidin, A.; Kumar, S.-V. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biol* **2015**, *15*, 196. [CrossRef] [PubMed]
21. Yao, X.-H.; Tang, P.; Li, Z.-Z.; Li, D.-W.; Liu, Y.-F.; Huang, H.-W. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. *PLoS ONE* **2015**, *10*, e2109347. [CrossRef]
22. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; Depamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J. Methods for Obtaining and Analyzing Whole Chloroplast Genome Sequences. *Methods Enzymol.* **2005**, *395*, 348–384. [PubMed]
23. Khan, A.; Khan, I.-A.; Asif, H.M.; Azim, M. Current trends in chloroplast genome research. *Afr. J. Biotechnol.* **2010**, *9*, 3494–3500.
24. Twyford, A.D.; Ness, R.W. Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* **2017**, *17*, 858–868. [CrossRef] [PubMed]
25. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18. [CrossRef]
26. Ankenbrand, M.J.; Pfaff, S.; Terhoeven, N.; Qureischi, M.; Gundel, M.; Weis, C.L.; Hackl, T.; Forster, F. chloroExtractor: Extraction and assembly of the chloroplast genome from whole genome shotgun data. *J. Open Sour. Softw.* **2018**, *3*, 464. [CrossRef]
27. Jin, J.; Yu, W.; Yang, J.; Song, Y.; Yi, T.; Li, D. GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv* **2018**, 256479. [CrossRef]
28. Ivanova, Z.; Sablok, G.; Daskalova, E.; Zahmanova, G.; Apostolova, E.; Yahubyan, G.; Baev, V. Chloroplast Genome Analysis of Resurrection Tertiary Relict *Haberlea rhodopensis* Highlights Genes Important for Desiccation Stress Response. *Front. Plant Sci.* **2017**, *8*, 204. [CrossRef] [PubMed]
29. Song, Y.; Yao, X.; Tan, Y.-H.; Gan, Y.; Corlett, R.T. Complete chloroplast genome sequence of the avocado: Gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Can. J. For. Res.* **2016**, *46*, 1293–1301. [CrossRef]
30. Yang, J.-B.; Yang, S.-X.; Li, H.-T.; Yang, J.; Li, D.-Z. Comparative Chloroplast Genomes of Camellia Species. *PLoS ONE* **2013**, *8*, e73053. [CrossRef]
31. Wang, S.; Yang, C.; Zhao, X.; Chen, S.; Qu, G. Complete chloroplast genome sequence of Betula platyphylla: Gene organization, RNA editing, and comparative and phylogenetic analyses. *BMC Genom.* **2018**, *19*, 950. [CrossRef]
32. Wickens, M.; Cox, M.M. Critical reviews in biochemistry and molecular biology. Introduction. *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44*, 2. [CrossRef] [PubMed]
33. Li, X.; Li, Y.-F.; Zang, M.-Y.; Li, M.-Z.; Fang, Y.-M. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Quercus acutissima*. *Int. J. Mol. Sci.* **2018**, *19*, 2443. [CrossRef] [PubMed]
34. Asaf, S.; Khan, A.L.; Khan, A.R.; Waqas, M.; Kang, S.; Khan, M.A.; Lee, S.; Lee, I. Complete Chloroplast Genome of *Nicotiana otophora* and its Comparison with Related Species. *Front. Plant Sci.* **2016**, *7*, 843. [CrossRef] [PubMed]
35. Batnini, M.A.; Bourguiba, H.; Trififarah, N.; Krichen, L. Molecular diversity and phylogeny of Tunisian *Prunus armeniaca* L. by evaluating three candidate barcodes of the chloroplast genome. *Sci. Hortic.* **2019**, *245*, 99–106. [CrossRef]
36. Norouzi, M.; Talebi, M.; Sayedtabatabaei, B. Chloroplast microsatellite diversity and population genetic structure of Iranian pomegranate (*Punica granatum* L.) genotypes. *Sci. Hortic.* **2012**, *137*, 114–120. [CrossRef]
37. Feng, C.; Xu, C.-J.; Wang, Y.; Liu, W.-L.; Yin, X.-R.; Li, X.; Chen, M.; Chen, K.-S. Codon usage patterns in Chinese bayberry (*Myrica rubra*) based on RNA-Seq data. *BMC Genom.* **2013**, *14*, 732. [CrossRef] [PubMed]
38. Qin, Z.; Cai, Z.-Q.; Xia, G.-M.; Wang, M.-C. Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genom.* **2013**, *14*, 56. [CrossRef] [PubMed]
39. Wang, L.-Y.; Xing, H.-X.; Yuan, Y.-C.; Wang, X.-L.; Saeed, M.; Tao, J.-C.; Feng, W.; Zhang, G.-H.; Song, X.-L.; Sun, X.-Z. Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS ONE* **2018**, *13*, e0194372. [CrossRef]
40. Ning, L.I.; Sun, M.-H.; Jiang, Z.-S.; Shu, H.-R.; Zhang, S.-Z. Genome-wide analysis of the synonymous codon usage patterns in apple. *J. Integr. Agric.* **2016**, *15*, 983–991.
41. Suzuki, H.; Morton, B.R. Codon Adaptation of Plastid Genes. *PLoS ONE* **2016**, *11*, e0154306. [CrossRef]

42. Zhang, R.-Z.; Zhang, L.; Wang, W.; Zhang, Z.; Du, H.-H.; Qu, Z.; Li, X.-Q.; Xiang, H. Differences in Codon Usage Bias between Photosynthesis-Related Genes and Genetic System-Related Genes of Chloroplast Genomes in Cultivated and Wild Solanum Species. *Int. J. Mol. Sci.* **2018**, *19*, 3142. [CrossRef] [PubMed]
43. Nie, X.-J.; Deng, P.-C.; Feng, K.-W.; Liu, P.-X.; Du, X.-H.; You, F.M.; Song, W.-N. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Mol. Biol. Rep.* **2014**, *32*, 828–840. [CrossRef]
44. Wright, F. The effective number of codons used in a gene. *Gene* **1990**, *87*, 23–29. [CrossRef]
45. Das, S.; Paul, S.; Dutta, C. Synonymous codon usage in adenoviruses: Influence of mutation, selection and protein hydropathy. *Virus Res.* **2006**, *117*, 227–236. [CrossRef]
46. Harris, R.S.; Petersenmahrt, S.K.; Neuberger, M.S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **2002**, *10*, 1247–1253. [CrossRef]
47. Bentolila, S.; Oh, J.; Hanson, M.R.; Bukowski, R. Comprehensive High-Resolution Analysis of the Role of an Arabidopsis Gene Family in RNA Editing. *PLoS Genet.* **2013**, *9*. [CrossRef] [PubMed]
48. Saina, J.K.; Li, Z.; Gichira, A.W.; Liao, Y. The Complete Chloroplast Genome Sequence of Tree of Heaven (Ailanthus altissima (Mill.) (Sapindales: Simaroubaceae), an Important Pantropical Tree. *Int. J. Mol. Sci.* **2018**, *19*, 929. [CrossRef]
49. Pinard, D.; Myburg, A.A.; Mizrachi, E. The plastid and mitochondrial genomes of Eucalyptus grandis. *BMC Genom.* **2019**, *20*, 132. [CrossRef]
50. Rabah, S.O.; Lee, C.; Hajrah, N.H.; Makki, R.M.; Alharby, H.F.; Alhebshi, A.M.; Sabir, J.; Jansen, R.K.; Ruhlman, T.A. Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew. *Plant Genome* **2017**, *10*. [CrossRef]
51. Li, Y.; Zhang, J.; Li, L.; Gao, L.; Xu, J.; Yang, M. Structural and Comparative Analysis of the Complete Chloroplast Genome of Pyrus hopeiensis—"Wild Plants with a Tiny Population"—and Three Other Pyrus Species. *Int. J. Mol. Sci.* **2018**, *19*, 3262. [CrossRef]
52. Li, B.; Zheng, Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci. Rep.* **2018**, *8*, 9285. [CrossRef] [PubMed]
53. Dong, W.; Wang, R.; Zhang, N.; Fan, W.; Fang, M.; Li, Z. Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. *Int. J. Mol. Sci.* **2018**, *19*, 716. [CrossRef] [PubMed]
54. Hao, D.C.; Chen, S.L.; Xiao, P.G. Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *J. Plant Res.* **2010**, *123*, 241–247. [CrossRef] [PubMed]
55. Gitzendanner, M.; Soltis, P.; Wong, G.; Ruhfel, B.; Soltis, D. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* **2018**, *105*, 291–301. [CrossRef] [PubMed]
56. Wang, Y.-Q.; Song, F.-H.; Zhu, J.-W.; Zhang, S.-S.; Yang, Y.-D.; Chen, T.-T.; Tang, B.; Dong, L.-L.; Ding, N.; Zhang, Q. GSA: Genome Sequence Archive. *Genom. Proteom. Bioinform.* **2017**, *15*, 14–18. [CrossRef] [PubMed]
57. Zhang, Z.; Zhao, W.-M.; Xiao, J.-F.; Bao, Y.-M.; Wang, F.; Hao, L.-L.; Zhu, J.-W.; Chen, T.-T.; Zhang, S.-S.; Chen, X. Database Resources of the BIG Data Center in 2018. *Nucleic Acids Res.* **2018**, *46*, D14–D20.
58. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
59. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbrichtjones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef]
60. Kearse, M.; Moir, R.; Wilson, A.; Stoneshavas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
61. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [CrossRef]
62. Chen, X.; Jing, D.; Tong, C.-F.; Gong, X.-D.; Qiang, W.; Qiang, Z.-G. Analysis of Synonymous Codon Usage Patterns in Seven Different Citrus Species. *Evol. Bioinform. Online* **2013**, *9*, 215–228.
63. Mower, J.P. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, 253–259. [CrossRef] [PubMed]
64. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]

65. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **2004**, *20*, 289–290. [CrossRef] [PubMed]
66. Amiryousefi, A.; Hyvonen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]
67. Ankenbrand, M.J.; Hohlfeld, S.; Hackl, T.; Forster, F. AliTV—Interactive visualization of whole genome comparisons. *PeerJ* **2017**, *3*. [CrossRef]
68. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evolut.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
69. Nguyen, L.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
70. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556. [CrossRef]
71. Gao, F.; Chen, C.; Arab, D.A.; Du, Z.; He, Y.; Ho, S.Y.W. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evolut.* **2019**, *9*, 3891–3898. [CrossRef]
72. Gao, F.; Du, Z.; Shen, J.; Yang, H.; Liao, F. Genetic diversity and molecular evolution of Ornithogalum mosaic virus based on the coat protein gene sequence. *PeerJ* **2018**, *6*. [CrossRef]
73. Bi, G.; Mao, Y.; Xing, Q.; Cao, M. HomBlocks: A multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics* **2017**, *110*, 18–22. [CrossRef]
74. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evolut.* **2000**, *17*, 540–552. [CrossRef]
75. Kalyaanamoorthy, S.; Minh, B.; Wong, T.K.F.; Von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
76. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef]
77. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evolut.* **2017**, *8*, 28–36. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Complete Chloroplast Genomes and Comparative Analysis of Sequences Evolution among Seven *Aristolochia* (Aristolochiaceae) Medicinal Species

Xiaoqin Li [1,2,3], Yunjuan Zuo [1], Xinxin Zhu [4], Shuai Liao [5] and Jinshuang Ma [1,*]

1 Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai Chenshan Botanical Garden, Shanghai 201602, China; lxq@sibs.ac.cn (X.L.); zuo.yunjuan@gmail.com (Y.Z.)
2 Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences, Shanghai 201602, China
3 University of Chinese Academy of Sciences, Shanghai 201602, China
4 College of Life Sciences, Xinyang Normal University, Xinyang 464000, Henan, China; huamixinhua1@gmail.com
5 School of Life Sciences, East China Normal University, Shanghai 200241, China; shuai_liao@bjfu.edu.cn
* Correspondence: jinshuangma@gmail.com; Tel.: 86-21-6765-7816

Received: 18 January 2019; Accepted: 21 February 2019; Published: 28 February 2019

Abstract: Aristolochiaceae, comprising about 600 species, is a unique plant family containing aristolochic acids (AAs). In this study, we sequenced seven species of *Aristolochia*, and retrieved eleven chloroplast (cp) genomes published for comparative genomics analysis and phylogenetic constructions. The results show that the cp genomes had a typical quadripartite structure with conserved genome arrangement and moderate divergence. The cp genomes range from 159,308 bp to 160,520 bp in length and have a similar GC content of 38.5%–38.9%. A total number of 113 genes were identified, including 79 protein-coding genes, 30 tRNAs and four rRNAs. Although genomic structure and size were highly conserved, the IR-SC boundary regions were variable between these seven cp genomes. The *trnH*-GUG genes, are one of major differences between the plastomes of the two subgenera *Siphisia* and *Aristolochia*. We analyzed the features of nucleotide substitutions, distribution of repeat sequences and simple sequences repeats (SSRs), positive selections in the cp genomes, and identified 16 hotspot regions for genomes divergence that could be utilized as potential markers for phylogeny reconstruction. Phylogenetic relationships of the family Aristolochiaceae inferred from the 18 cp genome sequences were consistent and robust, using maximum parsimony (MP), maximum likelihood (ML), and Bayesian analysis (BI) methods.

Keywords: *Aristolochia*; chloroplast genome; molecular evolution; compare analysis; phylogeny

1. Introduction

Aristolochia sensu lato, comprising about 500 species, is the largest genus of Aristolochiaceae, with a broad distribution range from tropical to subtropical, extending to temperate regions [1,2]. Several species of *Aristolochia*, such as *Aristolochia moupinensis*, *Aristolochia tagala*, and *Aristolochia mollissima*, have been reported as traditional Chinese medicines [3,4]. Aristolochiaceae is a unique plant family containing aristolochic acids (AAs), and their derivatives are widely implicated in liver cancers [5,6]. However, current studies have demonstrated that AAs are of nephrotoxicity, carcinogenicity, and mutagenicity [7,8]. The sale and use of AA-containing herbal preparations have been restricted in many countries [9].

The monophyly of Aristolochiaceae was well supported in most analysis, and was divided into two subfamilies, Asaroideae and Aristolochioideae [10,11]. The studies recognized two genera *Saruma* and *Asarum* in Asaroideae [10–13]. The genus *Aristolochia* of subfamily Aristolochioideae

was classified into two major lineages, as indicated by previous studies based on morphological characters and molecular phylogenetic methods [10,14–16]. In the past years, the nuclear *ITS*2, *phyA* gene, and several plastid genome regions (such as *matk*, *rbcL*, *trnK*, and *trnL-trnF*) or their combinations have been frequently used in molecular systematics of Aristolochiaceae [11,15,17,18]. The inter-simple sequence repeat (ISSR) markers were also used to identify diverse genetic stocks and understand the evolutionary relationships of *Aristolochia* [19,20]. The selected loci failed to provide sufficient phylogenetic information to elucidate the evolutionary relationships among *Aristolochia* species. A universal barcode either using whole chloroplast (cp) genomes or hyper-variable regions are urgently needed, which may significantly improve the low resolution in plant relationships and contribute to the conservation, domestication, and utilization of *Aristolochia* plants.

The chloroplast is the key organelle for photosynthesis and carbon fixation in green plants [21]. Their genomes could provide valuable information for taxonomic classification and phylogenetic reconstruction among species of land plants [22–25]. Typical cp genomes in angiosperms have a generally conserved quadripartite circular structure with two copies of inverted repeat (IR) regions that are separated by a large single copy (LSC) region and a small single copy (SSC) region, and encode 120–135 genes with sizes in the range of 120–170 kb [26,27]. In recent years, the cp genomes of *Aristolochia debilis*, *Aristolochia contorta*, *Saruma henryi*, and nine species of *Asarum* within the Aristolochiaceae family have been reported [28–31]. Those sequenced cp genomes of Aristolochiaceae, except for those of *Asarum* species, were conserved in length, gene and GC content, from which no rearrangement event had been detected.

With the rapid development of next-generation sequencing (NGS), it is now more convenient and cheaper to obtain cp genome sequences, feasible to compare analysis of sequences evolution among different individuals. In this study, we reported seven complete cp genomes of *Aristolochia* and conducted comparative genomic analyses, focused on gene size, content, patterns of nucleotide substitutions, and variable sites. Another 12 published cp genome sequences of Magnoliids downloaded from the National Center for Biotechnology Information (NCBI) organelle genome database (https://www.ncbi.nlm.nih.gov) [32] were also used to detect selective sites, repeat sequences, simple sequence repeats (SSRs), and phylogenetic constructions. We performed these comparative genomes analysis to obtain comprehensive understanding the structure of plastomes within *Aristolochia* and to provide genetic resources for future research in the genus.

2. Results

2.1. The Chloroplast Genome Structures of Species

All the species of *Aristolochia* we sequenced had a typical quadripartite structure, with a circular molecule of 159,308 bp to 160,520 bp in length. The complete cp genomes of involved species comprise an LSC region (88,652–89,859 bp) and an SSC region (19,322–19,799 bp), separated by a pair of IRs ranging from 25,242 bp to 25,700 bp in length (Figure 1, Table 1). GC content of the plastomes of the seven *Aristolochia* species varies slightly from 38.5% to 38.9% (Table 1). The GC content within coding sequence (CDS) of the two species (*A. tagala* and A. *tubiflora*) of subgenus *Aristolochia* and five species (*A. kunmingensis*, *A. moupinensis*, *A. macrophylla*, *A. kaempferi*, and *A. mollissima*) of subgenus *Siphisia* was 38.9% and 39.2%, respectively. GC% content of the first position was higher compared to those of the second and third positions (Figure 2, Table S1). A total of 113 unique genes were identified in the seven cp genomes, including 79 protein coding genes, 30 tRNAs and four rRNAs, 19 or 18 genes of which duplicated in the IR region (Tables 1 and 2).

A

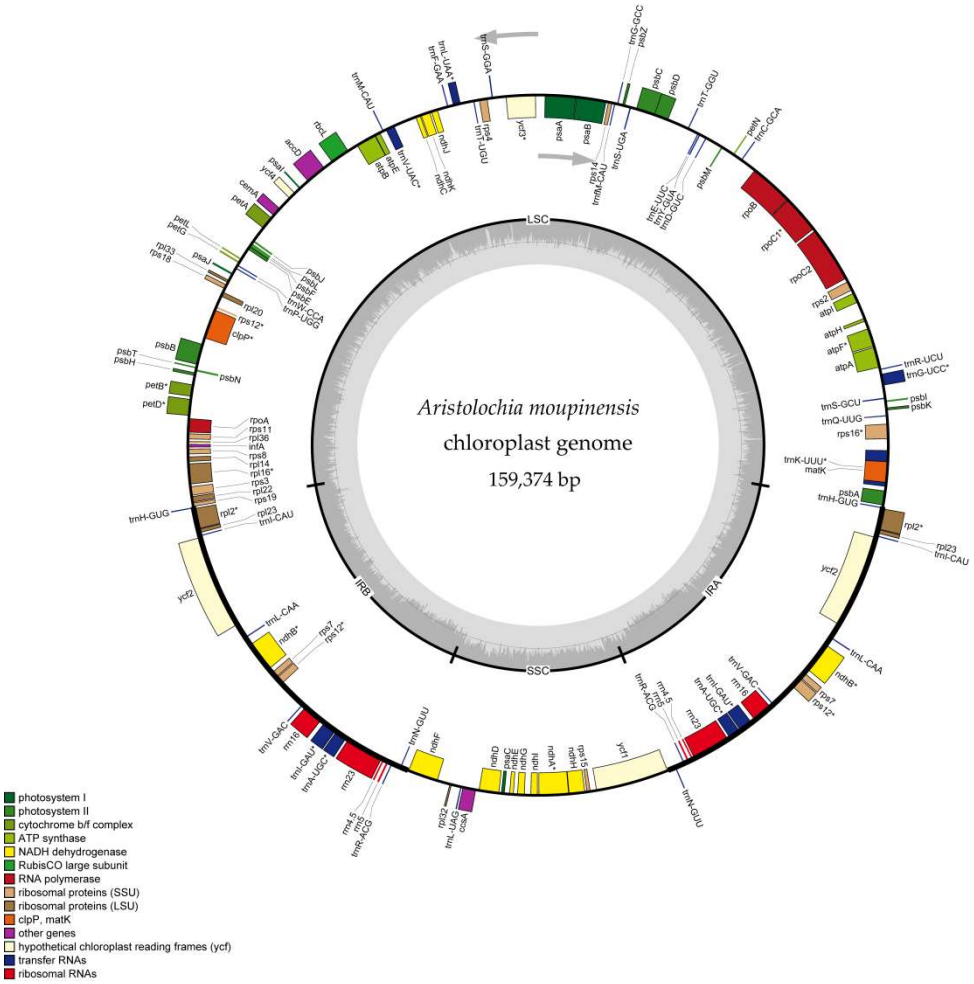


Figure 1. *Cont*.

B

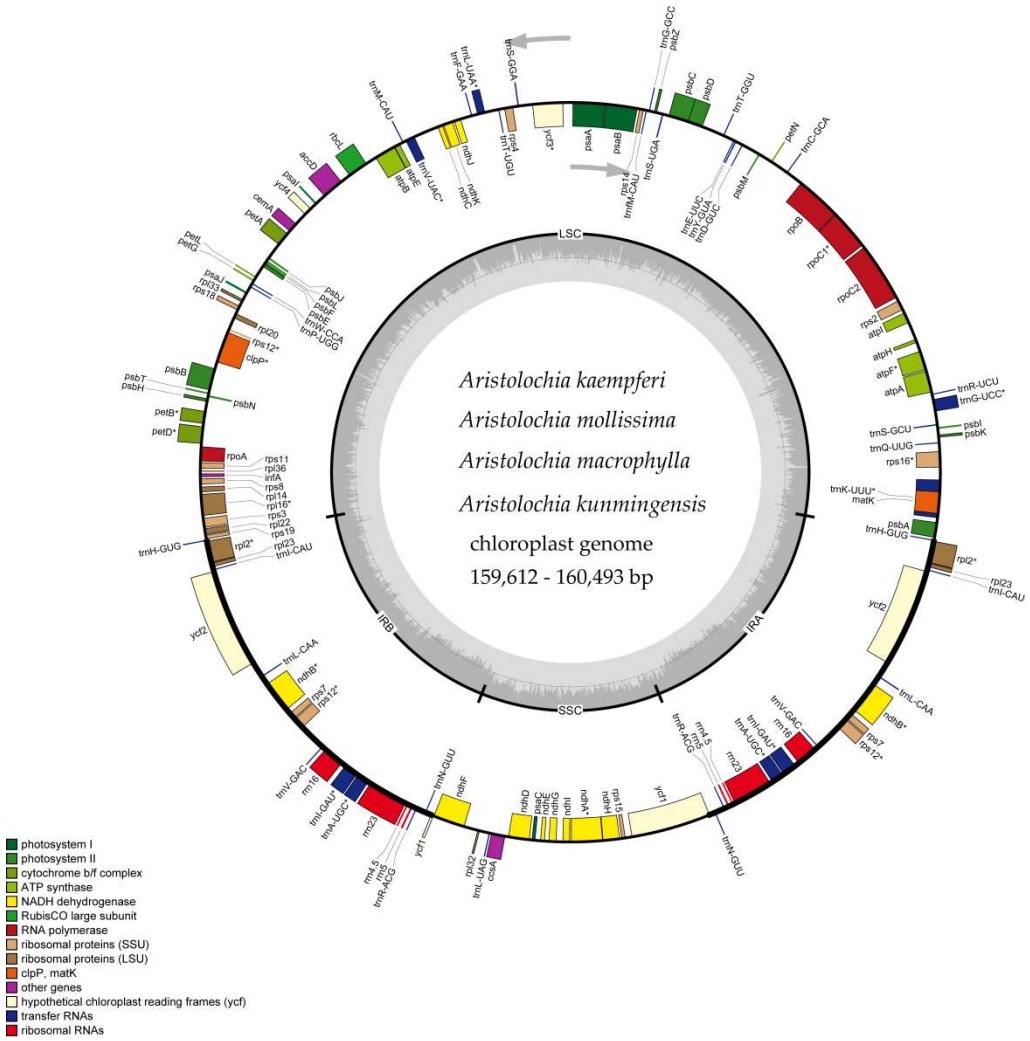


Figure 1. *Cont*.

C

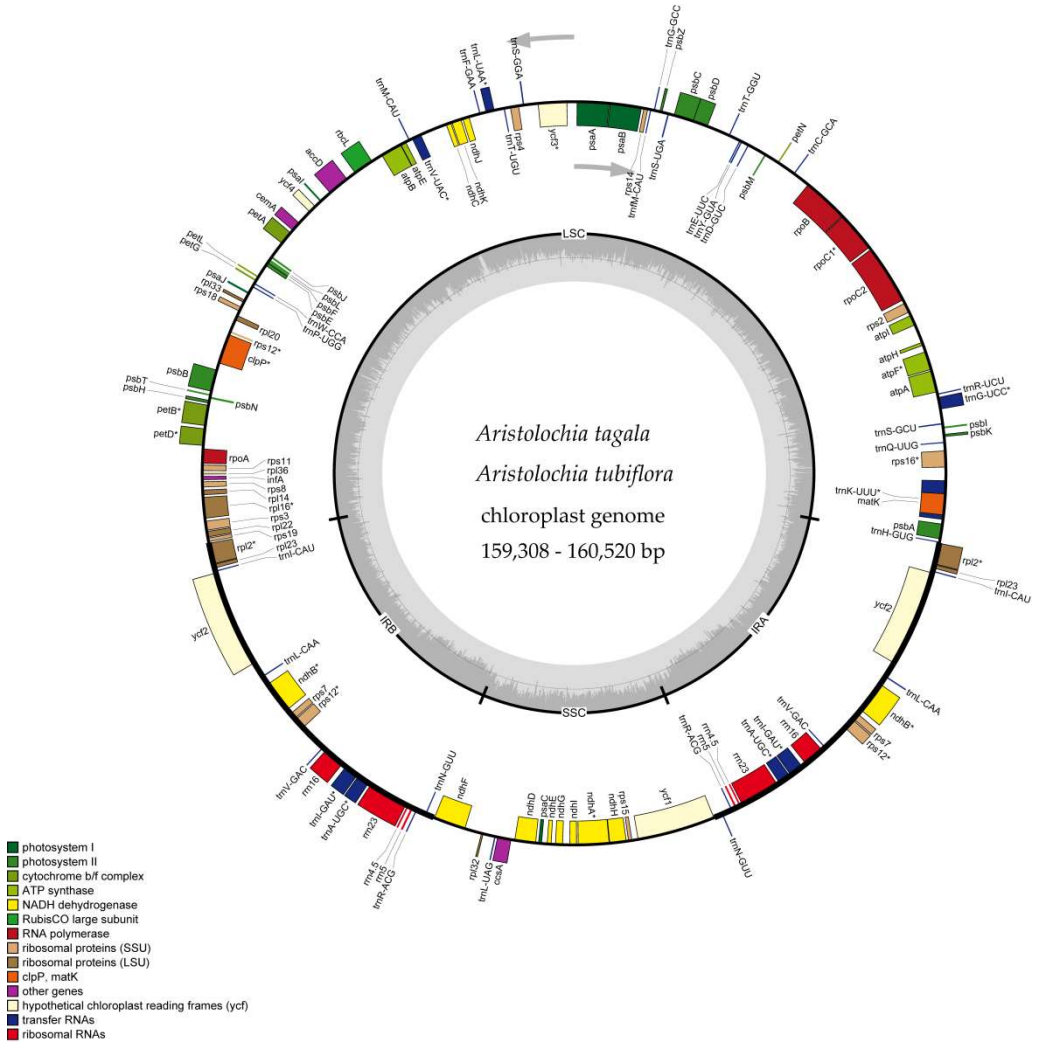


Figure 1. Gene maps of the complete cp genome of seven species of *Aristolochia*. Gene map of cp genome of (**A**) *Aristolochia manshuriensis*; (**B**) *Aristolochia kaempferi*, *Aristolochia macrophylla*, *Aristolochia mollissima* and *Aristolochia kunmingensis*; (**C**) *Aristolochia tagala* and *Aristolochia tubiflora*. Genes on the inside of the circle are transcribed clockwise, while those outside are transcribed counter clockwise. The darker gray in the inner circle corresponds to GC content, whereas the lighter gray corresponds to AT content.

Table 1. Summary of complete chloroplast (cp) genomes of *Aristolochia* species.

Species	Total	LSC	IR	SSC	CDS	Total	Protein Coding Genes	tRNA	rRNA	GC%
A. kaempferi	159,612	88,890	25,681	19,360	79191.0	113	79	30	4	38.8
A. kunmingensis	160,051	89,308	25,698	19,347	79143.0	113	79	30	4	38.7
A. macrophylla	160,493	89,788	25,664	19,377	79116.0	113	79	30	4	38.6
A. mollissima	159,653	88,948	25,681	19,338	79194.0	113	79	30	4	38.8
A. moupinensis	159,374	88,652	25,700	19,322	78753.0	113	79	30	4	38.7
A. tagala	159,308	89,414	25,242	19,410	78582.0	113	79	30	4	38.5
A. tubiflora	160,520	89,859	25,431	19,799	78624.0	113	79	30	4	38.8

Table 2. Gene contents in the cp genomes of *Aristolochia* species.

No.	Group of Genes	Genes Names	Amount
1	Photosystems I	*psaA, psaB, psaC, psaI, psaJ*	5
2	Photosystems II	*psbA, psbB, psbC, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*	15
3	Cytochrome b/f complex	*petA, petB *, petD *, petG, petL, petN*	6
4	ATP synthase	*atpA, atpB, atpE, atpF *, atpH, atpI*	6
5	NADH dehydrogenase	*ndhA *, ndhB *(x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12 (1)
6	Rubisco large subunit	*rbcL*	1
7	RNA polymerase	*rpoA, rpoB, rpoC1 *, rpoC2*	4
8	Ribosomal proteins(SSU)	*rps2, rps3, rps7(x2), rps8, rps11, rps12 **(x2), rps14, rps15, rps16 *, rps18, rps19*	14 (2)
9	Ribosomal proteins(LSU)	*rpl2 *(x2), rpl14, rpl16 *, rpl20, rpl22, rpl23(x2), rpl32, rpl33, rpl36*	11 (2)
10	Assembly/stability of photosystem I	*ycf3 **, ycf4*	2
11	Transfer RNAs	*37/38 tRNAs (6 contain an intron, 7/8 in the IRs)*	37 (7)/38(8)
12	Ribosomal RNAs	*rrn4.5(x2), rrn5(x2), rrn16(x2), rrn23(x2)*	8 (8)
13	RNA processing	*matK*	1
14	Carbon metabolism	*cemA*	1
15	Cytochrome c synthesis	*ccsA*	1
16	Proteins of unknown function	*ycf1, ycf2(x2)*	3 (1)
17	Other genes	*accD, clpP **, infA*	3

* Gene contains one intron; ** gene contains two introns; (x2) indicates the number of the repeat unit is 2.

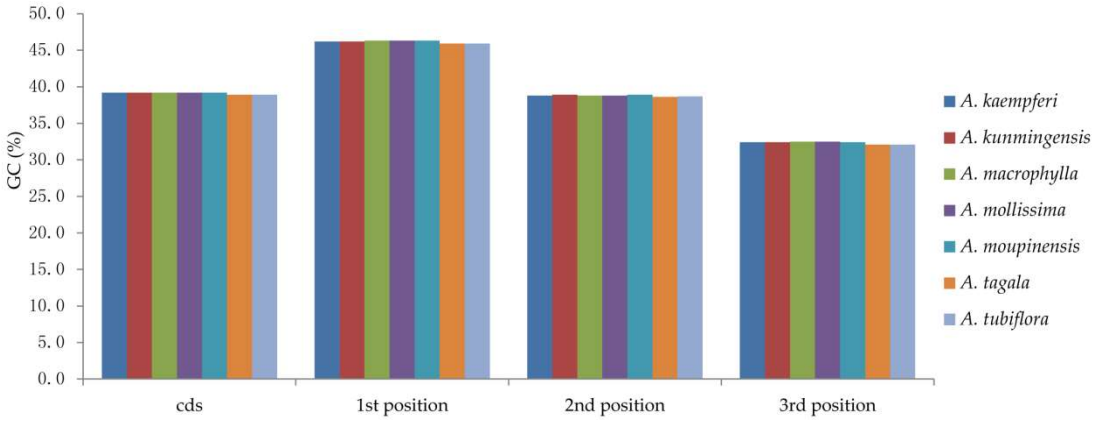


Figure 2. The GC (%) composition in different positions of coding sequence (CDS) region of species within *Aristolochia*.

Introns play an important role in the regulation of some gene expressions [33]. Eighteen genes of seven plastomes contain one intron, including *atpF*, *rpoC1*, *ycf3*, *rps12*, *rpl2*, *rpl16*, *clpP*, *petB*, *petD*, *rps16*, *ndhA*, *ndhB*, and six tRNA genes, while three genes (*clpP*, *ycf3*, and *rps12*) contain two introns. The longest intron occurred in the *trnK*-UUU gene is 2552–2687 bp of seven plastomes, and had been used to the inter- and intra-species of *Aristolochia* [2,16]. In addition, the length of *rpl2* intron in species of subgenus *Siphisia* and subgenus *Aristolochia* is 700 bp and 659 bp, respectively (Table 3).

Table 3. Genes with introns in the seven cp genomes of *Aristolochia* as well as the lengths of the exons and introns.

Taxon	Gene	Location	Exon I	Intron I	Exon II	Intron II	Exon III
A. kunmingensis	*atpF*	LSC	144	792	411		
	clpP	LSC	71	912	288	674	250
	ndhA	SSC	551	1095	541		
	ndhB	IR	777	703	756		
	petB	LSC	6	215	642		
	petD	LSC	6	702	477		
	rpl16	LSC	9	1092	402		
	rpl2	IR	391	700	431		
	rpoC1	LSC	432	762	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	842	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	763	48		
	trnI-GAU	IR	37	936	35		
	trnK-UUU	LSC	37	2574	35		
	trnL-UAA	LSC	35	454	50		
	trnV-UAC	LSC	37	587	36		
	ycf3	LSC	126	871	226	745	155
A. mollissima	*atpF*	LSC	144	772	411		
	clpP	LSC	71	897	295	672	243
	ndhA	SSC	551	1097	541		
	ndhB	IR	777	702	756		
	petB	LSC	6	215	642		
	petD	LSC	6	702	477		
	rpl16	LSC	9	1098	399		
	rpl2	IR	391	700	431		
	rpoC1	LSC	432	762	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	848	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	764	47		
	trnI-GAU	IR	37	936	35		
	trnK-UUU	LSC	37	2562	35		
	trnL-UAA	LSC	35	455	50		
	trnV-UAC	LSC	37	587	36		
	ycf3	LSC	126	820	226	753	155
A. kaempferi	*atpF*	LSC	144	780	411		
	clpP	LSC	71	900	295	671	243
	ndhA	SSC	551	1097	541		
	ndhB	IR	777	702	756		
	petB	LSC	6	215	642		
	petD	LSC	6	702	477		
	rpl16	LSC	9	1101	399		
	rpl2	IR	391	700	431		
	rpoC1	LSC	432	765	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	842	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	764	48		
	trnI-GAU	IR	37	936	35		
	trnK-UUU	LSC	37	2552	35		
	trnL-UAA	LSC	50	455	35		
	trnV-UAC	LSC	37	587	36		
	ycf3	LSC	126	812	226	752	155

Table 3. *Cont.*

Taxon	Gene	Location	Exon I	Intron I	Exon II	Intron II	Exon III
A. moupinensis	*atpF*	LSC	144	789	411		
	clpP	LSC	71	909	288	669	250
	ndhA	SSC	551	1101	541		
	ndhB	IR	777	703	756		
	petB	LSC	6	211	646		
	petD	LSC	6	708	477		
	rpl16	LSC	399	1100	9		
	rpl2	IR	391	700	431		
	rpoC1	LSC	432	764	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	839	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	758	48		
	trnI-GAU	IR	37	936	35		
	trnK-UUU	LSC	37	2567	35		
	trnL-UAA	LSC	35	462	50		
	trnV-UAC	LSC	37	587	36		
	ycf3	LSC	126	920	226	746	155
A. macrophylla	*atpF*	LSC	144	778	411		
	clpP	LSC	71	928	288	664	250
	ndhA	SSC	551	1084	541		
	ndhB	IR	777	702	756		
	petB	LSC	6	215	642		
	petD	LSC	6	706	477		
	rpl16	LSC	9	1095	402		
	rpl2	IR	391	700	431		
	rpoC1	LSC	432	788	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	836	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	755	47		
	trnI-GAU	IR	37	936	35		
	trnK-UUU	LSC	37	2558	35		
	trnL-UAA	LSC	35	475	50		
	trnV-UAC	LSC	37	589	36		
	ycf3	LSC	126	892	226	757	155
A.tubiflora	*atpF*	LSC	144	751	411		
	clpP	LSC	71	819	288	671	250
	ndhA	SSC	551	1079	541		
	ndhB	IR	777	705	756		
	petB	LSC	6	214	642		
	petD	LSC	6	693	477		
	rpl16	LSC	9	1077	399		
	rpl2	IR	391	657	431		
	rpoC1	LSC	432	780	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	889	221		
	trnA-UGC	IR	38	809	35		
	trnG-UCC	LSC	24	768	47		
	trnI-GAU	IR	37	937	35		
	trnK-UUU	LSC	37	2635	35		
	trnL-UAA	LSC	35	514	50		
	trnV-UAC	LSC	37	594	36		
	ycf3	LSC	126	764	226	752	149

Table 3. *Cont.*

Taxon	Gene	Location	Exon I	Intron I	Exon II	Intron II	Exon III
A. tagala	*atpF*	LSC	144	778	408	671	250
	clpP	LSC	71	802	288		
	ndhA	SSC	551	1101	541		
	ndhB	IR	777	704	756		
	petB	LSC	6	219	642		
	petD	LSC	6	488	477		
	rpl16	LSC	9	1071	399		
	rpl2	IR	391	657	431		
	rpoC1	LSC	432	785	1617		
	rps12	LSC	114		232	536	23
	rps16	LSC	46	848	221		
	trnA-UGC	IR	38	804	35		
	trnG-UCC	LSC	24	768	48		
	trnI-GAU	IR	37	743	35		
	trnK-UUU	LSC	37	2687	35		
	trnL-UAA	LSC	35	490	50		
	trnV-UAC	LSC	37	595	36		
	ycf3	LSC	126	830	226	763	149

2.2. IR Contraction and Expansion

The IR regions are expanded in five species of subgenus *Siphisia* compare with other two species (*A. tagala* and *A. tubiflora*) of subgenus *Aristolochia*, indicated by different duplication genes in the IR regions, where eight or seven tRNA genes were duplicated, respectively (Figure 1, Table 2). The size of the IR region of subgenus *Siphisia* varies from 25,664 bp to 25,700 bp, and is 25,242 bp and 25,431 bp in the two plastomes of subgenus *Aristolochia* (Table 1).

Fluctuation of IR-SC borders, together with the adjacent genes, were examined among seven *Aristolochia* species and six plastomes retrieved from GenBank (including *Aristolochia contorta*: NC_036152.1, *Aristolochia debilis*: NC_036153.1, *Asarum canadense*: MG544845-MG544851, *Saruma henryi*: MG520100, *Piper auritum*: NC_034697.1, and *Drimys granadensis*: NC_008456.1) (Figure 3). The LSC-IRb border, was located within the genic spacer of *rps19-trnH* for *A. kaempferi*, *A. macrophylla*, and *A. mollissima* (Type I), within the *rps19* gene for *A. kunmingensis* and *A. moupinensis* (Type II), while in the *rps19-rpl2* spacer for *A. tagala* and *A. tubiflora* (Type III). There were two types of SSC-IRa border among 13 detected species. In the three plastomes (*A. moupinensis*, *A. tubiflora*, and *A. tagala*), which *ycf1* gene was fully located in the SSC region, and 25-43 bp apart from the SSC-IRa border. The SSC-IRa border was situated in the coding region *ycf1* gene in the other 10 sequenced species, which spanned into the IRa region. Among the 10 detected species, the pseudogene *ycf1* in the IRb region with the same length as far as the IRa expanded into *ycf1* gene, and the length ranged from 153 bp to 2271 bp. The *ndhF* gene was entirely located in the SSC region in 10 species of Aristolochiaceae, but varied in distance (11-80 bp) from the IRb-SSC border. The LSC-IRa border in the species of subgenus *Aristolochia* was situated in the *trnH* gene with 10 bp into the IRa region (Type III), while the border was located in the *trnH-psbA* spacer in subgenus *Siphisia* species (Type I and II) (Figure 3).

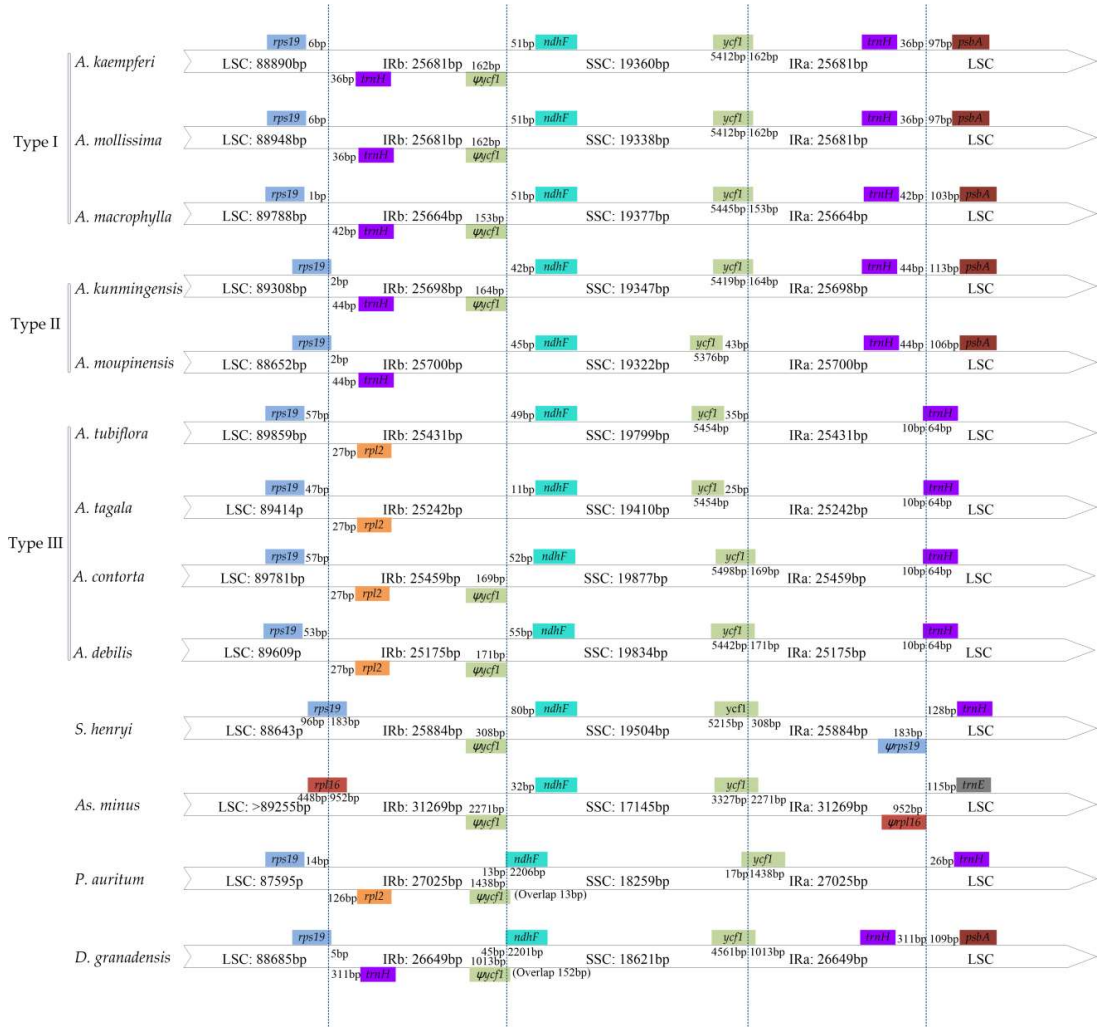


Figure 3. Comparison of the borders of large single copy (LSC), small single copy (SSC) and inverted repeat (IR) regions among 13 cp genomes. Number above the gene features means the distance between the ends of genes and the borders sites. These features are not to scale.

2.3. Codon Usage

All the protein-coding genes were composed of 26,194–26,398 codons in the cp genomes of the seven species of *Aristolochia*. The codon usages of protein-coding genes in the cp genomes are summarized in Figure 4 and Table S2. Among these codons, the most common amino acid in the protein-coding genes is leucine, which appears 2775 times in *A. kaempferi* and *A. mollissima*. The relative synonymous codon usage (RSCU) value analysis showed that almost all amino acids have more than one synonymous codon, except methionine and tryptophan. Nearly all of the protein-coding genes of *Aristolochia* species had the standard ATG start codon (RSCU = 1). About half of codons have RSCU > 1, and most of those (29/31, 93.5%) end with base A or T. About half of the codons have RSCU < 1, and most of those (28/31, 90.3%) end with base C or G.

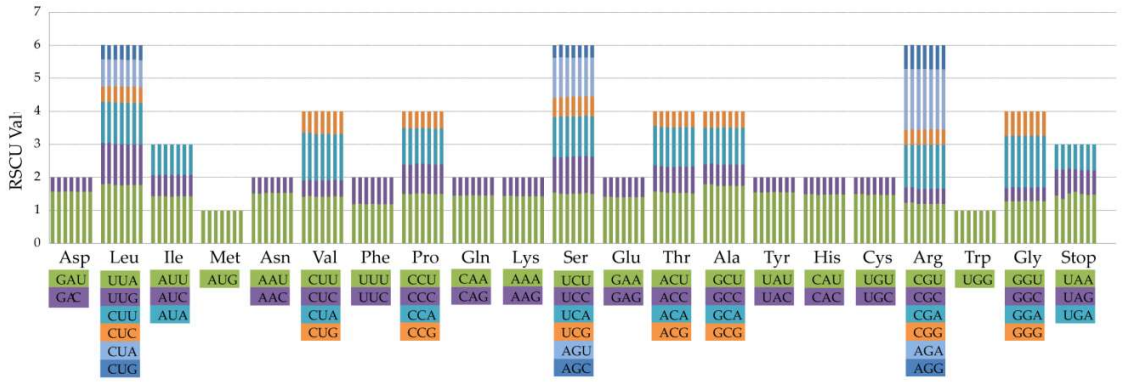


Figure 4. Codon content of 20 amino acid and stop codons in all protein-coding genes of the seven cp genomes. The histogram from the left-hand side of each amino acid shows codon usage within *Aristolochia* (From left to right: *A. tagala*, *A. tubiflora*, *A. moupinensis*, *A. kunmingensis*, *A. kaempferi*, *A. macrophylla*, and *A. mollissima*).

2.4. Positive Selection Analysis

We compared the ratio of non-synonymous (dN) and synonymous (dS) substitution for 79 protein-coding genes among seven species, including *A. kunmingensis*, *A. kaempferi*, *A. tagala*, *A. debilis*, *As. canadense*, *S. henryi*, and *P. auritum* within Piperales. The statistical neutrality test showed that five genes in the seven cp genomes are under significant positive selection, and these genes are involved in the synthesis of ribosomal small and large subunit protein (*rps12*, *rps18*, and *rpl20*) or unknown function (*ycf1* and *ycf2*) (Table 4). Likelihood ratio tests (M1a vs. M2a, M7 vs. M8) supported the presence of positively selected codon sites ($p < 0.05$) (Table S3). According to the M2a and M8 models, the *rpl20* harbored three or four sites under positive selection. The gene *ycf1* harbored one or three sites under positive selection based on two models, respectively. In addition, we identified *rps12* gene with one positive selection site.

Table 4. Positive selected sites detected in the cp genome of the Piperales.

Gene Name	M2a		M8	
	Selected Sites	Pr (w > 1)	Selected Sites	Pr (w > 1)
rpl20	71A	0.918	71A	0.967 *
	72L	0.999 **	72L	1.000 **
	105R	0.963 *	105R	0.984 *
	116H	0.963 *	116H	0.988 *
rps12	79M	0.966 *	79M	0.987 *
rps18	4S	0.937	4S	0.975 *
	99T	0.921	99T	0.967 *
ycf1	206S	0.914	206S	0.967 *
	211V	0.975 *	211V	0.989 *
	1412N	0.922	1412N	0.971 *
ycf2	2036W	0.932	2036W	0.950 *

* $p < 0.05$; ** $p < 0.01$.

2.5. Repeat Structure and Simple Sequence Repeats Analyses

Repeats in ten cp genomes were analyzed using REPuter, including seven species of *Aristolochia*, *S. henryi*, *P. auritum*, and *D. granadensis* (Figure 5, Table S4). The results showed that *A. macrophylla* had the greatest number of repetitive elements in cp genome, comprised of 25 forward, 26 palindromic, 21 reverse, and eight complement repeats. The size of the most repeats were 30–39 bp, and the repeats with the length > 49 bp only occurred in cp genomes of *S. henryi* and *P. auritum*. The longest repeats,

with a length of 1591 bp, was detected in *S. henryi*. The total numbers of SSRs were also identified in the cp genomes of the ten species (Figure 6 and Table S5). Mononucleotide repeats were the largest in a number of these SSRs, with 88% and 85% found in A. *tubiflora* and *A. tagala*, respectively. A/T repeats were the most common of mononucleotides, while AT/TA repeats are the majority of dinucleotide repeat sequences (96.3%–100%). The trinucleotide in the five species of subgenus *Siphisia* were only comprised of AAT/ATT repeats (100%), while *A. tubiflora* and *A. tagala* of subgenus *Aristolochia* also comprised AAC/GTT and AAG/CTT repeats.

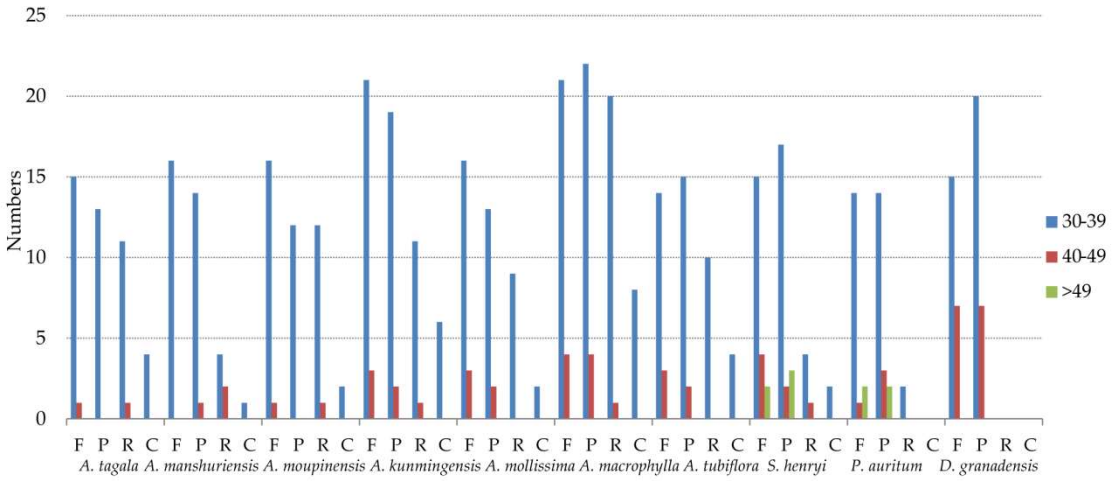


Figure 5. Repeat sequences in ten cp genomes. REPuter was used to identify repeat sequences with length $\geq$ 30 bp and sequence identity $\geq$ 90% in the cp genomes. F, P, R, and C indicate the repeat types F (forward), P (palindrome), R (reverse), and C (complement), respectively. Repeats with different lengths are indicated in different colors.

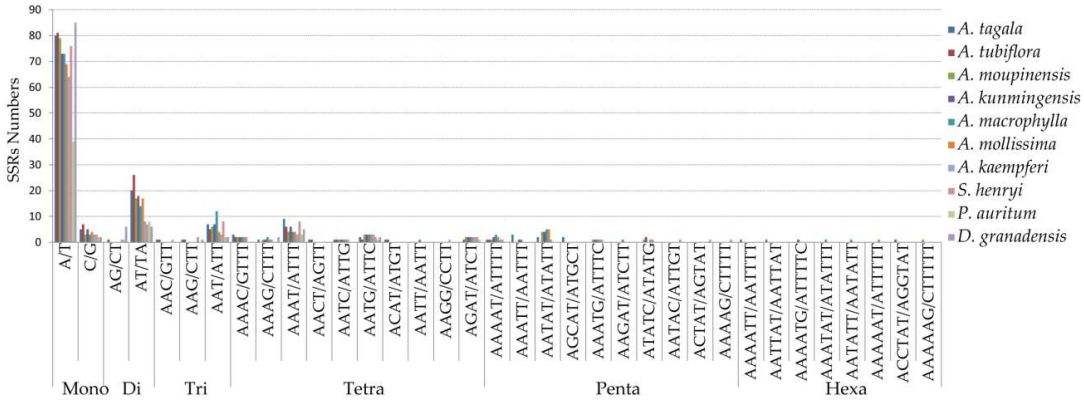


Figure 6. Frequency of simple sequence repeats (SSRs) in the ten cp genomes.

2.6. Comparative Genomic Divergence and Hotspots Regions

The SC and IR regions of cp genomes of the seven species (including *A. moupinensis*, *A. kunmingensis*, *A. tagala*, *A. contorta*, *S. henryi*, *As. canadense*, and *P. auritum*) were compared using the mVISTA program to detect hyper-variable regions (Figure 7). The alignment revealed high sequence conservatism across the cp genomes of *A. moupinensis* and *A. kunmingensis* of subgenus *Siphisia*. The comparison among seven cp genomes showed that the IR region was more conserved than the SC regions. The most divergent regions located in the intergenic spacers, and the most divergent coding regions were *ndhF* and *ycf1*.

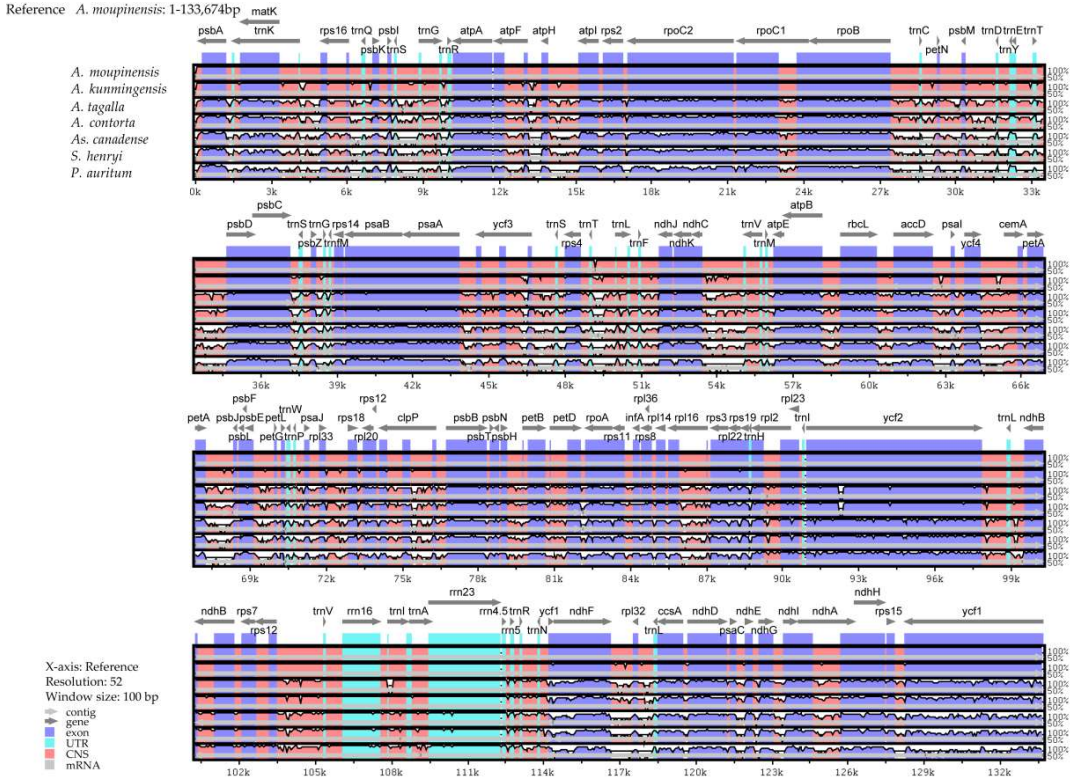


Figure 7. Sequence identity plot compared seven cp genomes with *A. moupinensis* as a reference by using mVISTA. Grey arrows and thick black lines above the alignment indicate genes with their orientation and the position of the IRs, respectively. A cut-off of 70% identity was used for the plots, and the *Y*-scale represents the percent identity from 50% to 100%.

Comparative analysis among our seven sequenced species within *Aristolochia* was conducted of the entire cp genomes, LSC, SSC, IR, and CDS regions, respectively (Table 5). The nucleotide diversity (Pi) value was also calculated to evaluate the sequence divergence among these cp genomes, and their values varied from 0 to 0.07746 (Figure 8). The analysis revealed that the SSC region, compared with other regions, exhibited the highest levels of divergence (Pi = 0.03114). These values of the LSC region, varied from 0.00175 to 0.07746, with the mean value of 0.02182. The IR region exhibited the lowest Pi values varying from 0 to 0.01056, with the mean of 0.00411, indicating that IR region was the most conserved one. Furthermore, we identified 16 hotspot regions (Pi > 0.04, the mean value = 0.05413) with the full length of 20,296 bp, including *rps16-trnQ-psbK*, *psbI-trnS-trnG*, *atpH-atpI*, *psbM-trnD*, *rps4-trnT-trnL*, *trnF-ndhJ*, *ndhC-trnV*, *accD-psaI*, *petA-psbJ*, *rps18-rpl20*, *trnN-ndhF*, *rpl32-trnL-ccsA*, and four regions of *ycf1* coding gene (Table 6). Ten of these (*rps16-trnQ-psbK*, *psbI-trnS-trnG*, *atpH-atpI*, *psbM-trnD*, *rps4-trnT-trnL*, *trnF-ndhJ*, *ndhC-trnV*, *accD-psaI*, *petA-psbJ*, and *rps18-rpl20*) are located in the LSC, and six (*trnN-ndhF*, *rpl32-trnL-ccsA* and *ycf1*) in the SSC region, which could be utilized as potential markers for the phylogeny reconstruction and species identification of this subgenus in further studies.

Table 5. Variable sites analyses in the seven *Aristolochia* cp genomes.

Regions	Number of Sites	Variable Sites	Parsimony Informative Sites	Nucleotide Diversity
LSC	94,564	4430	2315	0.02182
SSC	20,451	1433	804	0.03114
IR	25,884	253	154	0.00411
Complete	166,113	6422	3461	0.01717
CDS	79,365	2528	1376	0.01337

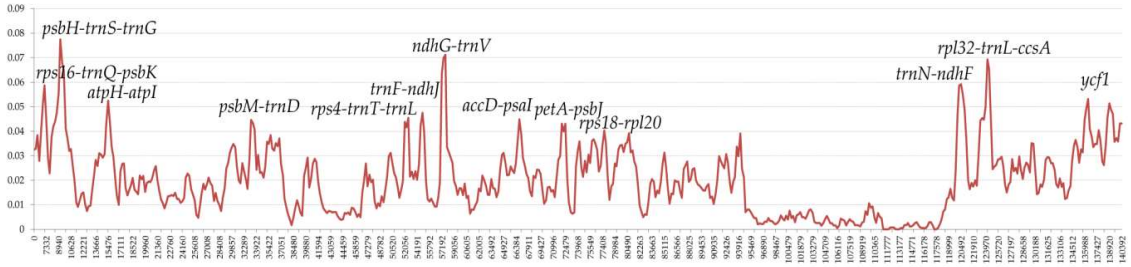


Figure 8. Sliding window analysis of the entire cp genome of seven *Aristolochia* species (window length: 600 bp; step size: 200 bp). *X*-axis: position of the midpoint of a window; *Y*-axis: nucleotide diversity of each window.

Table 6. Sixteen regions of highly variable sequences (Pi > 0.04) of *Aristolochia*.

High Variable Marker	Length	Variable Sites	Parsimony Informative Sites	Nucleotide Diversity
rps16-trnQ-psbK	1301	104	58	0.04278
psbI-trnS-trnG	2357	257	159	0.05364
atpH-atpI	1160	104	69	0.04439
psbM-trnD	1119	152	75	0.05888
rps4-trnT-trnL	1572	105	50	0.04178
trnF-ndhJ	920	85	51	0.04216
ndhC-trnV	1402	152	97	0.06311
accD-psaI	637	61	35	0.04220
petA-psbJ	1130	106	65	0.04444
rps18-rpl20	682	58	36	0.04155
trnN-ndhF	1492	161	86	0.04758
rpl32-trnL-ccsA	2679	202	113	0.04608
ycf1a	1225	126	74	0.04285
ycf1b	652	56	32	0.04053
ycf1c	1228	134	84	0.04611
ycf1d	740	70	39	0.04278
Combine	20296	2216	1349	0.05413

2.7. Phylogenetic Analyses

The phylogenetic relationships of Aristolochiaceae were constructed based on six datasets (entire cp genome sequences except a copy of IR, LSC, SSC, IR, and CDS regions and combining 16 hotspots) of 18 samples, using three methods of ML, MP, and BI, respectively (Figure 9). The robust topologies were consistent for most clades of cp genomes, LSC, SSC, CDS, and hotspots datasets, with the high bootstrap values for most of the branches (Figure 9A). From these six different datasets, the phylogenetic analysis showed that the genera *Asarum* and *Saruma* represented by seven species formed a clade with posterior probabilities (PP) = 1 based on BI, bootstrap values (%) (BS) =100 based on ML and BS =100 based on MP methods. However, the tree constructed using sequences of the IR region failed to resolve the phylogeny position of *Asarum epigynum* and *As. canadense* (Figure 9B), maybe due to inadequate information sites in the IR region. These nine species of *Aristolochia* species formed another strongly supported monophyletic group (PP = 1; [ML] BS = 100; [MP] BS = 100), and were divided into two subclades with strong support, corresponding to the taxonomic division of subgenus *Siphisia* (PP = 1; [ML] BS = 100; [MP] BS = 100) and subgenus *Aristolochia* (PP = 1; [ML] BS = 100; [MP] BS = 100). Within the subgenus *Siphisia*, the species *A. macrophylla* from North America was sister to the rest of four species from Asian region (PP = 1; [ML] BS = 100; [MP] BS = 100).

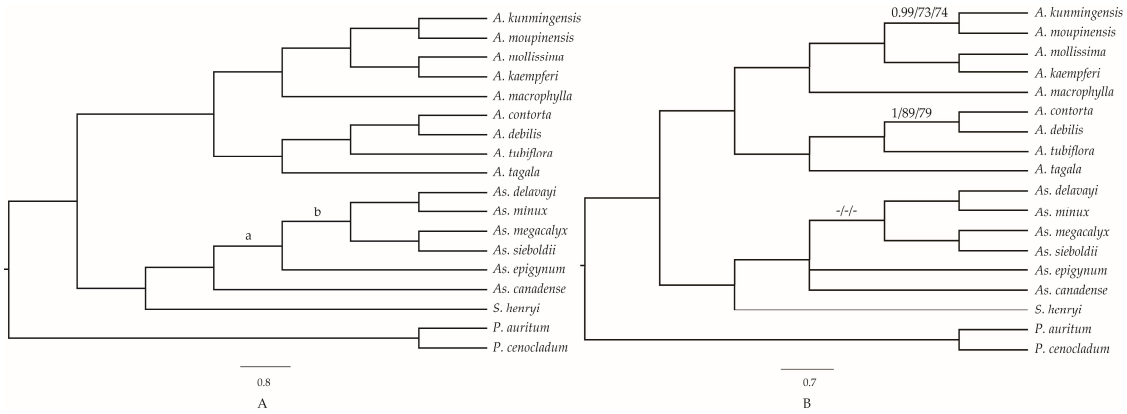


Figure 9. Phylogenetic relationships of the 18 species inferred from maximum parsimony (MP), maximum likelihood (ML), and Bayesian (BI) analyses. (**A**) The topology was constructed by cp genomes, LSC, SSC, CDS, and hotspots regions; (**B**) tree constructed by IR region. Bayesian posterior probability values < 0.95 or Bootstrap values < 90 were marked on the branches. The support values in node (a): 1/86/93 (using LSC region), 0.97/78/84 (SSC), 0.82/-/- (CDS), and 1/81/80 (hotspots); (b): 1/90/79 (SSC) and 1/73/71 (hotspots). Numbers above nodes are support values with Bayesian posterior probabilities values on the left, ML bootstrap values in the middle, and MP bootstrap values on the right. " - " indicates the value < 70.

3. Discussion

3.1. IR Contraction and Expansion

Taken another two reported species (*A. debilis* and *A. contorta*) of subgenus *Aristolochia* into account, although genomic structure and size were highly conserved, the IR-SC boundary regions were variable between these nine cp genomes of *Aristolochia* (Figure 3). In general, contraction and expansion at the borders of IR regions are common evolutionary events and may cause IR size variation of plastomes [29,34–36]. The length of the IR regions of five *Siphisia* species, varying in the range of 25,664–25,700 bp, was longer than those of the four species of subgenus *Aristolochia*, which varied from 25,175 bp to 25,459 bp (Table 1) [28]. We identified three types of the IR-SC junctions from the nine *Aristolochia* species, according to the organization of genes (Figure 3). Within five detected species of subgenus *Siphisia*, its patterns were Type I and II, while the Type III only occurred in the four species of subgenus *Aristolochia*. Type I was found in *A. mollissima*, *A. macrophylla*, and *A. kaempferi*, and was characterized by *trnH* gene in IR region and LSC-IRb border located in *rps19-trnH* spacer. Type II was only found in *A. moupinensis* and *A. kunmingensis* and refers to LSC-IRb border within the *rps19* gene. The *trnH* gene is intact and located upstream of *rpl2* in IRb region for type I and II. Type III pattern was found in the four species of subgenus *Aristolochia*, characterized by LSC-IRb and SSC-IRa border in the *rps19-rpl2* spacer and *trnH* gene, respectively. The *trnH* gene spanned the junction between IR-LSC regions in the four species of subgenus *Aristolochia*.

The shift of IR-LSC borders, caused by contraction and expansion of the gene *trnH*, is one of major differences between the plastomes of the subgenera *Siphisia* and *Aristolochia*. The whole gene duplication of *trnH* was detected in most monocots (e.g., *Acorus*, *Phalaenopsi* and *Dioscorea*), *D. granadensis* (Winteraceae) of magnoliids, and basal eudicots (*Ranunculus japonica* and *Ranunculus macranthus*) [34,37–41]. Wang et al. (2008) conducted RT-PCR assays and deduced that the duplicated *trnH* genes in most of non-monocots and monocots were regulated by different expression levels of promoters, and had distinct fates [37]. Within the family Aristolochiaceae, the *trnH* gene was located in the LSC region of *S. henryi*, 128 bp away from the border of LSC-IR, and was also a single copy in the six cp genomes of *Asarum*, but not sure the positions of the gene [29]. Furthermore, the study proposed that the low-complexity *trnH* region and ultimately inversion of a portion of the LSC were due to an AAT repeat. For inversion of a large portion of the LSC region, there were genes rearranged

in SC-IR borders of sequenced species of *Asarum*, the IR boundaries of cp genomes of *Asarum* were highly variable and experienced positional shifts at borders. Such as there was an entirety of the SSC of *As. canadense* and *As. sieboldii* has been incorporated into the IR, and the boundary of the LSC-IR was found within *rpl2* or *rpl14* gene [29]. Within the species of *S. henryi*, *rps19* pseudogene existed in the IRa region, with the length of 183 bp. The *trnH-rps19* gene cluster had been used to distinguish monocots from other angiosperm for the organization of gene flanking the IR-SC junction [37,39]. The events of contraction or expansion of the IR regions also can be used to distinguish the species within Aristolochiaceae.

3.2. Inferring the Phylogeny and Species Identification of Aristolochia

Chloroplast genomes provide abundant resources significant for evolutionary, taxonomic, and phylogenetic studies [42–44]. The whole cp genomes and protein-coding genes have been successfully used to resolve phylogenetic relationships at multiple taxonomic levels during the past decade [45,46]. Repeats can lead to changes in genomic structure, and can be investigated to population genetics of allied taxa [47–50]. Repeats in ten cp genomes revealed that the repeats had a great number, comprised of 38–80 repeats (Figure 5 and Table S4), 66 and 138 repeats were respectively detected in *A. debilis* and *A. contorta* [28]. Given the variability of these repeats between lineages, they can be informative regions for developing genomic markers for phylogenetic analysis. SSRs, known as microsatellites, are tandemly repeated DNA sequences that consist of one–six nucleotide repeat units and are ubiquitous throughout the genomes [51]. A total number of 95–142 SSRs were identified in the seven cp genomes detected (Figure 6 and Table S5). According to the analysis of high variable regions, the hotspot regions within seven cp genomes also provide sufficient information sites to reveal phylogeny structure among species of family Aristolochiaceae, especially for the spacer *ycf1* and *rpl20*, with high nucleotide diversity and under positive selection (Table 4). The *ycf1* gene could be served as the barcode of land plants, and was also recognized as the most variable regions in plastid genome [50,52]. The gene *rpl20* is an important part of protein synthesis, and is involved in translation [53]. This study will also provide a reference for phylogenomic studies of closely related lineages among *Aristolochia* and other genera.

Furthermore, we can design effective markers for clarifying the phylogenetic relationships of *Aristolochia* and elucidating the evolutionary history of species complex of *Aristolochia* at the population level, based on the analysis of SSR and SNP sites. Understanding genetic variation within and between populations plays an important role in improving genetic diversity and is essential for future adaptive changes, reproduction patterns, and its conservation [20,54,55]. The cpDNA and B-class gene PISTILLATA (*PI*) have been used to investigate taxonomy at the species complex, such as *Aristolochia kaempferi* group, and these studies revealed that its DNA barcoding and taxonomy are difficult to assess for multiple hybridization and introgression events in the group [56,57]. More genes under selection and neutral markers should be used to clarify those multiple diversification events. It will better to apply the full genome information or hyper-variation regions to elucidate the species diversity of *Aristolochia*.

4. Materials and Methods

4.1. Plant Material, DNA Extraction, and Sequencing

We selected seven species according to their potential medicinal uses, including *A. kaempferi*, *A. kunmingensis*, *A. macrophylla*, *A. mollissima*, and *A. moupinensis* from subgenus *Siphisia* and *A. tagala* and *A. tubiflora* of subgenus *Aristolochia* (Table 7). Genomic DNA was isolated from silica-gel dried leaf tissue or herbarium specimens using Plant Genomic DNA Kit (TIANGEN, Beijing, China). DNA integrity was examined by electrophoresis in 1% (w/v) agarose gel and their concentration was measured using a NanoDrop spectrophotometer 2000 (Thermo Scientific; Waltham, MA, USA). The

DNA was used to construct PE libraries with insert sizes of 150 bp and sequenced according to the manufacturer's manual for the Illumina Hiseq X.

Table 7. Sampled species and their voucher specimens used in this study.

Species	Samples	Voucher	locality
A. kaempferi	E2265	Yuan Wang	Japan, Tokyo
A. kunmingensis	E754	Zhanghua Wang	China, Yunnan
A. macrophylla	E2111	Jinshuang Ma	North America, North Carolia
A. mollissima	E1016	Xinxin Zhu & Zhixiang Hua	China, Guangdong
A. moupinensis	E1086	Xinxin Zhu & Zhixiang Hua	China, Sichuan
A. tagala	E1071	Yuan Wang	China, Hongkong
A. tubiflora	E2239	Shuwan Li	China, Guangxi

4.2. Chloroplast Genome Assembly and Annotation

We used the software Trimmomatic version 0.36 (Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany) [58] to trim the low-quality reads. We retrieved the plastome sequence of *A. contorta* (NC_036152.1), *A. debilis* (NC_036153.1), *Asarum costatum* (AP018513.1), *Asarum minamitanianum* (AP018514.1), and *Asarum sakawanum* (AP017908.1) from GenBank and used these sequences as the references [28,30,31]. The plastome was assembled using mapping to reference genome and de novo methods as implemented in Geneious R11 (Biomatters, Auckland, New Zealand) [59].

The cp genomes of the seven species was annotated using the online program Dual Organellar GenoMe Annotator (DOGMA) (University of Texas at Austin, Austin, TX, USA) [60], Annotation of Organellar Genomes (GeSeq) [61] and Chloroplast Genome Annotation, Visualization, Analysis, and GenBank Submission (CPGAVAS) (Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China) [62]. The tRNA genes were confirmed using tRNAscan-SE software (v2.0, University of California, Santa Cruz, CA, USA) [63]. Plastome annotations were manually corrected with the software Artemis [64]. The gene map was drawn using the Organelle Genome DRAW (OGDRAW) [65,66] with default settings and checked manually. The complete cp genome sequences of the seven species were deposited in GenBank, accession numbers are MK503927-MG503933 (Table S6).

4.3. Genome Structure Analyses

The distribution of codon usage was investigated using the software CodonW (University of Texas, Houston, TX, USA) with the RSCU value [67]. GC content was analyzed using Molecular Evolutionary Genetics Analysis (MEGA v6.0, Tokyo Metropolitan University, Tokyo, Japan) [68]. REPuter program (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (University of Bielefeld, Bielefeld, Germany) [69] was used to identify the size and location of repeat sequences, including forward, palindromic, reverse, and complement repeats in the seven cp genomes. For all repeat types, the minimal size was set as 30 bp and the two repeat copies had at least 90% similarity. Perl script MISA (https://webblast.ipk-gatersleben.de/misa/) [70] was used to detect microsatellites (mono-, di-, tri-, tetra-, penta-, hexanucleotide repeats) with the following thresholds (unit size, min repeats): ten repeat units for mononucleotide SSRs, five repeat units for dinucleotide SSRs, four repeat units for trinucleotide SSRs, and three repeat units each for tetra-, penta-, and hexanucleotide SSRs.

4.4. Positive Selection Analysis

To identify the genes under selection, we scanned the cp genomes of seven species within Piperales using codeml of the package PAML4 [71,72]. The software was used for calculating the non-synonymous (dN) and synonymous (dS) substitution rates, along with their ratios (ω = dN/dS). The analyses of selective pressures were conducted along the ML tree in Newick format (S7), which

based on the whole CDS region was used to determine the phylogenetic relationships of these seven species. Each single-copy CDS sequences was aligned according to their amino acid sequence. We used the site-specific model with five site models (M0, M1a & M2a, M7 & M8) were employed to identify the signatures of adaptation across cp genomes. This model allowed the ω ratio to vary among sites, with a fixed ω ratio in all the branches. Comparing the site-specific model, M1a (nearly neutral) vs. M2a (positive selection) and M7 (β) vs. M8 (β & ω) were calculated in order to detect positive selection [73]. Likelihood ratio test (LRT) of the comparison (M1a vs. M2a and M7 vs. M8) was used respectively to evaluate of the selection strength and the p value of Chi square (χ^2) smaller than 0.05 is thought as significant. The Bayes Empirical Bayes (BEB) inference [74] was implemented in site models M2a and M8 to estimate the posterior probabilities and positive selection pressures of the selected genes.

4.5. Genome Comparison and Nucleotide Variation Analysis

The whole-genome (minus a copy of IR region) alignment for the cp genomes of the seven species including our *A. moupinensis*, *A. kunmingensis*, A. *tubiflora* and four reported species (*A. contorta*, *As. canadense*, *S. henryi* and *P. cenocladum*) of Piperales, was performed and plotted by the mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml) in Shuffle-LAGAN model [75,76], and with *A. moupinensis* as the reference. The seven cp genomes of *Aristolochia* were first aligned using MAFFT v7 [77] and then manually adjusted using BioEdit v7.0.9 [78]. Variable sites and nucleotide variability across complete cp genomes, LSC, IR, SSC, and CDS regions of seven species were calculated using DnaSP v5 [79]. Furthermore, for the seven cp genomes minus a copy IR region, a sliding window analysis was conducted to evaluate the nucleotide variability using DnaSP software. The step size was set to 200 base pairs, and the window length was set to 600 base pairs.

4.6. Phylogenetic Analyses

To estimate phylogenetic relationships within the Aristolochiaceae, plastomes of 18 taxa were compared, including nine samples from *Aristolochia*, six and one cp genomes from *Asarum* and *Saruma*, respectively (Table S5). A total of 11 cp genomes were downloaded from the NCBI database. In the phylogenetic analyses, *P. auritum* and *P. cenocladum* of *Piper* were used as outgroup. Phylogenetic trees were constructed by MP, ML and BI methods using the cp genomes, LSC, SSC, IR, CDS and hotspots regions. The sequences of the involved regions were aligned using MAFFT v7. MP analysis was performed with PAUP*4.0b10 [80], using a heuristic search performed 1000 replications and tree bisection-reconnection (TBR) branch swapping. BI was conducted using the program MrBayes v3.2 [81] with the GTR+I+G model at the CIPRES Science Gateway website (http://www.phylo.org/) [82]. The Markov Chain Monte Carlo (MCMC) analysis was run for 2,000,000 generations, sampling every 1000 generations. The posterior probabilities (PP) of the phylogeny and its branches were determined from the combined set of trees, discarding the first 25% trees of each run as burn-in, as determined by Tracer v1.7 [83]. Maximum likelihood phylogenies were constructed by a fast and effective stochastic algorithm using IQ-TREE v1.6.2 [84] with the Best-fit model by ModelFinder [85] according to Bayesian information criterion (BIC) and the robustness of the topology was estimated using 2000 bootstrap replicates. Figtree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) [86] was used to visualize and annotate trees.

5. Conclusions

The complete cp genomes of *A. kaempferi*, *A. kunmingensis*, *A. macrophylla*, *A. mollissima*, and *A. moupinensis* of the subgenus *Siphisia*, and *A. tagala* and *A. tubiflora* of the subgenus *Aristolochia* were reported in this study. The cp genomes length and gene content of the genus *Aristolochia* were comparatively conserved. Although genomic structure and size were highly conserved, the IR-SC boundary regions were variable between these nine cp genomes of *Aristolochia*. The whole duplicated *trnH* gene within five species of *Siphisia* is one of major differences between the plastomes of the subgenera *Siphisia* and *Aristolochia*. We also identified SSR sites, five positive selection sites and 16

variable regions, which provide a reference for developing tools to further study *Aristolochia* species. Furthermore, the phylogenetic constructions with six datasets of 18 cp genomes illustrated robust and consistent relationships with high supports.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/xxx/s1.

Author Contributions: X.L. performed the experiments, analyzed the data, and wrote the manuscript; Y.Z. assembled sequences and revised the manuscript; X.Z. and S.L. collected, identified plant materials and gave suggestions to the manuscript; J.M. revised the manuscript. All authors have read and approved the final manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 31370225).

Acknowledgments: The authors give special thanks to Shuwan Li, Yuan Wang, Zhanghua Wang, and Zhixiang Hua for collecting plant material. We acknowledged someone for their assistance with fieldwork, for data analysis, for giving comments on the manuscript paper. Our sincere thanks are also to the anonymous reviewers for their comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

RSCU	Relative synonymous codon usage
NGS	Next-generation sequencing
TLA	Three letter acronym
SSR	Simple sequence repeats
ATP	Adenosine triphosphate
MP	Maximum parsimony
ML	Maximum likelihood
Pi	Nucleotide diversity
BI	Bayesian Inference
AAs	Aristolochic acids
LSC	Large single copy
SSC	Small single copy
CDS	Coding sequence
IR	Inverted repeat
SC	Single copy
CP	Chloroplast

References

1. Neinhuis, C.; Wanke, S.; Hilu, K.W.; Müller, K.; Borsch, T. Phylogeny of Aristolochiaceae based on parsimony, likelihood, and Bayesian analyses of *trnL-trnF* sequences. *Plant Syst. Evol.* **2005**, *250*, 7–26. [CrossRef]
2. Wagner, S.T.; Isnard, S.; Rowe, N.P.; Samain, M.S.; Neinhuis, C.; Wanke, S. Escaping the lianoid habit: Evolution of shrub-like growth forms in *Aristolochia* subgenus *Isotrema* (Aristolochiaceae). *Am. J. Bot.* **2012**, *99*, 1609–1629. [CrossRef] [PubMed]
3. Feng, Y.X.; Lin, S.Q.; Zhang, X.Q. Botanical and pharmacognostical studies of Chinese *Aristolochia*: Resource utilization. *Acta Pharm. Sin.* **1983**, *18*, 291–298.
4. Ma, J.S. Medicinal resources of *Aristolochia* in China. *Chin. Wild Plant Resour.* **1988**, *3*, 31–34.
5. Chen, C.H.; Dickman, K.G.; Moriya, M.; Zavadil, J.; Sidorenko, V.S.; Edwards, K.L.; Gnatenkod, D.V.; Wu, L.; Turesky, R.J.; Wu, X.R.; et al. Aristolochic acid-associated urothelial cancer in Taiwan. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8241–8246. [CrossRef] [PubMed]
6. Wu, L.; Sun, W.; Wang, B.; Zhao, H.; Li, Y.; Cai, S.; Li, X.; Zhu, Y.J.; Yao, H.; Song, J.Y.; et al. An integrated system for identifying the hidden assassins in traditional medicines containing aristolochic acids. *Sci. Rep.* **2015**, *5*, 11318. [CrossRef] [PubMed]
7. Tsai, D.M.; Kang, J.J.; Lee, S.S.; Wang, S.Y.; Tsai, I.; Chen, G.Y.; Liao, H.W.; Chu, L.W.; Kuo, C.H.; Tseng, Y.J. Metabolomic analysis of complex Chinese remedies: Examples of induced nephrotoxicity in the mouse from a series of remedies containing aristolochic acid. *Evid.-Based Complement. Altern. Med.* **2013**, *2013*, 263757. [CrossRef] [PubMed]

8. Ng, A.W.T.; Poon, S.L.; Huang, M.N.; Lim, J.Q.; Boot, A.; Yu, W.; Suzuki, Y.; Thangaraju, S.; Ng, C.C.Y.; Tan, P.; et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci. Transl. Med.* **2017**, *9*, eaan6446. [CrossRef] [PubMed]
9. Martena, M.J.; Van der Wielen, J.C.A.; Van de Laak, L.F.J.; Konings, E.J.M.; de Groot, H.N.; Rietjens, I.M.C.M. Enforcement of the ban on aristolochic acids in chinese traditional herbal preparations on the dutch market. *Anal. Bioanal. Chem.* **2007**, *389*, 263–275. [CrossRef] [PubMed]
10. Lawrence, M.K.; González, F.A. Phylogenetic Relationships in Aristolochiaceae. *Syst. Bot.* **2003**, *28*, 236–249. [CrossRef]
11. Wanke, S.; Jaramillo, M.A.; Borsch, T.; Samain, M.S.; Quandt, D.; Neinhuis, C. Evolution of Piperales-*matK* gene and *trnK* intron sequence data reveal lineage specific resolution contrast. *Mol. Phylogenet. Evol.* **2007**, *42*, 477–497. [CrossRef] [PubMed]
12. Lawrence, M.K. A cladistic analysis of *Asarum* and implications for the evolution of Herkogamy. *Am. J. Bot.* **1997**, *84*, 1752–1765.
13. Lawrence, M.K. Phylogenetic relationships in *Asarum* (Aristolochiaceae) based on morphology and *ITS* sequences. *Am. J. Bot.* **1998**, *85*, 1454–1467.
14. González, F.A.; Stevenson, D.W. A phylogenetic analysis of the subfamily Aristolochioideae (Aristolochiaceae). *Rev. Acad. Colomb. Cienc.* **2002**, *66*, 59–65.
15. Ohi-Toma, T.; Sugawara, T.; Neihuis, C.; Murata, H.; Wanke, S.; Murata, J. Molecular phylogeny of *Aristolochia* sensu lato(Aristolochiaceae) based on sequences of *rbcL*, *matK*, and *phyA* genes, with special reference to differentiation of chromosome numbers. *Syst. Bot.* **2006**, *31*, 481–492. [CrossRef]
16. Wanke, S.; Samain, M.S.; Vanderschaeve, L.; Mathieu, G.; Goetghebeur, P.; Neinhuis, C. Phylogeny of the genus Peperomia (Piperaceae) inferred from the trnK/matK region (cpDNA). *Plant Biol.* **2006**, *8*, 93–102. [CrossRef] [PubMed]
17. González, F.; Wagner, S.T.; Salomo, K.; Symmank, L.; Samain, M.-S.; Isnard, S.; Rowe, N.P.; Neinhuis, C.; Wanke, S.; Carine, M. Present trans-Pacific disjunct distribution of *Aristolochia* subgenus *Isotrema* (Aristolochiaceae) was shaped by dispersal, vicariance and extinction. *J. Biogeogr.* **2014**, *41*, 380–391. [CrossRef]
18. Dechbumroong, P.; Aumnouypol, S.; Denduangboripant, J.; Sukrong, S. DNA barcoding of *Aristolochia* plants and development of species-specific multiplex PCR to aid HPTLC in ascertainment of *Aristolochia* herbal materials. *PLoS ONE* **2018**, *13*, e0202625. [CrossRef] [PubMed]
19. Sarma, B.; Tanti, B. Analysis of genetic diversity of certain species of *Aristolochia* using ISSR-based molecular markers. *Curr. Life Sci.* **2017**, *3*, 47–53. [CrossRef]
20. Yang, Z.Y.; Yi, T.S.; Zeng, L.Q.; Gong, X. The population genetic structure and diversification of *Aristolochia delavayi* (Aristolochiaceae), an endangered species of the dry hot valleys of the Jinsha River, southwestern China. *Botany* **2014**, *92*, 579–587. [CrossRef]
21. Douglas, S.E. Plastid evolution: Origins, diversity, trends. *Curr. Opin. Genet. Dev.* **1998**, *8*, 655–661. [CrossRef]
22. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Am. Sci. USA* **2007**, *104*, 19363–19368. [CrossRef] [PubMed]
23. Yang, J.B.; Yang, S.X.; Li, H.T.; Jing, Y.; Li, D.Z. Comparative chloroplast genomes of *Camellia* species. *PLoS ONE* **2013**, *8*, e73053. [CrossRef] [PubMed]
24. Huang, H.; Shi, C.; Liu, Y.; Mao, S.; Gao, L. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Boil.* **2014**, *14*, 151. [CrossRef] [PubMed]
25. Lei, W.; Ni, D.; Wang, Y.; Shao, J.; Wang, X.; Yang, D.; Wang, J.S.; Chen, H.M.; Liu, C. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus Membranaceus*. *Sci. Rep.* **2016**, *6*, 21669. [CrossRef] [PubMed]
26. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; Depamphilis, C.W.; Chumley, T.W.; Haberle, R.C. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **2005**, *395*, 348–384. [PubMed]
27. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [CrossRef] [PubMed]

28. Zhou, J.G.; Chen, X.L.; Cui, Y.X.; Sun, W.; Li, Y.H.; Wang, Y.; Song, J.Y.; Yao, H. Molecular Structure and Phylogenetic Analyses of Complete Chloroplast Genomes of Two *Aristolochia* Medicinal Species. *Int. J. Mol. Sci.* **2017**, *18*, 1839. [CrossRef] [PubMed]
29. Sinn, B.T.; Sedmak, D.D.; Kelly, L.M.; Freudenstein, J.V. Total duplication of the small single copy region in the angiosperm plastome: Rearrangement and inverted repeat instability in *Asarum*. *Am. J. Bot.* **2018**, *105*, 71–84. [CrossRef] [PubMed]
30. Takahashi, D.; Sakaguchi, S.; Isagi, Y.; Setoguchi, H. Comparative chloroplast genomics of series Sakawanum in genus *Asarum* (Aristolochiaceae) to develop single nucleotide polymorphisms (SNPs) and simple sequence repeat (SSR) markers. *J. For. Res.* **2018**, *23*, 387–392. [CrossRef]
31. Sakaguchi, S.; Ueno, S.; Tsumura, Y. Application of a simplified method of chloroplast enrichment to small amounts of tissue for chloroplast genome sequencing. *Appl. Plant Sci.* **2017**, *5*, 1700002. [CrossRef] [PubMed]
32. NCBI. Genome. Available online: https://www.ncbi.nlm.nih.gov/ (accessed on 14 January 2019).
33. Xu, J.W.; Feng, D.J.; Song, G.S.; Wei, X.L.; Chen, L.; Wu, X.L.; Li, X.G.; Zhu, Z. The first intron of rice epsp synthase enhances expression of foreign gene. *Sci. China Life Sci.* **2003**, *46*, 561–569. [CrossRef] [PubMed]
34. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus Macranthus*. *BMC Genom.* **2007**, *8*, 174. [CrossRef] [PubMed]
35. Yang, M.; Zhang, X.; Liu, G.; Yin, Y.; Chen, K.; Yun, Q.; Zhao, D.J.; Al-Mssallem, I.S.; Yu, J. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* **2010**, *5*, e12762. [CrossRef] [PubMed]
36. Kim, K.J.; Lee, H.L. Complete chloroplast genome sequences from Korean Ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [CrossRef] [PubMed]
37. Wang, R.J.; Cheng, C.L.; Chang, C.C.; Wu, C.L.; Su, T.M.; Chaw, S.M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Boil.* **2008**, *8*, 36. [CrossRef] [PubMed]
38. Huotari, T.; Korpelainen, H. Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene* **2012**, *508*, 96–105. [CrossRef] [PubMed]
39. Luo, Y.; Ma, P.F.; Li, H.T.; Yang, J.B.; Wang, H.; Li, D.Z. Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. *Genome Biol. Evol.* **2016**, *8*, 932–945. [CrossRef] [PubMed]
40. Cai, Z.Q.; Penaflor, C.; Kuehl, J.V.; Leebens-Mack, J.; Carlson, J.E.; Pamphilis, C.W.D.; Boore, J.L.; Jansen, R.K. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: Implications for the phylogenetic relationships of magnoliids. *BMC Evol. Boil.* **2006**, *6*, 77. [CrossRef]
41. Chang, C.C.; Lin, H.C.; Lin, I.; Chow, T.Y.; Chen, H.H.; Chen, W.H.; Cheng, C.H.; Lin, C.Y.; Liu, S.M.; Chang, C.C.; et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* **2006**, *23*, 279–291. [CrossRef] [PubMed]
42. Borsch, T.; Quandt, D. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Syst. Evol.* **2009**, *282*, 169–199. [CrossRef]
43. Dong, W.P.; Liu, J.; Yu, J.; Wang, L.; Zhou, S.L. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [CrossRef] [PubMed]
44. Tong, W.; Kim, T.S.; Park, Y.J. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. *Rice* **2016**, *9*, 57. [CrossRef] [PubMed]
45. Dong, W.P.; Liu, H.; Xu, C.; Zuo, Y.J.; Chen, Z.J.; Zhou, S.L. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 138. [CrossRef] [PubMed]
46. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Chen, X.Q.; Xue, J.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 5751. [CrossRef] [PubMed]
47. Timme, R.E.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **2007**, *94*, 302–312. [CrossRef] [PubMed]

48. Nie, X.J.; Lv, S.Z.; Zhang, Y.X.; Du, X.H.; Wang, L.; Biradar, S.S.; Tan, X.F.; Wan, F.H.; Song, W.N. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* **2012**, *7*, e36869. [CrossRef] [PubMed]
49. Huang, J.; Chen, R.H.; Li, X.G. Comparative analysis of the complete chloroplast genome of four known *Ziziphus* species. *Genes* **2017**, *8*, 340. [CrossRef] [PubMed]
50. Gu, C.H.; Tembrock, L.R.; Zheng, S.Y.; Wu, Z.Q. The complete chloroplast genome of *Catha edulis*: A comparative analysis of genome features with related species. *Int. J. Mol. Sci.* **2018**, *19*, 525. [CrossRef] [PubMed]
51. Powell, W.; Morgante, M.; Mcdevitt, R.; Vendramin, G.G.; Rafalski, J.A. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7759–7763. [CrossRef] [PubMed]
52. Dong, W.P.; Xu, C.; Li, C.H.; Sun, J.H.; Zuo, Y.J.; Shi, S.; Cheng, T.; Guo, J.J.; Zhou, S.L. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [CrossRef] [PubMed]
53. Muto, A.; Ushida, C. Transcription and translation. *Methods Cell Biol.* **1995**, *48*, 483.
54. Koren, O.G.; Nakonechnaya, O.V.; Zhuravlev, Y.N. Genetic structure of natural populations of the relict species *Aristolochia manshuriensis* (Aristolochiaceae) in disturbed and intact habitats. *Russ. J. Genet.* **2009**, *45*, 678–684. [CrossRef]
55. Nakonechnaya, O.V.; Kholina, A.B.; Koren, O.G.; Zhuravlev, Y.N. Genetic diversity of a rare species *Aristolochia contorta* Bunge (Aristolochiaceae) in Primorsky krai. *Russ. J. Genet.* **2012**, *48*, 152–162. [CrossRef]
56. Watanabe, K.; Kajita, T.; Murata, J. Chloroplast DNA and variation and geographical structure of the *Aristolochia kaempferi* group (Aristolochiaceae). *Am. J. Bot.* **2006**, *93*, 442–453. [CrossRef] [PubMed]
57. Watanabe, K.; Ohi-Toma, T.; Murata, J. Multiple hybridization in the *Aristolochia kaempferi* group (Aristolochiaceae): Evidence from reproductive isolation and molecular phylogeny. *Am. J. Bot.* **2008**, *95*, 885–896. [CrossRef] [PubMed]
58. Bolger, A.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
59. Drummond, A.J.; Ashton, B.; Buxton, S.; Cheung, M.; Cooper, A.; Duran, C.; Field, M. *Geneious*; Version 5.3+; Biomatters: Auckland, New Zealand, 2011.
60. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organelle genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [CrossRef] [PubMed]
61. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbrichtjones, E.S.; Fischer, A.; Bock, R.; Greiner, S. Geseq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef] [PubMed]
62. Liu, C.; Shi, L.C.; Zhu, Y.; Chen, H.M.; Zhang, J.H.; Lin, X.H.; Guan, X.J. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom.* **2012**, *13*, 715. [CrossRef] [PubMed]
63. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [CrossRef] [PubMed]
64. Carver, T.; Harris, S.R.; Berriman, M.; Parkhill, J.; Mcquillan, J.A. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **2012**, *28*, 464–469. [CrossRef] [PubMed]
65. Lohse, M.; Drechsel, O.; Bock, R. Organellargenomedraw (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [CrossRef] [PubMed]
66. Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, W575–W581. [CrossRef] [PubMed]
67. Sharp, P.M.; Li, W.H. The codon adaptation index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [CrossRef] [PubMed]
68. Koichiro, T.; Glen, S.; Daniel, P.; Alan, F.; Sudhir, K. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729.
69. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. Reputer: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef] [PubMed]

70. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [CrossRef] [PubMed]
71. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556. [CrossRef]
72. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef] [PubMed]
73. Yang, Z.; Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [CrossRef] [PubMed]
74. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [CrossRef] [PubMed]
75. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *19*, i54–i62. [CrossRef] [PubMed]
76. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. Vista: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, w273–w279. [CrossRef] [PubMed]
77. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
78. Hall, T.A. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
79. Rozas, J.; Sánchez-DelBarrio, J.C.; Messeguer, X.; Rozas, R. Dnasp, dna polymorphism analyses by the coalescent and other methods. *Bioinformatics* **2003**, *19*, 2496–2497. [CrossRef] [PubMed]
80. Swofford, D.L. *PAUP*: Phylogenetic Analysis Using Parsimony, Distance, and Likelihood*; Version 4; Sinauer: Sunderland, MA, USA, 2002.
81. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef] [PubMed]
82. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *GCE* **2010**, *14*, 1–8.
83. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *10*, 901–904. [CrossRef] [PubMed]
84. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
85. Kalyaanamoorthy, S.; Bui, M.Q.; Wong, T.K.F.; Von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
86. Figtree v1.4. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 14 January 2019).

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Gene Expression Profiling Reveals Enhanced Defense Responses in an Invasive Weed Compared to Its Native Congener During Pathogenesis

Bharani Manoharan [1,2], Shan-Shan Qi [1,2,3], Vignesh Dhandapani [4], Qi Chen [2], Susan Rutherford [5], Justin SH Wan [3,5], Sridharan Jegadeesan [6], Hong-Yu Yang [2], Qin Li [2], Jian Li [1,2], Zhi-Cong Dai [1,2,3,7,*] and Dao-Lin Du [1,2,*]

1 Institute of Environment and Ecology, Academy of Environmental Health and Ecological Security, Jiangsu University, Zhenjiang 212013, China; biobharani@gmail.com (B.M.); qishanshan1986120@163.com (S.-S.Q.); lj060404002@126.com (J.L.)
2 School of the Environment and Safety Engineering, Jiangsu University, Zhenjiang 212013, China; chen1455526947@163.com (Q.C.); younghy524@163.com (H.-Y.Y.); liyunduo1024@163.com (Q.L.)
3 School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia; wanjustinsh@gmail.com
4 Environmental Genomics Group, School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK; v.dhandapani@bham.ac.uk
5 The Royal Botanic Gardens and Domain Trust, Mrs Macquaries Road, Sydney, NSW 2000, Australia; susan.rutherford@rbgsyd.nsw.gov.au
6 The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 761001, Israel; jsri.86@gmail.com
7 Institute of Agricultural Engineering, Jiangsu University, Xuefu Road 301, Zhenjiang 212013, China.
* Correspondence: daizhicong@163.com (Z.-C.D.); ddl@ujs.edu.cn (D.-L.D.)

Received: 23 August 2019; Accepted: 1 October 2019; Published: 3 October 2019

Abstract: Invasive plants are a huge burden on the environment, and modify local ecosystems by affecting the indigenous biodiversity. Invasive plants are generally less affected by pathogens, although the underlying molecular mechanisms responsible for their enhanced resistance are unknown. We investigated expression profiles of three defense hormones (salicylic acid, jasmonic acid, and ethylene) and their associated genes in the invasive weed, *Alternanthera philoxeroides*, and its native congener, *A. sessilis*, after inoculation with *Rhizoctonia solani*. Pathogenicity tests showed significantly slower disease progression in *A. philoxeroides* compared to *A. sessilis*. Expression analyses revealed jasmonic acid (JA) and ethylene (ET) expressions were differentially regulated between *A. philoxeroides* and *A. sessilis*, with the former having prominent antagonistic cross-talk between salicylic acid (SA) and JA, and the latter showing weak or no cross-talk during disease development. We also found that JA levels decreased and SA levels increased during disease development in *A. philoxeroides*. Variations in hormonal gene expression between the invasive and native species (including interspecific differences in the strength of antagonistic cross-talk) were identified during *R. solani* pathogenesis. Thus, plant hormones and their cross-talk signaling may improve the resistance of invasive *A. philoxeroides* to pathogens, which has implications for other invasive species during the invasion process.

Keywords: invasive weed; *Rhizoctonia solani*; gene expression; phytohormone signaling; systemic signal; plant defense

1. Introduction

Numerous plant species have been directly or indirectly introduced to new habitats as ornamentals, or as sources of food and fiber. However, many of these species have become invasive and pose a

serious threat to agriculture, biodiversity and ecosystem function [1]. Following initial introduction, the spread of an invasive plant species could be further enhanced by global climate change (e.g., increasing CO_2 emissions) [2,3]. Much of the previous research in invasion biology has focused on the ecological and evolutionary factors that contribute to plant invasions [4,5]. Due to the unavailability of genomic resources [6], the genetic factors underlying invasion success are still not well understood.

Invasive plants may encounter novel abiotic and biotic stresses across the introduction–naturalization–invasion continuum. These stresses can affect their survival and reproduction, and can act as barriers to plant invasion. Successful invasive plant species possess various attributes (such as rapid adaptation, fast growth and spread, or high fecundity), and have effective defenses against natural enemies, which allow them to overcome barriers to invasion [7,8]. Higher resistance against generalist herbivores and pathogens may benefit invasive plants more than non-invasive species in new regions [8,9]. The enemy release hypothesis (ERH) and the biotic resistance hypothesis (BRH) have been proposed to explain the success and limitations that invasive species experience based on studies of novel plant–natural enemy interactions [10,11]. The ERH proposes that exotic plants will be less impacted by natural enemies compared to their native range because they have escaped their specialist herbivores or pathogens [12]. Therefore, the success of invasive plants in new environments can be attributed to the allocation of resources from defense to growth to outcompete other plants (evolution of increased competitive ability hypothesis or EICA) [13]. There has been a lot of support for the ERH, including reduced impacts by herbivores on invasive plants [14–16], as well as reduced attacks by above ground fungal and soil pathogens [14,15,17–20]. In contrast, the BRH states that native enemies should limit the growth of exotic plants in new ranges [21]. Many studies support the BRH [11,22–25], while some studies support both the ERH and BRH [26,27].

The improved performance of invasive plants in a competitive environment to enemy attack may be due to genetic changes acquired during the invasion process [6,28]. The use of genomic and transcriptomic technologies could identify the genetic architecture underlying the success of invasive plant species. A recent transcriptomic study was used to compare the gene expression profiles of introduced (North American) and native (European) populations of the Canada thistle, *Cirsium arvense*, in response to nutrient deficiency and shading [29]. This study identified significant differences in R-protein mediated defense and expression pattern between introduced and native populations of *C. arvense* [29]. Similarly, in the common ragweed (*Ambrosia artemisiifolia*), an invasive species to Europe, candidate genes were identified using oligonucleotide microarrays under light and nutrient stress conditions that were thought to contribute to invasiveness [30]. In addition, weedy sunflower genotypes of *Helianthus annuus* naturalized in the USA were tested for variations in gene expression compared to wild non-weedy species [31]. This study found extensive genetic differentiation between the two species [31]. However, until now, only a few investigations have been undertaken to elucidate the genomic mechanisms responsible for the adaptation of invasive plants to biotic stresses.

Generally, a multitude of plant defense pathways are activated in response to microbial pathogens [32]. The first line of active defense occurs at the plant cell surface, when generalist microbe elicitors (the microbe or pathogen-associated molecular patterns, i.e., MAMP or PAMPs) are produced by pathogens such as, flagellin (from bacteria) [33], chitin (from fungal pathogens) [34], β-glucan (from oomycete pathogens) [35], or effectors (from specialized pathogens) [32]. Plants detect these elicitors by pattern-recognition receptors (PRRs) within the cell membrane, which leads to PAMP-triggered immunity (i.e., PTI) [36]. In the case of effectors, receptors with nucleotide-binding domains and leucine-rich repeats (NLRs or *R* genes) [37] are used, which lead to effector-triggered immunity (i.e., ETI) [38,39]. The evolutionary development of the plant immune system is represented as a Zig-zag model [40], where specialist pathogens often co-evolve with their host (see Han [41] for evolutionary dynamics between plants and pathogens). Invasive plants generally lack the need for defense against specialist pathogens in their new ranges (due to lack of co-evolution) [42] and therefore can invest more energy into growth or reproduction [12,13].

The timely recognition of the invading pathogen and a rapid, effective induction of defense responses are required for resistance to disease in plants [43]. Plant hormones play a key regulatory role in inducing defense responses shortly after the perception of a pathogen, through an extensive transcriptional reprogramming of genes involved in hormonal signaling [44]. In addition, the plant defense hormones play a critical role in response to adverse environmental conditions [45–48]. Abscisic acid (ABA) plays a central role in plant defense to abiotic stresses, such as salt and drought stress [49]. Salicylic acid (SA) is a major plant defense hormone induced by infections from biotrophic and hemi-biotrophic pathogens [50,51]. SA, jasmonic acid (JA), and ethylene (ET) also play key roles under biotic stresses [52]. These induced defenses and hormone signaling networks have been well characterized in model organisms and crop plants, such as *Arabidopsis* [53,54], tobacco [55] and tomato [56], but are not well known in invasive plants. For example, pyrosequencing was used to identify molecular signaling networks linked with paradormancy in underground vegetative buds of invasive *Cirsium arvense*. Interestingly, the plant hormone auxin and ABA-signaling was found to regulate paradormancy, allowing plants to resist weed control methods (e.g., chemical and biological controls), thereby enhancing their invasiveness [57]. Currently, there is a broad interest among invasion biologists to unravel the genetic mechanisms of resistance and defense responses of invasive plants. In this study, we hypothesize that invasive plant species have higher or enhanced resistance to microbial pathogens compared to native congeners. We also predict that the endogenous defense mechanism and signaling (especially the defense hormones SA, JA, and ET) play a crucial role in resistance, thereby benefiting the invasion success of invasive plants.

The alligator weed, *Alternanthera philoxeroides* (Martius) Grisebach, is an amphibious stoloniferous perennial herb [58]. It is native to South America [59] and was first introduced to China in the late 1930s as a forage crop from Japan [60]. It is the most noxious invasive plant in China [61,62] where it is a significant weed in rice farms, causing an estimated agricultural loss of $75 million per year [63]. Populations of *A. philoxeroides* growing in China have extremely low genetic diversity, which is attributed to the predominance of a single genotype (likely due to a single recent introduction) and extensive vegetative propagation by cuttings since being introduced [60,64].

Alternanthera philoxeroides is highly vulnerable to insect herbivore attack. For example, more than 15 generalist insects were found to feed on *A. philoxeroides* in China [11]. A specialist beetle, *Agasicles hygrophila* (alligator weed flea beetle) from South America was introduced to the USA [59] and China [11] to control *A. philoxeroides*. However, only a few disease incidences in *A. philoxeroides* have been reported. For example, species of *Nimbya* have been found to cause leaf and stem spot on *A. philoxeroides* in Australia [65]. *Nimbya alternantherae* has been identified as a biocontrol agent in Brazil [66], while species of *Fusarium* have been used as a biocontrol for alligator weed in China [67]. *Rhizoctonia solani* has been shown to be pathogenic to *A. philoxeroides*, and has also been found to infect a related species, *Alternanthera sessilis*, in the USA [68] and India [69]. Earlier preliminary pathogenicity screening tests showed that *R. solani* (ACCC 30374) is more virulent on *A. philoxeroides* compared to *Fusarium oxysporum* f.sp *cubense* (ACCC 36369) (SS Qi, unpublished data). Therefore, *A. philoxeroides* represents a good model to study pathogen resistance and defenses in an invasive plant species following invasion.

In this study, we aimed to isolate the defense hormones (SA, JA, and ET) and associated genes in invasive *A. philoxeroides* and its native congener *A. sessilis* to test for differences in gene expression between the species against a generalist necrotrophic fungus, *Rhizoctonia solani*. Although signaling of these defense hormones and their cross-talk in response to pathogens are well documented in *Arabidopsis* and other model plant species [53–56], our study is the first to examine these phenomena in wild populations of a co-occurring invasive and native congener pair. Furthermore, divergence in gene expression between the species may allow us to identify patterns of defense signaling that may be responsible for enhanced resistance in invasive *A. philoxeroides*. Specifically, we ask the following questions: (1) Is invasive *A. philoxeroides* less susceptible to the pathogen (*R. solani*) compared to its native congener *A. sessilis*? (2) Are there differences in gene expression between the native and invasive

species after inoculation with *R. solani*? (3) Are there differences in resistance between infected and un-infected neighboring leaves? To address these questions, we performed in vitro and in planta leaf inoculations using *R. solani*. Six hormones and their responsive genes (three JA, two SA, and one ET) were successfully isolated from both invasive *A. philoxeroides* and native *A. sessilis* for expression analysis using RT-qPCR in response to three treatments. The first included *R. solani* inoculation for susceptibility tests, the second comprised un-infected samples for systemic resistance tests, and the third included hormone pretreatments using SA, JA, and ET (prior to inoculation with *R. solani*).

2. Results

2.1. Disease Development in A. philoxeroides Was Delayed Compared with A. sessilis

The pathogenicity of *R. solani* was tested on invasive *A. philoxiroides* and native *A. sessilis*. Disease symptoms and necrotic lesions developed on leaves of both plant species (Figure 1a). However, *A. philoxiroides* showed slower disease progression compared to *A. sessilis* (Figure 1a,b). *Alternanthera sessilis* exhibited prominent disease symptoms at 24 hpi, whereas *A. philoxiroides* had no visible symptoms and mycelium was only observed at the site of inoculation under the microscope, indicating minimal fungal colonization (Figure 1c). Necrotic lesions were noticeable only at 48 hpi in *A. philoxiroides* (Figure 1a).

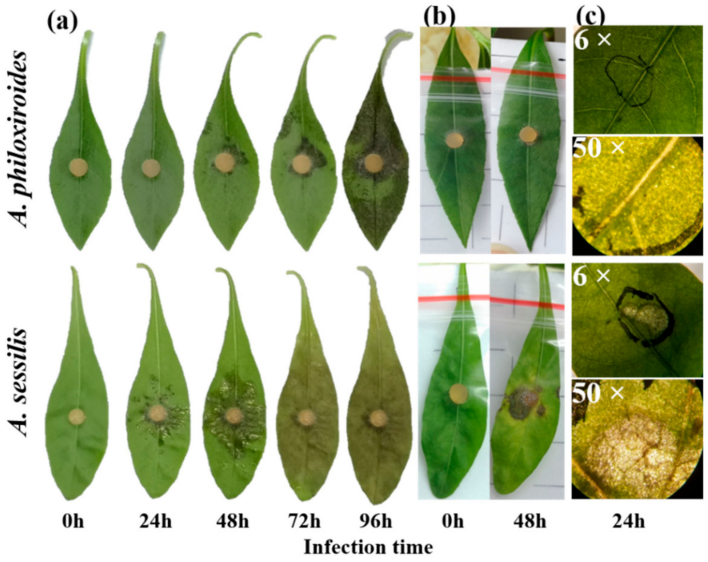


Figure 1. Disease symptoms caused by *R. solani* on invasive *A. philoxiroides* and native *A. sessilis* showing (**a**) in vitro detached leaf assay; (**b**) in planta pathogenicity test; and (**c**) microscopic observation of the infected area on leaves (at 24 hpi). Four-week-old plant leaves were inoculated with mycelium plugs of *R. solani* and disease symptoms were observed at various time intervals (24, 48, 72 and 96 hpi). For in planta tests, plastic bags were placed on the leaves after inoculation to avoid drying of plugs and moisture development for better infections. In (**c**), the inoculated area is circled with a black marker. After removing plugs, threads of fungal mycelium can be seen on the leaf surface in *A. philoxiroides*, indicating fungal colonization. This is in contrast to the clear necrotic disease lesions in *A. sessilis*.

Infected leaves stained with trypan blue clearly highlighted the region of cell death on the leaves (Figure 2a). *Alternanthera philoxiroides* had a significantly lower cell death area compared to *A. sessilis* across all time intervals post-inoculation with *R. solani* (Figure 2b). Furthermore, the amount of ion leakage was significantly higher at earlier time periods (<24 hpi) in *A. sessilis* compared to *A. philoxiroides*, which showed higher leakages at a later time interval (>48 hpi; Figure 2c). Overall, these results suggest that, although both plants are susceptible to *R. solani*, the invasive *A. philoxiroides* showed slower disease progression compared to the native *A. sessilis*.

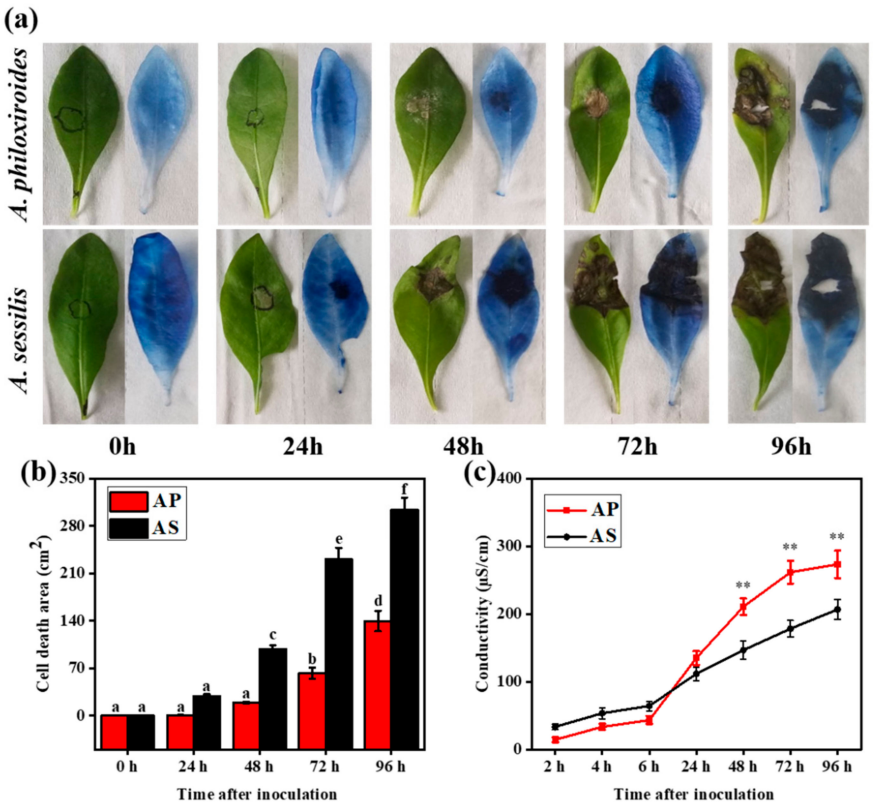


Figure 2. Trypan blue staining and ion leakage tests of *R. solani* infected leaves from invasive *A. philoxiroides* (AP) and native *A. sessilis* (AS) at different time intervals after inoculation. (**a**) Before and after staining with trypan blue for visualizing dead plant leave tissues. Circled areas on leaves are at 0 h (un-inoculated control) and 24 h (inoculated) before staining. (**b**) Cell death areas represented by means ± standard errors (SE) from three biological replicates with different letters representing the groups that were significantly different from other groups as determined by a one-way analysis of variance (ANOVA), followed by a multiple comparison using Duncan's method ($p < 0.05$). (**c**) Electrolyte leakage of leaf discs infected by *R. solani* from 2 h to 96 h measured using an electrolytes' conductivity meter. Electrolytic conductivity increased in native AS during earlier time intervals; however, it increased in invasive AP at later time intervals compared to the un-inoculated control (0 h). Error bars indicate means ± SE (number of disks = 12 for each species) and two discs from each plant representing a total of six biological replicates for each species. Asterisks indicate significantly different from native plants using Duncan's method ($p < 0.01$).

2.2. Expression Divergence of Defense Hormone Genes between *A. philoxiroides* and *A. sessilis*

Expressions of all six genes (*PAL*, *PR3*, *LOX*, *JAR1*, *PR6*, and *EIN3*) were quantified in both the local and systemic leaves of the two species against *R. solani* (the fold–change ratio of defense hormones and their responsive genes are presented in Table S1). Expression differences were observed in both infected local (Figure 3) and systemic leaves (Figure 4), as well as the hormone content (Figure S1) between *A. philoxiroides* and *A. sessilis* during disease development caused by *R. solani* (Table 1 and Table S1). Furthermore, expression differences were also observed in the hormone pre-treatment group between the two species (Figure 5). The key difference was that JA and ET-signaling was differentially regulated in *A. sessilis*. JA was partially suppressed in *A. sessilis*, whereas there was a consistent reduction in expression in *A. philoxiroides* (Figure 3a,b). ET-*EIN3* expression was reduced in *A. sessilis* but was induced in *A. philoxiroides* (Figure 3e). SA level was induced in both species (Figure 3c,d). In addition, weaker or no antagonistic cross-talk was observed between SA (Figure 3c,d) and JA-signaling (Figure 3a,b) in *A. sessilis*. In contrast, there was strong cross-talk in *A. philoxiroides* (Figure 3 and Table 1).

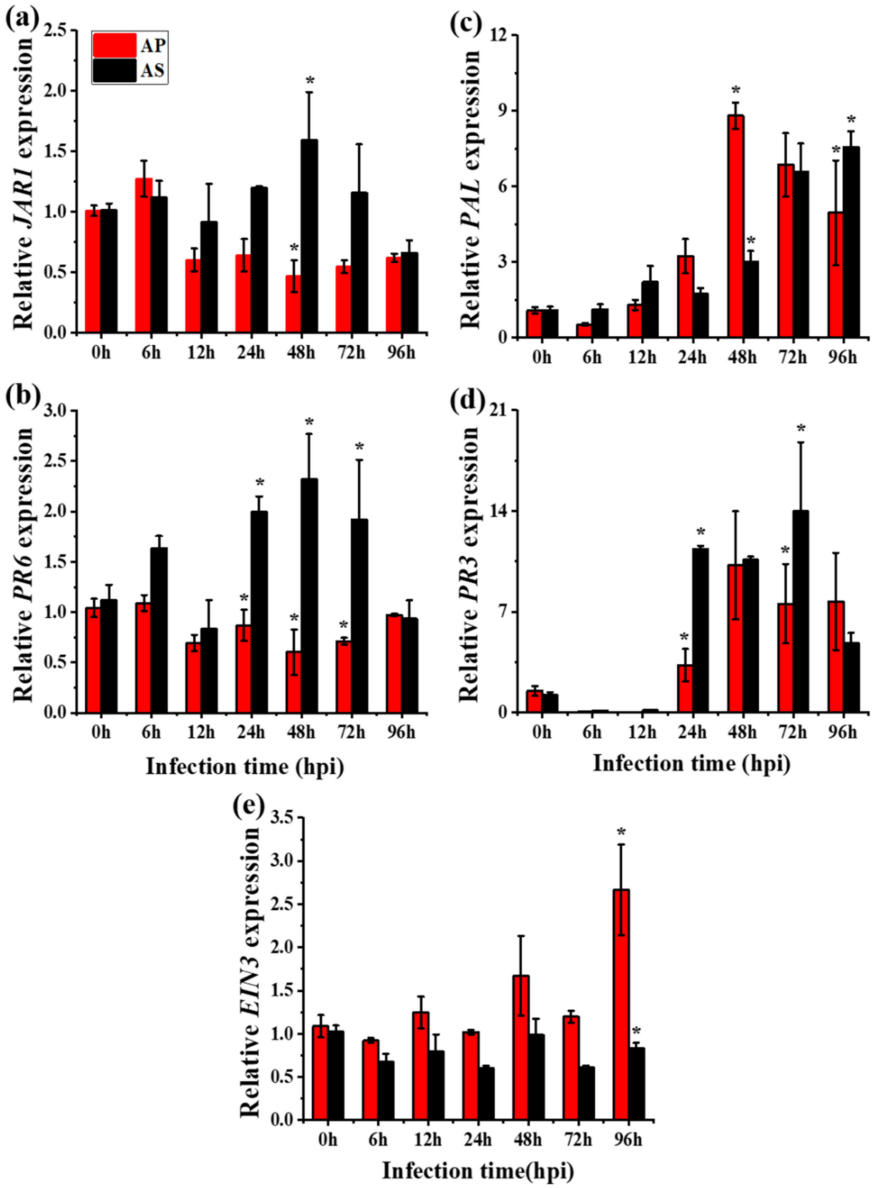


Figure 3. qPCR analysis of salicylic acid (SA), jasmonic acid (JA) and ethylene (ET)- responsive gene expressions between invasive *A. philoxiroides* (AP) and native *A. sessilis* (AS). Four-week-old plants were infected with *R. solani* and the samples were harvested for RNA extractions at the indicated time intervals (0 to 96 hours) after inoculations. Un-inoculated leaves were used as a control (0 h). Specific primers were used for *JAR1* (**a**), *PR6* (**b**), *PAL* (**c**), *PR3* (**d**) and *EIN3* (**e**) with *Actin* (control) as shown in Supplementary Table S2. Values represent means ± SE from three biological replicates. The asterisks (*) represent significantly different levels at each time period and were determined using one-way ANOVA, followed by a multiple comparison using Duncan's method ($p < 0.05$).

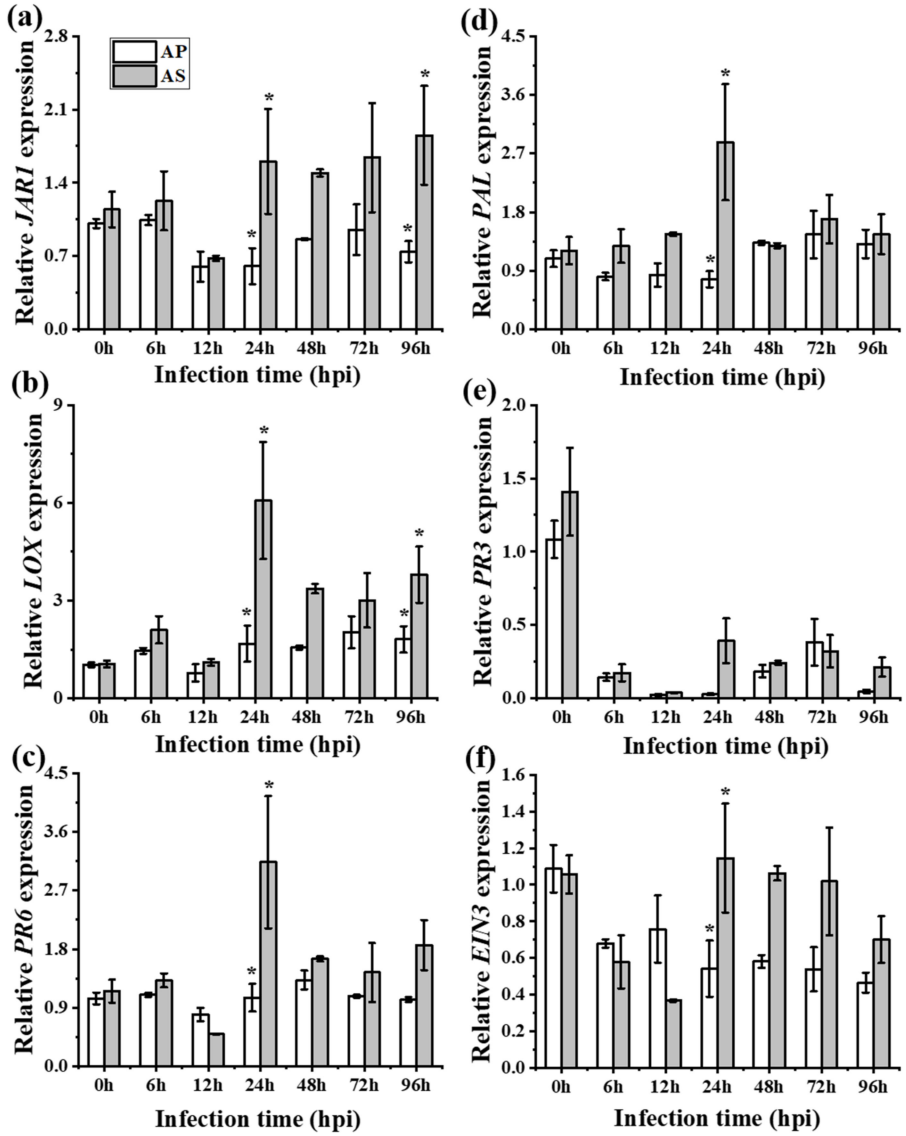


Figure 4. Expression analysis of SA, JA and ET-responsive genes for systemic acquired resistance tests between *Alternanthera philoxiroides* (AP) and *A. sessilis* (AS). Four-week-old plants were infected with *R. solani* and samples of healthy un-inoculated leaves were sampled for RNA extractions at the indicated time intervals (0 to 96 hours). 0 h is from the samples of plant completely un-infected for control. Relative expression of *JAR1* (**a**), *LOX* (**b**), *PR6* (**c**), *PAL* (**d**), *PR3* (**e**), *EIN3* (**f**) and *Actin* (control) were tested using gene specific primers at the indicated time intervals. Error bars show ± SE from three biological replicates and the asterisks (*) represent significantly different levels, which were determined via a one-way ANOVA, followed by a multiple comparison using Duncan's method ($p < 0.05$).

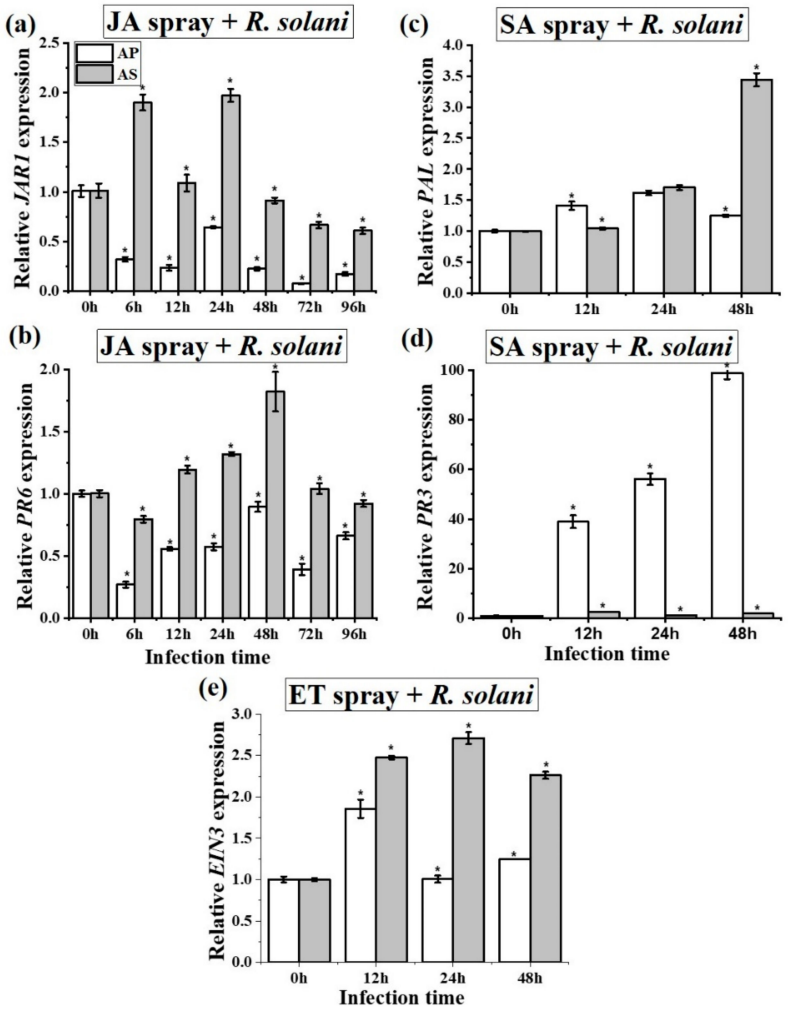


Figure 5. Expression analysis of SA, JA and ET-responsive genes for hormone pre-treatment tests between *Alternanthera philoxiroides* (AP) and *A. sessilis* (AS). Four-week-old plants were sprayed with MeJA pretreatment (**a**,**b**) SA pretreatment (**c**,**d**), and ET pretreatment (**e**) before being inoculated with *R. solani*. Samples were harvested for RNA extractions at the indicated time intervals after inoculations. Un-inoculated leaves were used as a control (0 h). Specific primers were used for *JAR1* (**a**), *PR6* (**b**), *PAL* (**c**), *PR3* (**d**) and *EIN3* (**e**) with *Actin* (control) being tested using gene specific primers at the indicated time intervals. Values represent means ± SE from three biological replicates. Asterisk (*) represent significantly different levels at each time period and was determined via a one-way ANOVA, followed by a multiple comparison using Duncan's method ($p < 0.05$).

Table 1. Expression differences in defense hormones and their responsive genes in the invasive *A. philoxeroides* and native *A. sessilis* during *R. solani* pathogenesis. See Table S1 for fold–change ratios of each gene for each treatment in both the invasive and native species.

Treatment	*Alternanthera philoxeroides*	*Alternanthera sessilis*
Rhizoctonia solani (Infected local leaves)	- JA was suppressed (*JAR1* (Figure 3a) and *PR6* (Figure 3b)) - SA was induced (*PAL* (Figure 3c) and *PR3* (Figure 3d)) - ET was induced (*EIN3* (Figure 3e)) - SA may promote disease development by suppressing JA and by inducing ET (Figure 3) - Strong antagonistic cross-talk between SA and JA - Stronger antagonistic resistance cross-talk between SA and JA may be reason for the delayed disease development (Figures 1 and 2)	- JA was partially suppressed (*JAR1* (Figure 3a) and *PR6* (Figure 3b)) - SA was induced (*PAL* (Figure 3c) and *PR3* (Figure 3d)) - ET was suppressed (*EIN3* (Figure 3e)) - SA may promote disease development but by partially suppressing JA and by completely suppressing ET (Figure 3) - No strict or strong antagonistic cross-talk between SA and JA - Lack of stronger antagonistic resistance effect between SA and JA may be reason for the enhanced disease development (Figures 1 and 2)
Rhizoctonia solani (Systemic leaves)	- JA-*JAR1* was suppressed (Figure 4a) - JA-*LOX* (Figure 4b) and *PR6* (Figure 4c) was induced but lower than *A. sessilis* - SA-*PAL* was induced from 48 hpi (Figure 4d) - SA-*PR3* was suppressed at all time (Figure 4e) - ET-*EIN3* was suppressed all the time (Figure 4f) - An antagonism between SA-*PAL* and JA-*JAR1* was observed (Figure 4)	- JA was induced (*JAR1* (Figure 4a), *LOX* (Figure 4b) and *PR6* (Figure 4c)) - SA-*PAL* was induced at all the indicated time (Figure 4d) - SA-*PR3* was completely suppressed (Figure 4e) - ET-*EIN3* was partially suppressed (Figure 4f) - No antagonism between SA-*PAL* and JA-*JAR1* was observed (Figure 4)
Rhizoctonia solani (Hormone contents)	- Endogenous contents of SA, JA and ET were higher. For example, at 96 hpi, SA, JA and ET contents were significantly higher than control samples (Figure S1)	- Endogenous contents of SA, JA and ET were comparatively lower. For example, at 96 hpi, SA, JA and ET contents were significantly higher than control samples (Figure S1)

2.3. Functional Analysis of Defense Hormone Gene Sequences between A. philoxiroides and A. sessilis

Six hormones and responsive gene sequences of both invasive and native plants from our study (Table S3) were searched in the NCBI nucleotide database and their identity was checked against other plant species in the Amaranthaceae (Table S4). The output of each gene sequence was obtained to predict a high confidence protein coding sequence from six reading frames (Table S4). Furthermore, a gene from each hormone (*PAL*, *JAR1*, and *EIN3*) was selected on the basis of core signaling component for detailed functional analysis in both *A. philoxiroides* and *A. sessilis*. Each sequence was assembled to a single long contig to predict protein coding genes along with mRNA and the amino acid translation (Table S5). We aimed to compare each gene sequence with other closely related species in Amaranthaceae, and also to compare between *A. philoxiroides* and *A. sessilis* (Figures S2–S5). A comparative phylogenetic analysis of the three hormone genes revealed high conservation within Amaranthaceae (including *Beta vulgaris* and *Spinacia oleracea*, Figure S2). We predicted the conserved domains and motifs in each gene sequence and analyzed how conserved they were across species (see Figure S3 for motifs and Figure S4 for domains). We found that multiple sequence alignments of each amino acid sequence showed high levels of gene conservation between *A. philoxiroides* and *A. sessilis*, and high conservation between the two study species and other related species (Figure S5).

2.4. R. solani Suppresses Jasmonic Acid Signaling for Disease Promotion in A. philoxiroides

To examine the role of JA-signaling, expression levels of three JA-dependent transcripts, *LOX* (Lipoxygenase), *JAR1* (JA amido synthetase 1), and *PR6* (Proteinase inhibitor), were tested in *A. philoxiroides* following inoculation with *R. solani*. We found the expressions of *JAR1* and *PR6* were reduced in *A. philoxiroides*, whereas *LOX* expression was inconsistent at each time after inoculation (Figure 6a–c). At 6 hpi, all three (*LOX*, *JAR1*, and *PR6*) levels had increased (Figure 6a–c; Table S1). Plants subjected to the MeJA hormone pretreatment also showed reduced expressions across all three JA genes at all time intervals (Figure 6d–f). Overall, our results suggest that JA signaling may be responsible for resistance to *R. solani* in *A. philoxiroides* because JA was induced earlier (up to 6 hpi). During disease progression (after 48 hpi), *R. solani* may overcome JA-resistance signaling by suppressing the defensive responses in *A. philoxiroides* (Figure 6).

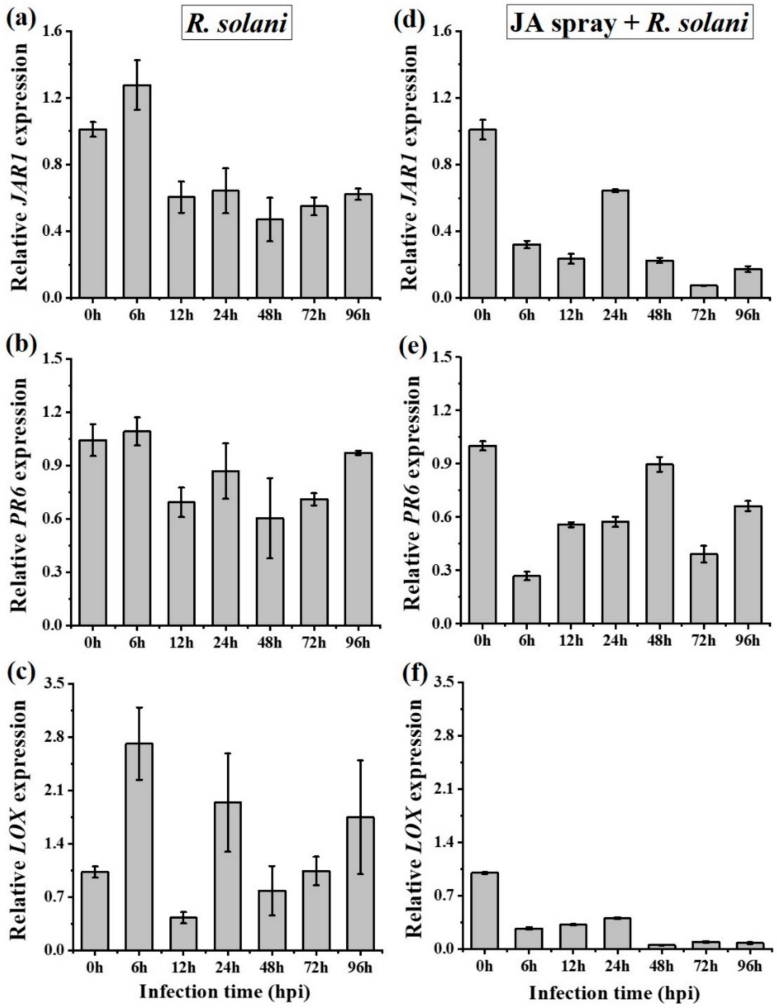


Figure 6. qPCR analysis of JA-dependent gene expression in *A. philoxiroides*. Four-week-old plants were infected with *R. solani* (**a**–**c**), or sprayed with MeJA pretreatment before being inoculated with *R. solani* (**d**–**f**). Samples were harvested for RNA extractions at the indicated time points (0 to 96 hours). 0 h is the un-inoculated control. Relative expression of JA responsive genes: *JAR1* (**a** and **d**), *PR6* (**b** and **e**), *LOX* (**c**,**f**) were tested with specific primers for *A. philoxiroides* as described in the Methods and Supplementary Table S2. Values represent means ± SE from three biological replicates.

2.5. Salicylic Acid and Ethylene Signaling Enhances Disease Susceptibility in A. philoxiroides

To determine the role of SA in plant defense response, we tested expression levels of two SA transcripts in *A. philoxiroides*: *PAL* and *PR3* (Supplementary Table S3). We found that the expression of both transcripts consistently increased at each time interval, with the exception of 6 hpi (where expression decreased, Figure 7a,b). The reduced expression at 6 hpi was in contrast to the positive induction of JA transcripts (Figure 6a–c). After inoculation, both *PAL* and *PR3* were also induced in SA pretreated plants (Figure 7d, e). However, *PR3* showed very high levels of expression (up to 100-fold at 48 hpi, Figure 7e and Table S1), suggesting that SA promotes disease development in *A. philoxiroides* following inoculation with *R. solani*.

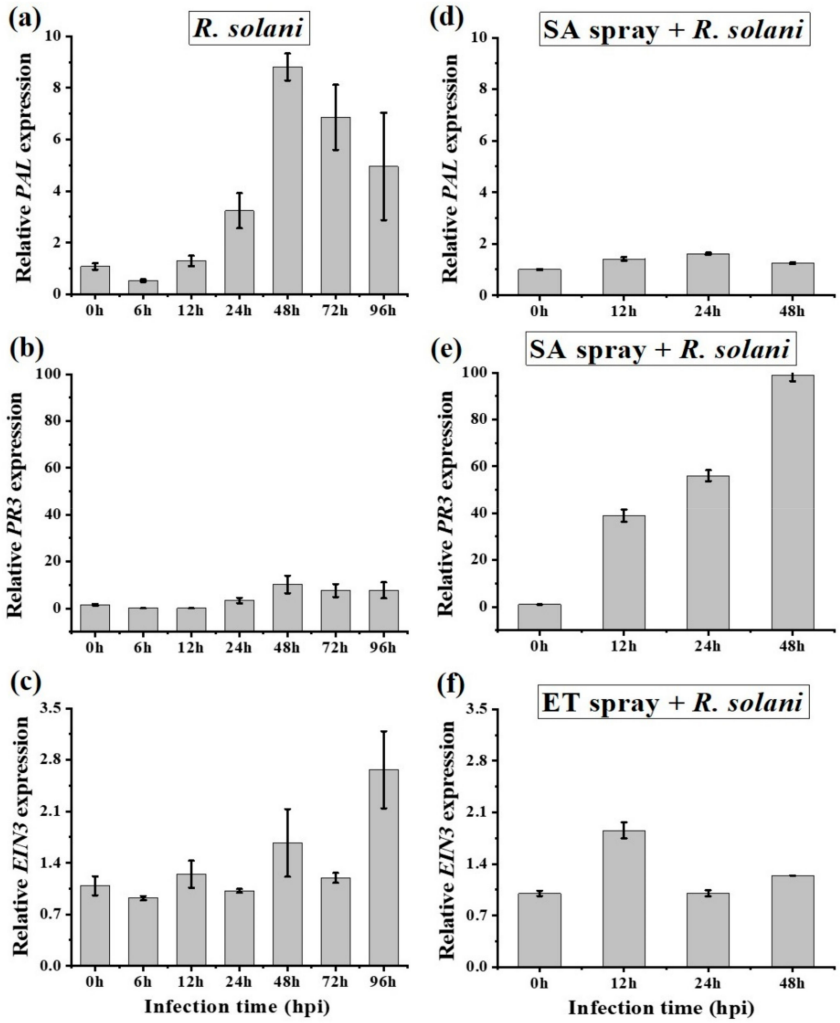


Figure 7. qPCR analysis of SA and ET-responsive gene expression in *A. philoxiroides*. Four-week-old plants were infected with *R. solani* (**a**–**c**), or were sprayed prior to *R. solani* inoculations with SA pretreatment (**d**,**e**) or ET pretreatment (**f**). Samples were harvested for RNA extractions at the indicated time points after the inoculations. 0 h is the un-inoculated control. qPCR was performed with specific primers for SA-*PAL* (**a**,**d**), *PR3* (**b**,**e**), ET-*EIN3* (**c**,**f**) and *Actin* (control) as described in the Methods and shown in Supplementary Table S2. Values represent means ±SE from three biological replicates.

The ET transcript, *EIN3*, displayed a minor increase in expression over time in both the *R. solani* infected and ET-pretreatment samples (Figure 7c,f). Specifically, the expression of *EIN3* was reduced at 6 hpi (similar to SA, Figure 7a). Overall, the results suggest ET-signaling may act synergistically with SA, thereby promoting disease susceptibility in *A. philoxiroides*.

2.6. Signaling Cross-Talk between SA, JA and ET in A. philoxiroides during Interactions with R. solani

We investigated the cross-talk between hormone signaling pathways in response to *R. solani* infection and hormone pretreatments. We observed that JA (*LOX*, *JAR1*, and *PR6*) and SA (*PAL* and *PR3*) transcripts had prominent antagonistic cross-talk in *A. philoxiroides* at both earlier (at 6 hpi) and later (>24 hpi) time intervals following *R. solani* inoculations (Figures 6a–c and 7a,b). In addition, MeJA pretreated *A. philoxiroides* showed decreased expressions in all three JA transcripts (Figure 6d–f), and also displayed higher expression levels of SA transcripts (*PAL* and *PR3*) (Figure 8a,b and Table S1). These results indicate a clear antagonistic cross-talk between SA and JA pathways in *A. philoxiroides* after inoculation with *R. solani*.

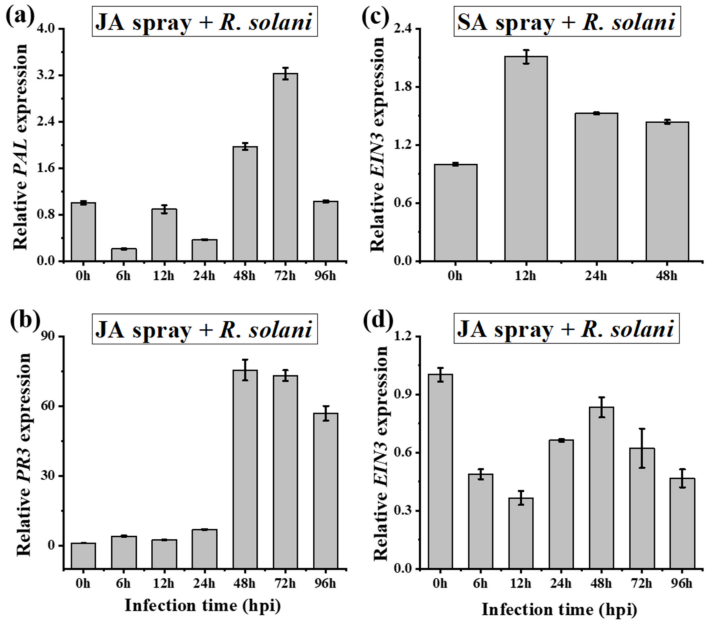


Figure 8. Signaling cross-talk analysis between SA, JA and ET-responsive gene expressions during hormone pretreatments, before *R. solani* inoculations in *A. philoxiroides*. Four-week-old plants were sprayed with MeJA (**a**, **b** and **d**) and SA (**c**). Samples were harvested for RNA extractions at the indicated time points. Un-inoculated leaves were used as a control (0 h). qRT-PCR was performed with specific primers for SA-*PAL* (**a**), *PR3* (**b**), ET-*EIN3* (**c**,**d**) and *Actin* (control) as described in the Methods and shown in Supplementary Table S2. Values represent means ± SE from three biological replicates.

As described above for ET-*EIN3*, gene expression was induced in both treatments in *A. philoxiroides* (Figure 4c, f). In addition, we also tested the expression of *EIN3* with other hormone pretreatments in *A. philoxiroides*. For example, SA pretreated plants induced *EIN3* expression similar to the SA-*PAL* and *PR3* transcripts (Figure 8c). In contrast, MeJA pretreated plants showed a decrease in *EIN3* expression, similar to the JA (*LOX*, *JAR1*, and *PR6*) transcripts (Figure 8d). The differential expression of *EIN3* to SA and MeJA suggests that ET may be regulated by both hormones depending on the type of infection and treatment. Other combinations of hormone signaling and their cross-talk gene expressions are presented in Supplementary Table S1.

To further test whether *R. solani* influences the hormones (SA, JA and ET) during pathogenesis in *A. philoxiroides*, we quantified the endogenous contents of each hormone in *R. solani* infected plants at each time interval (0, 6, 12, 24, 48, 72 and 96 hpi) using ELISA. We detected a 6.9-fold higher SA content (Figure S6a) and a 2.2-fold higher JA content (Supplementary Figure S6b) at 96 hpi compared to the control un-inoculated samples. ET was at a moderate level (about 3.2-fold higher) compared to both SA and JA in the infected plants (Supplementary Figure S6c). The levels of each hormone increased, as time since infection increased (Supplementary Figure S6).

2.7. R. solani Induced Resistance Trade-Offs between SA and JA-Signaling in A. philoxiroides

To identify and correlate whether hormonal cross-talk provides signals to adjacent leaves from local infected tissues, we examined the resistance trade-offs between SA and JA-signaling in *A. philoxiroides*. Our findings suggest that SA was induced during *R. solani* pathogenesis, which initiated antagonistic cross-talk to JA at the local infected leaves in *A. philoxiroides* (Figures 6 and 7). Investigating neighboring leaves for this cross-talk may provide information regarding how trade-off signals are regulated during infection in invasive *A. philoxiroides*. Therefore, we quantified the expression levels of all six genes in neighboring un-inoculated leaves across all time intervals. We found that JA-*JAR1* showed reduced expressions at each time interval (Figure 9a). In contrast, increased expression of SA-*PAL* from 48 hpi was detected (Figure 9b). The cross-talk between these transcripts in the neighboring un-inoculated

leaf was similar to the infected leaves. However, other transcripts of JA-*LOX* and *PR6* were induced in the same plants (Figure 9c,d), while SA-*PR3* was much reduced (Figure 9e). These findings suggest that the key signaling component of SA (*PAL*) and JA (*JAR1*) engage in antagonistic cross-talk during pathogenesis. The ET-*EIN3* expression was reduced at each time interval (Figure 9f).

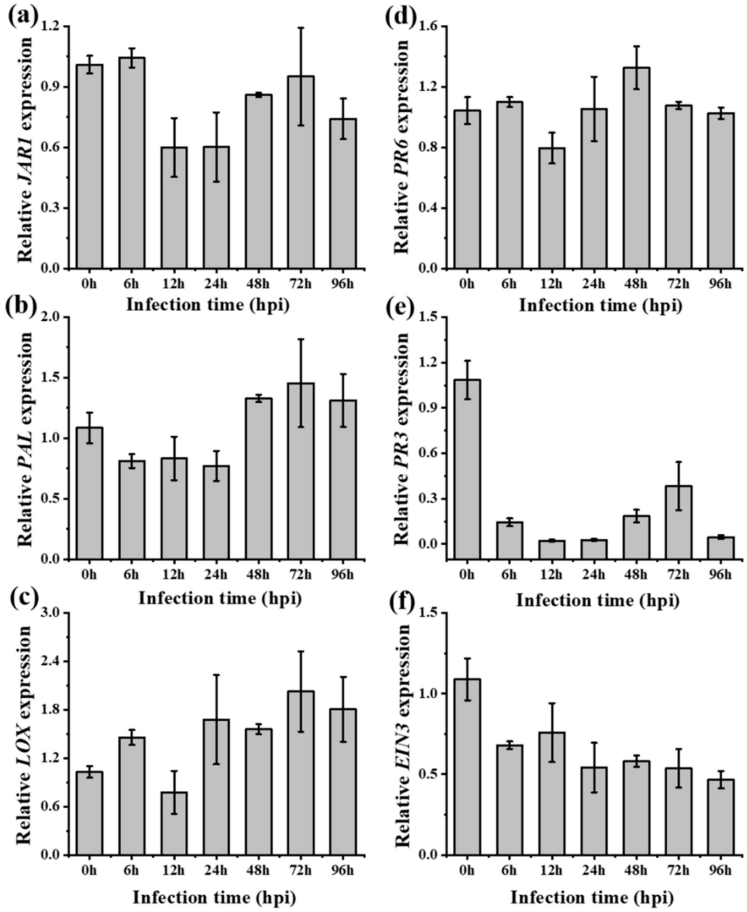


Figure 9. Expression analysis of SA, JA and ET-responsive genes for systemic acquired resistance tests in *A. philoxiroides*. Four-week-old *A. philoxiroides* were infected with *R. solani* and samples of healthy un-inoculated leaves were collected for RNA extractions (at time intervals from 0 to 96 hours). 0 h is from the samples of plant completely un-infected for control. Relative expression of *JAR1* (**a**), *PAL* (**b**), *LOX* (**c**), *PR6* (**d**), *PR3* (**e**) and *EIN3* (**f**) with *Actin* (control) genes were tested using specific primers at the indicated time intervals. Values represent means ± SE from three biological replicates.

3. Discussion

Plants are exposed to numerous biotic and abiotic stresses in their environment. Invasive plants may have evolved to adapt to these stresses and to fluctuating environmental conditions via plasticity in growth and development [6,28,70]. Studies investigating the molecular mechanisms responsible for invasion success are scarce in non-model plant species, but can now be investigated due to the availability of genomic technologies. In our study, we isolated partial gene sequences of key defense hormones in two non-model plant species (Table S6). These novel template gene sequences allowed us to investigate hormone signaling and their cross-talk within and between species to better understand the resistance of an invasive species to a globally distributed necrotrophic fungal pathogen. There were three major findings in our study: (1) *R. solani* successfully infected the invasive *A. philoxeroides* and its native congener *A. sessilis*, with disease development being much slower in *A. philoxeroides* compared to *A. sessilis*; (2) there were interspecific differences in hormone gene expression (including hormone

signaling and cross-talk) following inoculation by *R. solani*; and (3) there were differences in the hormone signaling and their cross-talk between infected local and the un-infected systemic leaves.

3.1. Differential Expression between Invasive A. philoxiroides and Native A. sessilis during Disease Development

We found that there were interspecific differences in JA and ET-signaling, as well as in the degree of antagonistic cross-talk between SA and JA (which was stronger in the invasive species and weaker in the native species). This variation in the expression of defense hormone genes between species suggests *R. solani* may affect host plant defense hormones, specifically by inducing the major defense hormone SA (which may regulate other hormones, such as JA and ET, differently between plants for successful infection of plant tissue, Table 1). Our findings also suggest that the infection of *R. solani* was enhanced in the native *A. sessilis*, perhaps due to the weaker antagonistic cross-talk between SA and JA (which may allow faster or enhanced disease development to occur in the native species). Furthermore, our results indicate that the stronger antagonistic cross-talk between SA and JA in the invasive *A. philoxiroides* may delay disease progression. To our knowledge, this is the first study that demonstrates a necrotrophic pathogen influencing plant defense hormone pathways in an invasive clonal weed differentially compared to a native congener (Figure 10). A previous study [30] identified several genes that may be involved in the introduction success of invasive *A. artemisiifolia*. Of the 180 genes identified in this earlier study, some genes were found to be involved in the metabolism of plant hormone signaling and biosynthesis (e.g., lipoxygenases and cytokinins-zeatin o-glucosyltransferase) [30]. Gene expression differences between the invasive and native species suggested that invasive species may have evolved to stressful conditions during the invasion process [29–31].

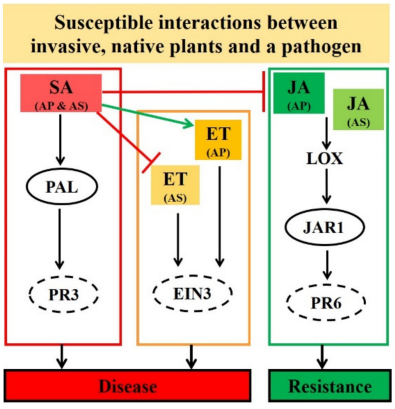


Figure 10. Proposed model showing response of defense hormone genes, salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) during pathogenesis in an invasive *Alternanthera philoxeroides* (AP) and native *A. sessilis* (AS). We investigated the molecular interactions between *R. solani* compared to two species (focusing on defense hormone signaling). Two key differences in hormone gene expressions were identified between species during pathogenesis. Firstly, the JA and ET-signaling was differentially regulated between species (partially suppressed in *A. sessilis*, whereas there was a consistent reduction in expression in *A. philoxiroides*, shown as difference in height and color of JA). Reduction in JA-dependent *LOX*, *JAR1* and *PR6* expressions during disease development suggested that JA-signaling is responsible for resistance to *R. solani*. Secondly, ET-*EIN3* expression was reduced in *A. sessilis*, but was induced in *A. philoxiroides* (shown as difference in height and color of ET). SA level was induced in both species. The elevated levels of SA in both species during disease development suggest that the unknown virulence factor from pathogenic *R. solani* may potentially target SA. This in turn may affect the antagonistic effect between SA and JA/ET differentially between species (shown as a different arrow between SA and ET). Circled SA-*PAL* and JA-*JAR1* show antagonistic effects in both the local inoculated and neighboring systemic leaves. Dotted circles (*PR3*, *EIN3* and *PR6*) represent differential regulation in both inoculated and un-inoculated systemic sites.

The differences in defense hormone signaling between the invasive and native species observed in our study during pathogenesis may be the result of interspecific genetic variation. It would be interesting to study how differences in genetic variation between the invasive *A. philoxeroides* (which has previously been found to have low genetic diversity in China due to predominantly clonal vegetative propagation, [60,64]) and native *A. sessilis* (which may have higher genetic diversity in its native range, although this needs to be confirmed) may influence the resistance response of each species to pathogens. The data from our study provide an appropriate baseline for investigating this line of inquiry in the future.

3.2. Changes in Defense Hormone Gene Expression during Pathogenesis in A. philoxeroides

Our study demonstrates how a common and widespread pathogen regulates plant defense hormones (SA, JA, and ET) allowing for the successful infection of an invasive species, *A. philoxeroides* (using RT-qPCR). Our results showed that both JA-*JAR1* and *PR6* transcript expressions decreased during disease development following inoculation with *R. solani* AG4 HGI (Figure 6a,b), indicating that JA-signaling pathway is likely to be involved in plant resistance to the pathogen. There is much evidence to suggest that endogenous JA is triggered against necrotrophic pathogens [71,72]. For example, gene expression changes induced by *R. solani* AG1 IA in resistant and susceptible rice plants have been reported and it has also been shown that JA plays an important role in disease resistance [73]. In the present study, we found that the JA biosynthetic *LOX* gene was inconsistent in expression (Figure 6c). This was consistent with the findings of a previous study [74], where it was found that *LOX* expression increased approximately 6-fold in response to virulent and avirulent strains of *Pseudomonas syringae*.

In our study, *R. solani* induced SA (*PAL* and *PR3*), which may antagonize JA during pathogenesis in invasive *A. philoxeroides* (Figure 6a,b and Figure 7a,b). Earlier studies have shown cross-talk between hormone signaling pathways can greatly help the plant to regulate defense responses to a wide range of pathogens [75–77]. The predominant cross-talk observed between SA and JA is antagonistic [50,52]. For example, an exogenous application of SA was found to inhibit JA-induced proteinase inhibitor expression in tomato [54,78], whereas MeJA treatment inhibited SA-induced acidic *PR* gene expression in tobacco [55]. Like in our study, previous studies have found that phytopathogens can take advantage of the cross-talk between SA and JA allowing them to successfully infect plants [79,80]. For example, the hemi-biotrophic pathogen *P. syringae* was shown to manipulate antagonism between SA and JA by producing coronatine, which is a phytotoxin that mimics plant JA to suppress SA. This mechanism promotes disease infection in *Arabidopsis* and tomato [81]. The necrotrophic *Botrytis cinerea* produces exopolysaccharide (EPS), a virulence factor that elicits SA to activate antagonism to JA, thereby enhancing the ability of the fungus to infect tomato [56]. Our results are consistent with the findings of El Oirdi et al. [56], although the virulence factor from *R. solani* needs to be investigated further. A follow-up study by El Oirdi et al. [56] showed SA-signaling can also contribute to disease development (caused by another necrotrophic pathogen, *Alternaria solani*) in tomato [82]. It has also been found that the SA pathway might contribute to disease development (i.e., a *SAR8.2k* gene induced by SA may be involved in disease susceptibility caused by *R. solani* in tobacco [83]).

PR genes are often associated with specific signaling pathways and their expression can be regulated by different plant hormones [84]. In our study, at 48 hpi, the SA-inducible *PR3* showed a 10-fold higher expression to *R. solani* infection (Figure 7b), a 100-fold higher expression for SA pretreatment (Figure 7e), and a 75-fold higher expression for JA pretreatment (Figure 8b). This increase in SA-inducible *PR3* supports the hypothesis that SA enhances disease development by suppressing JA in *A. philoxeroides*.

Our results indicate that, as with SA, *EIN3* (an active form of the ET transcription factor) [77] enhances disease development in *A. philoxeroides* (Figure 7c). Although *EIN3* was induced at moderate levels, its expression was consistently induced over time during inoculation (Figure 7c,f). Previous studies have shown that ethylene is a potential modulator of plant pathogen defenses [85]. JA and ET

together form an effective defense against necrotrophs [86], and can act positively or negatively with SA, depending on the specific pathogen interactions [87]. For example, ET modulates the antagonism between SA and JA pathways, and is mediated by NPR1 [53]. ET induces SA-responsive *PR1* gene expression in *Arabidopsis* [88]. In contrast, ET (*EIN3* and EIL1) represses the SA biosynthesis gene *ICS/SID2*, thereby reducing SA accumulation [89].

3.3. Hormonal Cross-Talk and Systemic Resistance in A. philoxeroides during R. solani Pathogenesis

Invasive plants may encounter multiple pathogens with different infection strategies in their non-native range. Hormonal cross-talk provides signals to adjacent leaves from local infected tissues to resist a forthcoming infection [90–92]. This systemic defense is effective against pathogens with a similar attacking strategy [93]. For pathogens with a different infection mode, hormonal cross-talk between pathways (specifically SA and JA) plays a critical role in plant resistance [90,94]. The biotrophic pathogen inducing SA can activate antagonism to JA in infected local and systemic tissues, which in turn favors insect herbivores or necrotrophic pathogens [91].

In our study, trade-offs between the SA and JA pathways was clearly observed in infected local leaves (Figures 6 and 7). This trade-off was lower in neighboring leaves (Figure 9). Only the core signaling component of SA-*PAL* and JA-*JAR1* showed antagonism in the neighboring leaves, whereas other gene transcriptions (i.e., JA-*LOX* and *PR6*) were not suppressed (Figure 9). Our results are consistent with the finding of Spoel et al. [90], who reported that *P. syringae* suppressed JA-mediated resistance to *Alternaria brassicicola* at the infected site. However, this antagonism was at modest levels in neighboring tissues in *Arabidopisis* [90]. It was suggested that, while antagonism between SA and JA was moderate, their cross-talk expression was detected in systemic tissues [90]. However, spatial separation (local, adjacent or systemic), time (immediate or delayed), pathogen type (biotroph or necrotroph), and inoculation dosage are important factors determining systemic resistance trade-offs in plants [91].

4. Materials and Methods

4.1. Plant Species and Pathogens

Alternanthera philoxeroides (Amaranthaceae) is one of the 100 worst invasive species in the world [59]. Climate modeling suggests that many regions around the world are suitable habitat for the growth of *A. philoxeroides*, including areas in Southeast Asia, Southern Africa, and Southern Europe [95]. In China, it has successfully invaded 19 provinces since its introduction (usually on roadsides and lakeshores) [96].

Alternanthera sessilis (L.) DC. is native to China, and was used for comparison with *A. philoxeroides* in our study [22]. Ramets of *A. philoxeroides* and *A. sessilis* were randomly collected in August (summer) 2016 from naturally occurring sites in Fuzhou National Forest Park (119°17′12′′E, 26°9′35′′N, Figure S7) and propagated in a greenhouse at Jiangsu University, Zhenjiang, China. Healthy stems with two nodes and without roots were planted in 250 g of sterilized sand in plastic pots (8 × 7 × 5 cm). Fixed volumes (30 mL) of distilled water or full-strength Hoagland liquid nutrient solution [97] were supplemented alternatively every two days to maintain optimal growth conditions.

Rhizoctonia solani is a ubiquitous soil-borne necrotroph that causes significant yield loss in many economically important crops globally [98]. The most common symptom is the 'damping-off' of seedlings or failure of infected seeds to germinate. The *R. solani* AG4-HGI (accession ACCC 30374) used in this study was obtained from the Agriculture Culture Collection of China (Agricultural and Microbial Culture Collection Management Center, Beijing, China). This strain was further sequenced to identify the specific anastomosis group (AG4-HGI), which was confirmed by PCR amplification of AG common and subgroup specific primers [99], and internal transcribed spacer (ITS) primers [100]. The fungus was grown and sub-cultured on potato dextrose agar (PDA) and incubated for five days at 28 °C.

4.2. Isolation of Putative Hormone-Responsive Genes for RT-qPCR

Partial gene sequences of six hormones and responsive genes (one for ET, two for SA and three for JA) were successfully isolated from both *A. philoxiroides* and *A. sessilis*. Although there is currently a lack of hormone specific sequence information for many invasive and native plant species (including *Alternanthera* species), the species from our study are in the Amaranthaceae family (which was used for sequence isolation). Twenty-one hormones and their responsive gene sequences (i.e., eight for SA, seven for JA, and six for ET) were retrieved from species in the Amaranthaceae (Table S6). Each gene sequence from species in the Amaranthaceae was aligned using Clustal Omega [101] for primer design in the conserved region (using Primer3) [102] (for initial screening). We also performed a phylogenetic analysis of these sequences (Figure S2) to compare with other species.

Plant leaf material was ground using liquid nitrogen, and 50–100 mg of ground tissue was used for DNA extraction using the Rapid Plant Genomic DNA Isolation Kit (Sangon Biotech, Shanghai, China). Genomic DNA was further precipitated by adding sodium acetate and ethanol and purified using the method of Dellaporta et al. [103]. PCR was performed, and DNA was purified using the UNIQ-10 Column MicroDNA Gel Extraction Kit (Sangon Biotech, Shanghai, China) prior to sequencing. A total of six genes (i.e., *PAL* and *PR3* for SA; *LOX*, *JAR1* and *PR6* for JA; *EIN3* for ET) were successfully isolated based on the PCR amplification of similar amplicon lengths obtained from both *A. philoxiroides* and *A. sessilis* (Table S3). Furthermore, the coding (exonic) region of each isolated gene sequence was confirmed by cDNA PCR. We performed a NCBI nucleotide search of each of the isolated sequences and found that they showed closest homology to sequences of species in the Amaranthaceae. RT-qPCR primers (amplicon length < 200 bp) were designed using web-based Integrated DNA Technologies (IDT)-Primer Quest [104] from initially screened sequences of each gene. Primers with similar amplification efficiency in the cDNA of both invasive and native plants were used in RT-qPCR. All the primer sequences used for RT-qPCR are listed in Supplementary Table S2.

4.3. Plant Inoculations with R. solani for RT-qPCR

Pathogenicity tests consisted of detached (in vitro) and in planta leaves from four-week old stem cuttings of *A. philoxeroides* and *A. sessilis* inoculated with 5-mm agar plugs (*R. solani* culture). Both detached and in planta leaf assays were performed using the method used by El Oirdi and Bouarab [105]. Inoculated leaves of in planta inoculations with mycelium plugs were covered with clear zip-lock plastic bags to maintain high humidity. Detached and in planta disease symptoms were photographed every day, up to five days after inoculation.

In planta inoculations of *A. philoxeroides* and *A. sessilis* were used for gene expression experiments by quantitative RT-qPCR. Inoculated leaf samples were harvested at different time intervals, including 0, 6, 12, 24, 48, 72, and 96 h post-inoculation (hpi). There were three biological replicates per time interval (i.e., one plant per replicate each for *A. philoxeroides* and *A. sessilis*). The 0 hpi time interval refers to un-inoculated control samples. All inoculated plants were grown at 25 °C and 70% humidity with a 16 h photoperiod. Harvested leaf material at different time intervals was quickly frozen in liquid nitrogen for RNA extractions. For the systemic acquired resistance (SAR) tests, samples of un-inoculated neighboring leaves (i.e., younger leaves just above the inoculated leaf) were investigated. For SAR gene expression experiments, samples of three biological replicates were taken per time interval (0, 6, 12, 24, 48, 72, and 96 hpi). That is, one plant per replicate for each species was investigated. The experiments of *R. solani* infection and SAR were repeated twice.

The effect of each hormone (SA, JA, and ET) on resistance to *R. solani* was tested using RT-qPCR by spraying each hormone two days before inoculation with *R. solani*. Hormone pretreatments were performed on four-week old plants of *A. philoxeroides* and *A. sessilis*. There were three treatments in this experiment group for each species (SA, JA, and ET pretreatments). Hormones SA (0.5 mM, BBI Life Sciences, Shanghai, China), methyl jasmonate (MeJA, 0.1 mM, Sigma-Aldrich, St. Louis, MO, USA), and ethephon (ET, 0.5 mM, BBI Life Sciences, Shanghai, China) were dissolved in water. Hormones were sprayed directly on plant leaves (in each treatment group) until drenched (surface run-off) once

per day, for three consecutive days. All treatment plants were grown at 25 °C and 70% humidity with a 16 h photoperiod until they were harvested. Each of the hormone pretreated samples was harvested at different time intervals: 0, 6, 12, 24, 48, 72, and 96 hpi with *R. solani*. The hormone pre-treatment samples for SA and ET spray were unable to be collected after 72 and 96 h. This was because the ET-spray samples wilted early and could not be sampled (i.e., healthy leaves could not be obtained). The control treatment (0 hpi) consisted of only water pretreated and un-inoculated. There were three biological replicates for each time interval (i.e., one plant per replicate) for each species. All harvested samples were quickly frozen in liquid nitrogen for RNA extractions. The experiment was repeated twice.

4.4. RT-qPCR Analysis

Total RNA was isolated from leaves of invasive and native plants in each treatment using the TaKaRa MiniBEST Plant RNA Extraction Kit according to the manufacturer's instructions (Takara, Shiga, Japan). First-strand cDNA was synthesized from 500 μg total RNA using PrimeScript RT Master Mix (Takara, Japan). The targets were amplified using primers that are listed in Table S2. RT-qPCR was performed using SYBR Premix Ex-Taq (Takara, Japan) on CFX96 Real-Time PCR Detection System (Bio-rad). Melt-curve analysis was performed after each PCR run to ensure specific amplification of each gene-specific primer. To identify the suitable housekeeping gene, we tested three reference genes, *β-tub* (β-tubulin), *EF1α* (Elongation factor 1-alpha), and *Act* (Actin), in *R. solani* infected samples based on Cycle threshold (Ct) difference and the coefficient of variance method [106]. Primers are listed in Table S2. *Actin* had the lowest Ct difference in *R. solani* infected samples across different time intervals (Manoharan B. et al., unpublished data) and was used for normalizing the expression level of each target gene. Relative gene expression was calculated using the comparative $2^{-\Delta\Delta CT}$ method [107]. Three biological replicates were used for each gene and each reaction was independently replicated three times.

4.5. Hormone Content Quantification

The endogenous hormone contents (SA, JA, and ET) was determined for *R. solani* infected *A. philoxiroides* and *A. sessilis*. Fungal inoculations were performed on four-week old plants (as described above). Samples collected at different time intervals (0, 6, 12, 24, 48, 72, and 96 hpi) were subjected to hormone quantification by Lengton Bio. Tec. Co., Ltd. (Shanghai, China) using the ELISA method [97]. The 0 hpi time interval refers to the un-inoculated control samples.

4.6. Quantification of Plant Cell Death by Trypan Blue Staining and Ion Leakage Assay

To visualize dead plant cells and to measure the cell death area of infected leaves, trypan blue staining was performed on the leaves of four-week old infected *A. philoxiroides* and *A. sessilis*. Leaves were removed at each time interval (0, 24, 48, 72, and 96 hpi) and stained with lactophenol-trypan blue, followed by de-staining with saturated chloral hydrate using the method of Koch and Slusarenko [108]. Samples were photographed every day before and after staining to measure the diameter of the infected area using ImageJ [109]. To further confirm cell death, ion leakage assay was performed on infected leaves of both species using an electrolytic conductivity meter (model P772) following the procedure outlined in Hatsugai and Katagiri [110].

4.7. Flanking Sequence Isolation and Bioinformatic Analyses

Six isolated DNA sequences from both *A. philoxeroides* and *A. sessilis* (Table S3) were searched in the NCBI non-redundant protein database using BLAST+ (version 2.2.31) to compare with other plant species. BLASTx output of each sequence was submitted to OrfPredictor [111] to identify the best matching open reading frame (ORF) sequence [110] (Table S4). Furthermore, one gene from each of the hormones, *PAL* (SA), *JAR1* (JA), and *EIN3* (ET) was selected for additional flanking sequence isolation for both *A. philoxiroides* and *A. sessilis* (for detailed function and bioinformation analysis). Flanking

(5′ and 3′) sequences were isolated from a known sequence region (i.e., initially isolated sequences, Table S2) using 5′ Genome walking and 3′ RACE (Rapid Amplification of cDNA Ends) techniques, following the methods outlined in the respective kit instructions (Genome Walking Kit, code 6108 and 3′-Full RACE Core Set with PrimeScript RTase, code 6106, Takara, Shiga, Japan). Isolated sequences from both techniques were assembled for each gene with the corresponding initial sequence using CAP3 sequence assembly program with the following parameters: base quality cut-off for clipping value of 12, overlap length cut-off value ≥20, overlap percent identity ≥75, and overlap similarity score ≥500 [112]. The assembled single long contig was selected for annotating protein coding gene with ab-initio method and *Beta vulgaris* gene-specific parameters using the FGENESH online tool [113]. Predicted CDS and peptide sequences were reconfirmed by aligning to the NCBI RefSeq nucleotide and protein database using mega BLAST and BLASTp search tools. Clustal-W from MEGA 7.0.26 was used for multiple sequence alignment of each peptide sequence with other related species. A maximum likelihood phylogenetic tree was constructed with 1000 bootstrap replicates (Figure S2) [114]. In addition, motifs were searched for each peptide sequences using MEME suite [115] with the default options. The conserved domain was searched using the NCBI database with an expected value of 0.010000 [116].

4.8. Data Availability

The partial gene sequences isolated in this study were deposited in the GenBank under accession IDs: MK790145 to MK790156. The *Rhizoctonia solani* was sequenced to identify the specific anastomosis group and was deposited under GenBank accession ID: MK801228.

5. Conclusions

During the invasion process, an invasive species from a small founding population may face difficulties due to adverse abiotic and biotic stresses, impacting its survival and reproduction. Invasive species can undergo genetic changes to overcome these natural barriers. Our findings are consistent with previous studies, which identified that many stress induced genes are differentially expressed between invasive and native plant species [29–31]. Our study advances our current understanding of hormone resistance to a widespread pathogen in an invasive species compared to its native congener. In addition, our findings provide insights into the significance of invasive plant defense pathway genes that may have evolved during the invasion process. The manipulation of host defense hormones in favor of pathogen colonization in native plant species may be a mechanism where invasive species gain an advantage over co-occurring native congeners, and this should be the focus of future research. Further studies are required to identify the currently unknown virulence factors (i.e., pathogen effectors or toxins that mimic plant hormone coronatine from *R. solani* and other pathogens) that may affect the signaling pathways differentially between invasive and native species [56]. More specifically, research is needed on resistance (*R*) genes, such as the non-expresser of PR genes 1 (NPR1) of hormone SA, which is a major transcriptional activator that activates antagonism to JA or ET during disease development [117].

Supplementary Materials: All supplementary material from this study is available online at http://www.mdpi.com/1422-0067/20/19/4916/s1, including Figure S1: Endogenous hormone salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) contents in both *Alternanthera philoxiroides* and *A. sessilis* after *Rhizoctonia solani* inoculations. Figure S2: Comparative phylogenetic analysis of the selected hormone genes in both the invasive and native species, as well as other closely related species. Figure S3: Identification of conserved motifs in selected genes from both invasive and native species. Figure S4: Identification and alignment of the conserved domain in each of the selected genes from both invasive and native species. Figure S5: Multiple sequence alignments of predicted amino acid sequences of both *Alternanthera philoxeroides* and *A. sessilis* along with other closely related plants. Figure S6: Endogenous hormone Salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) contents in *Alternanthera philoxiroides* after infected with *Rhizoctonia solani*. Figure S7: Sampling site of invasive *Alternanthera philoxeroides* and native *A. sessilis*. Table S1: Fold-change ratio of defense hormones and their responsive genes under different treatment conditions in invasive *Alternanthera philoxeroides* compared to native *A. sessilis*. Table S2: Primers used for quantitative RT-qPCR. Table S3: List of putative defense hormones and responsive genes isolated in both *Alternanthera philoxeroides* and *A. sessilis*. Table S4: Sequence information of six defense hormones and responsive

genes isolated from *Alternanthera philoxeroides* and *A. sessilis*. Table S5: Functional analysis of three hormone genes in both *Alternanthera philoxeroides* and *A. sessilis*. Table S6: Screening and isolation of defense hormones and responsive gene sequences from both invasive *Alternanthera philoxeroides* and native *A. sessilis*.

Author Contributions: Z.C.D., S.S.Q., D.L.D. and B.M. designed the research; Q.C., Q.L., Y.H.Y. and J.L. performed plant growth, sample collection and pathogenicity tests; B.M. and V.D. performed gene isolation and bioinformatic analyses; B.M., V.D., S.R., J.S.H.W. and S.J. analyzed the data; B.M. performed the research and wrote the manuscript; Z.C.D., S.S.Q., S.R., J.S.H.W. and S.J. reviewed the manuscript.

Funding: This study was funded by the State Key Research Development Program of China (2017YFC1200100), the National Natural Science Foundation of China (31700342, 31600326, 31770446, 31570414), the Natural Science Foundation of Jiangsu (BK20150503), and the China Postdoctoral Science Foundation (2017T100329). Part of the funding for this research was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the Jiangsu Collaborative Innovation Center of Technology and Material of Water Treatment, and the Study Abroad Scholarship of Jiangsu Province and Jiangsu University.

Acknowledgments: We thank the Jiangsu University postdoctoral fellowship program.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Vila, M.; Espinar, J.L.; Hejda, M.; Hulme, P.E.; Jarosik, V.; Maron, J.L.; Pergl, J.; Schaffner, U.; Sun, Y.; Pysek, P. Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. *Ecol. Lett.* **2011**, *14*, 702–708. [CrossRef] [PubMed]
2. Bradley, B.A.; Blumenthal, D.M.; Wilcove, D.S.; Ziska, L.H. Predicting plant invasions in an era of global change. *Trends Ecol. Evol.* **2010**, *25*, 310–318. [CrossRef] [PubMed]
3. Chown, S.L.; Hodgins, K.A.; Griffin, P.C.; Oakeshott, J.G.; Byrne, M.; Hoffmann, A.A. Biological invasions, climate change and genomics. *Evol. Appl.* **2015**, *8*, 23–46. [CrossRef] [PubMed]
4. Pyšek, P.; Richardson, D.M. Invasive Species, Environmental Change and Management, and Health. *Annu. Rev. Environ. Resour.* **2010**, *35*, 25–55. [CrossRef]
5. Stewart, C.N.; Tranel, P.J.; Horvath, D.P.; Anderson, J.V.; Rieseberg, L.H.; Westwood, J.H.; Mallory-Smith, C.A.; Zapiola, M.L.; Dlugosch, K.M. Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics. *Weed Sci.* **2009**, *57*, 451–462. [CrossRef]
6. Dlugosch, K.M.; Parker, I.M. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecol. Lett.* **2008**, *11*, 701–709. [CrossRef] [PubMed]
7. Richardson, D.M.; Pyšek, P. Naturalization of introduced plants: Ecological drivers of biogeographical patterns. *N. Phytol.* **2012**, *196*, 383–396. [CrossRef] [PubMed]
8. Van Kleunen, M.; Dawson, W.; Schlaepfer, D.; Jeschke, J.M.; Fischer, M. Are invaders different? A conceptual framework of comparative approaches for assessing determinants of invasiveness. *Ecol. Lett.* **2010**, *13*, 947–958. [CrossRef]
9. Engelkes, T.; Morriën, E.; Verhoeven, K.J.F.; Bezemer, T.M.; Biere, A.; Harvey, J.A.; McIntyre, L.M.; Tamis, W.L.M.; van der Putten, W.H. Successful range-expanding plants experience less above-ground and below-ground enemy impact. *Nature* **2008**, *456*, 946. [CrossRef]
10. Cushman, J.H.; Lortie, C.J.; Christian, C.E. Native herbivores and plant facilitation mediate the performance and distribution of an invasive exotic grass. *J. Ecol.* **2011**, *99*, 524–531. [CrossRef]
11. Fan, S.; Yu, H.; Dong, X.; Wang, L.; Chen, X.; Yu, D.; Liu, C. Invasive plant Alternanthera philoxeroides suffers more severe herbivory pressure than native competitors in recipient communities. *Sci. Rep.* **2016**, *6*, 36542. [CrossRef] [PubMed]
12. Keane, R.M.; Crawley, M.J. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* **2002**, *17*, 164–170. [CrossRef]
13. Blossey, B.; Notzold, R. Evolution of Increased Competitive Ability in Invasive Nonindigenous Plants: A Hypothesis. *J. Ecol.* **1995**, *83*, 887–889. [CrossRef]
14. DeWalt, S.J.; Denslow, J.S.; Ickes, K. Natural-enemy Release Facilitates Habitat Expansion of The Invasive Tropical Shrub Clidemia Hirta. *Ecology* **2004**, *85*, 471–483. [CrossRef]
15. Roy, B.A.; Hudson, K.; Visser, M.; Johnson, B.R. Grassland fires may favor native over introduced plants by reducing pathogen loads. *Ecology* **2014**, *95*, 1897–1906. [CrossRef] [PubMed]

16. Vila, M.; Maron, J.L.; Marco, L. Evidence for the enemy release hypothesis in Hypericum perforatum. *Oecologia* **2005**, *142*, 474–479. [CrossRef] [PubMed]
17. Callaway, R.M.; Thelen, G.C.; Rodriguez, A.; Holben, W.E. Soil biota and exotic plant invasion. *Nature* **2004**, *427*, 731–733. [CrossRef]
18. Reinhart, K.O.; Callaway, R.M. Soil Biota Facilitate Exotic Acer Invasions in Europe and North America. *Ecol. Appl.* **2004**, *14*, 1737–1745. [CrossRef]
19. Reinhart, K.O.; Tytgat, T.; Van der Putten, W.H.; Clay, K. Virulence of soil-borne pathogens and invasion by Prunus serotina. *N. Phytol.* **2010**, *186*, 484–495. [CrossRef]
20. Wolfe, L.M. Why alien invaders succeed: Support for the escape-from-enemy hypothesis. *Am. Nat.* **2002**, *160*, 705–711. [CrossRef]
21. Parker, J.D.; Hay, M.E. Biotic resistance to plant invasions? Native herbivores prefer non-native plants. *Ecol. Lett.* **2005**, *8*, 959–967. [CrossRef]
22. Fan, S.; Yu, D.; Liu, C. The invasive plant Alternanthera philoxeroides was suppressed more intensively than its native congener by a native generalist: Implications for the biotic resistance hypothesis. *PLoS ONE* **2013**, *8*, e83619. [CrossRef] [PubMed]
23. Levine, J.M.; Adler, P.B.; Yelenik, S.G. A meta-analysis of biotic resistance to exotic plant invasions. *Ecol. Lett.* **2004**, *7*, 975–989. [CrossRef]
24. Lombardero, M.J.; Alonso-Rodríguez, M.; Roca-Posada, E.P. Tree insects and pathogens display opposite tendencies to attack native vs. non-native pines. *For. Ecol. Manag.* **2012**, *281*, 121–129. [CrossRef]
25. Pearse, I.S.; Hipp, A.L. Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18097–18102. [CrossRef] [PubMed]
26. Knevel, I.C.; Lans, T.; Menting, F.B.; Hertling, U.M.; van der Putten, W.H. Release from native root herbivores and biotic resistance by soil pathogens in a new habitat both affect the alien Ammophila arenaria in South Africa. *Oecologia* **2004**, *141*, 502–510. [CrossRef]
27. Mitchell, C.E.; Power, A.G. Release of invasive plants from fungal and viral pathogens. *Nature* **2003**, *421*, 625–627. [CrossRef]
28. Colautti, R.I.; Eckert, C.G.; Barrett, S.C.H. Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proc. R. Soc. B Biol. Sci.* **2010**, *277*, 1799–1806. [CrossRef]
29. Guggisberg, A.; Lai, Z.; Huang, J.; Rieseberg, L.H. Transcriptome divergence between introduced and native populations of Canada thistle, Cirsium arvense. *N. Phytol.* **2013**, *199*, 595–608. [CrossRef]
30. Hodgins, K.A.; Lai, Z.; Nurkowski, K.; Huang, J.; Rieseberg, L.H. The molecular basis of invasiveness: Differences in gene expression of native and introduced common ragweed (Ambrosia artemisiifolia) in stressful and benign environments. *Mol. Ecol.* **2013**, *22*, 2496–2510. [CrossRef]
31. Lai, Z.; Kane, N.C.; Zou, Y.; Rieseberg, L.H. Natural variation in gene expression between wild and weedy populations of Helianthus annuus. *Genetics* **2008**, *179*, 1881–1890. [CrossRef] [PubMed]
32. Andersen, E.J.; Ali, S.; Byamukama, E.; Yen, Y.; Nepal, M.P. Disease Resistance Mechanisms in Plants. *Genes* **2018**, *9*, 339. [CrossRef] [PubMed]
33. Zipfel, C.; Robatzek, S.; Navarro, L.; Oakeley, E.J.; Jones, J.D.; Felix, G.; Boller, T. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **2004**, *428*, 764–767. [CrossRef]
34. Kaku, H.; Nishizawa, Y.; Ishii-Minami, N.; Akimoto-Tomiyama, C.; Dohmae, N.; Takio, K.; Minami, E.; Shibuya, N. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11086–11091. [CrossRef] [PubMed]
35. Fliegmann, J.; Mithofer, A.; Wanner, G.; Ebel, J. An ancient enzyme domain hidden in the putative beta-glucan elicitor receptor of soybean may play an active part in the perception of pathogen-associated molecular patterns during broad host resistance. *J. Biol. Chem.* **2004**, *279*, 1132–1140. [CrossRef]
36. Zipfel, C. Plant pattern-recognition receptors. *Trends Immunol.* **2014**, *35*, 345–351. [CrossRef] [PubMed]
37. Kourelis, J.; van der Hoorn, R.A.L. Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *Plant Cell* **2018**, *30*, 285–299. [CrossRef]
38. Dodds, P.N.; Rathjen, J.P. Plant immunity: Towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **2010**, *11*, 539. [CrossRef]
39. Dangl, J.L.; Horvath, D.M.; Staskawicz, B.J. Pivoting the plant immune system from dissection to deployment. *Science* **2013**, *341*, 746–751. [CrossRef]
40. Jones, J.D.G.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [CrossRef]

41. Han, G.Z. Origin and evolution of the plant immune system. *N. Phytol.* **2019**, *222*, 70–83. [CrossRef] [PubMed]
42. Dai, Z.-C.; Qi, S.-S.; Miao, S.-L.; Liu, Y.-T.; Tian, Y.-F.; Zhai, D.-L.; Huang, P.; Du, D.-L. Isolation of NBS-LRR RGAs from invasive Wedelia trilobata and the calculation of evolutionary rates to understand bioinvasion from a molecular evolution perspective. *Biochem. Syst. Ecol.* **2015**, *61*, 19–27. [CrossRef]
43. Bari, R.; Jones, J.D. Role of plant hormones in plant defence responses. *Plant Mol. Biol.* **2009**, *69*, 473–488. [CrossRef] [PubMed]
44. De Vos, M.; Van Oosten, V.R.; Van Poecke, R.M.; Van Pelt, J.A.; Pozo, M.J.; Mueller, M.J.; Buchala, A.J.; Metraux, J.P.; Van Loon, L.C.; Dicke, M.; et al. Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack. *Mol. Plant Microbe Interact. MPMI* **2005**, *18*, 923–937. [CrossRef] [PubMed]
45. Jegadeesan, S.; Beery, A.; Altahan, L.; Meir, S.; Pressman, E.; Firon, N. Ethylene production and signaling in tomato (Solanum lycopersicum) pollen grains is responsive to heat stress conditions. *Plant Reprod.* **2018**, *31*, 367–383. [CrossRef] [PubMed]
46. Ku, Y.S.; Sintaha, M.; Cheung, M.Y.; Lam, H.M. Plant Hormone Signaling Crosstalks between Biotic and Abiotic Stress Responses. *Int. J. Mol. Sci.* **2018**, *19*, 3206. [CrossRef]
47. Mine, A.; Seyfferth, C.; Kracher, B.; Berens, M.L.; Becker, D.; Tsuda, K. The Defense Phytohormone Signaling Network Enables Rapid, High-Amplitude Transcriptional Reprogramming during Effector-Triggered Immunity. *Plant Cell* **2018**, *30*, 1199–1219. [CrossRef] [PubMed]
48. Pieterse, C.M.; Van der Does, D.; Zamioudis, C.; Leon-Reyes, A.; Van Wees, S.C. Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Biol.* **2012**, *28*, 489–521. [CrossRef]
49. Cutler, S.R.; Rodriguez, P.L.; Finkelstein, R.R.; Abrams, S.R. Abscisic acid: Emergence of a core signaling network. *Annu. Rev. Plant Biol.* **2010**, *61*, 651–679. [CrossRef]
50. Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **2005**, *43*, 205–227. [CrossRef]
51. Vlot, A.C.; Dempsey, D.A.; Klessig, D.F. Salicylic Acid, a multifaceted hormone to combat disease. *Annu. Rev. Phytopathol.* **2009**, *47*, 177–206. [CrossRef] [PubMed]
52. Pieterse, C.M.; Leon-Reyes, A.; Van der Ent, S.; Van Wees, S.C. Networking by small-molecule hormones in plant immunity. *Nat. Chem. Biol.* **2009**, *5*, 308–316. [CrossRef] [PubMed]
53. Leon-Reyes, A.; Spoel, S.H.; De Lange, E.S.; Abe, H.; Kobayashi, M.; Tsuda, S.; Millenaar, F.F.; Welschen, R.A.; Ritsema, T.; Pieterse, C.M. Ethylene modulates the role of NONEXPRESSOR OF PATHOGENESIS-RELATED GENES1 in cross talk between salicylate and jasmonate signaling. *Plant Physiol.* **2009**, *149*, 1797–1809. [CrossRef] [PubMed]
54. Leon-Reyes, A.; Van der Does, D.; De Lange, E.S.; Delker, C.; Wasternack, C.; Van Wees, S.C.; Ritsema, T.; Pieterse, C.M. Salicylate-mediated suppression of jasmonate-responsive gene expression in Arabidopsis is targeted downstream of the jasmonate biosynthesis pathway. *Planta* **2010**, *232*, 1423–1432. [CrossRef]
55. Niki, T.; Mitsuhara, I.; Seo, S.; Ohtsubo, N.; Ohashi, Y. Antagonistic Effect of Salicylic Acid and Jasmonic Acid on the Expression of Pathogenesis-Related (PR) Protein Genes in Wounded Mature Tobacco Leaves. *Plant Cell Physiol.* **1998**, *39*, 500–507. [CrossRef]
56. El Oirdi, M.; El Rahman, T.A.; Rigano, L.; El Hadrami, A.; Rodriguez, M.C.; Daayf, F.; Vojnov, A.; Bouarab, K. Botrytis cinerea manipulates the antagonistic effects between immune pathways to promote disease development in tomato. *Plant Cell* **2011**, *23*, 2405–2421. [CrossRef]
57. Anderson, J.V.; Dogramaci, M.; Horvath, D.P.; Foley, M.E.; Chao, W.S.; Suttle, J.C.; Thimmapuram, J.; Hernandez, A.G.; Ali, S.; Mikel, M.A. Auxin and ABA act as central regulators of developmental networks associated with paradormancy in Canada thistle (Cirsium arvense). *Funct. Integr. Genom.* **2012**, *12*, 515–531. [CrossRef]
58. Schooler, S.S. Alternanthera philoxeroides (Martius) Grisebach. In *A Handbook of Global Freshwater Invasive Species*; Francis, R.A., Ed.; Routlegde: Abingdon, UK, 2012.
59. Buckingham, G.R. Biological Control of Alligatorweed, Alternanthera philoxeroides, the World's First Aquatic Weed Success Story. *Castanea* **1996**, *61*, 232–243.
60. Wang, B.; Li, W.; Wang, J. Genetic diversity of Alternanthera philoxeroides in China. *Aquat. Bot.* **2005**, *81*, 277–283. [CrossRef]
61. Feng, J.; Zhu, Y. Alien invasive plants in China: Risk assessment and spatial patterns. *Biodivers. Conserv.* **2010**, *19*, 3489–3497. [CrossRef]

62. Wan, L.-Y. Clonal integration is beneficial for resource sharing in a creeping amphibian herb (Alteranthera philoxeroides). *Folia Geobot.* **2017**, *52*, 10–432. [CrossRef]
63. Shen, J.; Shen, M.; Wang, X.; Lu, Y. Effect of environmental factors on shoot emergence and vegetative growth of alligatorweed (Alternanthera philoxcroides). *Weed Sci.* **2005**, *53*, 471–478. [CrossRef]
64. Geng, Y.-P.; Pan, X.-Y.; Xu, C.-Y.; Zhang, W.-J.; Li, B.; Chen, J.-K.; Lu, B.-R.; Song, Z.-P. Phenotypic plasticity rather than locally adapted ecotypes allows the invasive alligator weed to colonize a wide range of habitats. *Biol. Invasions* **2007**, *9*, 245–256. [CrossRef]
65. Gilbert, R.L.; Auld, B.A.; Hennecke, B.R. Leaf and stem spot of Alternanthera philoxeroides (alligatorweed) in Australia caused by Nimbya sp. *Plant Pathol.* **2005**, *54*, 585. [CrossRef]
66. Pomella, A.W.V.; Barreto, R.W.; Charudattan, R. Nimbya alternantherae a potential biocontrol agent for alligatorweed, Alternanthera philoxeroides. *BioControl* **2007**, *52*, 271–288. [CrossRef]
67. Tan, W.Z.; Li, Q.J.; Qing, L. Biological control of alligatorweed (Alternanthera philoxeroides) with a Fusarium sp. *BioControl* **2002**, *47*, 463–479. [CrossRef]
68. Joyner, B.G.; Freeman, T.E. Pathogenicity of Rhizoctonia solani to aquatic plants. *Phytopathology* **1972**, *63*, 681–685. [CrossRef]
69. Saveinai, R.; Baiswar, P.; Kumar, R.; Rajesh, T.; Behere, G.T. Pathogenicity of Rhizoctonia solani AG 1-IA on major weeds prevalent in rice and maize ecosystem in Meghalaya. *Indian Phytopathol.* **2017**, *70*, 91–97. [CrossRef]
70. Prentis, P.J.; Wilson, J.R.; Dormontt, E.E.; Richardson, D.M.; Lowe, A.J. Adaptive evolution in invasive species. *Trends Plant Sci.* **2008**, *13*, 288–294. [CrossRef]
71. Berens, M.L.; Berry, H.M.; Mine, A.; Argueso, C.T.; Tsuda, K. Evolution of Hormone Signaling Networks in Plant Defense. *Annu. Rev. Phytopathol.* **2017**, *55*, 401–425. [CrossRef]
72. Mengiste, T. Plant immunity to necrotrophs. *Annu. Rev. Phytopathol.* **2012**, *50*, 267–294. [CrossRef] [PubMed]
73. Zhang, J.; Chen, L.; Fu, C.; Wang, L.; Liu, H.; Cheng, Y.; Li, S.; Deng, Q.; Wang, S.; Zhu, J.; et al. Comparative Transcriptome Analyses of Gene Expression Changes Triggered by Rhizoctonia solani AG1 IA Infection in Resistant and Susceptible Rice Varieties. *Front. Plant Sci.* **2017**, *8*, 1422. [CrossRef] [PubMed]
74. Melan, M.A.; Dong, X.; Endara, M.E.; Davis, K.R.; Ausubel, F.M.; Peterman, T.K. An Arabidopsis thaliana lipoxygenase gene can be induced by pathogens, abscisic acid, and methyl jasmonate. *Plant Physiol.* **1993**, *101*, 441–450. [CrossRef] [PubMed]
75. Beckers, G.J.; Spoel, S.H. Fine-Tuning Plant Defence Signalling: Salicylate versus Jasmonate. *Plant Biol.* **2006**, *8*, 1–10. [CrossRef] [PubMed]
76. Robert-Seilaniantz, A.; Grant, M.; Jones, J.D. Hormone crosstalk in plant disease and defense: More than just jasmonate-salicylate antagonism. *Annu. Rev. Phytopathol.* **2011**, *49*, 317–343. [CrossRef] [PubMed]
77. Shigenaga, A.M.; Argueso, C.T. No hormone to rule them all: Interactions of plant hormones during the responses of plants to pathogens. *Semin. Cell Dev. Biol.* **2016**, *56*, 174–189. [CrossRef] [PubMed]
78. Doares, S.H.; Narvaez-Vasquez, J.; Conconi, A.; Ryan, C.A. Salicylic Acid Inhibits Synthesis of Proteinase Inhibitors in Tomato Leaves Induced by Systemin and Jasmonic Acid. *Plant Physiol.* **1995**, *108*, 1741–1746. [CrossRef] [PubMed]
79. Ma, K.W.; Ma, W. Phytohormone pathways as targets of pathogens to facilitate infection. *Plant Mol. Biol.* **2016**, *91*, 713–725. [CrossRef] [PubMed]
80. Tanaka, S.; Han, X.; Kahmann, R. Microbial effectors target multiple steps in the salicylic acid production and signaling pathway. *Front. Plant Sci.* **2015**, *6*, 349. [CrossRef] [PubMed]
81. Uppalapati, S.R.; Ishiga, Y.; Wangdi, T.; Kunkel, B.N.; Anand, A.; Mysore, K.S.; Bender, C.L. The phytotoxin coronatine contributes to pathogen fitness and is required for suppression of salicylic acid accumulation in tomato inoculated with Pseudomonas syringae pv. tomato DC3000. *Mol. Plant Microbe Interact.* **2007**, *20*, 955–965. [CrossRef]
82. Rahman, T.A.; Oirdi, M.E.; Gonzalez-Lamothe, R.; Bouarab, K. Necrotrophic pathogens use the salicylic acid signaling pathway to promote disease development in tomato. *Mol. Plant Microbe Interact. MPMI* **2012**, *25*, 1584–1593. [CrossRef] [PubMed]
83. Portieles, R.; Ochagavia, M.E.; Canales, E.; Silva, Y.; Chacon, O.; Hernandez, I.; Lopez, Y.; Rodriguez, M.; Terauchi, R.; Borroto, C.; et al. High-throughput SuperSAGE for gene expression analysis of Nicotiana tabacum-Rhizoctonia solani interaction. *BMC Res. Notes* **2017**, *10*, 603. [CrossRef] [PubMed]

84. Derksen, H.; Rampitsch, C.; Daayf, F. Signaling cross-talk in plant disease resistance. *Plant Sci.* **2013**, *207*, 79–87. [CrossRef]
85. Broekaert, W.F.; Delaure, S.L.; De Bolle, M.F.; Cammue, B.P. The role of ethylene in host-pathogen interactions. *Annu. Rev. Phytopathol.* **2006**, *44*, 393–416. [CrossRef] [PubMed]
86. Penninckx, I.A.; Thomma, B.P.; Buchala, A.; Metraux, J.P.; Broekaert, W.F. Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. *Plant Cell* **1998**, *10*, 2103–2113. [CrossRef]
87. Adie, B.A.; Perez-Perez, J.; Perez-Perez, M.M.; Godoy, M.; Sanchez-Serrano, J.J.; Schmelz, E.A.; Solano, R. ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. *Plant Cell* **2007**, *19*, 1665–1681. [CrossRef]
88. De Vos, M.; Van Zaanen, W.; Koornneef, A.; Korzelius, J.P.; Dicke, M.; Van Loon, L.C.; Pieterse, C.M.J. Herbivore-Induced Resistance against Microbial Pathogens in Arabidopsis. *Plant Physiol.* **2006**, *142*, 352–363. [CrossRef]
89. Chen, H.; Xue, L.; Chintamanani, S.; Germain, H.; Lin, H.; Cui, H.; Cai, R.; Zuo, J.; Tang, X.; Li, X.; et al. ETHYLENE INSENSITIVE3 and ETHYLENE INSENSITIVE3-LIKE1 repress SALICYLIC ACID INDUCTION DEFICIENT2 expression to negatively regulate plant innate immunity in Arabidopsis. *Plant Cell* **2009**, *21*, 2527–2540. [CrossRef]
90. Spoel, S.H.; Johnson, J.S.; Dong, X. Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18842–18847. [CrossRef]
91. Spoel, S.H.; Dong, X. Making sense of hormone crosstalk during plant immune responses. *Cell Host Microbe* **2008**, *3*, 348–351. [CrossRef]
92. Spoel, S.H.; Dong, X. How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.* **2012**, *12*, 89–100. [CrossRef] [PubMed]
93. Ton, J.; Van Pelt, J.A.; Van Loon, L.C.; Pieterse, C.M. Differential effectiveness of salicylate-dependent and jasmonate/ethylene-dependent induced resistance in Arabidopsis. *Mol. Plant Microbe Interact. MPMI* **2002**, *15*, 27–34. [CrossRef] [PubMed]
94. Heil, M.; Ton, J. Long-distance signalling in plant defence. *Trends Plant Sci.* **2008**, *13*, 264–272. [CrossRef] [PubMed]
95. Julien, M.H.; Skarratt, B.; Maywald, G.F. Potential geographical distribution of alligator weed and its biological control by Agasicles hygrophila. *J. Aquat. Plant Manag.* **1995**, *33*, 55–60.
96. Weber, E.; Sun, S.-G.; Li, B. Invasive alien plants in China: Diversity and ecological insights. *Biol. Invasions* **2008**, *10*, 1411–1429. [CrossRef]
97. Dai, Z.C.; Fu, W.; Qi, S.S.; Zhai, D.L.; Chen, S.C.; Wan, L.Y.; Huang, P.; Du, D.L. Different Responses of an Invasive Clonal Plant Wedelia trilobata and its Native Congener to Gibberellin: Implications for Biological Invasion. *J. Chem. Ecol.* **2016**, *42*, 85–94. [CrossRef]
98. Paulitz, T.C.; Okubara, P.A.; Schillinger, W.F. First Report of Damping-Off of Canola Caused by Rhizoctonia solani AG 2-1 in Washington State. *Plant Dis.* **2006**, *90*, 829. [CrossRef] [PubMed]
99. Matsumoto, M. Trials of direct detection and identification of Rhizoctonia solani AG 1 and AG 2 subgroups using specifically primed PCR analysis. *Mycoscience* **2002**, *43*, 185–189. [CrossRef]
100. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*; Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J., Eds.; Academic Press: San Diego, CA, USA, 1990; pp. 315–322.
101. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41*, W597–W600. [CrossRef]
102. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]
103. Dellaporta, S.L.; Wood, J.; Hicks, J.B. A plant DNA minipreparation: Version II. *Plant Mol. Biol. Rep.* **1983**, *1*, 19–21. [CrossRef]
104. Thornton, B.; Basu, C. Real-time PCR (qPCR) primer design using free online software. *Biochem. Mol. Biol. Educ. A Bimon. Publ. Int. Union Biochem. Mol. Biol.* **2011**, *39*, 145–154. [CrossRef] [PubMed]
105. El Oirdi, M.; Bouarab, K. Plant signalling components EDS1 and SGT1 enhance disease caused by the necrotrophic pathogen Botrytis cinerea. *N. Phytol.* **2007**, *175*, 131–139. [CrossRef] [PubMed]

106. Lin, L.; Han, X.; Chen, Y.; Wu, Q.; Wang, Y. Identification of appropriate reference genes for normalizing transcript expression by quantitative real-time PCR in Litsea cubeba. *Mol. Genet. Genom. Mgg* **2013**, *288*, 727–737. [CrossRef]
107. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef]
108. Koch, E.; Slusarenko, A. Arabidopsis is susceptible to infection by a downy mildew fungus. *Plant Cell* **1990**, *2*, 437–445.
109. Abràmoff, M.D.; Magalhães, P.J.; Ram, S.J. Image Processing with ImageJ. *Biophoton. Int.* **2004**, *11*, 36–42.
110. Hatsugai, N.; Katagiri, F. Quantification of Plant Cell Death by Electrolyte Leakage Assay. *Bio Protoc.* **2018**, *8*, e2758. [CrossRef]
111. Min, X.J.; Butler, G.; Storms, R.; Tsang, A. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **2005**, *33*, W677–W680. [CrossRef]
112. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999**, *9*, 868–877. [CrossRef]
113. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **2006**, *7*, S10.1–S12. [CrossRef]
114. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
115. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]
116. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *4*, D200–D203. [CrossRef]
117. Ding, Y.; Sun, T.; Ao, K.; Peng, Y.; Zhang, Y.; Li, X.; Zhang, Y. Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity. *Cell* **2018**, *173*, 1454–1467. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Protective Role of Leaf Variegation in *Pittosporum tobira* under Low Temperature: Insights into the Physio-Biochemical and Molecular Mechanisms

Zhilu Zhang *, Zhonghua Liu, Haina Song, Minghui Chen and Shiping Cheng

College of Chemistry and Environmental Engineering, Pingdingshan University, Pingdingshan 467000, China; hphy226@sohu.com (Z.L.); zhanglaoqi1234@sina.com (H.S.), zzl1111@tom.com (M.C.); chengship@163.com (S.C.)

* Correspondence: zzl1111@pdsu.edu.cn

Received: 9 September 2019; Accepted: 21 September 2019; Published: 30 September 2019

Abstract: Leaf variegation has been demonstrated to have adaptive functions such as cold tolerance. *Pittosporum tobira* is an ornamental plant with natural leaf variegated cultivars grown in temperate regions. Herein, we investigated the role of leaf variegation in low temperature responses by comparing variegated "Variegatum" and non-variegated "Green Pittosporum" cultivars. We found that leaf variegation is associated with impaired chloroplast development in the yellow sector, reduced chlorophyll content, strong accumulation of carotenoids and high levels of ROS. However, the photosynthetic efficiency was not obviously impaired in the variegated leaves. Also, leaf variegation plays low temperature protective function since "Variegatum" displayed strong and efficient ROS-scavenging enzymatic systems to buffer cold (10 °C)-induced damages. Transcriptome analysis under cold conditions revealed 309 differentially expressed genes between both cultivars. Distinctly, the strong cold response observed in "Variegatum" was essentially attributed to the up-regulation of *HSP70/90* genes involved in cellular homeostasis; up-regulation of *POD* genes responsible for cell detoxification and up-regulation of *FAD2* genes and subsequent down-regulation of *GDSL* genes leading to high accumulation of polyunsaturated fatty acids for cell membrane fluidity. Overall, our results indicated that leaf variegation is associated with changes in physiological, biochemical and molecular components playing low temperature protective function in *P. tobira*.

Keywords: cold response; *Pittosporum tobira*; leaf variegation; linoleic acid; ROS scavenging enzyme; heat shock protein

1. Introduction

Leaf variegation has been observed in many species of higher plants [1–3] and this special attractive trait has become a focus of plant breeding as it increases the economic value of ornamental plants [4]. There are two categories of leaf variegation in plants: structural-related variegation and pigment-related variegation [5,6]. Two different types of structural variegation have been described, including the air-space type and epidermis type of variegation, which play adaptive roles to varying light conditions [7]. Pigment-leaf variegation is most common in ornamental plants because of the chlorophyll-deficiency [6]. It is marked by the existence of sections that contain abnormal plastids [5]. The leaf color variegation in plants are divided into several types based on color classification for instance green, yellow and albino (white) sectors on leaves [1,2,8,9].

Nuclear and plastid mutations or changes in expression of several genes which contribute to chloroplast biogenesis and chlorophyll biosynthesis induce the leaf variegation [10,11]. The white sectors of variegated leaves lack photosynthetic activity, therefore, leaf variegation may affect photosynthetic efficiency [12]. Previously, a transcriptome study of the *Arabidopsis* white-green

variegated mutant *immutans* (*im*) and an *Arabidopsis FtsH2* mutant line (*var2*) revealed that the genes related to photosynthesis were down-regulated in the white sectors of leaves [13,14]. Furthermore, the chlorophyll-deficient leaf-mutant also showed the expressional repression of transcriptional factors *GLK1*, *Ftsz* and *MinD* that regulate chloroplast development and division [15]. Recently, a mutation in the transcription factor mitochondrial transcription termination factor (*mTERF*) has been found to induce colorlessness in leaves of variegated fig [3]. Although these studies have provided a deep understanding of the variegation mechanism in plants, the advantage of this trait for the good fitness or for the plant physiology is still poorly understood.

Several potential physiological advantages of variegation have been proposed in plants. For example, it was reported that leaf variegation is involved in plant defense from enemies including aposematic coloration, mimicry of dead or infested plants, masquerade and camouflage [16–19]. It can also play physiological roles such as improved water or gas transport [20], mitigation of UV radiation [21] and thermoregulation [22]. Investigations led on forest trees displaying variegated leaves hinted that the trait might be a strategy to prevent the attack of herbivores [23,24]. Later, studies by Mwafongo et al. [25] on leaf variegation patterns in *Ledebouria revoluta* highlighted two possible functions including the photoprotection role and the aposematic role. Very recently, Shelef et al. [22] demonstrated that under lower temperatures, variegated wild type *Silybum marianum* leaves were significantly warmer than all-green mutants, conferring cold stress tolerance. These studies showed that variegation in plants is not just a color mutation but has some physiological advantages.

Pittosporum tobira (Thunb.) Aiton belonging to the family Pittosporaceae originated from East Asia and at present is being widely cultivated as an ornamental flowering plant in temperate and subtropical regions around the world [26]. Typically, *P. tobira* plants are about 2–3 m high with thick, rubbery and dark green colored leaves. The fragrant flowers of *P. tobira* have been well studied for their antimicrobial and anti-oxidant activities [27,28]. Importantly, some cultivars exhibit leaf variegation with yellowish or creamy white leaf margins and green interior, which have a greater aesthetic appeal and ornamental value compared to the typical all-green *P. tobira*. These particular variegated cultivars are spread to temperate regions. However, besides the aesthetic advantage, the intrinsic physiological importance of leaf variegation for *P. tobira* is unknown.

In the present work, we studied two *P. tobira* cultivars namely, "Variegatum" and "Green Pittosporum" with distinct leaf coloration features. To thoroughly understand the role of leaf variegation in *P. tobira* under cold condition, we investigated the physio-biochemical characteristics at different temperature gradients and profiled leaf transcriptome of the two cultivars under cold stress. Our findings elucidate the leaf variegation mechanism in *P. tobira* and provide novel insights into the thermo-protective function of this important trait.

2. Results

2.1. Characteristics of Variegated Leaves in Pittosporum Tobira

A naturally occurring leaf variegated cultivar of *Pittosporum tobira* named "Variegatum" was collected from Pingdingshan, Henan province in China. The cultivar "Variegatum" bears yellowish margins and green interior leaves, whereas, the typical cultivar "Green Pittosporum" exhibits dark green colored leaves (Figure 1). The phenotypic characteristics such as leaf thickness and shape were found to be similar for both cultivars except for the variegation. It is well documented that the leaves of variegated plants having green/yellow sectors have impaired chloroplast biogenesis, less photosynthetic pigments in the yellow sectors and also accumulate excessive levels of reactive oxygen species (ROS) [29,30]. To verify these observations in *P. tobira*, we analyzed the chloroplast ultrastructure in the yellow sector compared to the green sector of the variegated leaf. As shown in Figure 2A,B, the green sector contained well-developed chloroplasts with stacked grana. In contrast, in the white sector of the leaf, plastids did not contain stacked grana but contained large starch granules and many plastoglobuli (Figure 2C,D). Next, we assessed the photosynthetic parameters

and malonaldehyde (MDA) in both cultivars in August when the ambient temperature is around 20 °C (Figure 1). The net photosynthetic rate (*Pn*), the intercellular CO_2 concentration (*Ci*) and the transpiration (*Tr*) rate were found similar between leaves from both cultivars (Figure 3A–C), showing that the photosynthetic efficiency is not significantly impaired in "Variegatum" as compared to "Green Pittosporum". Next, we compared the content of photosynthesis-related pigments such as total chlorophyll (chlT) and carotenoids (Ca) in both leaf types. The results revealed that the chlT contents were significantly lower ($p < 0.05$) in "Variegatum" compared to the "Green Pittosporum" (Figure 3D), while Ca was higher in "Variegatum" compared to the "Green Pittosporum" (Figure 3E), indicating that the yellowish phenotype in "Variegatum" is underlined by a reduced chlorophyll content and a stronger accumulation of carotenoids. We further measured the MDA content, which is associated with lipid peroxidation via an increased generation of ROS [31]. The MDA was significantly ($p < 0.01$) and highly accumulated in "Variegatum" leaves compared to "Green Pittosporum" leaves (Figure 3F), implying a high level of ROS in the variegated leaves.

Taken together, our results showed that leaf variegation trait in *P. tobira* is associated with defected chloroplast biogenesis in the yellow sector, reduced chlorophyll content, strong accumulation of carotenoids and high level of ROS.

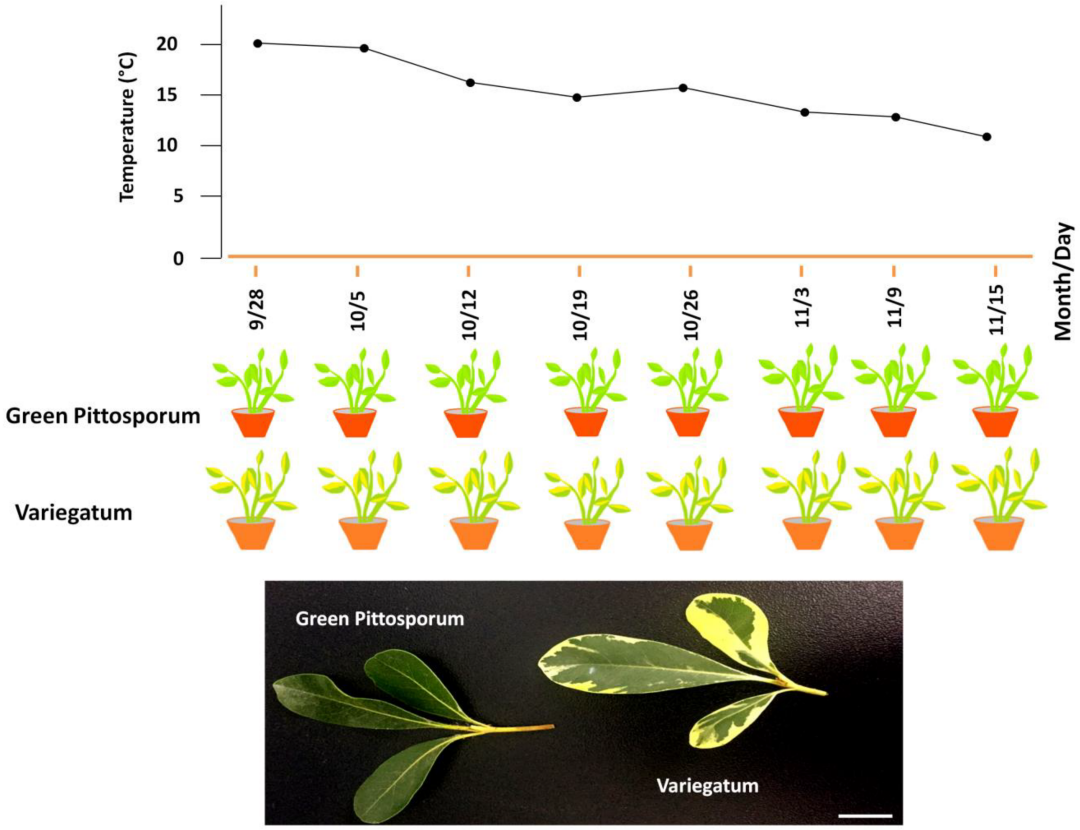


Figure 1. Overview of the experiment design and phenotypes of the two *Pittosporum tobira* cultivars, namely "Variegatum" with green/yellowish variegated leaf and "Green Pittisporum" with complete dark green leaf. Leaf samples were harvested at different dates following decrease of ambient temperature. The bar = 2 cm.

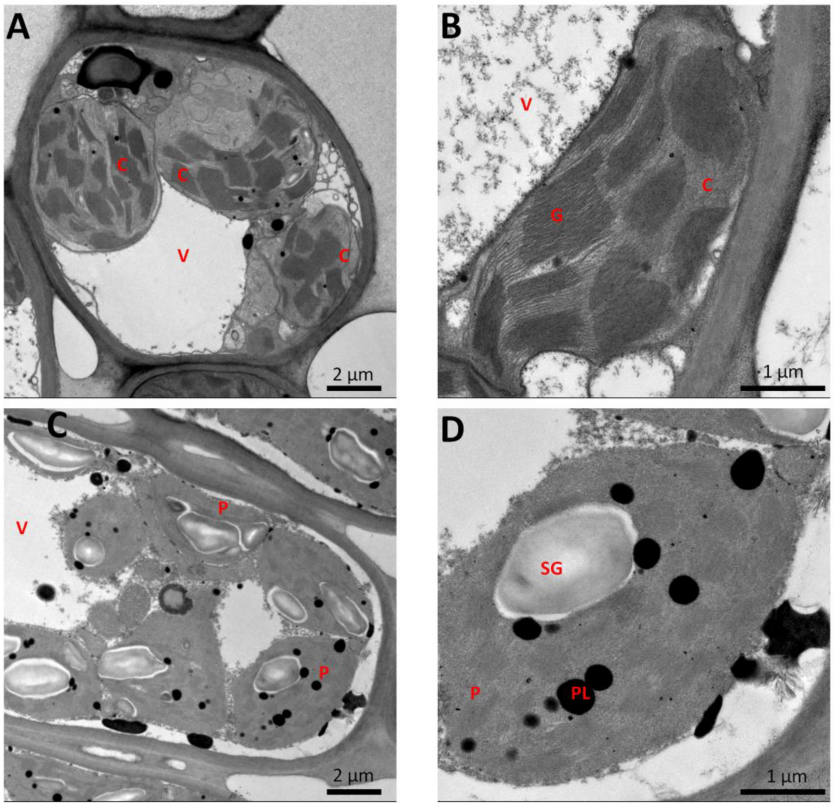


Figure 2. Chloroplast ultrastructure of the green (**A**,**B**) and yellow (**C**,**D**) sectors in variegated leaves of *Pittosporum tobira* cultivar "Variegatum". C = chloroplast; P = plastid; SG = starch granule; G = grana; V = vacuole, PL = plastoglobuli.

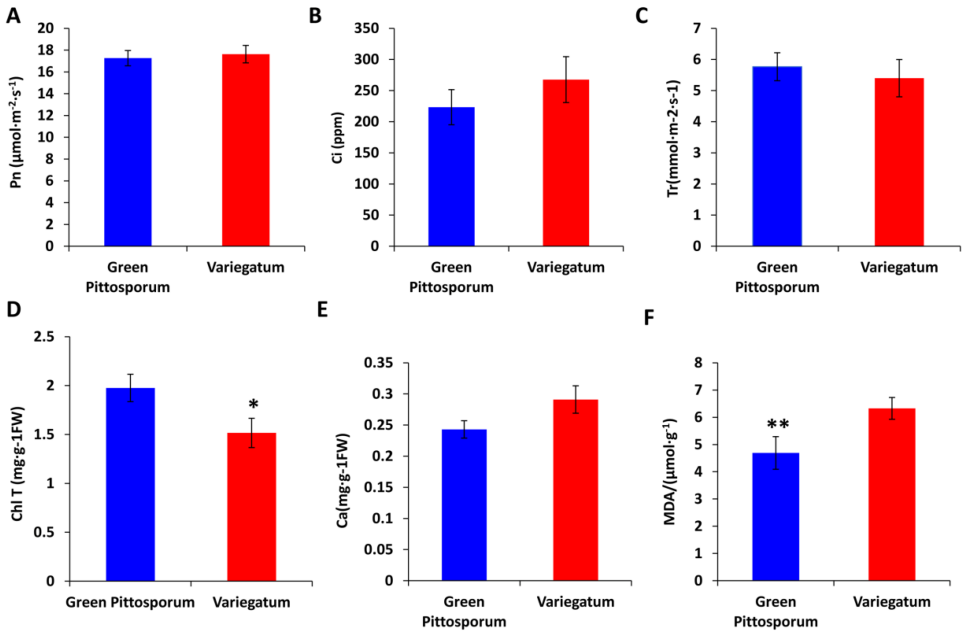


Figure 3. Physio-biochemical comparison of leaf from variegated "Variegatum" and non-variegated "Green Pittosporum" cultivars. (**A**) net photosynthetic rate (*Pn*), (**B**) intercellular CO_2 concentration (*Ci*), (**C**) transpiration rate (*Tr*), (**D**) total chlorophyll content (ChlT), (**E**) carotenoids content (Ca) and (**F**) malonaldehyde content (MDA). *, ** above the bars represent significant difference between the two cultivars at $p < 0.05$ and $p < 0.001$, respectively, using Tukey's honestly significant difference (HSD) test.

2.2. *Effect of Temperature Decrease on ROS-Scavenging Enzyme Activities in P. tobira Cultivars*

The natural occurrence of leaf variegation in plants suggests that the trait might have adaptive functions [32]. In line with this, we investigated the enzymatic changes with respect to cold stress response in leaves of "Variegatum" and "Green Pittosporum" cultivars over a period of three months (from August to November) when the ambient temperature decreases from optimal condition (20 °C) to cold condition (10 °C). The results showed that the two cultivars respond similarly to the temperature decrease (Figure 4). The activities of all the three ROS-scavenging enzymes including peroxidase (POD), catalase (CAT), superoxide dismutase (SOD), were increased over the assayed period. Notably, CAT and POD displayed a sharp increase in response to the temperature decrease. At the lowest temperature (10 °C, November 15th), POD and CAT activities were significantly higher ($P < 0.01$) in "Variegatum" as compared to "Green Pittosporum" (Figure 4A–C), denoting a stronger response to cold in "Variegatum". We extended the investigation on the MDA contents in both cultivars in order to record the stress levels induced by the temperature decrease. As expected, MDA levels also increased in both cultivars with the temperature decrease, but "Variegatum" seems to suffer less from cold stress. This is evidenced by the significantly higher ($p < 0.01$) MDA in "Green Pittosporum" when the temperature reached 10 °C (Figure 4D).

Overall, our results indicated that "Variegatum" is endowed with an efficient ROS-scavenging enzymatic system, which is mainly triggered under low temperature. Hence, leaf variegation trait plays a low temperature protective function in *P. tobira*.

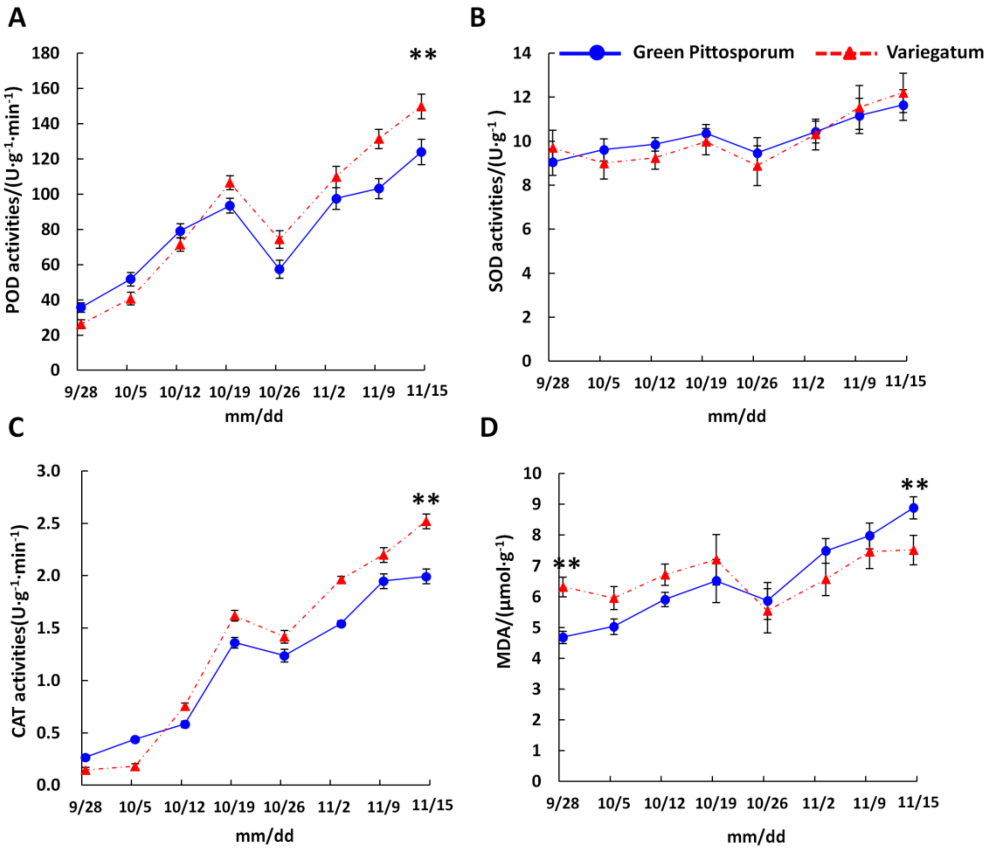


Figure 4. Antioxidant enzymatic activities during temperature decrease. (**A**) Peroxidase activity (POD), (**B**) Superoxide dismutase activity (SOD), (**C**) Catalase activity (CAT), (**D**) malonaldehyde content (MDA). ** above the lines represents significant difference between the two cultivars at $p < 0.001$, using Tukey HSD test.

2.3. Transcriptome Sequencing in Leaves of "Variegatum" and "Green Pittosporum" and Functional Annotation of Unigenes

To get an insight into the molecular pathways and genes conferring the strong response to low temperature in "Variegatum", we synthesized six cDNA libraries from leaves collected from "Variegatum" and "Green Pittosporum" plants under cold conditions (10 °C) and generated *de novo* RNA-sequencing data for the first time in *P. tobira*.

The RNA-seq yielded a total of 40.88 Gb clean data with 92.78% of bases scoring Q30 and above (Table 1). The assembly was performed using the Trinity software and a total of 112,875 unigenes were obtained with N50 length about 1,017 bp. The assembly integrity was high and specific statistics are shown in Table 2. A total of 51,718 unique genes were functionally annotated based on various databases (Table 3, Table S1). The clean data of each sample was serialized with the assembled unigene libraries and the mapping result statistics are presented in Table 4. Of these genes, 19,677 genes were expressed with the number of fragments per kilobase of exon per million fragments mapped (FPKM) values ranging from 0.04 to 22578.37 (Table S2, Figure 5A).

Hierarchical clustering of the samples based on FPKM showed that all the biological replicates clustered together, suggesting a high reliability of our RNA-sequencing data (Figure 5B). Moreover, a clear separation of the two leaf sample types was observed, implying that a large number of genes may be differentially expressed between the two cultivars to explain the relative stronger response to cold stress in "Variegatum".

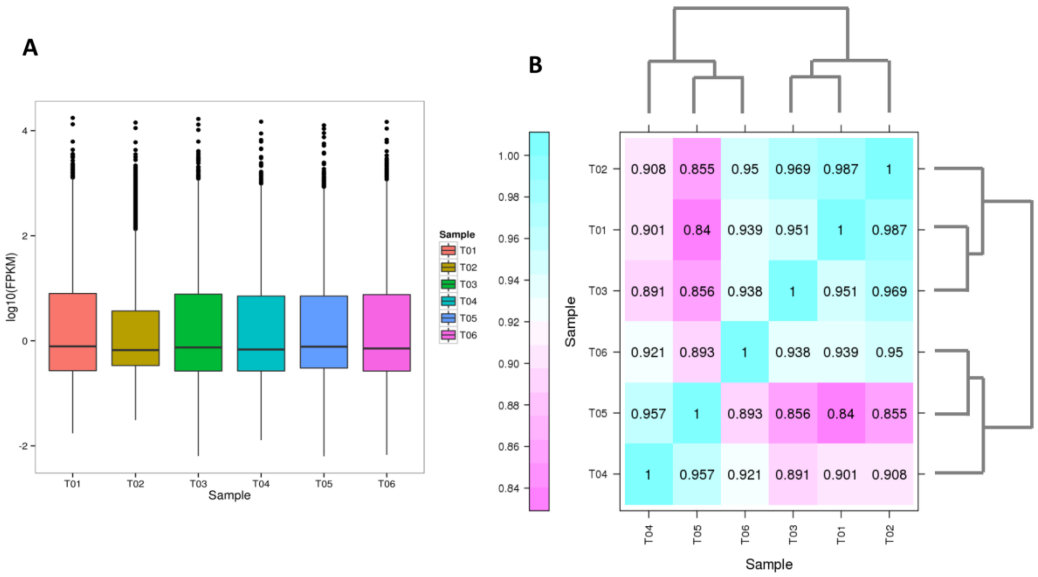


Figure 5. Overview of the transcriptome sequencing. (**A**) Gene expression profile in the 6 libraries. T01-T03 represent the three replicates libraries of the cultivar "Green Pittosporum" and T04-T06 represent the three replicates libraries of the cultivar "Variegatum". (**B**) Heatmap clustering showing correlation among *P. tobira* different samples based on global expression profiles. Numbers in the heatmap represent the Pearson correlation value.

Table 1. Overview of the transcriptome sequencing dataset and quality check.

Cultivar	Library-ID	Read Number	Base Number	GC (%)	% ≥ Q30
"Green Pittosporum"	T01	22,115,602	6,612,391,012	45.27	92.78
"Green Pittosporum"	T02	23,965,370	7,164,119,518	45.01	92.97
"Green Pittosporum"	T03	22,120,563	6,612,379,716	45.25	92.96
"Variegatum"	T04	23,413,612	6,996,233,542	44.79	93.60
"Variegatum"	T05	22,179,179	6,627,936,224	44.64	93.36
"Variegatum"	T06	22,956,424	6,863,508,574	45.19	92.86

Table 2. Statistics of the assembly results.

Length Range	Transcript	Unigene
200–300	53,860(25.64%)	46,279(41.00%)
300–500	38,333(18.25%)	27,916(24.73%)
500–1000	38,190(18.18%)	20,115(17.82%)
1000–2000	39,154(18.64%)	11,929(10.57%)
2000	40,551(19.30%)	6,636(5.88%)
Total number	210,088	112,875
Total length (bp)	241,105,749	72,533,944
N50 length (bp)	2,137	1,017
Mean length (bp)	1147.64	642.60

Table 3. Functional annotation statistics of the unigenes.

#Anno_Database	Annotated_Number	300 <= length < 1000	Length >= 1000
COG_Annotation	17,065	6,007	6,317
GO_Annotation	28,283	10,378	9,062
KEGG_Annotation	19,595	7,857	6,277
KOG_Annotation	30,496	11,470	10,018
Pfam_Annotation	33,826	12,283	13,569
Swissprot_Annotation	28,074	10,455	11,043
eggNOG_Annotation	48,410	18,186	15,546
nr_Annotation	47,309	17,751	15,596
All_Annotated	51,718	19,526	15,881

Table 4. Statistics of the mapping of sequencing data with assembly results.

Cultivar	Library-ID	Clean Reads	Mapped Reads	Mapped Ratio
"Green Pittosporum"	T01	22,115,602	17,518,471	79.21%
"Green Pittosporum"	T02	23,965,370	19,005,314	79.30%
"Green Pittosporum"	T03	22,120,563	17,599,921	79.56%
"Variegatum"	T04	23,413,612	19,153,207	81.80%
"Variegatum"	T05	22,179,179	17,804,804	80.28%
"Variegatum"	T06	22,956,424	18,290,888	79.68%

2.4. Differential Gene Expression Analysis between "Variegatum" and "Green Pittosporum" under Cold Condition

The differential gene expression analysis was performed on all expressed genes by comparing their expression levels between "Variegatum" and "Green Pittosporum". As shown in Figure 6A, a total of 309 differentially expressed genes (DEG) were obtained, including 156 up-regulated and 153 down-regulated genes in the variegated leaves. To validate our differential expressed gene result, we selected five up-regulated genes and five down-regulated genes (Table S3) and performed qRT-PCR using the cDNAs from leaves of the two cultivars. The qRT-PCR results were strongly correlated with the RNA-seq data (R^2 = 0.89, Figure S1). This result confirms well the high reliability of the RNA-seq data obtained in the present study.

We performed gene ontology (GO) enrichment analysis of these DEGs based on three ontologies: biological process, cellular component and molecular functions. In the biological process components, metabolic and cellular process was found to be the most dominant group (Figure 6B). Within the cellular components, cell and cell part represented the most dominant functional groups. Meanwhile, the catalytic activity and binding were the most abundant functional groups among the molecular functions, showing that enzymes and transcription factors encoding genes may play key roles in the differential cold response. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the DEGs showed that the biosynthesis of unsaturated fatty acids, sesquiterpenoid and triterpenoid biosynthesis, fatty acid metabolism, phenylalanine metabolism and protein processing in endoplasmic

reticulum were the main pathways contributed by the DEGs (Figure 6C). The diversity of these molecular pathways highlights the complex mechanism of the improved cold response in relation with leaf variegation in *P. tobira*.

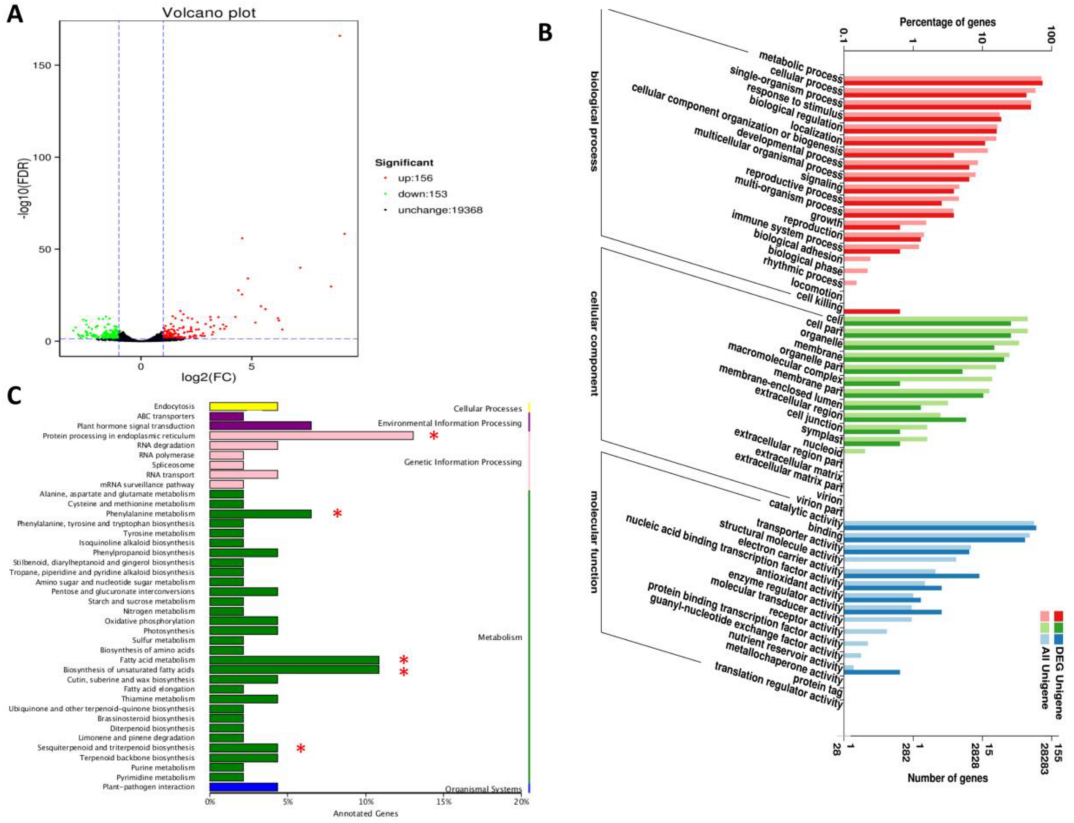


Figure 6. Differentially expressed genes (DEG) analysis between "Green Pittosporum" and "Variegatum". (**A**) Volcano plot depicting the up-, down- and no- regulated genes between the two cultivars. (**B**) Gene ontology enrichment analysis of the DEGs. (**C**) KEGG enrichment analysis of the DEGs. * represent the significantly enriched pathways.

2.5. Major Transcription Factors Differentially Regulated between "Variegatum" and "Green Pittosporum" under Cold Conditions

Since the GO enrichment showed that differential binding activity was important for the cold responses in "Variegatum", we extended our study over the major transcription families (TF) present within the DEGs. In total, 11 down-regulated and 14 up-regulated TFs in "Variegatum" were detected. Among the down-regulated TFs, AP2-ERF, bHLH and MADS-box TFs were enriched (Figure 7A). Distinct TF families were enriched in the up-regulated genes and included NAC, WRKY, HSF and MYB (Figure 7B). Expression fold change of these TFs showed that two NAC genes (*c67871.graph_c0* and *c63655.graph_c1*) were strikingly up-regulated in "Variegatum" (Figure 7C) and may play prominent positive regulatory roles for cold stress endurance.

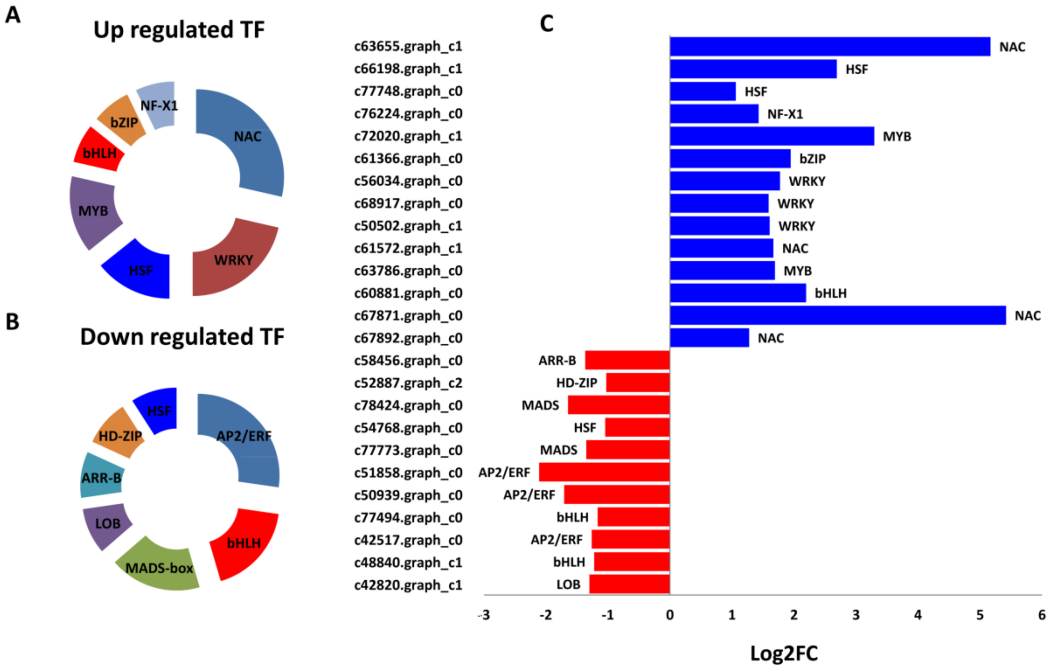


Figure 7. Major transcription factors (TF) families regulating cold response in *P. tobira*. (**A**) up-regulated TFs families in "Variegatum", (**B**) down-regulated TFs families in "Variegatum", (**C**) Log2 Fold change of the expression of TF genes.

2.6. DEGs Related to the Biosynthesis of Unsaturated Fatty Acids and Fatty Acids Metabolism

An increase in polyunsaturated fatty acids has also been reported to play a crucial role in the chilling tolerance of plants [33]. In this study, ten DEGs were mapped to the pathways related to the biosynthesis of unsaturated fatty acids and fatty acids metabolism. Interestingly, all of these DEGs were annotated as endoplasmic reticulum omega-6 fatty acid desaturase (FAD2) and were all up-regulated in "Variegatum" (Table 5). Since the microsomal enzyme FAD2 principally acts on the desaturation of C18:1 to C18:2 [34], we deduce that "Variegatum" strongly accumulates C18:2 in leaf as a protective molecule under cold conditions. Besides, we also detected nine GDSL-Lipase involved in lipid biosynthesis in plants. All the GDSL genes were down regulated in "Variegatum" (Table 5), suggesting a probable opposite function of GDSL and FAD2 genes during cold endurance in "Variegatum".

Table 5. Key DEGs related to the enriched KEGG pathways involved in the cold responses in variegated *P. tobira*.

Pathway	KO	Gene ID	Log2 Fold Change	Gene Description
Phenylalanine metabolism				
	K00815	*c55523.graph_c0*	1.565	Aminotransferase TAT2
	K00430	*c68309.graph_c0*	1.057	Peroxidase
	K00430	*c74970.graph_c1*	1.157	Peroxidase
Sesquiterpenoid and triterpenoid biosynthesis				
	K15472	*c29794.graph_c0*	1.488	Premnaspirodiene oxygenase, Cytochrome P450
	K15472	*c43399.graph_c0*	1.198	Premnaspirodiene oxygenase, Cytochrome P450
Biosynthesis of unsaturated fatty acids and fatty acid metabolism				
	K10256	*c45880.graph_c0*	1.973	FAD2
	K10256	*c72696.graph_c0*	1.287	FAD2
	K10256	*c74068.graph_c0*	1.493	FAD2
	K10256	*c74296.graph_c0*	1.385	FAD2
	K10256	*c75682.graph_c0*	1.284	FAD2
	K10256	*c45880.graph_c0*	1.973	FAD2
	K10256	*c72696.graph_c0*	1.287	FAD2

Table 5. *Cont.*

Pathway	KO	Gene ID	Log2 Fold Change	Gene Description
	K10256	*c74068.graph_c0*	1.493	FAD2
	K10256	*c74296.graph_c0*	1.385	FAD2
	K10256	*c75682.graph_c0*	1.284	FAD2
	––	*c52980.graph_c1*	−1.179	GDSL
	––	*c52915.graph_c0*	−1.582	GDSL
	––	*c62735.graph_c0*	−2.134	GDSL
	––	*c28930.graph_c0*	−1.998	GDSL
	––	*c79198.graph_c0*	−1.068	GDSL
	––	*c63752.graph_c0*	−1.502	GDSL
	––	*c28455.graph_c0*	−1.267	GDSL
	––	*c71358.graph_c0*	−1.635	GDSL
	––	*c76303.graph_c0*	−1.713	GDSL
		Protein processing in endoplasmic reticulum		
	K13993	*c55937.graph_c0*	−2.839	22.7kDa HSP IV
	K13993	*c60091.graph_c0*	−1.125	15 kDa HSP
	K13993	*c60491.graph_c0*	−1.961	17.9 kDa HSP II
	K13993	*c70233.graph_c1*	−1.292	18.1 kDa HSP I
	K04079	*c74041.graph_c0*	1.283	90 kDa HSP
	K03283	*c75081.graph_c0*	1.095	70 kDa HSP
	K09489	*c68688.graph_c0*	1.266	70 kDa HSP
	K03283	*c28616.graph_c1*	1.018	70 kDa HSP
	K03283	*c69810.graph_c6*	1.408	70 kDa HSP

2.7. Disturbance of Protein Processing in Endoplasmic Reticulum under Cold Conditions

The endoplasmic reticulum is a subcellular compartment where proteins and lipids are folded with the help of chaperones. The enrichment of this pathway (Figure 6C) indicates a disturbance of proteins and lipids synthesis under cold conditions. Nine DEGs, all being heat shock proteins (HSP) were detected within this pathway. Notably, we observed that all the small HSP genes (15–22 kDa) were down-regulated while the high molecular weight HSP genes (70–90 kDa) were up-regulated in the variegated leaves (Table 5). This result highlights the weight dependent roles of HSP genes for a stout cold response in variegated *P. tobira*.

2.8. DEGs in the Phenylalanine Metabolism

In this important pathway, we found three DEGs including two POD genes (*c68309.graph_c0* and *c74970.graph_c1*) and *c55523.graph_c0* annotated as an aminotransferase TAT2. Interestingly, all these genes were up-regulated in "Variegatum", showing that they contribute positively to the enhanced cold response (Table 5). More importantly, the activation of these genes correlates well with the strong enzymatic activity of POD detected through our biochemical assay in "Variegatum" when the temperature reached 10 °C (Figure 4A).

3. Discussion

3.1. Characteristics of Leaf Variegation in P. tobira

Leaf variegated plants have green/white (or yellow) sectors and cells in the green sectors contain normal appearing chloroplasts, while cells in the white sectors have impaired chloroplast biogenesis and lack photosynthetic pigments [12]. Moreover, it has been shown that leaf variegated plants accumulate high levels of ROS [29,30]. Although these mechanisms are commonly found in variegated plants, a recent study of the rice *z3* mutant leaves showed a new mechanism of variegation, which was caused by an unbalanced distribution of citrate in a transverse pattern in leaf tissues [34]. In our study, we also noted a defected chloroplast development in the yellow sector, reduced chlorophyll content and a high level of ROS in the variegated cultivar (Figures 2 and 3). We also observed an abundance of starch granules in the yellow sector as compared to the green sector, suggesting that the yellow sectors are nutrient sinks because they are unable to perform photosynthesis. Similar

conclusions were previously reported in different species including variegated *Arabidopsis* [1,35,36], tobacco [37], begonia [6] and fig [3]. However, the photosynthetic efficiency was not obviously affected in "Variegatum" (Figure 3), contrasting with the reports that leaf variegation affects photosynthetic efficiency [12]. In fact, the yellowish area on "Variegatum" leaves is located on the margin and has a very low surface coverage. So, an explanation to this observation can be that the green part of the leaf is large enough to ensure the photosynthetic activity. Leaf variegation has been attributed to a deficiency or a significant reduction of photosynthetic pigments including carotenoids. In the *Arabidopsis* white-green variegated mutant *immutans* (*im*), an inhibition of carotenoids formation was observed [38]. Similarly, the white section in leaf of variegated *Epipremnum aureum* contains 10-fold less carotenoids than the green section [39]. In *Cyclamen purpurascens*, the light green leaf stripes were found with reduced carotenoids and chlorophyll contents [40]. In green/yellow patterns variegated species, similar observations were also noticed in *Aucuba japonica* [41] and *Coleus bluemei* [42]. Intriguingly, we observed a higher concentration of carotenoids in the variegated leaves of *P. tobira* as compared to the complete green leaves (Figure 3), a phenomenon which has not yet been reported in variegated plants. Since carotenoids function as accessory light-harvesting pigments, broadening the spectral range over which light can support photosynthesis in plants [43], we deduce that the high carotenoids content in "Variegatum" may compensate the reduced chlorophyll to maintain similar photosynthetic activity as in leaves of "Green Pittosporum".

3.2. Protective Role of Leaf Variegation in P. tobira under Cold Condition

The natural occurrence of variegation in plants suggests that the trait might play some adaptive functions beyond their aesthetic value [32]. It has been suggested that leaf variegation plays several physiological and ecological functions such as defense from enemies, adaptations to light, temperature, etc. [16–25]. We tested the hypothesis that leaf variegation plays a low temperature protective function in *P. tobira*, which is an ornamental shrub widely grown in temperate climate and therefore is annually subjected to cold stress. It is well known that increased activities of antioxidant enzymes such as POD, CAT, SOD under abiotic stress conditions including drought, salt, chilling, heat, etc., promote enhanced stress tolerance in plants [44]. Our results demonstrated that "Variegatum" has much more efficient ROS-scavenging machinery compared to "Green Pittosporum" and accumulates less MDA, an indicator of limited cellular membrane damage due to lipid peroxidation. Hence, "Variegatum" better tolerates low temperature stress than "Green Pittosporum" (Figure 4). We further sequenced the transcriptomes of both leaf types under cold condition (10 °C). Differential gene expression (DEG) analysis resulted in 309 DEGs between the two cultivars, enriched in biological pathways related to the biosynthesis of unsaturated fatty acids, sesquiterpenoid and triterpenoid biosynthesis, fatty acid metabolism, phenylalanine metabolism and protein processing in endoplasmic reticulum, which may be crucial pathways involved in cold stress alleviation (Figure 6).

Cell membrane structure, integrity and fluidity are affected by lipid composition and the degree of fatty acid (FA) desaturation in plants [45]. It has been documented that changes in unsaturated fatty acids content can improve plant tolerance to environmental stresses such as cold, heat and drought [46–51], since modification of membrane fluidity results in an environment suitable for the function of critical integral proteins, such as the photosynthetic machinery, during stresses [52]. In this study, we detected ten FAD2 genes all significantly up-regulated in "Variegatum" leaves under cold condition (Table 5). The microsomal enzyme FAD2 principally acts on the desaturation of C18:1 (monounsaturated FA) to C18:2 (polyunsaturated FA) [53], suggesting that "Variegatum" tends to increase polyunsaturated FA (PUFA) level, a mechanism to maintain cell membrane fluidity under low temperature [54,55]. This skill of adjusting membrane fluidity by varying the unsaturated fatty acid content is characteristic of cold-responsive plants [52]. Cold acclimating potato (*Solanum commersonii*) was found to accumulate linoleic acid (18:2) in the membrane glycerolipids of the leaves, whereas commercial, non-acclimating potato (*Solanum tuberosum*) did not show this trait during cold stress [56]. Our findings are in perfect accordance with reports of Liu et al. [51], who showed that over-expression of

tomato FAD2 gene alleviates the photoinhibition of photosystems 2 and 1 and improves tolerance under chilling stress. Similar observations were reported in various plants such as cotton [57], *A. thaliana* [58], *Olea europaea* [59], *Synechocystis* sp. [60], etc., under low temperature conditions.

Membrane fatty acid composition is, to a great extent, determined by the activities of complexly regulated integral fatty acid desaturases and lipases [52]. GDSL-lipase participates in fatty acid catabolism and studies have shown that the linoleic acid and other PUFAs contents are significantly decreased when GDSL genes are over-expressed [61–64]. Here, we detected nine GDSL-lipase genes all down-regulated in "Variegatum" under low temperature stress (Table 5), denoting a strategy to keep the high level of PUFA for the maintenance of cell membrane stability. We deduce that down-regulation of GDSL genes and up-regulation of FAD2 genes is therefore an integrated and efficient mechanism to cope with cold stress in *P. tobira* cv. "Variegatum".

Another group of genes detected within the DEGs between "Variegatum" and "Green Pittosporum" under cold condition are heat shock proteins (HSP) (Table 5). HSPs are molecular chaperones that are constantly present in cells to correctly fold proteins involved in routine cellular processes such as translocation, cell-signaling and metabolism [65]. However, HSPs become abundant in most organisms in response to protein denaturation caused by environmental, metabolic and pathological stresses [66]. For example, *Arabidopsis*, grape, rice, *Brassicas* increase the production of HSPs to augment survival in cold environments [67–70]. On the other hand, it was reported that a complex coordination of HSPs underlines cold tolerance in plants [65]. In fact, some HSPs are either up- or down-regulated when heat shock factors (HSFs) bind to their promoter regions [71,72]. This suggests that not all HSPs positively participate in cold or stress tolerance in plants. Each group of these HSPs has a unique mechanism [65]. In our study, we observed that small HSPs were all down-regulated while high molecular weight HSP genes were up-regulated in "Variegatum", pointing out an opposite function of HSPs for cold response in *P. tobira* with respect to their molecular weights. For now, a clear explanation for this phenomenon is yet to be found, hence, an in-depth investigation of the role of HSP genes and their relation with the significantly altered HSF transcription factors under cold condition in *P. tobira* is necessary in order to clarify this intriguing finding.

Our transcriptome analysis also unveiled several peroxidase genes from the phenyalanine pathway as well as some cytochrome *P450* genes from the sesquiterpenoid and triterpenoid biosynthesis as candidate genes, which positively contribute to the enhanced cold responses in "Variegatum" (Table 5). Peroxidase genes have been extensively studied in plants for their ROS-scavenging activity under various biotic and abiotic stresses, including chilling [73–75]. Similarly, Liu et al. [76] recently investigated the prominent biological pathways engaged in wild banana tolerance to chilling. They observed significant changes in the sesquiterpenoid and triterpenoid biosynthesis, particularly cytochrome *P450* genes, a finding that supports well the results of our study.

Taken together, we showed that leaf variegation in *P. tobira* is associated to defected chloroplast development, reduced chlorophyll content, high content of carotenoids and a high level of ROS. The results of transcriptome analysis were consistent with the enzymatic activity under cold conditions. These results pointed out that the leaf variegation trait plays low temperature protective effect in *P. tobira* by inducing a strong ROS-scavenging activity through catalase and peroxidase enzymes, inducing heat shock proteins for cellular homeostasis and, more importantly, by maintaining high levels of PUFA for cell membrane stability and fluidity through a coordinated up-regulation of FAD2 and down-regulation of GDSL-lipase genes. The modulation of the expression levels of these key genes may be orchestrated by transcription factors from the families of NAC, WRKY, HSF and AP2/ERF. A proposed schematic model for the stronger cold response in "Variegatum" is summarized in Figure 8.

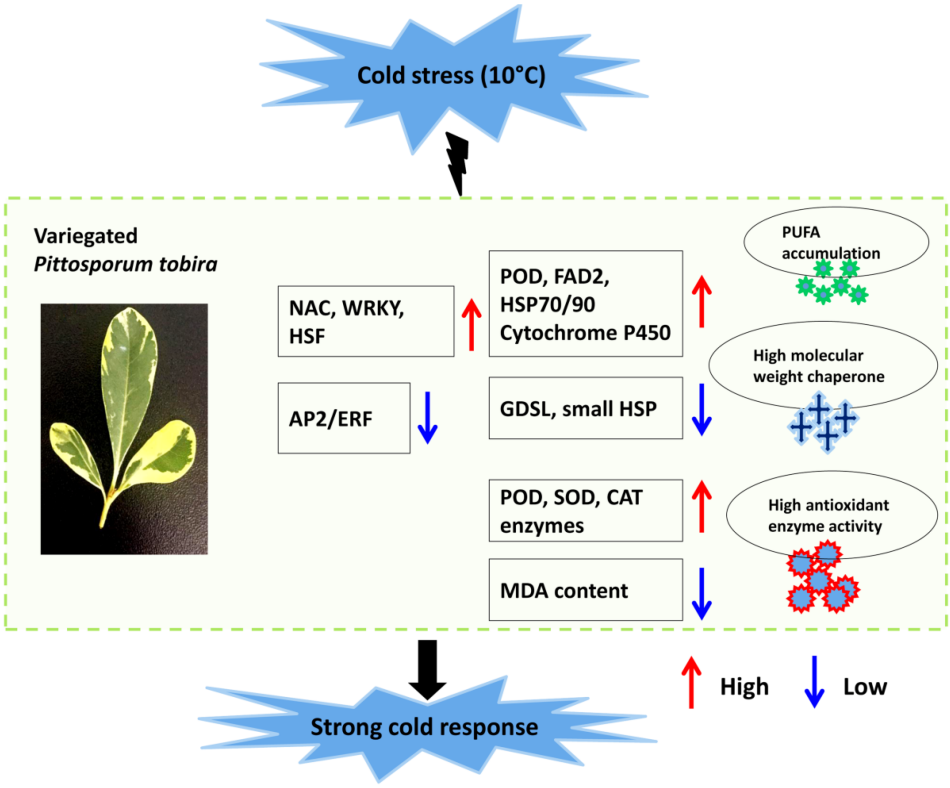


Figure 8. A schematic model of the proposed mechanism underlying the strong cold response in *P. tobira* cv. "Variegatum".

4. Materials and Methods

4.1. Plant Materials

The naturally occurring variegated cultivar of *Pittosporum tobira* "Variegatum" and the widely grown cultivar "Green Pittosporum" were originally collected from Pingdingshan, Henan in China and used as experimental materials. Biochemical data were recorded on three different plants of each cultivar at different dates corresponding to various temperature gradients (20–10 °C) from August to November (Figure 1).

4.2. Transmission Electron Microscopy (TEM)

TEM analysis was performed as described by Shih et al. [3]. Green and yellow sectors of leaves were cut into small cubes in the field and placed in a fixation solution containing 2.5% glutaraldehyde and 4% paraformaldehyde in 0.1 M sodium phosphate buffer (pH 7.0). Samples underwent 20 min of rinsing three times and were post-fixed in 1% osmium tetroxide for 2 h. After being dehydrated through an ethanol series, samples were infiltrated and embedded in Spurr's resin and then polymerized at 70 °C for 8 h. Ultrathin sections (~70–90 nm) were collected and stained with ethanol uranyl acetate and lead citrate. The morphology of plastids was observed with Tecnai F20S TEM (The Thermo Scientific™, Waltham, MA, USA) at 200 kV.

4.3. Measurement of Physio-Biochemical Parameters

A total of 50 mg fresh leaves were used to extract chlorophyll. The total chlorophyll content (ChlT, mg g^{-1}FW) and carotenoids content (Ca, mg g^{-1} FW) were determined as described by Wellburn [77]. The net photosynthetic rate (*Pn*, μmol m^{-2} s^{-1}), intercellular CO_2 concentration (*Ci*/ppm) and transpiration rate (*Tr*, mmol.m^{-2} s^{-1}) were determined with a portable L-6400XT (LI-COR, Lincoln, NB, USA). The measurements of photosynthetic parameters were taken at the

saturation irradiance with an incident photosynthetic photo flux density (PPFD) of 1200 µmo m^{-2} s^{-1} and an airflow rate at 500 µmol s^{-1}. The enzymatic activities of superoxide dismutase (SOD, U g^{-1}), catalase (CAT, U g^{-1}.min^{-1}), peroxidase (POD, U g^{-1}.min^{-1}) and the content of malonaldehyde (MDA, µmol g^{-1}) were calculated by following the manufacturer's instructions (Biological Engineering Institute of Nanjing Jiancheng, China). Means from three replicates were used for statistical analysis.

4.4. RNA Extraction, cDNA Library Construction, and Transcriptome Sequencing

The complete leaves from the cultivars "Green Pittosporum" and "Variegatum" were collected in replicates from three different plants under cold conditions at November 15th (Temperature = 10 °C), immediately frozen in liquid nitrogen and stored at −80 °C until further use. Total RNAs were extracted using Spin Column Plant total RNA Purification Kit following the manufacturer's protocol (Sangon Biotech, Shanghai, China). Purity of the extracted RNAs was assessed on 1% agarose gels followed by NanoPhotometer spectrophotometer (IMPLEN, Los Angeles, CA, USA). We quantified the RNA using Qubit RNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA, USA). RNA integrity was checked using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA).

Libraries preparation, and sequencing on Illumina HiSeq 4000 platform (Illumina Inc., San Diego, CA, USA) were performed as described by Zhuang et al. [78].

4.5. De novo Assembly, Functional Annotation, Classification and Metabolic Pathway Analysis

Raw transcriptome data were submitted to NCBI SRA, freely accessible at www.ncbi.nlm.nih.gov/bioproject/PRJNA553027. The clean reads were retrieved after trimming adapter sequences, removal of low quality (containing > 50% bases with a Phred quality score < 15) and reads with unknown nucleotides (more than 1% ambiguous residues N) using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The high-quality reads from all the six libraries were de novo assembled into transcripts using Trinity (Version r20140717) [79] by employing paired-end method. Next, the transcripts were realigned to construct unigenes. The assembled unigenes were then annotated by searching against various databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [80], Gene Ontology (GO) [81], Clusters of Orthologous Groups (COG) [82], Pfam [83], Swissprot [84], egNOG [85], NR [86], euKaryotic Orthologous Groups (KOG) [87] using BLAST [88] with a threshold of E-value <1.0 E^{-5}.

The software KOBAS2.0 [89] was employed to get the unigene KEGG orthology; the analogs of the unigene amino acid sequences were searched against the Pfam database [83] using HMMER tool [90] with a threshold of E-value < 1.0 E^{-10}. The sequenced reads were compared with the unigene library using Bowtie [91], and the level of expression was estimated in combination with RSEM [92]. The gene expression level was determined according to the fragments per kilobase of exon per million fragments mapped (FPKM).

4.6. Differential Expression and Enrichment Analysis

The read count was normalized and EdgeR Bioconductor package [93] was used to determine the differential expressed genes (DEGs) between the two cultivars with the fold change of > 2 [94] and false discovery rate correction (FDR) set at $p < 0.01$. GO enrichment analysis was performed using the topGO method [95] based on the wallenius non-central hypergeometric distribution with $p < 0.05$. KEGG pathway enrichment analysis of the DEGs was done using KOBAS2.0 [89]. The FDR correction was employed ($p < 0.05$) to reduce false positive prediction of enriched KEGG pathways.

4.7. Validation of Gene Expression Using Quantitative Real Time-PCR

The qRT-PCR was performed on RNA extracted from leaf samples of "Variegatum" and "Green Pittosporum" as described by Dossa et al. [96] using the *Actin* gene as the internal control. Specific

primer pairs of ten selected genes were designed using the Primer Premier 5.0 [97] (Table S3). Data are presented as relative transcript level based on the $2^{-\Delta\Delta Ct}$ method [98].

4.8. Statistical Analysis

Data were analyzed with the R software (www.r-project.org) using the one-way analysis of variance (ANOVA) for significant difference. The error bars were calculated with data from three replicates. ANOVA results were considered significant at $p < 0.05$ and mean comparisons were done using the Tukey HSD test.

Supplementary Materials: Supplementary Materials can be found at http://www.mdpi.com/1422-0067/20/19/4857/s1. Table S1. Full list of the unigenes annotated in *Pittosporum tobira* leaf; Table S2. Full list of the genes expressed in *Pittosporum tobira* leaf and their FPKM values. T01-T03 represent the three replicates libraries of the cultivar "Green Pittosporum" and T04-T06 represent the three replicates libraries of the cultivar "Variegatum"; Table S3. The primer sequences of genes used for real time quantitative PCR; Figure S1. qRT-PCR results of 10 selected genes and correlation between transcriptome data and real time PCR results.

Author Contributions: Z.Z. project design and paper writing; Z.L. Data processing and analysis; H.S. experimental test and data analysis; M.C. experimental test and data collection; S.C. project design and management. All authors have read and approved the final version of this manuscript.

Funding: This research was funded by the Project of Henan Science and Technology Research Program (Agricultural Field), grant number: 162120110070.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CAT	Catalase
DEG	Differentially expressed gene
FAD2	Fatty acid desaturase 2
FPKM	Fragments per kilobase of exon per million fragments mapped
GO	Gene ontology
HSF	Heat shock factor
HSP	Heat shock protein
KEGG	Kyoto Encyclopedia of Genes and Genomes
MDA	Malonaldehyde
POD	Peroxidase
PUFA	Poly-unsaturated fatty acids
ROS	Reactive oxygen species
SOD	Superoxide dismutase
TF	Transcription factor

References

1. Aluru, M.R.; Bae, H.; Wu, D.; Rodermel, S.R. The Arabidopsis *immutans* mutation affects plastid differentiation and the morphogenesis of white and green sectors in variegated plants. *Plant Physiol.* **2001**, *127*, 67–77. [CrossRef] [PubMed]
2. Kato, Y.; Kouso, T.; Sakamoto, W. Variegated tobacco leaves generated by chloroplast *FtsH* suppression: Implication of *FtsH* function in the maintenance of thylakoid membranes. *Plant Cell Physiol.* **2012**, *53*, 391–404. [CrossRef] [PubMed]
3. Shih, T.-H.; Lin, S.-H.; Huang, M.-Y.; Huang, W.-D.; Yang, C.-M. Transcriptome profile of the variegated *Ficus microcarpa* c.v. milkys fig leaf. *Int. J. Mol. Sci.* **2019**, *20*, 1338. [CrossRef] [PubMed]
4. Tsai, C.C.; Wu, Y.J.; Sheue, C.R.; Liao, P.C.; Chen, Y.H.; Li, S.J.; Liu, J.W.; Chang, H.T.; Liu, W.L.; Ko, Y.Z.; et al. Molecular basis underlying leaf variegation of a moth orchid mutant (*Phalaenopsis aphrodite* subsp. *formosana*). *Front. Plant Sci.* **2017**, *8*, 1333. [CrossRef] [PubMed]
5. Hara, N. Study of the variegated leaves with special reference to those caused by air spaces. *Jpn. J. Bot.* **1957**, *16*, 86–101.

6. Sheue, C.R.; Pao, S.H.; Chien, L.F.; Chesson, P.; Peng, C.I. Natural foliar variegation without costs? The case of Begonia. *Ann. Bot.* **2012**, *109*, 1065–1074. [CrossRef] [PubMed]
7. Tsukaya, H.; Okada, H.; Mohamed, M. A novel feature of structural variegation in leaves of the tropical plant *Schismatoglottis calyptrata*. *J. Plant Res.* **2004**, *117*, 477–480. [CrossRef]
8. Tilney-Bassett, R.A.E. Genetics of variegated plants. In *Genetics and Biogenesis of Mitochondria and Chloroplasts*; Birky, C.W., Perlman, P.S., Byers, T.J., Eds.; Ohio State University Press: Columbus, OH, USA, 1975; pp. 268–308.
9. Beardsell, D.; Norden, U. Ficus rubiginosa 'variegata', a chlorophyll-deficient chimera with mosaic patterns created by cell divisions from the outer meristematic layer. *Ann. Bot.* **2004**, *94*, 51–58. [CrossRef]
10. Yu, F.; Fu, A.; Aluru, M.; Park, S.; Xu, Y.; Liu, H.; Liu, X.; Foudree, A.; Nambogga, M.; Rodermel, S. Variegation mutants and mechanisms of chloroplast biogenesis. *Plant Cell Environ.* **2007**, *30*, 350–365. [CrossRef]
11. Li, X.; Kanakala, S.; He, Y.; Zhong, X.; Yu, S.; Li, R.; Sun, L.; Ma, J. Physiological characterization and comparative transcriptome analysis of white and green leaves of *Ananas comosus* var. bracteatus. *PLoS ONE* **2017**, *12*, e0169838. [CrossRef]
12. Lysenko, V. Fluorescence kinetic parameters and cyclic electron transport in guard cell chloroplasts of chlorophyll-deficient leaf tissues from variegated weeping fig (*Ficus benjamina* L.). *Planta* **2012**, *235*, 1023–1033. [CrossRef] [PubMed]
13. Aluru, M.R.; Zola, J.; Foudree, A.; Rodermel, S.R. Chloroplast photooxidation-induced transcriptome reprogramming in Arabidopsis *immutans* white leaf sectors. *Plant Physiol.* **2009**, *150*, 904–923. [CrossRef] [PubMed]
14. Miura, E.; Kato, Y.; Sakamoto, W. Comparative transcriptome analysis of green/white variegated sectors in Arabidopsis yellow variegated2: Responses to oxidative and other stresses in white sectors. *J. Exp. Bot.* **2010**, *61*, 2433–2445. [CrossRef] [PubMed]
15. Yang, Y.; Chen, X.; Xu, B.; Li, Y.; Ma, Y.; Wang, G. Phenotype and transcriptome analysis reveals chloroplast development and pigment biosynthesis together influenced the leaf color formation in mutants of Anthurium and raeanum 'Sonate'. *Front. Plant Sci.* **2015**, *6*, 139. [CrossRef] [PubMed]
16. Lev-Yadun, S.; Dafni, A.; Flaishman, M.A.; Inbar, M.; Izhaki, I.; Katzir, G.; Ne'eman, G. Plant coloration undermines herbivorous insect camouflage. *BioEssays* **2004**, *26*, 1126–1130. [CrossRef] [PubMed]
17. Lev-Yadun, S. Defensive (anti-herbivory) coloration in land plants. In *Anti- Herbivory Plant Coloration and Morphology*; Springer: Zug, Switzerland, 2016.
18. Lev-Yadun, S. Local loss of the zebra-like coloration supports the aposematic and other visual defense hypotheses in *Silybum marianum*. *Isr. J. Plant Sci.* **2017**, *64*, 170–178. [CrossRef]
19. Niu, Y.; Sun, H.; Stevens, M. Plant camouflage: Ecology, evolution, and implications. *Trends Ecol. Evol.* **2018**, *33*, 608–618. [CrossRef]
20. Fooshee, W.C.; Henny, R.J. Chlorophyll levels and anatomy of variegated and non variegated areas of *Aglaonema nitidum* leaves. *Proc. Fla. State Hortic. Soc.* **1990**, *103*, 170–172.
21. Roelfsema, M.; Konrad, K.R.; Marten, H.; Psaras, G.K.; Hartung, W.; Hedrich, R. Guard cells in albino leaf patches do not respond to photosynthetically active radiation, but are sensitive to blue light, CO_2 and abscisic acid. *Plant Cell Environ.* **2006**, *29*, 1595–1605. [CrossRef]
22. Shelef, O.; Summerfield, L.; Lev-Yadun, S.; Villamarin-Cortez, S.; Sadeh, R.; Herrmann, I.; Rachmilevitch, S. Thermal benefits from white variegation of *Silybum marianum* Leaves. *Front. Plant Sci.* **2019**, *10*, 688. [CrossRef]
23. Smith, A.P. Ecology of a leaf color polymorphism in a tropical forest species: Habitat segregation and herbivory. *Oecologia* **1986**, *69*, 283–287. [CrossRef] [PubMed]
24. Soltau, U.; Dötterl, S.; Liede-Schumann, S. Leaf variegation in *Caladium steudneriifolium* (Araceae): A case of mimicry? *Evol. Ecol.* **2009**, *23*, 503–512. [CrossRef]
25. Mwafongo, E.; Vollsnesb, A.V.; Bjorå, C.S.; Nordal, I.; Eriksen, A.B. Leaf mottling/variegation and shape in the *Ledebouria revoluta* complex—Development, stability and putative function. *Flora* **2017**, *236*, 33–43. [CrossRef]
26. Dib, R.A.E.; Eskander, J.; Mohamed, M.A.; Mohammed, N.M. Two new triterpenoid estersaponins and biological activities of *Pittosporum tobira*, 'variegata' (Thunb.) WT. Aiton leaves. *Fitoterapia* **2015**, *106*, 272–279. [CrossRef]

27. Burt, S. Essential oils: Their antibacterial properties and potential applications in foods-a review. *Int. J. Food Microbiol.* **2004**, *94*, 223–253. [CrossRef]
28. Eunsil, K.O.; Choi, M.R.; Choi, K.M.; Cha, J.D. The effect of *Pittosporum tobira* against anti-Helicobacter pylori and anti-oxidant activity. *IJVR* **2014**, *8*, 4.
29. Jiang, H.; Chen, Y.; Li, M.; Xu, X.; Wu, G. Overexpression of *SGR* results in oxidative stress and lesion-mimic cell death in rice seedlings. *J. Integr. Plant Biol.* **2011**, *53*, 375–387. [CrossRef]
30. Han, S.H.; Sakuraba, Y.; Koh, H.J.; Paek, N.C. Leaf variegation in the rice zebra2 mutant is caused by photoperiodic accumulation of tetra-Cis-lycopene and singlet oxygen. *Mol. Cells* **2012**, *33*, 87–97. [CrossRef]
31. Sheoran, S.; Thakur, V.; Narwal, S.; Turan, R.; Mamrutha, H.M.; Singh, V.; Tiwari, V.; Sharma, I. Differential activity and expression profile of antioxidant enzymes and physiological changes in wheat (*Triticum aestivum* L.) under drought. *Appl. Biochem. Biotechnol.* **2015**, *177*, 1282–1298. [CrossRef]
32. Esteban, R.; Fernandez-Marin, B.; Becerril, J.M.; Garcia-Plazaola, J.I. Photoprotective implications of leaf variegation in *E. denscanis* L. and *P. officinalis* L. *J. Plant Physiol.* **2007**, *165*, 1255–1263. [CrossRef]
33. Sui, N.; Li, M.; Zhao, S.J.; Li, F.; Liang, H.; Meng, Q.W. Overexpression of glycerol-3-phosphate acyltransferase gene improves chilling tolerance in tomato. *Planta* **2007**, *226*, 1097–1108. [CrossRef] [PubMed]
34. Kim, S.-H.; Kwon, C.-T.; Song, G.; Koh, H.-G.; An, G.; Paek, N.-C. The rice *zebra3* (*z3*) mutation disrupts citrate distribution and produces transverse dark-green/green variegation in mature leaves. *Rice* **2018**, *11*, 1. [CrossRef] [PubMed]
35. Wetzel, C.M.; Jiang, C.Z.; Meehan, L.J.; Voytas, D.F.; Rodermel, S.R. Nuclear-organelle interactions: The immutans variegation mutant of Arabidopsis is plastid autonomous and impaired in carotenoid biosynthesis. *Plant J.* **1994**, *6*, 161–175. [CrossRef] [PubMed]
36. Miura, E.; Kato, Y.; Matsushima, R.; Albrecht, V.; Laalami, S.; Sakamoto, W. The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in Arabidopsis yellow variegated mutants. *Plant Cell* **2007**, *19*, 1313–1328. [CrossRef] [PubMed]
37. Bae, C.H.; Abe, T.; Matsuyama, T.; Fukunishi, N.; Nagata, N.; Nakano, T.; Kaneko, Y.; Miyoshi, K.; Matsushima, H.; Yoshida, S. Regulation of chloroplast gene expression is affected in ali, a novel tobacco albino mutant. *Ann. Bot.* **2001**, *88*, 545–553. [CrossRef]
38. Aluru, M.R.; Yu, F.; Fu, A.; Rodermel, S. Arabidopsis variegation mutants: New insights into chloroplast biogenesis. *J. Exp. Bot.* **2006**, *57*, 1871–1881. [CrossRef] [PubMed]
39. Sun, Y.-H.; Hung, C.-Y.; Qiu, J.; Chen, J.; Kittur, F.S.; Oldham, C.E.; Henny, R.J.; Burkey, K.O.; Fan, L.; Xie, J. Accumulation of high OPDA level correlates with reduced ROS and elevated GSH benefiting white cell survival in variegated leaves. *Sci. Rep.* **2017**, *7*, 44158. [CrossRef]
40. Klančnik, K.; Levpušček, M.; Gaberščik, A. Variegation and red abaxial epidermis define the leaf optical properties of Cyclamen purpurascens. *Flora* **2016**, *224*, 87–95. [CrossRef]
41. Zhang, Q.; Zhang, M.; Ding, Y.; Zhou, P.; Fang, Y. Composition of photosynthetic pigments and photosynthetic characteristics in green and yellow sectors of the variegated *Aucuba japonica* Variegata leaves. *Flora* **2018**, *240*, 25–33. [CrossRef]
42. Borek, M.; Baczek-Kwinta, R.; Rapacz, M. Photosynthetic activity of variegated leaves of Coleus x hybridus hort. cultivars characterised by chlorophyll fluorescence techniques. *Photosynthetica* **2016**, *54*, 331–339. [CrossRef]
43. Hashimoto, H.; Uragami, C.; Cogdell, R.J. Carotenoids and Photosynthesis. In *Carotenoids in Nature*; Subcellular Biochemistry; Stange, C., Ed.; Springer: Cham, Switzerland, 2016; Volume 79. [CrossRef]
44. Farooq, M.; Hussain, M.; Wahid, A.; Siddique, K.H.M. Drought stress in plants: An overview. In *Plant Responses to Drought Stress: From Morphological to Molecular Features*; Aroca, R., Ed.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–33.
45. Mikami, K.; Murata, N. Membrane fluidity and the perception of environmental signals in cyanobacteria and plants. *Prog. Lipid Res.* **2003**, *42*, 527–543. [CrossRef]
46. Dakhma, W.S.; Zarrouk, M.; Cherif, A. Effects of drought stress on lipids in rape leaves. *Phytochemistry* **1995**, *40*, 1383–1386. [CrossRef]
47. Olsson, M. Alteration in lipid-composition, lipid-peroxidation and antioxidative protection during senescence in drought stressed plants of *Pisum sativum*. *Plant Physiol. Biochem.* **1995**, *33*, 547–553.

48. Matos, M.C.; Campos, P.S.; Ramalho, J.C.; Medeira, M.C.; Maia, M.I.; Semedo, J.M.; Marques, N.M.; Matos, A. Photosynthetic activity and cellular integrity of the Andean legume *Pachyrhizus ahipa* (Wedd.) Parodi under heat and water stress. *Photosynthetica* **2002**, *40*, 493–501. [CrossRef]
49. Sui, N.; Li, M.; Shu, D.F.; Zhao, S.J.; Meng, Q.W. Antisense mediated depletion of tomato chloroplast glycerol-3- phosphate acyltransferase affects male fertility and increases thermal tolerance. *Physiol. Plant.* **2007**, *130*, 301–314. [CrossRef]
50. Sui, N.; Li, M.; Li, K.; Song, J.; Wang, B.-S. Increase in unsaturated fatty acids in membrane lipids of *Suaeda salsa* L. enhances protection of photosystem II under high salinity. *Photosynthetica* **2010**, *48*, 623–629. [CrossRef]
51. Liu, X.-Y.; Li, B.; Yang, J.-H.; Sui, N.; Yang, X.-M.; Meng, Q.-W. Overexpression of tomato chloroplast omega-3 fatty acid desaturase gene alleviates the photoinhibition of photosystems 2 and 1 under chilling stress. *Photosynthetica* **2008**, *46*, 185–192. [CrossRef]
52. Upchurch, R.G. Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol. Lett.* **2008**, *30*, 967–977. [CrossRef] [PubMed]
53. Dar, A.; Choudhury, A.R.; Kancharla, P.K.; Arumugam, N. The FAD2 gene in plants: Occurrence, regulation, and role. *Front. Plant Sci.* **2017**, *8*, 1789. [CrossRef] [PubMed]
54. Los, D.A.; Murata, N. Structure and expression of fatty acid desaturases. *Biochem. Biophys. Acta* **1998**, *1394*, 3–15. [CrossRef]
55. Routaboul, J.M.; Fischer, S.F.; Browse, J. Trienoic fatty acids are required to maintain chloroplast function at low temperatures. *Plant Physiol.* **2000**, *124*, 1697–1705. [CrossRef] [PubMed]
56. Vega, S.E.; del Rio, A.H.; Bamberg, J.B.; Palta, J.P. Evidence for the up-regulation of stearoyl-ACP (D9) desaturase gene expression during cold acclimation. *Am. J. Potato Res.* **2004**, *81*, 125–135. [CrossRef]
57. Kargiotidou, A.; Deli, D.; Galanopoulou, D.; Tsaftaris, A.; Farmaki, T. Low temperature and light regulate delta 12 fatty acid desaturases (FAD2) at a transcriptional level in cotton (*Gossypium hirsutum*). *J. Exp. Bot.* **2008**, *59*, 2043–2056. [CrossRef] [PubMed]
58. Kreps, J.A.; Wu, Y.; Chang, H.S.; Zhu, T.; Wang, X.; Harper, J.F. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol.* **2002**, *130*, 129–2141. [CrossRef] [PubMed]
59. Matteucci, M.; D'Angeli, S.; Errico, S.; Lamanna, R.; Perrotta, G.; Altamura, M.M. Cold affects the transcription of fatty acid desaturases and oil quality in the fruit of *Olea europaea* L. genotypes with different cold hardiness. *J. Exp. Bot.* **2011**, *62*, 3403–3420. [CrossRef] [PubMed]
60. Los, D.A.; Ray, M.K.; Murata, N. Differences in the control of the temperature-dependent expression of four genes for desaturases in *Synechocystis* sp. PCC 6803. *Mol. Microbiol.* **1997**, *25*, 1167–1175. [CrossRef] [PubMed]
61. Huang, L.M.; Lai, C.P.; Chen LF, O.; Chan, M.T.; Shaw, J.F. Arabidopsis SFAR4 is a novel GDSL-type esterase involved in fatty acid degradation and glucose tolerance. *Botani Stud.* **2015**, *56*, 33. [CrossRef]
62. Xu, H.M.; Kong, X.D.; Chen, F.; Huang, J.X.; Lou, X.Y.; Zhao, J.Y. Transcriptome analysis of *Brassica napus* pod using RNA-Seq and identification of lipid-related candidate genes. *BMC Genom.* **2015**, *16*, 858. [CrossRef]
63. Duan, S.; Jin, C.; Li, D.; Gao, C.; Qi, S.; Liu, K.; Hai, J.; Ma, H.; Chen, M. MYB76 inhibits seed fatty acid accumulation in Arabidopsis. *Front. Plant Sci.* **2017**, *8*, 226. [CrossRef]
64. Zhang, Y.; Bai, B.; Lee, M.; Alfiko, Y.; Suwanto, A.; Yue, G.H. Cloning and characterization of *EgGDSL*, a gene associated with oil content in oil palm. *Sci. Rep.* **2018**, *8*, 11406. [CrossRef]
65. Al-Whaibi, M.H. Plant heat-shock proteins: A mini review. *J. King Saud Univ. Sci.* **2011**, *23*, 139–150. [CrossRef]
66. Vierling, E. The Roles of Heat Shock Proteins in Plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **1991**, *42*, 579–620. [CrossRef]
67. Krishna, P.; Sacco, M.; Cherutti, J.F.; Hill, S. Cold induced accumulation of Hsp90 transcripts in *Brassica napus*. *Plant Physiol.* **1995**, *107*, 915–923. [CrossRef] [PubMed]
68. Pareek, A.; Singla, S.L.; Grover, A. Plant Hsp90 family with special reference to rice. *J. Biosci.* **1998**, *23*, 361–367. [CrossRef]
69. Kilian, J.; Whitehead, D.; Horak, J.; Wanke, D.; Weinl, S.; Batistic, O.; Bauer, E.B.; D'Angelo; Kudla, J.; Harter, K. The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* **2007**, *50*, 347–363. [CrossRef]

70. Zhang, J.-H.; Wang, L.-J.; Pan, Q.-H.; Wang, Y.-Z.; Zhan, J.-C.; Huang, W.-D. Accumulation and subcellular localization of heat shock proteins in young grape leaves during cross-adaptation to temperature stresses. *Sci. Hortic.* **2008**, *117*, 231–240. [CrossRef]
71. Baniwal, S.K.; Bharti, K.; Chan, K.Y.; Fauth, M.; Ganguli, A.; Kotak, S.; Mishra, S.K.; Nover, L.; Port, M.; Scharf, K.; et al. Heat stress response in plants: A complex game with chaperones and more than 20 heat stress transcription factors. *J. Biosci.* **2004**, *29*, 471–487. [CrossRef]
72. Hu, W.; Hu, G.; Han, B. Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. *Plant Sci.* **2009**, *176*, 583–590. [CrossRef]
73. Gao, C.; Wang, Y.; Liu, G.; Wang, C.; Jiang, J.; Yang, C. Cloning of ten peroxidase (POD) genes from *Tamarix hispida* and characterization of their responses to abiotic stress. *Plant Mol. Biol. Report.* **2010**, *28*, 77. [CrossRef]
74. Kim, S.-H.; Choi, H.-S.; Cho, Y.-C.; Kim, S.-R. Cold-responsive regulation of a flower-preferential class III peroxidase gene, *OsPOX1*, in rice (*Oryza sativa* L.). *J. Plant Biol.* **2012**, *55*, 123–131. [CrossRef]
75. Wang, J.-E.; Liu, K.-K.; Li, D.-W.; Zhang, Y.-L.; Zhao, Q.; He, Y.-M.; Gong, Z.-H. A Novel Peroxidase *CanPOD* Gene of Pepper Is Involved in Defense Responses to *Phytophtora capsici* Infection as well as Abiotic Stress Tolerance. *Int. J. Mol. Sci.* **2013**, *14*, 3158–3177. [CrossRef] [PubMed]
76. Liu, W.; Cheng, C.; Chen, F.; Ni, S.; Lin, Y.; Lai, Z. High-throughput sequencing of small RNAs revealed the diversified cold-responsive pathways during cold stress in the wild banana (*Musa itinerans*). *BMC Plant Biol.* **2018**, *18*, 308. [CrossRef] [PubMed]
77. Wellburn, A.R. The spectral determination of chlorophylls a and b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. *J. Plant Physiol.* **1994**, *144*, 307–313. [CrossRef]
78. Zhuang, H.; Lou, Q.; Liu, H.; Han, H.; Wang, Q.; Tang, Z.; Ma, Y.; Wang, H. Differential Regulation of Anthocyanins in Green and Purple Turnips Revealed by Combined De Novo Transcriptome and Metabolome Analysis. *Int. J. Mol. Sci.* **2019**, *20*, 4387. [CrossRef] [PubMed]
79. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full length transcriptome assembly from RNA Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
80. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [CrossRef] [PubMed]
81. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
82. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A. The COG database: A tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [CrossRef] [PubMed]
83. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [CrossRef]
84. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119. [CrossRef]
85. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **2015**, *44*, D286–D293. [CrossRef] [PubMed]
86. Deng, Y.Y.; Li, J.Q.; Wu, S.F.; Zhu, Y.; Chen, Y.W.; He, F.C. Integrated nr Database in Protein Annotation System and Its Localization. *Comput. Eng.* **2006**, *32*, 71–74.
87. Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **2004**, *5*, R7. [CrossRef] [PubMed]
88. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res. Ital.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]

89. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [CrossRef] [PubMed]
90. Eddy, S.R. Profile hidden Markov models. *Bioinform. Ital.* **1998**, *14*, 755–763.
91. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef] [PubMed]
92. Li, B.; Colin, N.D. RSEM: Accurate transcript quantification from RNA Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef]
93. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
94. Anders, S.; McCarthy, D.J.; Chen, Y.; Okoniewski, M.; Smyth, G.K.; Huber, W.; Robinson, M.D. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **2013**, *8*, 1765–1786. [CrossRef]
95. Alexa, A.; Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology*; R Package Version 2.18.0; 2010; Volume 2, p. 2010.
96. Dossa, K.; Li, D.J.; Yu, J.; Wang, L.; Zhang, Y.; You, J.; Zhou, R.; Mmadi, M.A.; Li, A.; Fonceka, D.; et al. The genetic basis of drought tolerance in the high oil crop *Sesamum indicum*. *Plant Biotechnol. J.* **2019**, 1–16. [CrossRef] [PubMed]
97. Lalitha, S. Primer premier 5. *Biotechnol. Softw Internet Rep.* **2000**, *1*, 270–272. [CrossRef]
98. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genomic Organization of the B3-Domain Transcription Factor Family in Grapevine (*Vitis vinifera* L.) and Expression during Seed Development in Seedless and Seeded Cultivars

Bilal Ahmad [1,2], Songlin Zhang [1,2], Jin Yao [1,2], Mati Ur Rahman [1,2], Muhammad Hanif [1,2], Yanxun Zhu [1,2] and Xiping Wang [1,2,*]

1 State Key Laboratory of Crop Stress Biology in Arid Areas, College of Horticulture, Northwest A&F University, Yangling, Xianyang 712100, China
2 Key Laboratory of Horticultural Plant Biology and Germplasm Innovation in Northwest China, Ministry of Agriculture, Northwest A&F University, Yangling, Xianyang 712100, China
* Correspondence: wangxiping@nwsuaf.edu.cn; Tel.: +86-29-87082129

Received: 4 August 2019; Accepted: 11 September 2019; Published: 14 September 2019

Abstract: Members of the plant-specific B3-domain transcription factor family have important and varied functions, especially with respect to vegetative and reproductive growth. Although B3 genes have been studied in many other plants, there is limited information on the genomic organization and expression of B3 genes in grapevine (*Vitis vinifera* L.). In this study, we identified 50 B3 genes in the grapevine genome and analyzed these genes in terms of chromosomal location and syntenic relationships, intron–exon organization, and promoter *cis*-element content. We also analyzed the presumed proteins in terms of domain structure and phylogenetic relationships. Based on the results, we classified these genes into five subfamilies. The syntenic relationships suggest that approximately half of the genes resulted from genome duplication, contributing to the expansion of the B3 family in grapevine. The analysis of *cis*-element composition suggested that most of these genes may function in response to hormones, light, and stress. We also analyzed expression of members of the B3 family in various structures of grapevine plants, including the seed during seed development. Many B3 genes were expressed preferentially in one or more structures of the developed plant, suggesting specific roles in growth and development. Furthermore, several of the genes were expressed differentially in early developing seeds from representative seeded and seedless cultivars, suggesting a role in seed development or abortion. The results of this study provide a foundation for functional analysis of B3 genes and new resources for future molecular breeding of grapevine.

Keywords: B3 superfamily; transcription factor; ovule abortion; *Vitis vinifera*; expression analysis

1. Introduction

Deployment of specific regulatory genes at opportune times is key to the development and growth of a plant in a way optimized for its environment. Plants, like other higher organisms, use a cadre of DNA-binding transcriptional factors that act both alone and combinatorically to activate or repress regulatory genes. The plant-specific B3 DNA-binding domain was originally recognized in studies of the maize Vp1 (Viviparous-1) transcription factor, which carries out numerous developmental functions including repression of premature seed germination. The Vp1 protein is a member of a small protein family related by amino acid sequence that can be partitioned into five subfamilies: ABI3/VP1, HSI, RAV (Related to ABI3/VP1), ARF, and REM [1–6]; although, some researchers consider ABI3 and HSI as members of a single family designated as LAV (LEC2 (LEAFY COTYLEDON 2)/ABI3 (ABSCISIC ACID INSENSITIVE 3)/VAL (VP1/ABI3-LIKE)) [7]. The structure of the B3 domain comprises seven beta-strands

(1–7) and two short alpha-helices (1 and 2). The beta strands form an open beta-barrel-like structure without closing the sheet, whereas the two short alpha-helices are located at either end of the barrel [8,9].

Several ARF and LAV family genes are transcriptionally regulated in response to auxin and abscisic acid [5,10,11]. ARF genes have been characterized in the context of many auxin-mediated physiological processes such as apical dominance, tropic responses, lateral root formation, vascular differentiation, embryo patterning, and shoot elongation [12,13]. LAV genes have been best characterized for their role in seed development and maturation [1,2,14–16]. RAV family members have not been as well characterized; however, some of them have been implicated in plant growth, development, and flowering time [17–19]. To date, the REM subfamily has been the least studied or characterized among all the B3 subfamilies [6]. However, it is known that the Arabidopsis REM gene VRN1 (VERNALIZATION 1) is involved in flowering [20,21].

Grapevine, including European (*Vitis vinifera* L.), American (*V. labrusca* L.), and Muscadine (*V. rotundifolia* L.), is one of the earliest fruit crops [22]. Grapes are cultivated worldwide and are consumed as fresh fruit, dried (raisins), juice, and wine [23,24]. Recently, the demand for seedless grapes is rising rapidly, fueling the development of seedless cultivars [25]. Seedlessness in grapes results from two distinct mechanisms: stenospermocarpy and parthenocarpy. Stenospermocarpy has great value as a research tool because the trait is heritable, is not strongly influenced by environmental factors, and because berry size is not greatly affected [26–28]. A large body of research has focused on potential mechanisms of seed abortion in stenospermocarpic grapes [29,30]. However, the key genes mediating this process have not been identified.

Recently, several genes encoding transcription factors have been implicated in seed development and abortion in grapevine. Overexpression of grapevine *VvCEB1*, encoding a helix–loop–helix transcription factor, affected embryo development and increased cell size [31]. Likewise, *VvAGL11*, encoding a MADS-domain transcription factor, was reported to have a role in stenospermocarpy [32]. Additional MADS genes, as well as some genes encoding homeodomain (HD) transcription factors, have been shown to be differentially expressed during development in seeded versus seedless grape cultivars [33,34]. In addition, *HD-ZIP* gene family members participate in the regulation of embryo abortion in grapes [35]. However, little is known about the role of B3 transcription factors during vegetative and reproductive development in grapevine. The important role of B3 family members as regulators of plant growth and developmental processes in other plants rationalizes a detailed bioinformatics and expression analysis of this gene family in this plant. In this study, we performed detailed bioinformatics analysis of B3 genes in grapevine, including gene structure and chromosomal locations, sequence homology, evolutionary duplication history, and cis-regulatory elements. We also analyzed expression of B3 genes during seed development in seeded and seedless grape cultivars. The results of this study will facilitate further studies of the mechanism of seed abortion in grapes.

2. Results

2.1. Genome-Wide Identification of B3 Genes in Grapevine

To identify B3 genes in the grapevine genome, we used a Hidden Markov Model (HMM) algorithm for the conserved B3 domain (PFAM 02362) in combination with the HMM search tool HMMer. A total of 61 putative B3 genes were revealed. All genes corresponding protein sequences were compiled and evaluated for the presence of an intact B3 domain using the Simple Modular Architecture Research Tool SMART (http://smart.emblheidelberg.de/) with default parameters [36]. The proteins having, incomplete B3 domains, DUF (domain of unknown function) or with non significant E-values were discarded. Finally, the integrity and accuracy of all the gene sequences were verified in the Grape Genome Database (12X). This approach led to the designation of 50 B3-domain-encoding genes. These genes were named according to their family name and their positions on the chromosomes [37–39]. Detailed information about genes including locus ID, accession number, chromosomal position, and length of coding sequence and open reading frame are given in Table 1.

Table 1. Characteristics of grapevine B3 genes.

Gene Locus ID	Gene ID	Accession No.	Chr. No.	Start Site	End Site	CDS (bp)	ORF (aa)
GSVIVT01004942001	*VvARF1*	CBI35669	1	21717511	21724801	2304	767
GSVIVT01019566001	*VvARF2*	CBI34510	2	1653587	1657804	2049	682
GSVIVT01035204001	*VvARF3*	CBI27334	4	10402895	10446692	1461	486
GSVIVT01025198001	*VvARF4*	CBI16322	6	3442998	3447976	1641	546
GSVIVT01025159001	*VvARF5*	CBI16287	6	3879687	3888448	2397	798
GSVIVT01011008001	*VvARF6*	CBI32272	7	2256656	2262296	2091	696
GSVIVT01025691001	*VvARF7*	CBI32737	8	12924236	12928912	1743	580
GSVIVT01021128001	*VvARF8*	CBI30623	10	1695978	1704736	2106	701
GSVIVT01021553001	*VvARF9*	CBI30950	10	6957710	6965185	1596	531
GSVIVT01015035001	*VvARF10*	CBI27770	11	629189	640325	2637	878
GSVIVT01032251001	*VvARF11*	CBI24055	11	13984914	13992724	3105	1034
GSVIVT01020805001	*VvARF12*	CBI22060	12	1745863	1766246	2511	836
GSVIVT01023149001	*VvARF13*	CBI29640	12	21939959	21950048	2037	678
GSVIVT01027166001	*VvARF14*	CBI40565	15	17300322	17304725	1824	607
GSVIVT01008639001	*VvARF15*	CBI15749	17	224846	230250	2589	862
GSVIVT01009865001	*VvARF16*	CBI19831	18	11920498	11929437	2724	907
GSVIVT01037136001	*VvARF17*	CBI16989	18	28749592	28756756	1932	643
GSVIVT01011947001	*VvRAV1*	CBI27062	1	2751620	2752908	687	228
VIT_202s0234g00085	*VvRAV2*	XP_019081631	2	110088	111017	930	309
GSVIVT01019699001	*VvRAV3*	CBI 34625	2	2549060	2551460	1143	380
GSVIVT01033902001	*VvRAV4*	CBI30344	8	16861285	16864429	720	239
VIT_211s0037g00010	*VvRAV5*	XP_002276492	11	7603337	7604681	1074	357
GSVIVT01036447001	*VvRAV6*	CBI16858	14	22050931	22055841	750	249
GSVIVT01027463001	*VvRAV7*	CBI38731	15	16539622	16542157	1128	375
GSVIVT01028540001	*VvABI3-1*	CBI37258	7	9431080	9434223	1695	564
GSVIVT01034419001	*VvABI3-2*	CBI35396	10	16556697	16566825	918	305
GSVIVT01033007001	*VvABI3-3*	CBI21292	14	25028976	25030714	864	287
GSVIVT01024559001	*VvVAL1*	CBI15813	6	8848747	8857227	2127	708
GSVIVT01030182001	*VvVAL2*	CBI18036	8	10751372	10767781	2571	856
GSVIVT01014432001	*VvVAL3*	CBI20373	19	3676945	3687513	729	242
GSVIVT01003212001	*VvVAL4*	CBI23327	Un	7958360	7975590	1806	601
GSVIVT01031761001	*VvREM1*	CBI32428	3	4170184	4173696	672	223
GSVIVT01031762001	*VvREM2*	CBI32429	3	4176729	4178437	1092	363
GSVIVT01031763001	*VvREM3*	CBI32430	3	4180639	4183390	483	160
VIT_203s0063g01415	*VvREM4*	XP_010647801	3	4859928	4864681	1218	405
GSVIVT01031850001	*VvREM 5*	CBI32502	3	4874233	4875367	696	231
GSVIVT01031852001	*VvREM 6*	CBI32503	3	4879760	4881735	1068	355
VIT_203s0063g01455	*VvREM 7*	XP_010647883	3	4886639	4887736	930	309
GSVIVT01031853001	*VvREM 8*	CBI32504	3	4887975	4921315	3255	1085
GSVIVT01024536001	*VvREM 9*	CBI15797	6	9077586	9081132	945	314
GSVIVT01005085001	*VvREM 10*	CBI40975	7	14390670	14392515	1200	399
GSVIVT01005087001	*VvREM 11*	CBI40976	7	14403658	14404557	516	171
GSVIVT01022143001	*VvREM 12*	CBI21437	7	16919585	16924427	1527	508
VIT_211s0037g01365	*VvREM 13*	XP_019078510	11	10865702	10866963	981	326
GSVIVT01023108001	*VvREM 14*	CBI29609	12	22478661	22484790	1332	443
GSVIVT01019293001	*VvREM 15*	CBI39994	15	1791086	1795242	633	210
GSVIVT01007552001	*VvREM 16*	CBI14875	17	12251718	12257309	1626	541
GSVIVT01009519001	*VvREM 17*	CBI19534	18	9025103	9026342	654	217
VIT_218s0001g11355	*VvREM 18*	XP_010664615	18	9665889	9667070	858	285
GSVIVT01009586001	*VvREM19*	CBI19591	18	9669874	9677462	1815	604

Abbreviations: Chr: chromosome; CDS: coding sequence; ORF: open reading frame; Un: unknown chromosome.

2.2. Phylogenetic Analysis of Grapes, Arabidopsis, and Tomato

To depict the phylogenetic history of the B3 gene family in grapevine and to assist in their classification, a phylogenetic tree was constructed based on amino acid sequence alignment within the conserved B3 domain [8,9]. The analysis included 91, 50, and 82 B3-domain genes of Arabidopsis, grapes and tomato, respectively. According to the phylogenetic tree (Figure 1), we classified the proteins into five subfamilies: ARF, ABI3, HSI (VAL), RAV, and REM. The REM subfamily contained 49, 19 and 49 genes from Arabidopsis, grapevine and tomato respectively, whereas the LAV subfamily contained only 7 genes from grapevine and 6 from each Arabidopsis and tomato. This suggests that the REM family is more diverse and that the LAV subfamily has constant gene number in different species [39,40]. All 17 ARF subfamily genes could be further classified into 6 sibling pairs, whereas the remaining 5 VvARF were not matched. In the REM family, 19 genes were distributed into 4 sister pairs, whereas the remaining eleven genes were not matched with each other. In the RAV family, 7 genes were divided into 2 sister pairs, whereas the remaining three were not matched. However, in ABI3 and HSI subfamily each contained one sister pair of paralogous genes (Figure 2A). These results are consistent with previous findings that the REM subfamily is characterized by relative diversification and low bootstrap values [36]. The predicted length of proteins in the ARF, RAV, LAV, and REM subfamilies ranges from 486–1034, 228–380, 242–855, and 160–1085 amino acids, respectively.

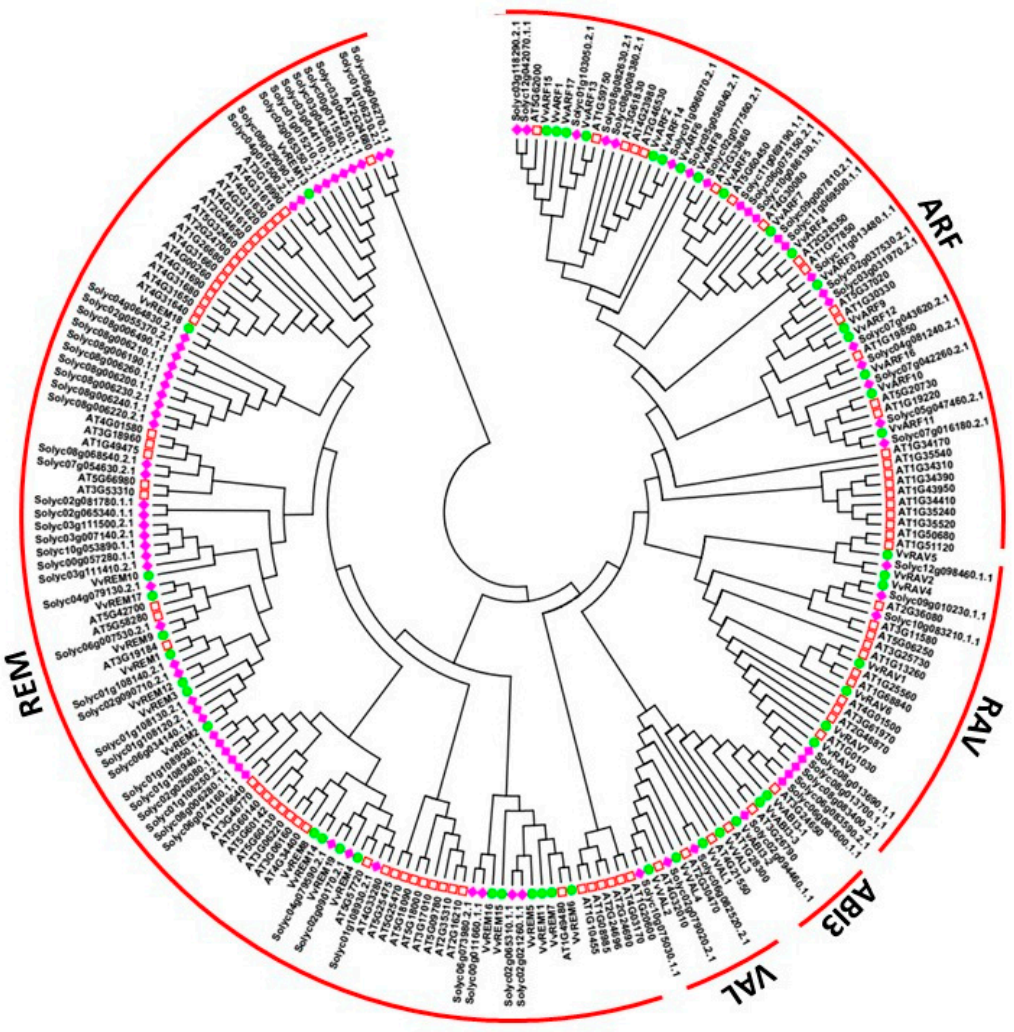


Figure 1. Phylogenetic analysis of B3 proteins from grapevine, tomato and Arabidopsis. Diamonds represent tomato protein, circles represent grapevine proteins, and squares represent Arabidopsis proteins.

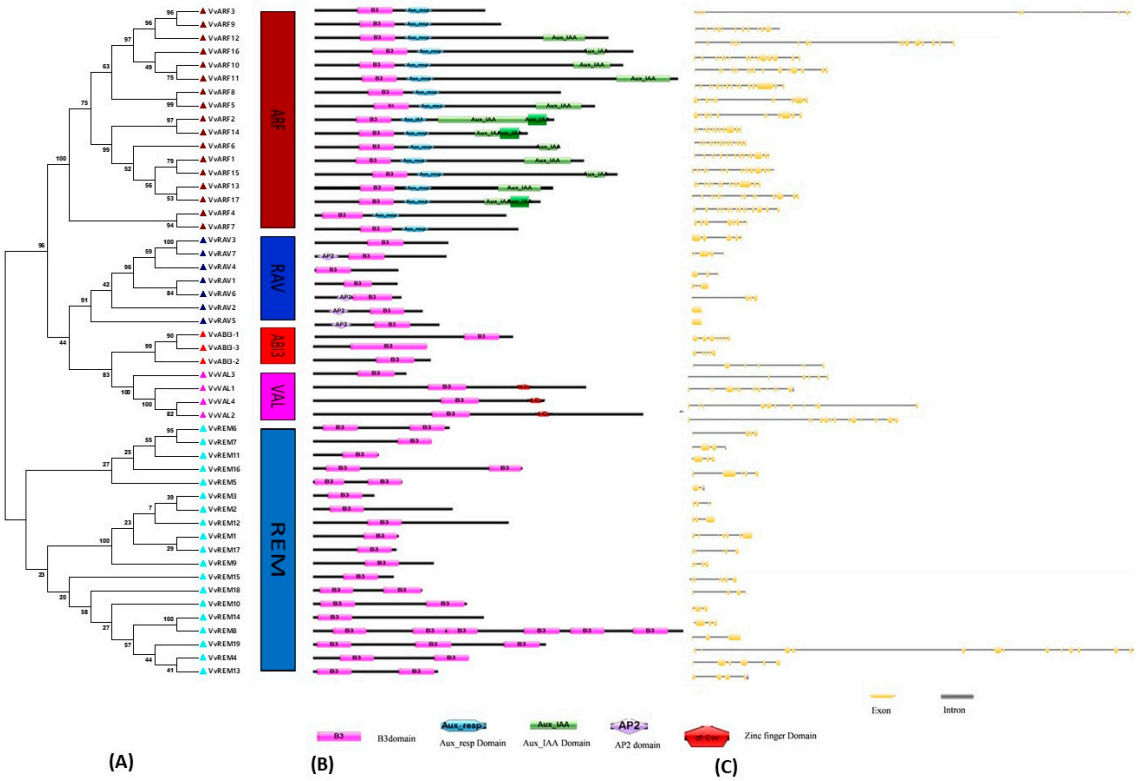


Figure 2. Structural analysis of grapevine B3 genes. (**A**) Phylogenetic analysis and classification. Different boxes are colored indicating different subfamilies. Numbers near the tree branches indicate bootstrap values. (**B**) Domain analysis. (**C**) Exon–intron structures. Exons are marked as yellow boxes, and introns are represented by black lines.

2.3. Analysis of Gene Structure and Conserved Nucleotide Motifs

To gain insight into the phylogenetic relationships of grapevine B3 genes, an unrooted tree was constructed based on the B3 domain (Figure 2A). Consistent with the phylogeny including the Arabidopsis B3 genes, this supported the classification into five subfamilies: ARF (17 genes), REM (19 genes), VAL (4 genes), RAV (7 genes), and ABI3 (3 genes). The number of exons/introns varied in each subfamily, ranging from 2–15, 6–9, and 1–13 in ARF, ABI3, and REM families, respectively (Figure 2C). Members of the RAV and VAL subfamilies showed less variation in number of introns, suggesting that they are more conserved. In general, genes with similar intron numbers, positions and lengths corresponded to closely related siblings identified through the phylogeny, supporting the validity of the phylogeny. Furthermore, we generally observed similar motif distribution patterns within subfamilies (Figure S1). The highly conserved Motif 1 (B3 DNA binding domain; IPR003340) was found in almost all members of the grapevine B3 family. Motif 6, which corresponds to the auxin response factor domain (Auxin_resp; IPR010515), was found in all members of the ARF subfamily (Figure S1). These results are consistent with the conserved domains shown in Figure 2B. The RAV and ABI3 subfamilies were more conserved with respect to motif number and distribution. RAV and ABI3 have the same motifs (motifs 1, 9, and 15) and differ only in the distribution of the motifs. Three of the four members of the VAL subfamily exhibit relatively strong homology. The REM subfamily showed the most divergence in motif number and distribution. VvREM8, 14, and 15 contained only one motif, whereas VvREM4, 5, 6, and 7 contain three motifs (motifs 1, 9, and 18). The remainder of the REM representatives contained only two motifs. Moreover, members of the same clade of a phylogenetic tree shared similar motif organization with respect to either gene length or motif number. These results provided support to the reliability of phylogenetic analysis and Exon–intron distribution for classification.

2.4. Domain Architecture Analysis of Grapevine B3 Proteins

To gain further insight into the phylogenetic relationships among grapevine B3 domain genes, we analyzed their presumed protein products (open reading frame translations) for recognized peptide domains. In the ARF subfamily, three conserved domains (B3, AUX_RESP, and AUX_IAA) were identified (Figure 2B). All three domains were found in 12 of the 17 ARF proteins, whereas only two domains (B3 and AUX_RESP) were found in the remaining five (VvARF3, 4, 7, 8, and 9). We noted that three ARF proteins (VvARF2, 14, and 17) contained two AUX_IAA domains. Within the RAV subfamily, four RAV (VvRAV2, 5, 6, and 7) proteins contained AP2 and B3 domains, whereas the other three members had only one B3 domain. All three members of the ABI3 subfamily contained only a single B3 domain, whereas proteins included in the VAL subfamily contained a Zinc-finger domain along with the B3 domain. With one exception, VAL subfamily proteins contained both zf-CW and B3 domains. The domain architecture in REM proteins was more complex than that of the B3 subfamilies. Nine members of the REM subfamily contained multiple B3 domains. The highest number of B3 domains was noted in REM8, which contained six B3 domains.

2.5. Expansion Patterns of B3 Genes in Grape

According to annotation of the grapevine genome, B3 genes were broadly distributed and found on 16 of the 20 chromosomes. Chromosome 3 contained the highest number of B3 genes, eight, all of which belong to the REM subfamily. Two or more genes found on the same chromosome within a 200 kb region are likely to have resulted from tandem duplication [41]. According to these criteria, we observed twelve B3 genes (Table S1) clustered by four tandem duplication events, on Chromosome 3 (two duplications), Chromosome 7 (one duplication), and Chromosome 18 (one duplication). Surprisingly, all of these genes belonged to the REM subfamily, indicating that REM has undergone more changes with the passage of time as compared with other B3 subfamilies. In addition to tandem duplication, we investigated eleven pairs of B3 genes apparently resulting from segmental duplication events (ARF13/ARF17, VAL1/VAL2, REM12/REM1, ARF4/ARF7, ARF5/ARF7, REM17/REM12, ARF2/ARF14, REM17/REM1, REM19/REM5, RAV3/RAV7, and RAV1/RAV6) (Figure 3, Table S2), suggesting that both tandem and segmental duplication generated the grapevine B3 family. Interestingly, three genes REM 1, REM17, and ARF7, each were paired with two genes. REM1 was paired with REM 12 and REM 17, and ARF7 was paired with ARF4 and ARF7; whereas, in REM17, segmental duplication was observed with REM1 and REM12. In conclusion, 30 out of 50 (60%) of the B3 genes contributed to duplication events (segmental or tandem), which may provide a reference for the evolutionary relationship and functional potential of B3 genes.

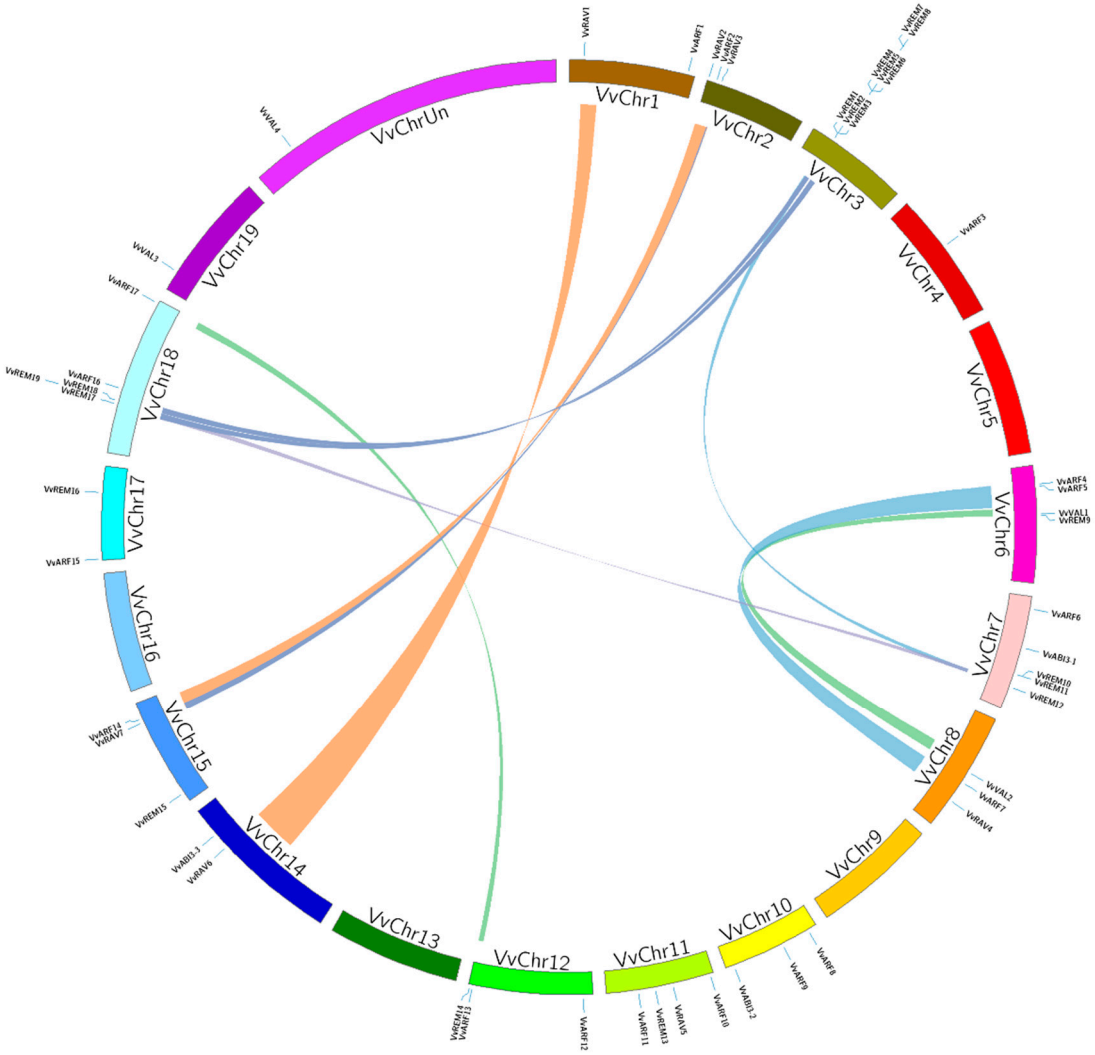


Figure 3. Synteny analysis and chromosomal distribution of grapevine B3 genes. Colored bars connecting two chromosomal regions denote syntenic regions; the corresponding genes on two chromosomes were regarded as segmental duplications. Chr: chromosomes.

2.6. Evolutionary Relationships among Grapevine, Tomato, and Arabidopsis B3 Genes

The function of B3 genes has been studied mainly in Arabidopsis. To further justify the origin, evolutionary history, and potential function of grapevine B3 genes, we examined genomic synteny of grape with tomato and Arabidopsis. A total of 16 pairs of syntenic relationships were identified between grapes and Arabidopsis, comprising 15 grapevine and 14 Arabidopsis genes (Table S3, Figure 4). There were two pairs (AT2G30470-VvVAL1/VvVAL2 and AT3G19184-VvREM17/VvREM1) where a single Arabidopsis gene paired with more than one grapevine gene. Moreover, in grapevine, we also found one gene (REM17-AT3G19184/AT5G42700) that paired with two Arabidopsis genes. In the case of grapes and tomato, a total of 19 pairs of segmental duplications events were identified, comprising of 17 tomato and 19 grapevine genes (Table S4). There were two pairs (Solyc08g013690.1-VvRAV3/VvRAV7 and Solyc02g079020.2-VvVAL3/VvVAL4) where a single tomato gene paired with more than one grapevine gene. In each syntenic block (grape and tomato or grape and Arabidopsis), both members belonged to the same subfamily and phylogenetic group. This suggests that they shared a common ancestor before evolution.

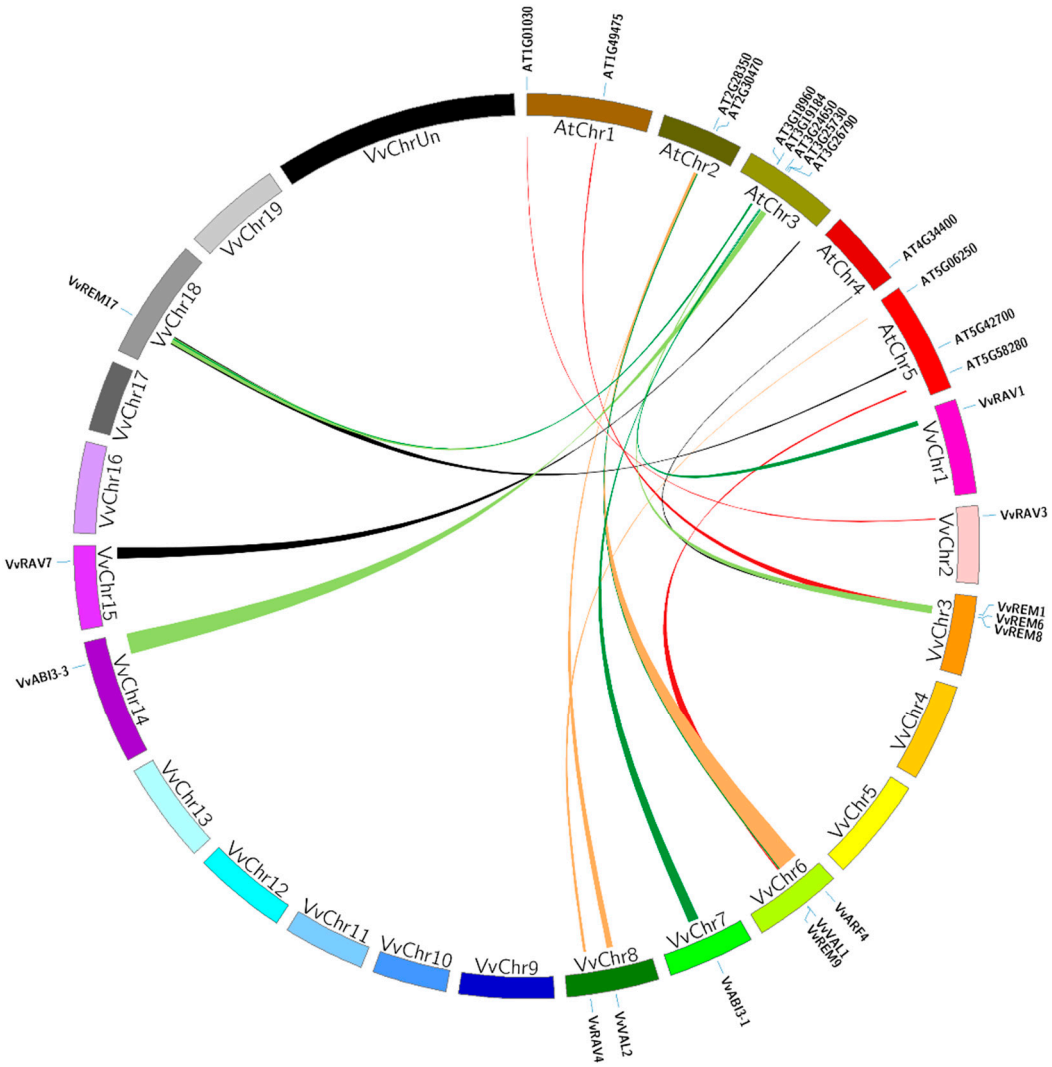


Figure 4. Synteny analysis of B3 genes between Arabidopsis and grapevine. Relative positions were depicted according to the grape and Arabidopsis chromosomes; colored lines represent syntenic regions.

2.7. Expression Patterns of Grape B3 Genes during Seed Development

To identify B3 genes with potential function in seed development or abortion, we used semiquantitative PCR to examine expression of the B3 genes during progressive stages of seed development. Expression of members of the ARF family was previously reported by Wan et al. [42]. As shown in Figure 5A, most of the genes were expressed to similar levels between the seeded cultivar ("Red Globe") and seedless cultivar ("Thompson Seedless"). However, some genes showed significantly different expression. For example, RAV3, RAV4, and REM2 were more highly expressed in Thompson Seedless than in Red Globe; in contrast, ABI3-1, ABI3-3, VAL1, and REM3 were more strongly expressed in Red Globe. To further support the results of semiquantitative PCR, nine genes determined to be differentially expressed between Red Globe and Thompson Seedless by semiquantitative PCR were analyzed by real-time, quantitative RT-PCR (Figure 5B). This supported that VvRAV3, VvRAV4, and VvREM2 were much more strongly expressed in Thompson Seedless as compared with Red Globe. The remaining six genes were more strongly expressed in Red Globe as compared to Thompson Seedless. For VvABI3-3 and VvVAL1, expression was almost undetectable in Thompson Seedless. These results suggest that these genes may have a role in seed development.

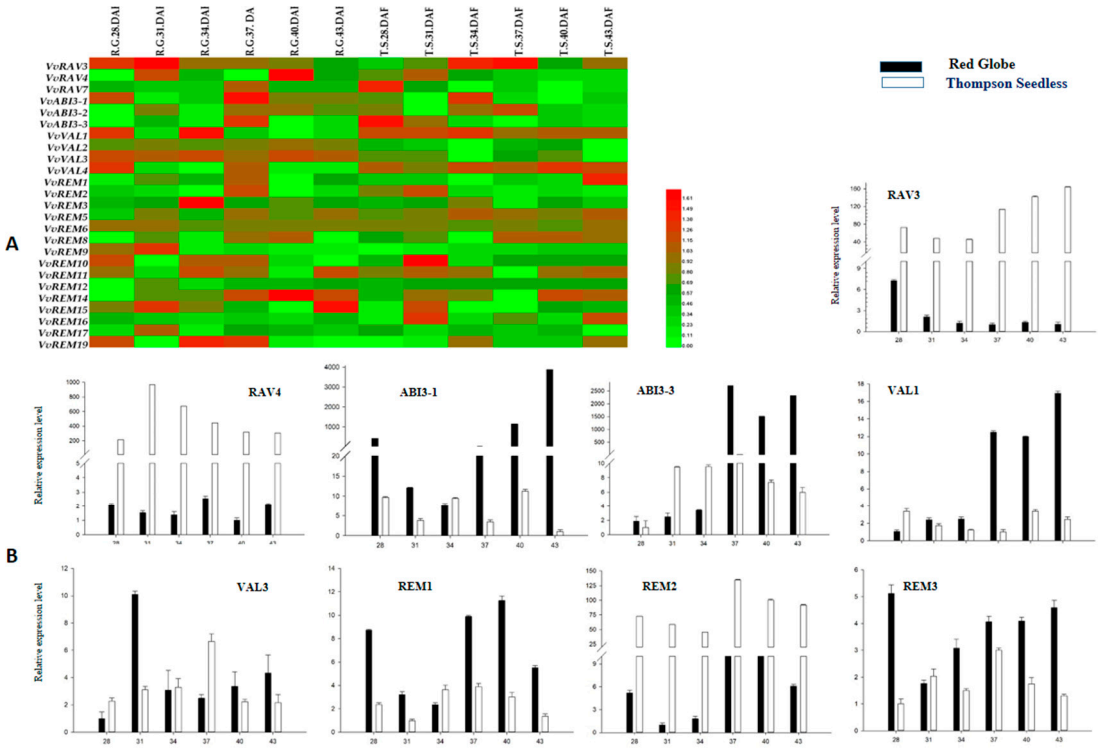


Figure 5. Expression profile analysis of B3 genes during progressive stages of seed development in seeded (Red Globe) and seedless (Thompson Seedless) cultivars. (**A**) Heat map of semiquantitative RT-PCR analysis. (**B**) Real-time PCR analysis. Numbers indicate the number of days after full bloom (DAF).

2.8. Developmental Regulation of Grapevine B3 Genes Outside of the Ovule

Analysis of the expression pattern of genes among various structures and organs is important for defining their functions. We used semiquantitative RT-PCR to analyze expression of the grapevine B3 genes in the root, stem, leaf, tendril and fruit of Thompson Seedless and Red Globe. We found that several of the genes, e.g., REM14, VAL2, VAL3, and RAV7, were expressed relatively ubiquitously, suggesting that they may have a general role in growth and development (Figure 6A). We noted that all the B3 genes were expressed in most of the grape structures. However, the level of expression of the B3 genes varied among the structures, or differed strikingly between Thompson Seedless and Red Globe. For example, ABI3-2 showed high expression levels in all structures of Red Globe, but a moderate level in roots, stems, leaves, and tendrils of Thompson Seedless and very low expression in fruits. In addition, ABI3-3 showed moderate expression levels in all structures of Thompson Seedless whereas no expression was detected in roots, leaves, or tendrils of Red Globe. To further elucidate the results of semiquantitative RT-PCR results, the nine genes identified as differentially expressed in developing seed (see above) were analyzed by real-time RT-PCR (Figure 6B). According to these results, ABI3-1, ABI3-3, RAV3, RAV4, REM1, and VAL3 were expressed more strongly in structures of Thompson Seedless as compared to Red Globe. The remaining three genes (VAL1, REM2, and REM3) were more strongly expressed in tissues of Red Globe as compared to Thompson Seedless.

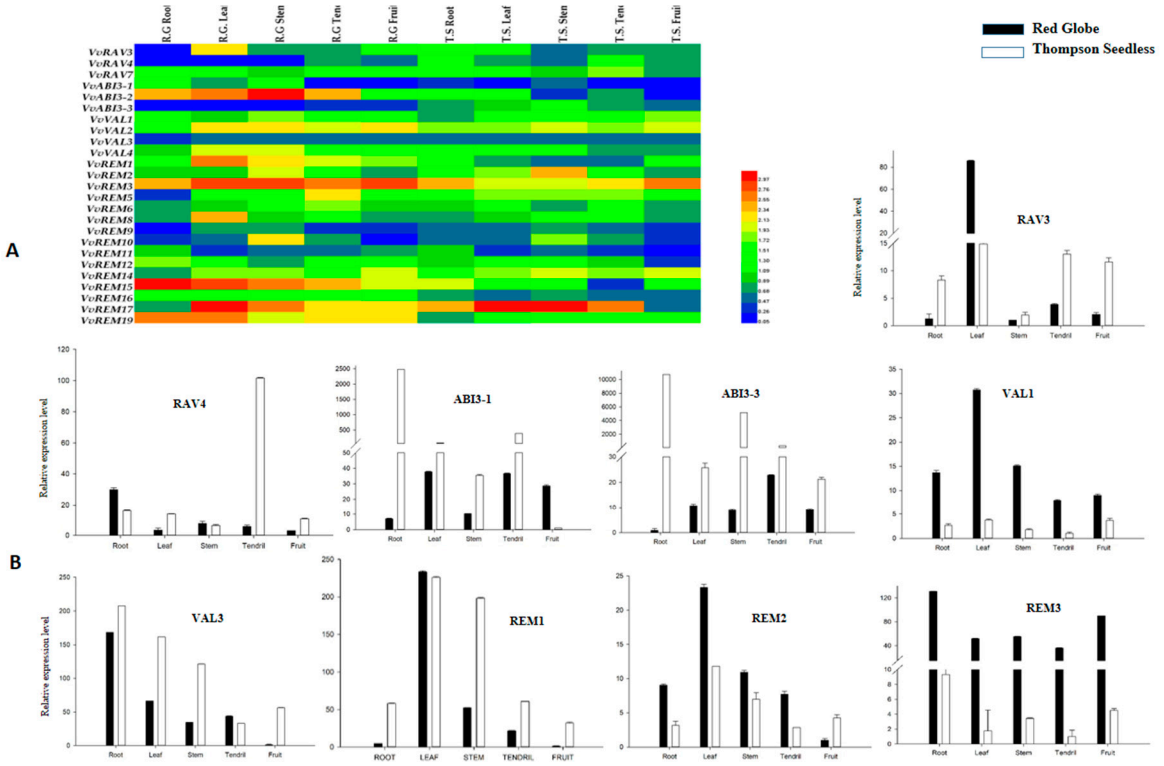


Figure 6. Expression analysis of grapevine B3 genes in various structures of seeded and seedless cultivars. (**A**) Heat map of Semiquantitative RT-PCR analysis. (**B**) Real-time PCR analysis.

2.9. Cis-Acting Elements and Gene Ontology (Go) Analysis of B3 Genes

To gain further insight into expression of the grapevine B3 genes, we carried out an in silico analysis of potential *cis*-elements conferring responsiveness to plant hormones (CGTCA, ERE, ABRE, P box, GARE, TGA-element, and AuxRR-core) within their promoters. *Cis*-elements related to disease resistance (TC-rich repeats and W-box) and stress response (TCA-motif, STRE, LTR) were also found in the promoter regions of 41 of the genes (Figure 7). Moreover, *cis*-acting elements involved in light responsiveness (G BOX, I BOX, BOX4, and GATA) and anaerobic respiration (ARE) were found in the promoter regions of almost all B3 genes.

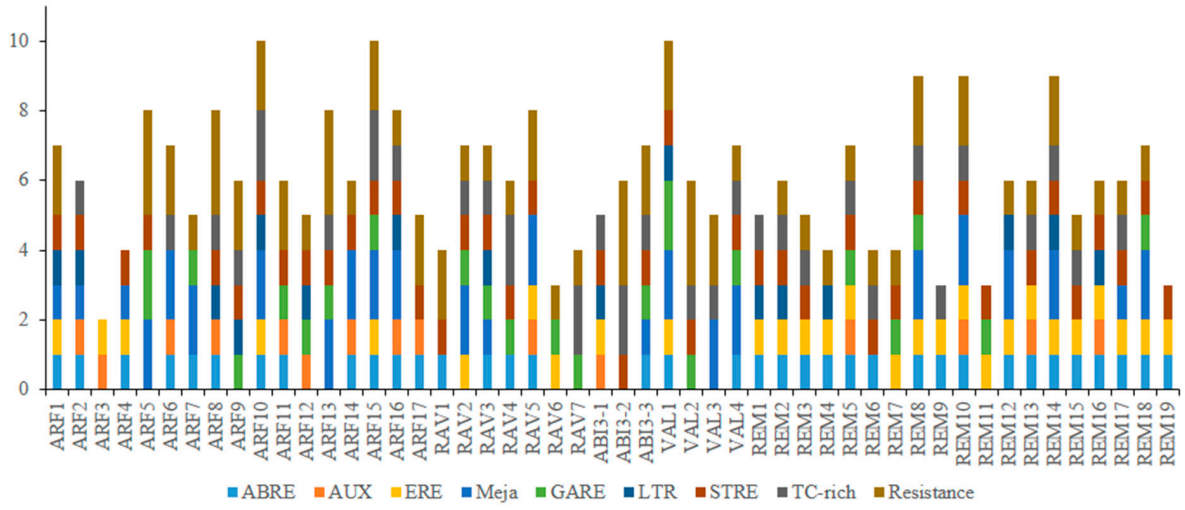


Figure 7. Predicted *cis*-elements in B3 gene promoters. The 1.5 kb sequences of 50 grape B3 genes were analyzed with the PlantCARE program.

We also carried out a gene ontology (GO) enrichment analysis of the grapevine B3 genes, considering biological process, molecular function, and cellular component, as previously described [37]. Prediction of biological process highlighted potential roles in transcription, fruit development, floral organ development, and response to abscisic acid (Figure 8). In addition, prediction of molecular functions of B3 proteins suggested that, as anticipated, most are involved in DNA binding. Analysis of cellular component suggested localization to various compartments including intracellular organelles and membrane-bound organelles.

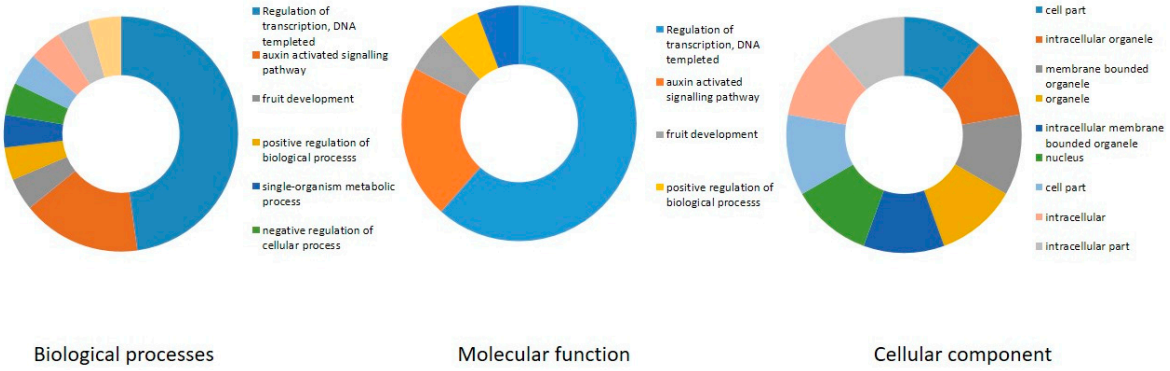


Figure 8. Gene ontology analysis of B3 proteins in three categories (Biological process, molecular function and cellular component) using Blast2Go program. Different colors which are indicated near the graphics represent distinct biological process, molecular function or cellular component.

3. Discussion

B3-domain transcription factors are known to have various roles in vegetative and reproductive development [36,43,44]. Members of the ARF and LAV B3 subfamilies have been well studied, whereas the RAV and REM subfamilies have remained less characterized [7]. Functional characterization of members of the B3 family has also implicated B3 domain genes in hormone signaling pathways including auxin, abscisic acid, brassinosteroid, and gibberellin. Genome-wide identification and expression analyses of the B3 gene family have been studied in several plants [40,45–49]. However, to our knowledge, this is the first comprehensive bioinformatics and expression analysis of B3 genes in grapevine.

In this study, we carried out comprehensive bioinformatics analysis of B3 genes in grapevine and compared their expression profiles between seeded and seedless cultivars during seed development and in various structures of the plant. We identified 50 B3 genes and classified these into five subfamilies based on phylogenetic analysis with Arabidopsis and tomato. As far as the number of genes in ARF and REM subfamilies in grapes is concerned, we observed that the previous trend, i.e., the relationship between ARF and REM gene numbers in one plant is generally, inversely proportional [39,40]. Similar to Arabidopsis and tomato, the REM subfamily comprises the largest group of B3 domain genes. Previously, 19 ARF genes were reported in grapevine [42]. In this study, we identified only 17, likely because we employed a more stringent E-value and other selection criteria. Within the ARF subfamily, twelve out of seventeen members contained three (B3, AUX_RESP and AUX_IAA) domains, with the remaining five (VvARF3, 4, 7, 8, and 9) containing only two domains (B3 and AUX_RESP). This suggests that the AUX_IAA domain has been less conserved than the B3 and AUX_RESP domains during the evolutionary process. The same observation have been reported in studies in other plants such as Arabidopsis, apple and papaya [12,45,47]. For the RAV subfamily our findings are concordant with previous results that most but not all proteins contain both B3 and AP2 domains [18]. According to the evolutionary relationships between the RAV proteins in Arabidopsis and grapevine, two RAV proteins (RAV3 and RAV4) are more closely related to NGA and NGAL in Arabidopsis, which do not have AP2 domains. In Arabidopsis, NGALs are involved in flower and seed development, whereas the NGA proteins are mainly involved in flower and leaf development [39,50]. Whereas, in tomato and

cassava, the RAV transcription factor has showed resistant against bacterial blight [51,52]. Therefore, the RAV genes in grapevine may have important roles in flower, seed and leaf development as well as disease resistant. All members of the ABI3 subfamily contained only one B3 domain, whereas proteins included in the VAL subfamily contained a Zinc-finger domain as well. All members of the VAL subfamily contain a zf-CW domain, with one exception (VAL3). We speculate that loss of the zf-CW and AP2 domain occurred during evolution. The domain architecture in REM proteins is more complex than that seen in other subfamilies. Nine members of REM subfamily contain multiple B3 domains, with VvREM8 containing the highest number (six). These multiple domains may have evolved from one or more domain duplication events. During evolution, domain duplications can generate new genes along with different functions, thereby increasing functional diversity [9]. However, the function of genes with domain duplications generally is not easy to determine, because many of the proteins are functionally redundant.

The arrangement of exons and introns within a gene family can provide indications to support the evolutionary relationships. For B3 domain genes, in general, we found that genes within the same group tended to have the same number of introns, but we noted some exceptions. Evolutionary transition in exon number in genes within the same family plays an important role in the evolution of multiple gene families [53]. Duplicated genes are initially redundant, and consequently can adopt new functions [54]. The same phenomenon has been observed in other plants, such as Arabidopsis, canola, chick pea, hot pepper, rice, and tobacco [12,37,48,55,56].

Gene duplication is the primary source of evolution of genes and gene families [57]. Tandem and segmental duplications appear to be the main source of gene family expansion in grapevine, as the grapevine genome has not been duplicated during evolution [58]. We observed highly variable numbers of B3 domain genes on different chromosomes. Previously, it was reported that REM genes are clustered in the genome in at least 12 species, including Arabidopsis, maize, rice, sorghum, and tobacco [47]. Likewise, we identified eight REM subfamily genes only on chromosomes 3. This suggests that duplication of B3 genes probably occurred on chromosome 3 during evolution of the gene families. Interestingly, we also found that genes of the ARF and REM subfamilies are also clustered together on chromosome 8. The extensive gene duplication events may be responsible for the clustering. However, there is, at present, no practical experimental evidence to explain this observation [48]. According to our results in grapevine, 58% of the genes are products of tandem or segmental duplication events. Tandem duplications appear to have occurred in only eleven genes of the REM subfamily. Our results are in accordance with previous findings [36] that REM genes have evolved quickly and are functionally less conserved. In most of the plants, segmental duplication happens more frequently than tandem duplication because of polyploidy, which conserves numerous duplicated chromosomal blocks in their genome [57]. Consistent with this, we observed segmental duplication for 17 B3 genes. Additionally these duplicated genes clustered together in the same phylogenetic group. Two genes of the REM subfamily (1 and 12) exhibited both tandem and segmental duplication. In the REM subfamily, 73% (14/19) of genes appear to have experienced duplication events with the passage of time, whereas no duplication event was noted for ABI3 genes. These results indicated that ABI3 genes are highly conserved, whereas REM genes have been diversified. These results are consistent with previous findings that ABI3 genes are highly conserved among many plant species, including maize, Arabidopsis, rice, oat, wheat and tomato [59,60]. In this study, we found that the grapevine B3 gene pairs generated from tandem and segmental duplications showed conserved domains (Figure 2B), showing that these pairs persisted after the gene duplication event. However, there were three exceptions where duplicated genes showed a change in conserved domains—VvARF5/VvARF7 and VvRAV3/VvRAV7 (segmentally) and VvREM10/VvREM11 (tandemly). The main difference within these three pairs is the change in domain arrangements. In addition to gene duplication, rearrangements of domain structure can contribute to expansion of gene families and functional diversity [61]. Similar observations have been made in studies of chickpea, in which CaARF4/CaARF5 and CaARF21/CaARF22 are tandemly

duplicated but structurally different [37]. These pairs might have been formed by ancient tandem duplication events.

Genomic synteny can provide clues to gene function, and can be an effective tool for gene family analyses where whole genome sequence is available [62,63]. We identified 16 gene pairs involved in segmental duplications between grapevine and Arabidopsis, and 19 pairs between grapes and tomato, suggesting they might have a common ancestor. A large number of synteny events suggests that many B3 genes arose before the divergence of the Arabidopsis, tomato and grapevine lineages. Although it is difficult to explain the evolutionary relationship between grapes, tomato and Arabidopsis based only on chromosomal syntenies, this analysis identified grapevine genes that share a common ancestor with their Arabidopsis and tomato B3 counterparts. Based on the reported functions of grapes orthologous genes in other species, we can predict possible functions of grapevine. For example, the At3g26790 (Fus3) orthologous of VvABI3-3, has role in seed development [15] and ectopic overexpression of SlABI3 resulted in the induction of seed-specific genes in vegetative tissues of tomato [64], while VvABI3-3 showed different levels of expression during seed development in seeded and seedless cultivars in this study (Figure 5B). These findings suggest that the grapes B3 genes endured a complicated evolutionary history. Taken together, these results enable further analyses of evolution and potential functions among these genes and will be helpful for a detailed study on the functions of homologous genes in various plant species.

The presence of *cis*-element motifs within promoter regions can provide clues for how genes might be regulated at the transcriptional level [63]. Our in silico analysis identified potential ABRE cis-elements, which participate in response to abscisic acid (ABA) in the promoter regions of 38 B3 genes. Interestingly, 17 of the 19 members of REM subfamily had both ABRE and ERE elements. This suggests that expression of these genes is influenced by ABA. In addition, GA-responsive elements (P box and GARE) and auxin-responsive elements (TGA-element and AuxRR-core) were found in some genes of the B3 family. Interestingly, all members of the RAV subfamily have GA-response elements, but only one member of RAV family contains an auxin-response element. This suggests that RAV family genes have a role in gibberellic acid signaling pathways. Presence of disease resistance and stress-responsiveness motifs suggests a role for B3 genes in disease resistance and stress signaling pathways. These results suggest that the regulation of B3 gene expression is complex and justifies the need for further research.

Gene expression patterns can provide important indication of gene function. In this study, we examined the expression of 25 B3 genes in five plant structures and six seed developmental stages in both seeded and seedless grape varieties. RAV3, RAV4 and REM2 were highly expressed in Thompson Seedless during various seed developmental stages. These genes might have a role in seed abortion of seedless grape cultivars. In contrast, ABI3-1, ABI3-3, and VAL1 showed high expression in the seeded cultivar Red Globe, suggesting that these three genes promote normal seed development. Our results are also supported by previous studies as AtLEC2 (Arabidopsis Leafy Cotyledon 2 gene) has been reported for its role in seed development and maturation [15,16], as well as its overexpression in cacao leaves changed the expression levels of several seed related genes [65].

Other B3 genes, including ABI3-2, VAL2, RAV3, REM5, REM6, and REM15, were expressed to similar levels during all stages of seed development in both cultivars, suggesting that these genes are functionally conserved in different grape cultivars and might have a general role in growth and development. For most of the genes, it was observed that, if a gene was highly expressed in fruit, it was also highly expressed in the seed. Other genes, such as ABI3-3, VAL3, and REM1, showed high expression levels in fruit of Thompson Seedless, but low expression during seed development. This same phenomenon was observed for the grapevine homeodomain family genes VvHB62, VvHB63, and VvHB55 [34]. In order to assess potential functional distinction of duplicated B3 genes, we compared their expression patterns in Red Globe and Thompson Seedless (Figure 6B). As projected, most of the duplicated genes were expressed in similar patterns between the seeded and seedless varieties. However, several examples of divergent expression patterns were also noticed. For example, VvRAV7

showed the same expression in all tissues and developmental stages of both cultivars, whereas VvRAV3 showed different expression profiles in both cultivars. The divergence in expression patterns between homologous and segmental duplication genes indicated that some of them might lose function or obtain new function after duplication in the evolutionary process.

4. Materials and Methods

4.1. Identification and Annotation of B3 Genes in Grapevine

To identify a complete list of B3 genes in the grapevine genome, annotated grapevine proteins were downloaded from three public databases: Grape Genome Database (http://www.genoscope.cns.fr), the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), and the Grapevine Genome CRIBI Biotech website (http://genomes.cribi.unipd.it/). The Hidden Markov Model (HMM) profile of the B3 domain (PFAM 02362) downloaded from the Pfam database (http://pfam.xfam.org/) was used to survey all grapevine proteins in the 12X coverage assembly of the *V. vinifera* PN40024 genome to identify B3 genes [58]. The sequence integrity of the B3 domain (PFAM 02362) was assessed using SMART (http://smart.embl-heidelberg.de) with default parameters and genes with incomplete domains or with non-significant E-value were omitted from consideration [66]. All non redundant protein sequences with a conserved B3 domain were identified as grapevine B3 family members. Additionally, manual annotation was performed to resolve any discrepancy between incorrectly predicted genes and the actual chromosomal locations of involved genes in question.

4.2. Multiple Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignments of 50 grapevine B3 sequences, 91 Arabidopsis B3 sequences and 82 tomato B3 sequences were performed using ClustalX 2.1 with default parameters [67]. The corresponding phylogenetic tree was constructed using the neighbor-joining (NJ) method and MEGA 6.0 software, using 1000 bootstrap replicates with the following parameters; "*p*-distance", "Complete Deletion", and gap setting [34]. Phylogeny among grapevine B3 sequences shown in Figure 2A was determined using MEGA 6 software by the neighbor-joining method with 1000 iterations and sequence alignments were performed using the 'W' approach.

4.3. Analyses of Exon–Intron Structure, Distribution of Conserved Motifs and Characteristic Domain Architecture

Exon–intron structures of the grapevine B3 genes were analyzed using aligned transcribed sequences and corresponding genomic sequences, and diagrams were created using the online Gene Structure Display Server 2.071 [62]. Conserved motifs and domains of proteins were analyzed using MEME 4.11.2 (http://meme-suite.org/tools/meme), searching up to 20 conserved motifs, and SMART (http://smart.embl-heidelberg.de). Protein domains were identified using the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

4.4. Synteny Analysis of Grape, Tomato and Arabidopsis B3 Genes

Tandem and segmental duplications of B3 genes in the grapevine genome were identified based on chromosomal locations. For synteny analysis, adjacent homologous grape B3 genes on a single chromosome without the presence of intervening genes was considered as a tandem duplication, whereas gene duplication events occurring on different chromosomes were defined as segmental duplications [68]. The list of grape B3 genes in duplicated genomic regions and a comparison of grape and Arabidopsis as well as grape and tomato genomes were retrieved from the Plant Genome Duplication Database [34]. Diagrams were generated using the program Circos version 0.63 (http://circos.ca/).

4.5. Analysis of Cis-Acting Elements and Gene Ontology

To examine the possible regulatory mechanisms of B3 genes, a 1.5-kb promoter region upstream of the start codon of each gene was retrieved from the Grape Genome Database (http://www.genoscope.cns.fr) and analyzed through the PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) online server. Gene ontology analysis of B3 protein sequences was acquired from Blast2GO (http://www.blast2go.com).

4.6. Plant Materials

This study used one seeded cultivar, Red Globe, and one seedless cultivar, Thompson Seedless (both *V. vinifera*). Plants were maintained in the grape germplasm resource orchard of Northwest A&F University, Yangling, China (34°20′ N 108°24′ E). The obtained plant structures were young roots, stems, leaves, tendrils, and fruits (42 days after full bloom, DAF). All samples were collected from plants under natural conditions. Developing seeds were dissected from fruits at 28, 31, 34, 37, 40, and 43 DAF. All samples were immediately frozen in liquid nitrogen and stored at −80 °C for RNA extraction and expression analysis.

4.7. RNA Extraction and Expression Analysis by PCR

Total RNA was extracted from samples using an EZNA Plant RNA Kit (R6827-01, OMEGA Biotek, Norcross, GA, USA), according to the manufacturer's guidelines. First-strand cDNA was synthesized by reverse transcription of 500 ng total RNA using Prime Script RTase (Trans Gen Biotech, Beijing, China). After this, cDNA was diluted six-fold and preserved at −40 °C for further analysis. The grapevine *ACTIN1* gene (Genbank Accession NC_012010) and *EF1-α* gene (Genbank Accession NC_012012) were used as internal controls; *ACTIN1* was amplified with the oligonucleotide primers (5′-GAT TCT GGT GAT GGT GTG AGT-3′) and (5′-GAC AAT TTC CCG TTC AGC AGT-3′), and *EF1-α* was amplified using (5′-AGG AGG CAG CCA ACT TCA CC-3′) and (5′-CAA ACC CTG CAT CAC CAT TC-3′). Gene-specific primers were designed for selected B3 genes using Primer Premier 6.0 (Table S5). The specificity of primers was checked in the NCBI (https://www.ncbi.nlm.nih.gov/) database, using the Primer-BLAST program. Semiquantitative RT-PCR assays were carried out in a volume of 20 µL per reaction containing 1 µL of cDNA template, 2 µL of gene-specific primers (1.0 µM), 7 µL sterile distilled water, and 10 µL PCR Master Mix (BIOSCI BIOTECH CO. LTD, Hangzhou, China). PCR conditions were 94 °C for 2 min, 35–38 cycles of 94 °C for 30 s, 51–63 °C for 30 s (depending on the specific gene), and 72 °C for 30 s, with a final extension of 72 °C for 7 min. In each case, 10 µL of the resulting product was resolved on a 1.5% (*w*/*v*) agarose gel and visualized using ethidium bromide and then imaged under ultraviolet light using GeneSnap software. Each assay was performed with three biological replicates. Semiquantitative RT-PCR expression data were visualized using GeneTools software. The mean expression value for each gene was determined in all tissues or seed developmental stages in all cultivars, and was then log2-transformed to generate heat maps using Multi Experiment Viewer software (Mev 4.8.1). Quantitative real-time PCR was performed for selected genes using SYBR Green (Trans Gen Biotech, Beijing, China) on an IQ5 real-time PCR machine (Bio-Rad, Hercules, CA, USA). *ACTIN1* was used as the internal reference gene and each reaction was performed with triplicate technical and biological repeats. The reaction mixture was 1 µL of cDNA template, 0.8 µL each primer (1.0 µM), 0.4 µL Rox reference dye, 7 µL sterile distilled water, and 10 µL of SYBR green. PCR was performed following the parameters 95 °C for 30 s, followed by 42 cycles of 95 °C for 5 s and 60 °C for 30 s. Relative expression levels were determined by the comparative CT method also referred to as $2^{-\Delta\Delta CT}$ method, where $\Delta\Delta CT$ = [(C_T target gene–C_T control gene) Sample A–(C_T target gene–C_T control gene) Sample B] [69]. After this, the RT-PCR values were used to create graphs using Sigma Plot 10.0 [33].

5. Conclusions

Bioinformatics analyses of 50 B3 genes in grapevine revealed that gene duplication played a role in the expansion of the B3 family in grapes. It was also found that B3 genes from grapevine and Arabidopsis shared some common ancestor. Studies of phylogeny, gene structure, and conserved motifs of the B3 genes have further provided insights on the evolutionary history of the grapevine B3 genes. Taken together, these results underscore the potential importance for B3 genes grapevine growth and development, particularly for seed development. Therefore, our systematic study of these genes will help in identifying candidate genes for functional studies. It will also serve as a foundation and reference for future research regarding the molecular mechanisms controlling the seedlessness trait of grapes.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/18/4553/s1. Data Figure S1. Conserved motifs analysis of grape B3 proteins according to the phylogenetic relationship. Box length corresponds to motif length. Each motif is represented by a number in a colored box. Table S1. Tandem duplication events in grapevine B3 genes. Table S2. Synteny blocks of B3 genes within grape genomes Table S3. Synteny blocks of B3 genes between grape and Arabidopsis Table S4. Segmental duplication pairs of B3 genes between grape and tomato. Table S5. Primer sequences used in expression analysis of B3 genes in grape. Table S6. Expression values of 25 grape B3 genes during different stages of ovule development in semiquantitative PCR. Red Globe is represented by 'RG' and Thompson Seedless is represented by 'TS'. Table S7. Expression values of 25 grape B3 genes in different plant organs in semiquantitative PCR. Red Globe is represented by 'RG' and Thompson Seedless is represented by 'TS'.

Author Contributions: X.W., B.A., and S.Z. designed the experiments. B.A. and S.Z. conducted the experiments. M.H. helped in RT-PCR. X.W. and M.U.R. supplied materials/analyses. X.W. provided guidance throughout the study. J.Y. and Y.Z. helped in bioinformatics work. B.A. and X.W. wrote the manuscript. All authors approved the final manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (U1603234), the Program for Innovative Research Team of Grape Germplasm Resources and Breeding (2013KCT-25), and the Fundamental Research Funds for the Central Universities (2452019170).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Suzuki, M.; Kao, C.Y.; McCarty, D.R. The Conserved B3 Domain of VIVIPAROUS1 has a Cooperative DNA Binding Activity. *Plant Cell* **1997**, *9*, 799–807. [PubMed]
2. Tsukagoshi, H.; Morikami, A.; Nakamura, K. Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2543–2547. [CrossRef] [PubMed]
3. Suzuki, M.; Wang, H.H.; McCarty, D.R. Repression of the Leafy Cotyledon 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC Acid Insensitive 3-LIKE B3 genes. *Plant Physiol.* **2007**, *143*, 902–911. [CrossRef] [PubMed]
4. Kagaya, Y.; Ohmiya, K.; Hattori, T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* **1999**, *27*, 470–478. [CrossRef] [PubMed]
5. Ulmasov, T.; Murfett, J.; Hagen, G.; Guilfoyle, T.J. Aux/IAA Proteins Repress Expression of Reporter Genes Containing Natural and Highly Active Synthetic Auxin Response Elements. *Plant Cell* **1997**, *9*, 1963–1971. [PubMed]
6. Franco-Zorrilla, J.M.; Cubas, P.; Jarillo, J.A.; Fernandez-Calvin, B.; Salinas, J.; Martínez-Zapater, J.M. AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. *Plant Physiol.* **2001**, *128*, 418–427. [CrossRef]
7. Swaminathan, K.; Peterson, K.; Jack, T. The plant B3 superfamily. *Trends Plant. Sci.* **2008**, *13*, 647–655. [CrossRef]
8. Yamasaki, K.; Kigawa, T.; Seki, M.; Shinozaki, K.; Yokoyama, S. DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. *Trends Plant Sci.* **2013**, *18*, 267–276. [CrossRef]
9. Waltner, J.K.; Peterson, F.C.; Lytle, B.L.; Volkman, B.F. Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci.* **2005**, *14*, 2478–2483. [CrossRef]

10. McCarty, D.R.; Carson, C.B.; Stinard, P.S.; Robertson, D.S. Molecular analysis of viviparous-1: An abscisic acid-insensitive mutant of maize. *Plant Cell* **1989**, *1*, 523–532. [CrossRef]
11. Giraudat, J.; Hauge, B.M.; Valon, C.; Smalle, J.; Parcy, F.; Goodman, H.M. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **1992**, *4*, 1251–1261. [PubMed]
12. Okushima, Y.; Mitina, I.; Quach, H.L.; Theologis, A. Auxin response factor 2 (ARF2): A pleiotropic developmental regulator. *Plant J.* **2005**, *43*, 29–46. [CrossRef] [PubMed]
13. Guilfoyle, T.J.; Hagen, G. Auxin response factors. *Curr. Opin. Plant. Biol.* **2007**, *10*, 453–460. [CrossRef] [PubMed]
14. Koornneef, M.; Reuling, G.; Karssen, C.M. The isolation and characterization of abscisic-acid insensitive mutants of Arabidopsis-thaliana. *Physiol. Plant.* **1984**, *61*, 377–383. [CrossRef]
15. Baumlein, H.; Misera, S.; Luerssen, H.; Kolle, K.; Horstmann, C.; Wobus, U.; Müller, A.J. The Fus3 Gene of Arabidopsis-Thaliana Is a Regulator of Gene-Expression During Late Embryogenesis. *Plant J.* **1994**, *6*, 379–387. [CrossRef]
16. Stone, S.L.; Braybrook, S.A.; Paula, S.L.; Kwong, L.W.; Meuser, J.; Pelletier, J.; Hsieh, T.F.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis Leafy Cotyledon2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3151–3156. [CrossRef] [PubMed]
17. Hu, Y.X.; Wang, Y.H.; Liu, X.F.; Li, J.Y. Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Res.* **2004**, *14*, 8–15. [CrossRef] [PubMed]
18. Alvarez, J.P.; Goldshmidt, A.; Efroni, I.; Bowman, J.L.; Eshed, Y. The NGATHA distal organ development genes are essential for style specification in Arabidopsis. *Plant Cell* **2009**, *21*, 1373–1393. [CrossRef] [PubMed]
19. Castillejo, C.; Pelaz, S. The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. *Curr. Biol.* **2008**, *18*, 1338–1343. [CrossRef]
20. Levy, Y.Y.; Mesnage, S.; Mylne, J.S.; Gendall, A.R.; Dean, C. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* **2002**, *297*, 243–246. [CrossRef]
21. Sung, S.; Amasino, R.M. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. *Nature* **2004**, *427*, 159–164. [CrossRef] [PubMed]
22. Salunkhe, D.K.; Kadam, S.S. *Handbook of Fruit Science and Technology: Production, Composition, Storage and Processing*; Food Science and Technology; CRC Press: Boca Raton, FL, USA, 1995; 611p.
23. Doligez, A.; Audiot, E.; Baumes, R.; This, P. QTLs for muscat flavor and monoterpenic odorant content in grapevine (*Vitis vinifera* L.). *Mol. Breed.* **2006**, *18*, 109–125. [CrossRef]
24. This, P.; Lacombe, T.; Thomas, M.R. Historical origins and genetic diversity of wine grapes. *Trends Genet.* **2006**, *22*, 511–519. [CrossRef] [PubMed]
25. FAO; OIV. *FAO-OIV Focus Table and Dried Grapes*; Non-Alcoholic Products of the Vitivini Cultural Sector Intended for Human Consumption; FAO: Rome, Italy, 2016; Volume I7042, p. 64.
26. Stout, A.B. Seedlessness in Grapes. In *New York State Agricultural Experiment Station Technical Bulletin*; New York State Agricultural Experiment Station: Geneva, NY, USA, 1936; Volume 238, p. 68.
27. Pratt, C. Reproductive Anatomy in Cultivated Grapes–A Review. *Am. J. Enol. Vitic.* **1971**, *22*, 92–109.
28. Ramnning, D.W.; Emershad, R.L.; Tarllo, R.A. Stenospermocarpic seedless *Vitis vinifera*×*Vitis rotundiyolia* hybrid developed by embryo rescue. *Hort. Sci.* **2000**, *35*, 732–734. [CrossRef]
29. Hanania, U.; Velcheva, M.; Flaishman, M.; Sahar, N.; Perl, A. Silencing of chaperonin 21, that was differentially expressed in inflorescence of seedless and seeded grapes, promoted seed abortion in tobacco and tomato fruits. *Tran. Res.* **2007**, *16*, 515–525. [CrossRef] [PubMed]
30. Royo, C.; Carbonell-Bejerano, P.; Torres-Perez, R.; Nebish, A.; Martinez, O.; Rey, M.; Aroutiounian, R.; Ibanez, J.; Martinez-Zapater, J.M. Developmental, transcriptome, and genetic alterations associated with parthenocarpy in the grapevine seedless somatic variant *Corinto bianco*. *J. Exp. Bot.* **2016**, *67*, 259–273. [CrossRef] [PubMed]
31. Nicolas, P.; Lecourieux, D.; Gomes, E.; Delrot, S.; Lecourieux, F. The grape berry-specific basic helix-loop-helix transcription factor VvCEB1 affects cell size. *J. Exp. Bot.* **2013**, *64*, 991–1003. [CrossRef]
32. Mejia, N.; Soto, B.; Guerrero, M.; Casanueva, X.; Houel, C.; Miccono Mde, L.; Ramos, R.; Le, C.L.; Boursiquot, J.M.; Hinrichsen, P. Molecular, genetic and transcriptional evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. *BMC Plant Biol.* **2011**, *11*, 57. [CrossRef]
33. Wang, L.; Yin, X.; Cheng, C.; Wang, H.; Guo, R.; Xu, X.; Zhao, J.; Zheng, Y.; Wang, X. Evolutionary and expression analysis of a MADS-box gene superfamily involved in ovule development of seeded and seedless grapevines. *Mol. Gen. Genet.* **2015**, *290*, 825–846. [CrossRef]

34. Li, Y.; Zhu, Y.; Yao, J.; Zhang, S.; Wang, L.; Guo, C.; Van, N.; Wang, X. Genome-wide identification and expression analyses of the homeobox transcription factor family during ovule development in seedless and seeded grapes. *Sci. Rep.* **2017**, *7*, 12638. [CrossRef] [PubMed]
35. Li, Z.; Chen, Z.; Yurui, G.; Weili, N.; Yuejin, W.; Yan, X. Evolution and expression analysis reveal the potential role of the HD-Zip gene family in regulation of embryo abortion in grapes (*Vitis vinifera* L.). *BMC Genom.* **2017**, *18*, 744. [CrossRef] [PubMed]
36. Romanel, E.A.; Schrago, C.G.; Couñago, R.M.; Russo, C.A.; Alves-Ferreira, M. Evolution of the B3 DNA binding superfamily: New insights into REM family gene diversification. *PLoS ONE* **2009**, *4*, e5791. [CrossRef] [PubMed]
37. Die, J.V.; Gil, J.; Millan, T. Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*. *BMC Genom.* **2018**, *19*, 301. [CrossRef] [PubMed]
38. Zhao, S.P.; Xu, Z.S.; Zheng, W.J.; Zhao, W.; Wang, Y.W.; Yu, T.F.; Chen, M.; Zhou, Y.B.; Min, D.H.; Ma, Y.Z.; et al. Genome-Wide Analysis of the RAV Family in Soybean and Functional Identification of GmRAV-03 Involvement in Salt and Drought Stresses and Exogenous ABA Treatment. *Front. Plant Sci.* **2017**, *8*, 905. [CrossRef] [PubMed]
39. Zouine, M.; Fu, Y.; Chateigner-Boutin, A.L.; Mila, I.; Frasse, P.; Wang, H.; Audran, C.; Roustan, J.P.; Bouzayen, M. Characterization of the tomato ARF gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS ONE* **2014**, *9*, e84203. [CrossRef] [PubMed]
40. Wang, Y.J.; Wang, S.X.; Deng, D.X.; Bian, Y.L.; Zhang, R.; Yin, Z.T. Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol. Biol. Rep.* **2012**, *39*, 16. [CrossRef] [PubMed]
41. Holub, E.B. The arms race is ancient history in Arabidopsis, the wildflower. *Nat. Rev. Genet.* **2001**, *2*, 516–527. [CrossRef]
42. Wan, S.; Li, W.; Zhu, Y.; Liu, Z.; Huang, W.; Zhan, J. Genome-wide identification, characterization and expression analysis of the auxin response factor gene family in *Vitis vinifera*. *Plant Cell Rep.* **2014**, *33*, 1365–1375. [CrossRef]
43. Yamasaki, K.; Kigawa, T.; Inoue, M.; Watanabe, S.; Tateno, M.; Seki, M.; Shinozaki, K.; Yokoyama, S. Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.* **2008**, *46*, 394–401. [CrossRef]
44. Agarwal, P.; Kapoor, S.; Tyagi, A.K. Transcription factors regulating the progression of monocot and dicotseed development. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **2011**, *33*, 189–202. [CrossRef] [PubMed]
45. Luo, X.C.; Sun, M.H.; Xu, R.R.; Shu, H.R.; Wang, J.W.; Zhang, S.Z. Genome wide identification and expression analysis of the ARF gene family in apple. *J. Genet.* **2014**, *93*, 785–797. [CrossRef] [PubMed]
46. Li, S.B.; Ouyang, W.Z.; Hou, X.J.; Xie, L.L.; Hu, C.G.; Zhang, J.Z. Genome-wide identification, isolation and expression analysis of auxin response factor (ARF) gene family in sweet orange (Citrus sinensis). *Front. Plant Sci.* **2015**, *6*, 119. [CrossRef] [PubMed]
47. Liu, K.; Yuan, C.; Li, H.; Lin, W.; Yang, Y.; Shen, C.; Zheng, X. Genome-wide identification and characterization of auxin response factor (ARF) family genes related to flower and fruit development in papaya (*Carica papaya* L.). *BMC Genom.* **2015**, *16*, 901. [CrossRef] [PubMed]
48. Xia, F.; Sun, T.; Yang, S.; Wang, X.; Chao, J.; Li, X.; Hu, J.; Cui, M.; Liu, G.; Wang, D.; et al. Insight into the B3Transcription Factor Superfamily and Expression Profiling of B3 Genes in Axillary Buds after Topping in Tobacco (*Nicotiana tabacum* L.). *Genes* **2019**, *10*, 164. [CrossRef] [PubMed]
49. Zhang, H.; Jin, J.; Tang, L.; Zhao, Y.; Gu, X.; Gao, G.; Luo, J. Plant TFDB 2.0: Update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res.* **2011**, *39*, D1114–D1117. [CrossRef] [PubMed]
50. Peng, F.Y.; Weselake, R.J. Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants. TAG. Theoretical and applied genetics. *Theor. Appl. Genet.* **2013**, *126*, 1305–1319. [CrossRef] [PubMed]
51. Li, C.W.; Su, R.C.; Cheng, C.P.; Sanjaya; You, S.J.; Hsieh, T.H.; Chao, T.C.; Chan, M.T. Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol.* **2011**, *156*, 213–227. [CrossRef] [PubMed]

52. Wei, Y.; Chang, Y.; Zeng, H.; Liu, G.; He, C.; Shi, H. Rav transcription factors are essential for disease resistance against cassava bacterial blight via activation of melatonin biosynthesis genes. *J. Pineal Res.* **2018**, *64*, e12454. [CrossRef]
53. Xu, G.; Guo, C.; Shan, H.; Kong, H. Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1187–1192. [CrossRef]
54. Dias, A.P.; Braun, E.L.; McMullen, M.D.; Grotewold, E. Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.* **2003**, *131*, 610–620. [CrossRef] [PubMed]
55. Wang, D.; Pei, K.; Fu, Y.; Sun, Z.; Li, S.; Liu, H.; Tang, K.; Han, B.; Tao, Y. Genome-wide analysis of the *auxin response factors (ARF)* gene family in rice (*Oryza sativa*). *Gene* **2007**, *394*, 13–24. [CrossRef] [PubMed]
56. Zhang, H.; Cao, N.; Dong, C.; Shang, Q. Genome-wide Identification and Expression of ARF Gene Family during Adventitious Root Development in Hot Pepper (*Capsicum annuum*). *Hortic. Plant J.* **2017**, *3*, 151–164. [CrossRef]
57. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol.* **2004**, *4*, 1–21. [CrossRef] [PubMed]
58. Jaillon, O.; Aury, J.M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467. [PubMed]
59. Finkelstein, R.R.; Gampala, S.S.L.; Rock, C.D. Abscisic acid signaling in seeds and seedlings. *Plant Cell* **2002**, *14*, S15–S45. [CrossRef] [PubMed]
60. Bassel, G.W.; Mullen, R.T.; Bewley, J.D. ABI3 expression ceases following, but not during, germination of tomato and Arabidopsis seeds. *J. Exp. Bot.* **2006**, *57*, 1291–1297. [CrossRef] [PubMed]
61. Wang, L.; Hu, X.Y.; Jiao, C.; Li, Z.; Fei, Z.J.; Yan, X.X.; Liu, C.H.; Wang, Y.J.; Wang, X.P. Transcriptome analyses of seed development in grape hybrids reveals a possible mechanism influencing seed size. *BMC Genom.* **2016**, *17*, 898. [CrossRef] [PubMed]
62. Zhang, Y.; Gao, M.; Singer, S.D.; Fei, Z.; Wang, H.; Wang, X. Genome-Wide Identification and Analysis of the TIFY Gene Family in Grape. *PLoS ONE* **2012**, *7*, e44465. [CrossRef] [PubMed]
63. Lyons, E.; Pedersen, B.; Kane, J.; Alam, M.; Ming, R.; Tang, H.; Wang, X.; Bowers, J.; Paterson, A.; Lisch, D.; et al. Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* **2008**, *148*, 1772–1781. [CrossRef] [PubMed]
64. Gao, Y.; Liu, J.; Zhang, Z.; Sun, X.; Zhang, N.; Fan, J.; Niu, X.; Xiao, F.; Liu, Y. Functional characterization of two alternatively spliced transcripts of tomato ABSCISIC ACID INSENSITIVE3 (ABI3) gene. *Plant Mol. Biol.* **2013**, *82*, 131–145. [CrossRef] [PubMed]
65. Zhang, Y.; Clemens, A.; Maximova, S.N.; Guiltinan, M.J. The Theobroma cacao B3 domain transcription factor TcLEC2 plays a duel role in control of embryo development and maturation. *BMC Plant Biol.* **2014**, *14*, 106. [CrossRef] [PubMed]
66. Bhattacharjee, A.; Ghangal, R.; Garg, R.; Jain, M. Genome-Wide Analysis of Homeobox Gene Family in Legumes: Identification, Gene Duplication and Expression Profiling. *PLoS ONE* **2015**, *10*, e01191982015. [CrossRef] [PubMed]
67. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]
68. Liu, Y.; Jiang, H.; Chen, W.; Qian, Y.; Ma, Q.; Cheng, B.; Zhu, S. Genome-wide analysis of the auxin response factor (ARF) gene family in maize (*Zea mays*). *Plant Growth Regul.* **2011**, *63*, 225–234. [CrossRef]
69. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Differential Regulation of Anthocyanins in Green and Purple Turnips Revealed by Combined De Novo Transcriptome and Metabolome Analysis

Hongmei Zhuang [1], Qian Lou [2], Huifang Liu [1], Hongwei Han [1], Qiang Wang [1], Zhonghua Tang [3,4], Yanming Ma [4] and Hao Wang [1,*]

1 Institute of Horticultural Crops, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China
2 College of Horticulture, Northwest A & F University, Yangling 712100, China
3 Key Laboratory of Plant Ecology, Northeast Forestry University, Harbin 150040, China
4 Institute of Genetic Resources, Xinjiang Academy of Agricultural Science, Urumqi 830091, China
* Correspondence: wanghao183@163.com; Tel.: +86-15009910026

Received: 17 August 2019; Accepted: 31 August 2019; Published: 6 September 2019

Abstract: Purple turnip *Brassica rapa* ssp. rapa is highly appreciated by consumers but the metabolites and molecular mechanisms underlying the root skin pigmentation remain open to study. Herein, we analyzed the anthocyanin composition in purple turnip (PT) and green turnip (GT) at five developmental stages. A total of 21 anthocyanins were detected and classified into the six major anthocynanin aglycones. Distinctly, PT contains 20 times higher levels of anthocyanins than GT, which explain the difference in the root skin pigmentation. We further sequenced the transcriptomes and analyzed the differentially expressed genes between the two turnips. We found that PT essentially diverts dihydroflavonols to the biosynthesis of anthocyanins over flavonols biosynthesis by strongly down-regulating one flavonol synthase gene, while strikingly up-regulating dihydroflavonol 4-reductase (DFR), anthocyanidin synthase and UDP-glucose: flavonoid-3-*O*-glucosyltransferase genes as compared to GT. Moreover, a nonsense mutation identified in the coding sequence of the DFR gene may lead to a nonfunctional protein, adding another hurdle to the accumulation of anthocyanin in GT. We also uncovered several key members of MYB, bHLH and WRKY families as the putative main drivers of transcriptional changes between the two turnips. Overall, this study provides new tools for modifying anthocyanin content and improving turnip nutritional quality.

Keywords: pigment; turnip; gene expression; antioxidant; nutritional quality

1. Introduction

Turnip (*Brassica rapa* ssp. rapa), belongs to the Cruciferae family and represents one of the most important leaf and root vegetable crops for human consumption and animal fodder in China and throughout East Asia. Turnip vegetables provide dietary fiber, vitamin C, high amounts of glucosinolates [1–3], and are also an important source of dietary phenolic and other bioactive compounds [4,5]. There are several turnip varieties with purple colored root skin, which are highly appreciated by consumers. Similar to turnip, there are various *B. rapa* subspecies enriched with purple pigments previously characterized as anthocyanins [6].

Anthocyanins are secondary metabolites with health-promoting virtues, such as anti-oxidation, anti-mutation, prevention of cardiovascular disease, liver protection, and inhibiting the metastasis of tumor cells [7–12]. Besides, they also play important fundamental physiological functions in plants, including UV protection, pigmentation of flowers and fruits to attract pollinators and for seed dispersal, and responses to biotic and environmental stresses [13–20]. Therefore, the identification, analysis and

genetic manipulation of anthocyanin metabolites have become an important topic in plant secondary metabolite research [21].

The biosynthesis and accumulation of anthocyanins are determined by metabolic networks correlated with the expression of several genes and regulatory factors [22]. During the past decades, extensive studies have been conducted to elucidate the biosynthetic pathway of anthocyanins in plants. Progressively, it has become evident that the anthocyanin biosynthetic pathway is a very well conserved network in plant species [23]. It starts with the chalcone synthase (CHS) mediated synthesis of naringenin chalcone from 4-coumaroyl-CoA and malonyl-CoA. Then, naringenin chalcone is isomerized by chalcone isomerase (CHI) to naringenin. Flavanone 3-hydroxylase (F3H) converts naringenin into dihydrokaempferol which can be further hydroxylated by flavonoid 3′-hydroxylase (F3′H) or flavonoid 3′,5′-hydroxylase (F3′5′H) into two other dihydroflavonols, dihydroquercetin and dihydrotricetin, respectively. Then, the three dihydroflavonols are converted into colorless leucoanthocyanidins by dihydroflavonol 4-reductase (DFR) and subsequently to colored anthocyanidins by anthocyanidin synthase (ANS). Anthocyanidins are glycolsylated to facilitate their accumulation in cells by the enzyme flavonoid 3-*O*-glucosyltransferase (UFGT) and might be further acylated with aromatic acyl groups by acyltransferases [18,22]. Although a well conserved biosynthetic pathway in plants, several studies have shown the species-specific peculiarity of anthocyanin regulation. For example, the numbers of structural genes (CHS, F3H, F3′H, CHI, DFR, ANS, UFGT, F3′5′H, etc.) vary considerably across species as do their expression levels [24–27]. Also, genetic mutations, microRNAs, transcription factors such as MYB, bHLH, WD40, WRKY, NAC, etc., have been linked to the regulation of anthocyanin biosynthetic structural genes through varying complex mechanisms among plants [14,26–41]. Therefore, in order to pinpoint the major players associated with quantitative and qualitative variations of anthocyanins in plant, a thorough investigation is necessary.

Herein, we investigated the anthocyanin compositions at different developmental stages in root skin of two turnip varieties (purple turnip and green turnip) widely grown in Xinjiang (China). In addition, we generated extensive transcriptome data and profiled the key genes involved in the differential pigmentation. The goal of this work was to elucidate the molecular and metabolic mechanisms underlying the differential pigmentation in turnips, as a foundation for the development of turnip varieties that are rich in anthocyanin compounds to meet the increasing demand for health-promoting components in our daily diet.

2. Results

2.1. De Novo Transcriptome Assembly and Gene Expression Profiles in the Two Turnips at Five Developmental Stages

In this work, two widely grown *Brassica rapa* ssp. rapa varieties in Xinjiang (China) including GT with green-colored root skin and PT with purple-colored root skin were studied. Skin samples were collected at five different developmental stages, namely seedling stage (S1, 15 days after sowing (DAS)), early stage of fleshy root expansion (S2, 30 DAS), full expansion stage of fleshy root (S3, 45 DAS), maturity stage of fleshy root (S4, 55 DAS) and harvest stage of fleshy root (S5, 65 DAS). The phenotypes of young and mature turnip roots for GT and PT are presented in Figure 1. The objective of the work was to elucidate the mechanisms underlying the differential skin coloration in these turnips with respect to the accumulation of anthocyanin compounds. First, we de novo sequenced and assembled the transcriptome from 30 samples of the two turnips.

The RNA-seq yielded a total of 232.22 Gb clean data, on average 6.20Gb for each sample with 90.74% of bases scoring Q30 and above (Table 1). A total of 76,152 unigenes were obtained after assembly using the Trinity software and 17,594 unigenes have length of more than 1 kb. The N50 length obtained was approximately 1443 bp (Table 2). The detected gene number in this study was much higher than the reported gene number (41,174 genes) in *Brassica rapa* ssp. pekinensis variety Chiifu-401-42 [42] or (40,708 genes) in *Brassica rapa* ssp. rapa [43]. We performed the functional annotation of the unigenes in various database, including NR, Swiss-Prot, KEGG, COG, KOG, GO and

Pfam databases, which resulted in 52,449 unigenes successfully annotated (Table 3). The clean data of each sample was serialized with the assembled unigene libraries and the mapping result statistics are presented in Table S1. Gene expression levels were estimated with the fragments per kilobase of exon per million fragments mapped (FPKM) values ranging from 0.04 to 5,566,185 (Figure 2A).

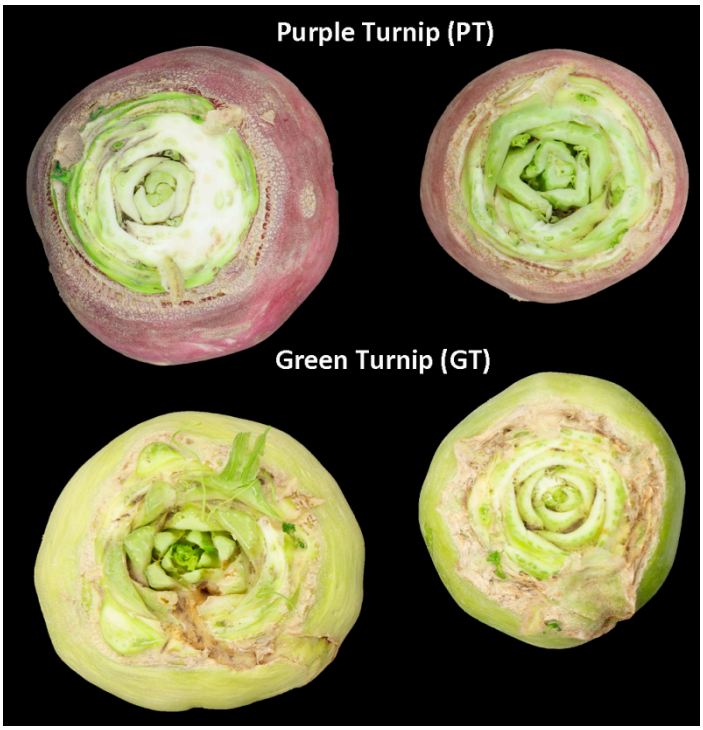


Figure 1. The phenotypes of young and mature purple-colored turnip and green-colored turnip roots.

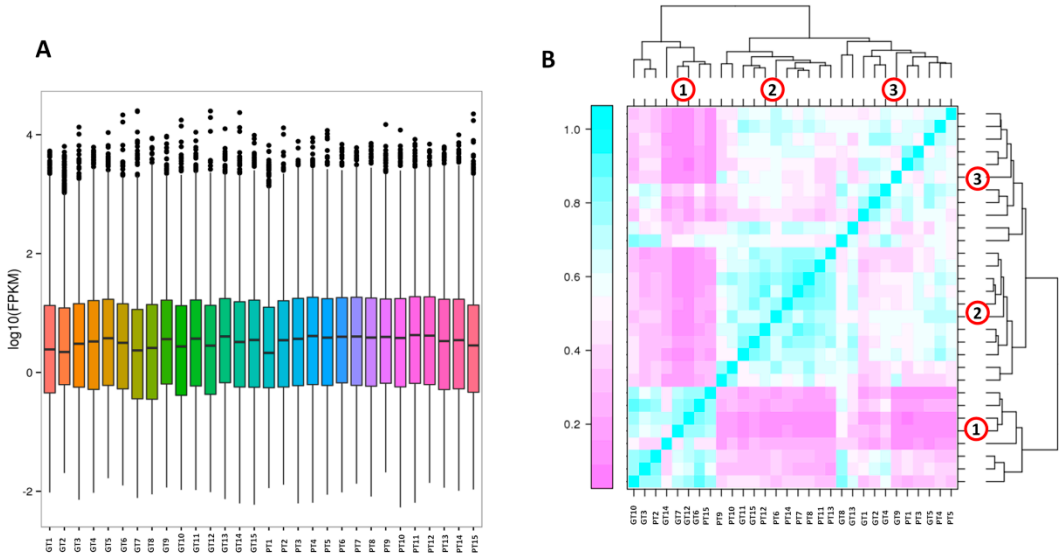


Figure 2. Overview of the transcriptome sequencing. (**A**) Gene expression profiles in the 30 libraries. PT represents the purple turnip while GT represents the green turnip; (**B**) heatmap clustering showing correlation among turnip samples based on global expression profiles. The number 1, 2 and 3 represent the 3 Clusters of samples.

Table 1. Overview of the transcriptome sequencing dataset and quality check.

Samples ID	Read Number	Base Number	GC Content	% ≥ Q30
GT1	29,185,016	8,726,667,568	47.42%	91.63%
GT2	20,717,765	6,196,426,170	47.42%	91.70%
GT3	21,156,763	6,319,382,840	47.27%	91.95%
GT4	27,588,012	8,249,989,480	47.55%	93.05%
GT5	21,971,588	6,574,089,374	47.05%	92.05%
GT6	23,582,547	7,053,356,366	46.92%	91.97%
GT7	37,452,770	11,183,417,788	46.90%	90.76%
GT8	32,252,399	9,647,120,456	47.09%	91.89%
GT9	20,867,338	6,235,980,998	47.36%	90.74%
GT10	32,765,445	9,782,653,688	47.16%	91.39%
GT11	23,201,289	6,937,494,090	47.56%	91.16%
GT12	31,341,050	9,370,066,168	47.21%	91.70%
GT13	20,754,387	6,205,716,562	47.56%	92.23%
GT14	24,312,743	7,274,160,626	47.35%	92.91%
GT15	26,794,514	8,013,487,372	47.48%	92.74%
PT1	24,145,033	7,223,733,604	47.23%	92.31%
PT2	27,681,645	8,273,851,708	47.04%	92.79%
PT3	26,077,522	7,801,143,252	47.51%	92.07%
PT4	25,070,137	7,501,916,476	47.29%	92.08%
PT5	24,021,197	7,187,109,432	47.42%	92.39%
PT6	22,053,479	6,589,934,434	47.30%	92.60%
PT7	24,707,235	7,377,482,668	47.25%	92.60%
PT8	25,937,942	7,760,138,900	47.43%	92.87%
PT9	21,983,093	6,562,135,072	47.19%	91.87%
PT10	28,772,133	8,603,239,086	47.33%	92.59%
PT11	24,124,411	7,219,966,492	47.46%	92.28%
PT12	25,846,725	7,730,236,716	47.35%	92.39%
PT13	26,784,282	8,007,970,564	47.24%	91.47%
PT14	26,748,564	7,995,602,902	47.50%	92.14%
PT15	28,865,986	8,619,275,022	47.07%	92.08%

Table 2. Statistics of the unigene assembly results.

Length Range (bp)	Transcript Number	Unigene Number
200–300	35,501 (12.87%)	27,206 (35.73%)
300–500	33,539 (12.15%)	17,833 (23.42%)
500–1000	55,072 (19.96%)	13,519 (17.75%)
1000–2000	93,093 (33.74%)	10,680 (14.02%)
2000+	58,724 (21.28%)	6914 (9.08%)
Total Number	275,929	76,152
Total Length	376,323,528	59,450,389
N50 Length	1921	1443
Mean Length	1363.84	780.68

Table 3. Functional annotation statistics of the unigenes.

Annotation Database	Annotated Number	300 ≤ Length < 1000	Length ≥ 1000
COG_Annotation	12,979	4215	6007
GO_Annotation	34,789	13,243	13,960
KEGG_Annotation	17,096	6870	5729
KOG_Annotation	28,073	11,024	9424
Pfam_Annotation	30,231	10,820	13,789
Swissprot_Annotation	29,952	11,517	11,893
eggNOG_Annotation	46,472	18,520	15,908
nr_Annotation	48,834	19,848	16,761
All_Annotated	52,449	21,373	16,967

Hierarchical clustering of the samples based on FPKM displayed 3 Clusters of samples. We did not observe a clear separation according to the developmental stages but to some extent, Clusters were related to the turnip variety. For example, Cluster 1 gathered mostly samples of GT, Cluster 2 grouped samples of PT, and Cluster 3 had two subgroups each mostly made of samples from a unique variety (Figure 2B). These results indicate that the global gene expression profile is quite uniform regardless of the developmental stages and only few differentially expressed genes between the two varieties may be associated with the difference in the turnip skin coloration.

2.2. Differentially Expressed Genes Between the Two Turnips and Analysis of Major Regulator Genes

To identify the differentially expressed genes (DEG) related to turnip skin coloration, we compared the FPKM values of each gene in PT to GT at the different developmental stages and retained DEGs with fold change > 2 and a false discovery rate (FDR) correction set at $p < 0.01$ [44]. We detected 242, 194, 807, 459 and 199 DEGs at S1, S2, S3, S4 and S5, respectively (Figure 3A). The marked change in gene expression between the two turnips observed at the S3, implies that S3 may represent a key stage for turnip skin coloration.

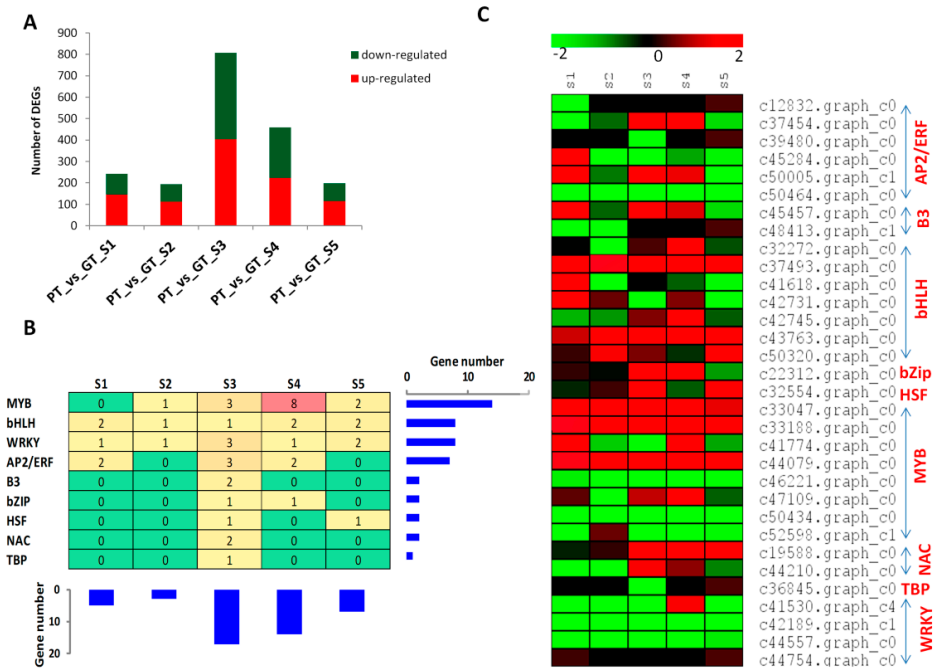


Figure 3. Transcription factors (TF) regulating the gene expression between the two turnips. (**A**) Number of up- and down-regulated genes between PT and GT at different developmental stages S1-S5. PT represents the purple turnip while GT represents the green turnip; (**B**) overview of the enriched TF family. The combined histograms showed the number of occurrence of genes belonging to each TF family or at each developmental stage; (**C**) Heatmap displaying the expression fold change (Log2 fold change) between PT and GT for the gene encoding transcription factors.

Transcription factors (TF) are the major regulators of gene expression profiles [45]. We therefore extended the study on the major transcription factor families differentially expressed between the two turnips. Our analysis showed that nine main TF families modulate the global gene expression levels among the two turnips (Figure 3B,C). In addition, the highest number of TF could be noticed at S3, which correlates well with the observed significant DEGs at this developmental stage (Figure 3A,B), showing that TFs are the main drivers of gene expression changes leading to the differential turnip skin coloration. Among the detected TF families, MYB, bHLH and WRKY families showed more active members involved in gene regulation (Figure 3B,C), therefore we deduce that these TF families may be crucial for the regulation of structural genes involved in turnip skin coloration. Distinctly, the genes

c42189.graph_c1 (WRKY) was strongly down-regulated over most of the developmental stages in PT while the genes *c33188.graph_c0* (MYB), *c44079.graph_c0* (MYB) and *c37493.graph_c0* (bHLH) exhibited the opposite trend. Given the role of WRKY, MYB and bHLH TFs in the regulation of structural genes involved in pigment (flavonoid-anthocyanin) biosynthesis in plants [33], it is tempting to speculate that these four key genes are the major regulators during turnip skin color formation.

2.3. Detection of Anthocyanin Compounds in the Two Turnips

Anthocyanins are the most important flavonoid colorants in plants [46]. We detected and determined 18, 18, 18 and 19 diverse anthocyanins at S1, S3, S4 and S5, respectively, resulting in 21 unique anthocyanin compounds in skin of the two turnips using the targeted-metabolomics approach (Table 4, Table S2). It is worth mentioning that the skin samples collected at the S2 from GT were deteriorated, so have not been used for this analysis. We did not detect procyanidin A1 and procyanidin A2 in PT, while GT did contain four anthocyanins: pelargonin and pelargonidin 3-*O*-beta-D-glucoside, pelargonidin *O*-acetylhexoside and cyanidin *O*-acetylhexoside (Table 4). With the determined quantities of all anthocyanins combined together, PT skin contained 20 times higher level of anthocyanins as compared to GT (Table S2). Also, the highest difference in total anthocyanin content between PT and GT during the developmental stages was observed at S3 (Figure 4A), which further supports the premise that S3 is the key for stage for turnip skin coloration. We investigated the differential accumulated metabolites (DAM) between the two turnips based on the variable importance in projection (VIP) ≥ 1 and fold change ≥ 2 or fold change ≤ 0.5 [47]). A total of 14, 14, 17 and 17 DAMs were recorded at S1, S3, S4 and S5, respectively (Figure 4A–E), resulting in all the 21 unique anthocyanin compounds differentially accumulated at least at one developmental stage between the two turnips. In addition, nine anthocyanins were constantly differentially accumulated in skin of the two turnips at all the four developmental stages (Figure 4F, Table S3). For most of these anthocyanin compounds, they showed presence/absence patterns among the two turnips skins, indicating that they are the key components conferring the differential pigmentation.

Table 4. Anthocyanins detected in two turnip varieties.

Index	KEGG ID	Compounds	Class
Bra18	-	Peonidin *O*-hexoside	Anthocyanins
Bra23	-	Rosinidin *O*-hexoside	Anthocyanins
Bra1	-	Delphinidin *O*-malonylhexoside	Anthocyanins
Bra8	C08604	Cyanidin 3-*O*-glucoside	Anthocyanins
Bra22	-	Delphinidin *O*-malonyl-malonylhexoside	Anthocyanins
Bra28	C08726	Peonidin	Anthocyanins
Bra32	-	Cyanidin *O*-acetylhexoside	Anthocyanins
Bra35	-	Pelargonidin *O*-acetylhexoside	Anthocyanins
Bra26	-	Procyanidin A1	Proanthocyanidins
Bra29	C10237	Procyanidin A2	Proanthocyanidins
Bra20	-	Procyanidin B2	Proanthocyanidins
Bra34	-	Procyanidin B3	Proanthocyanidins
Bra17	C05908	Delphinidin	Anthocyanins
Bra27	C05904	Pelargonidin	Anthocyanins
Bra5	C12138	Delphinidin 3-*O*-glucoside	Anthocyanins
Bra12	C08620	Cyanidin 3-*O*-rutinoside	Anthocyanins
Bra9	C08718	Malvidin 3,5-diglucoside	Anthocyanins
Bra10	C08725	Pelargonin	Anthocyanins
Bra7	C12139	Petunidin 3-*O*-glucoside	Anthocyanins
Bra11	-	Pelargonidin 3-*O*-beta-D-glucoside	Anthocyanins
Bra25	C05905	Cyanidin	Anthocyanins

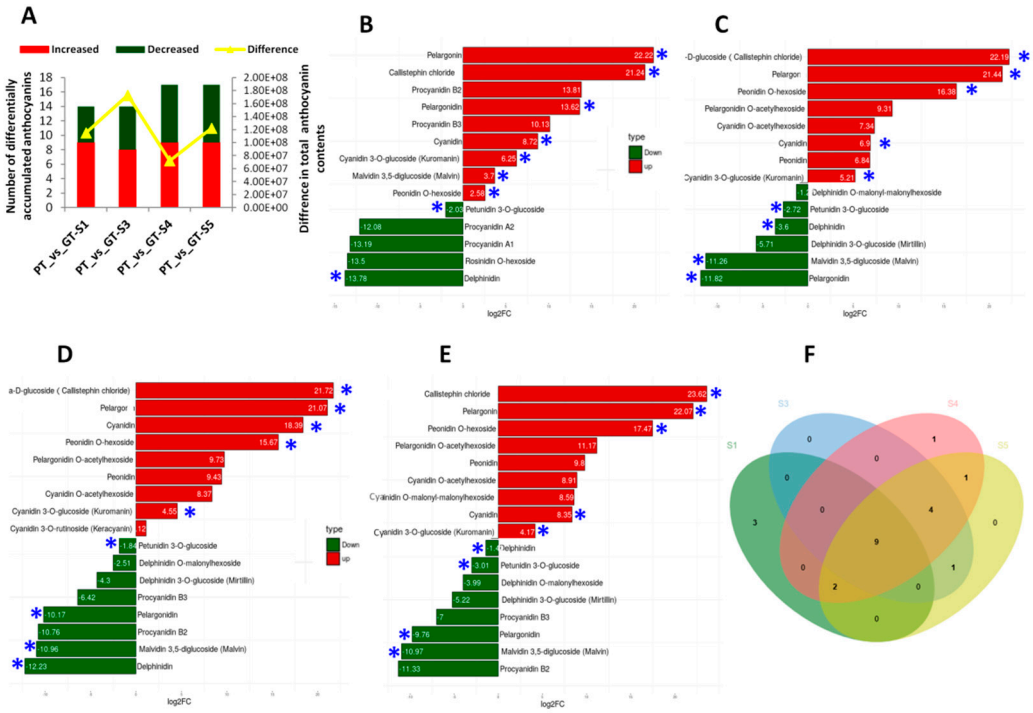


Figure 4. Temporal changes of anthocyanin content in two turnips. (**A**) Number of increased and decreased anthocyanins between PT and GT at different developmental stages S1–S5. The yellow line shows the difference in total anthocyanin content between PT and GT. PT represents the purple turnip while GT represents the green turnip; (**B**–**E**) differentially accumulated anthocyanins at S1, S3, S4, and S5 and their log2 fold change values between PT and GT. The asterisks mark the metabolites constantly differentially accumulated at the 4 developmental stages between PT and GT. (**F**) Venn diagram depicting the number of shared and unique differential accumulated anthocyanins between PT and GT at the 4 developmental stages.

2.4. Mapping of Differential Genes and Metabolites Related to Flavonoid-Anthocyanin Biosynthesis Pathway

The biosynthetic pathway of anthocyanins has been well-characterized in plants [22]. In order to predict the molecular mechanisms leading to the differential skin coloration in the two turnips, we have reconstructed the flavonoid-anthocyanin biosynthesis pathways (Figures 5 and 6). First, we searched among the DEGs, those coding for enzymes involved in the flavonoid-anthocyanin biosynthesis pathways. We obtained four genes namely, flavonol synthase [EC:1.14.11.23] (*c43941.graph_c0*, FLS), dihydroflavonol-4-reductase [EC:1.1.1.234] (*c39842.graph_c0*, DFR), anthocyanidin synthase [EC:1.14.11.19] (*c45741.graph_c0*, ANS), and UDP-flavonoid glucosyl transferase [EC:2.4.1.91] (*c48211.graph_c0*, UFGT). All these enzymes mainly participate in the late steps of the flavonoid-anthocyanin biosynthesis pathways.

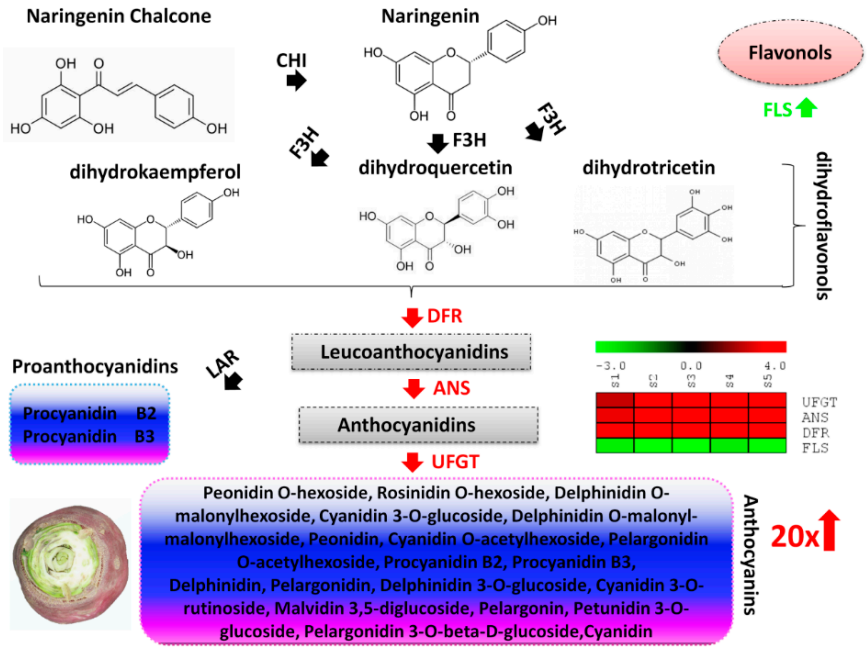


Figure 5. Proposed model of the molecular mechanism leading to the high anthocyanin content in the the purple turnip (PT). Naringenin chalcone is isomerized by chalcone isomerase (CHI) to naringenin. Flavanone 3-hydroxylase (F3H) converts naringenin into dihydroflavonols (dihydrokaempferol, dihydroquercetin or dihydrotricetin). Then, the three dihydroflavonols are converted into colorless leucoanthocyanidins by dihydroflavonol 4-reductase (DFR) and subsequently to colored anthocyanidins by anthocyanidin synthase (ANS). Anthocyanidins are glycolsylated to facilitate their accumulation in cells by the enzyme flavonoid 3-*O*-glucosyltransferase (UFGT). Proanthocyanidins are generated by the action of leucoanthocyanidin reductase (LAR) from leucoanthocyanidins. DFR, ANS and UFGT were found significantly up-regulated in PT leading to a high content of 17 anthocyanins compounds (more than 20 times compared to the green turnip). In contrast, FLS was found significantly down-regulated and may lead to a weak accumulation of flavonols. PT tends to prioritize anthocyanins accumulation by diverting dihydroflavonols to the anthocyanins biosynthesis pathway.

The flavonoid-anthocyanin biosynthesis pathways start from the key amino acid phenylalanine to produce 4-coumaroyl CoA by phenylalanine ammonia-lyase, cinnamic acid 4-hydroxylase and 4-coumarate CoA ligase [48]. The main precursors for flavonoids are 4-coumaroyl CoA and three molecules of malonyl CoA that produce chalcone by chalcone synthase (Dixon and Steele, 1999). Then, the pathway is catalyzed by a number of enzymes to yield flavanones (via chalcone isomerase), dihydroflavonols (via flavanone 3-hydroxylase) [49]. Dihydroflavonols are the keystone substrates for the biosynthesis of flavonols (via FLS) and anthocyanins (via DFR). In this study, we observed a constant down-regulation of one FLS in PT while a significant up-regulation of the expression level of one DFR during all the five developmental stages (Figures 5 and 6), indicating that PT tends to prioritize the anthocyanins biosynthesis over flavonols. Next, leucoanthocyanidins which are generated from DFR are converted into anthocyanidins (via ANS) [48]. Similar to DFR, we noticed a stout up-regulation of one ANS throughout PT growth pointing to high accumulation of anthocyanidins (Figures 5 and 6). Finally, anthocyanidins are converted into anthocyanins via UFGT [49]. We identified one UFGT significantly and constantly up-regulated in PT, showing a mechanism towards a strong accumulation of anthocyanins (Figures 5 and 6). Based on the metabolite detection and quantification, we confirm that the accumulated anthocyanins conferring the purple pigmentation in PT are mainly peonidin *O*-hexoside, cyanidin 3-*O*-glucoside (kuromanin), pelargonidin, malvidin 3,5-diglucoside (calvin), pelargonin, pelargonidin 3-*O*-beta-D-glucoside (callistephin chloride) and cyanidin (Figures 5 and 6).

On the opposite, delphinidin and petunidin 3-*O*-glucoside are enriched in the skin of GT and may confer the greenish coloration (Figures 5 and 6).

To confirm the differential expression levels of the four candidate structural genes together with the four key transcription factors detected by the RNA-seq analysis, we performed a quantitative real-time PCR (Table S4). The results showed that the genes *c43941.graph_c0* (FLS) and *c42189.graph_c1* (WRKY) were obviously down-regulated over the developmental stages in PT while the genes c39842.graph_c0 (DFR), *c45741.graph_c0* (ANS), *c48211.graph_c0* (UFGT), *c33188.graph_c0* (MYB), *c44079.graph_c0* (MYB) and *c37493.graph_c0* (bHLH) were all found clearly up-regulated over the developmental stages in PT (Figure 7A–H). The qRT-PCR results were therefore in perfect concordance with the RNA-seq report.

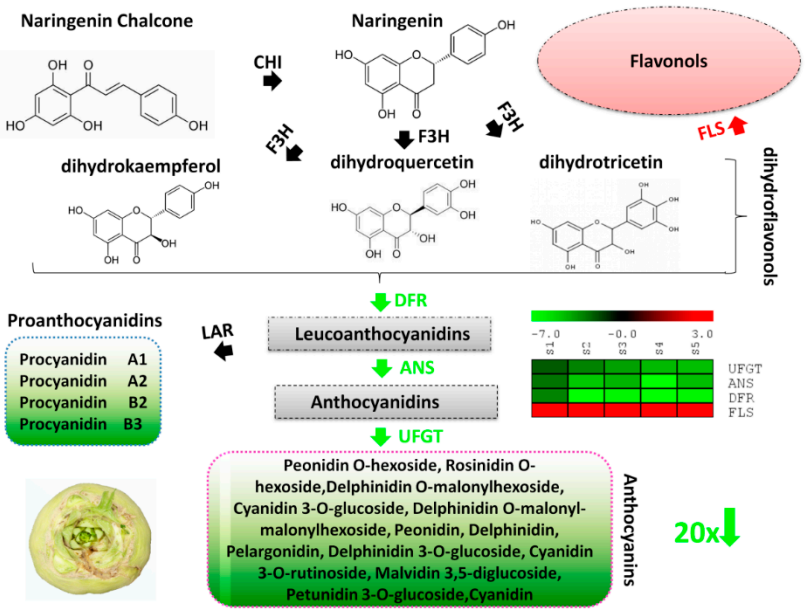


Figure 6. Proposed model of the mechanism leading to the low anthocyanin content in the green turnip (GT). Naringenin chalcone is isomerized by chalcone isomerase (CHI) to naringenin. Flavanone 3-hydroxylase (F3H) converts naringenin into dihydroflavonols (dihydrokaempferol, dihydroquercetin or dihydrotricetin). Then, the three dihydroflavonols are converted into colorless leucoanthocyanidins by dihydroflavonol 4-reductase (DFR) and subsequently to colored anthocyanidins by anthocyanidin synthase (ANS). Anthocyanidins are glycolsylated to facilitate their accumulation in cells by the enzyme flavonoid 3-*O*-glucosyltransferase (UFGT). Proanthocyanidins are generated by the action of leucoanthocyanidin reductase (LAR) from leucoanthocyanidins. DFR, ANS and UFGT were found significantly down-regulated in GT leading to a low content of 13 anthocyanins compounds (less than 20 times compared to the purple turnip). In contrast, FLS was found significantly up-regulated and may lead to a high accumulation of flavonols. PT tends to prioritize flavonol accumulation by diverting dihydroflavonols to the flavonols biosynthesis pathway.

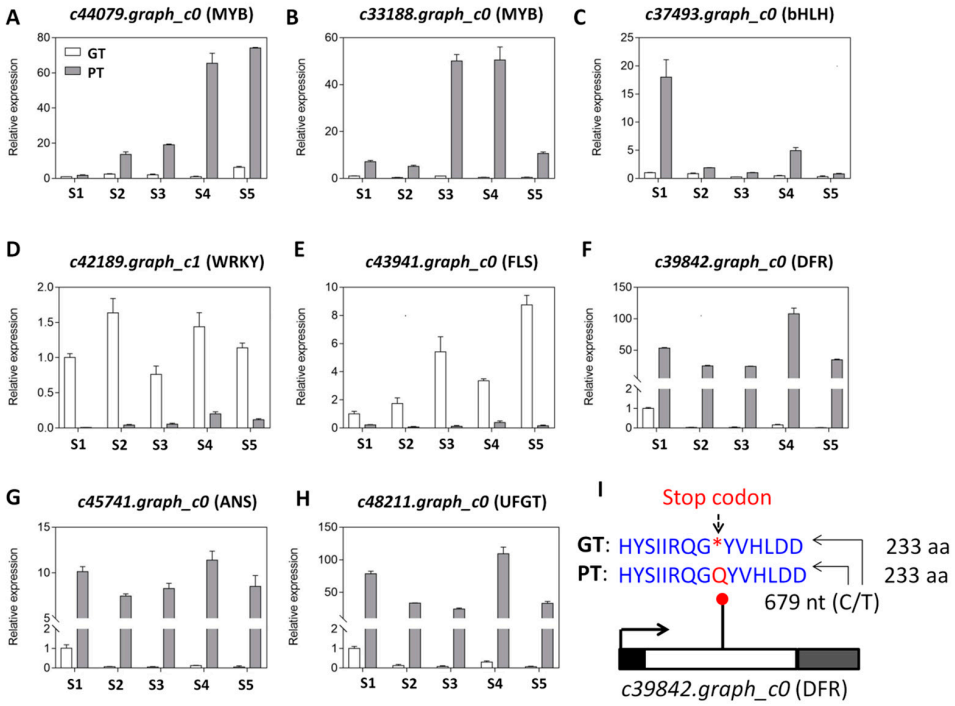


Figure 7. Quantitative real time PCR validation of selected candidate genes predicted to differentially affect the anthocyanin profiles in the two turnips. (**A–H**) Relative expression level of *c33188.graph_c0* (MYB), *c44079.graph_c0* (MYB), *c37493.graph_c0* (bHLH), *c42189.graph_c1* (WRKY), *c43941.graph_c0* (FLS), *c39842.graph_c0* (DFR), *c45741.graph_c0* (ANS) and *c48211.graph_c0* (UFGT) between PT and GT at five developmental stages (S1–S5). PT represents the purple turnip while GT represents the green turnip and are represented by the grey and white bars, respectively. The error bar represents the SD of biological replicates. The *Actin* gene was used as the internal reference gene for normalization; (**I**) identification of a non-sense mutation in the gene *c39842.graph_c0* (DFR) by comparing the sequences between PT and GT. The single nucleotide polymorphism (C/T) is located at the position 679 within the coding sequence of the gene and is predicted to generate an amino acid (aa) Q in PT while a stop codon in GT. The white box represents the exon while the black and gray boxes represent the UTR5'and UTR3', respectively. The arrow indicates the transcription start site and transcription orientation.

2.5. Detection of SNPs within the Four Candidate Structural Genes Regulating the Differential Skin Coloration in Turnips

Differential gene expression among individuals is not only caused by the modulation of transcription factors but could result from variations in the nucleotide sequences [50]. Herein, we investigated the single nucleotide polymorphisms (SNP) within the sequences of the four differentially expressed candidate structural genes (*c43941.graph_c0*, *c39842.graph_c0*, *c45741.graph_c0* and *c48211.graph_c0*) associated with the anthocyanin biosynthesis pathways, which were predicted to modulate the pigment formation in turnip skin. Sequence comparison of the unigenes between the two turnips unveiled a putative SNP in *c39842.graph_c0* (DFR). In the DFR gene which has a single coding sequence (CDS) of 1,158 nucleotides (nt), a point-nonsense mutation (C/T) at the position 679 nt was detected in the CDS. The SNP has an allele depth (number of reads) of 0/51 for GT and 113,774/0 for PT. This indicates that the allele T which induces a stop codon in the resulting protein is likely to be present only in GT while the allele C is apparently only present in PT (Figure 7I).

3. Discussion

Anthocyanin containing fruits and vegetables are an integral part of human diet without any known adverse effect [51]. In this study, we profiled the anthocyanin composition of two widely grown turnips (*Brassica rapa* ssp. rapa) in Xinjiang (China), with purple- (PT) and green-colored (GT) skins (Figure 1)

and investigated the underlying genetic basis. Common aglycones of anthocyanin are pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin [52]. Using the ultra-performance liquid chromatography and tandem mass spectrometry technologies, we determined 17 anthocyanin compounds in the turnip skins, classified into the common six aglycones of anthocynanins (Table 4, Table S2). In addition, we also detected four proanthocyanidins compounds, including procyanidin A1, procyanidin A2, procyanidin B2 and procyanidin B3 (Table 4, Table S2). There were no formal studies on the anthocyanins composition in *Brassica rapa* ssp. rapa, however, according to previous studies, cyanin glycosides are the major anthocyanin substances accumulated in Brassica crops [53–56]. Nine cyanidin anthocyanins were detected in purple cauliflower and purple cabbage [53]. Li [54] analyzed anthocyanin extracts of 'violet' purple cabbage and obtained eight different anthocyanin components, whose basic component is cyanidin-malonyl-glucoside. Later on, Guo et al. [55] examined purple seaweed sprouts, purple turnips, and purple cabbage and identified 23 anthocyanins compounds, composed of 17 cyanidins and six pelargonidin. Recently, Park et al. [56] also found 11 anthocyanins, predominantly cyanindin in purple Kohlrabi (*Brassica oleracea* var. gongylodes). We deduce that there is a large variation of the anthocyanin profiles among purple Brassica vegetables but *Brassica rapa* ssp. rapa has one of the most diversified anthocyanin metabolites in root, which may confer a superior health-promoting attribute. Moreover, comparative analysis of the anthocyanin contents in PT and GT turnip skins in this study showed that they have differential profiles (Figure 4, Table S2) but PT had 20 times more anthocyanin levels than GT, which may explain the difference in their skin coloration.

Variation in anthocyanin content in plants has been linked to the differential expression of key genes encoding structural enzymes involved in the anthocyanin biosynthesis pathways [57,58]. These genes have been classified as early biosynthesis genes and include chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone-3-hydroxylase (F3H), while others are classified as late biosynthesis genes, including dihydroflavonol-4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-flavonoid glucosyl transferase (UFGT) [22]. To uncover the key structural genes modulating the differential pigmentation in skin of PT and GT, we de novo sequenced and assembled the whole transcriptome from skin samples collected at five developmental stages by the RNA-sequencing technology (Tables 1–3). Although Lin et al. [43] reported the full sequencing of the genome of *Brassica rapa* ssp. rapa, the sequence is not publicly available and has propelled us for the de novo transcript assembly. A very high number of unigenes was identified in this study (~76,000 genes) as compared to previous reports in *Brassica rapa* genotypes (~40,000 genes) [42,43]. Our results will fuel further investigations on the genetic variation underlying the diverse morphotypes found in this species. Based on the differential gene expression (DEG) analysis and gene annotation, we searched for all DEGs related to the flavonoid-anthocyanin biosynthesis. Our results revealed four DEGs between PT and GT, all classified as late biosynthesis genes. Dihydroflavonols are at the halfway of anthocyanin biosynthesis from the end of the activity of early biosynthesis enzymes and the beginning of the activity of the late biosynthesis enzymes and represent the same substrate for both anthocyanin biosynthesis and flavonols biosynthesis. In various plants, it has been documented that the up-regulation of early biosynthesis genes increases the formation of dihydroflavonols, which later facilitates the high anthocyanin accumulation [30,34,35,37,40]. However, in turnip, we uncovered a different mechanism leading to the differential anthocyanin content (Figures 5 and 6). In fact, the purple turnip (PT) tends to prioritize the anthocyanins biosynthesis over the flavonols biosynthesis by strongly down-regulating one flavonol synthase (*c43941.graph_c0*, FLS) gene, which normally converts dihydroflavonols into flavonols. Then, we predicted that the dihydroflavonols are mainly diverted to the anthocyanins biosynthesis through a strong up-regulation of one DFR gene (*c39842.graph_c0*). This will result in a high accumulation of leucoanthocyanidins in PT. A similar mechanism has been recently discovered in *Mimulus lewisii* [39]. They demonstrated that the gene *LAR1*, encoding an R2R3-MYB transcription factor positively regulates FLS, essentially eliminating anthocyanin biosynthesis in the white region around the corolla throat of *M. lewisii* flowers by diverting dihydroflavonol into flavonol biosynthesis from the anthocyanin pigment pathway [39]. Interestingly, the putative nonsense mutation identified in

the coding sequence of the DFR gene could lead to a nonfunctional protein in GT (Figure 7I), which may impair the accumulation of anthocyanins in GT skin. Moreover, PT strongly activated one ANS gene (*c45741.graph_c0*) which will likely generate high level of anthocyanidins. All these genes (DFR, ANS) are essential for the formation of the higher content of anthocyanins in PT but without glucosylation, anthocyanins are unstable and do not accumulate in the cells to give the purple pigmentation [59]. In this regard, we detected one UFGT gene (*c48211.graph_c0*) strongly up-regulated in PT and will likely favor the high accumulation of purple anthocyanin pigments in PT (Figure 5). Analogically, we deduced that the green turnip (GT) prioritizes flavonols biosynthesis through a high activity of the FLS gene and strongly reduces anthocyanin accumulation by down-regulating DFR, ANS and UFGT genes (Figure 6). Nonetheless, it is still unclear how the different anthocyanin profiles were generated in both varieties. For example, how delphinidin and petunidin 3-*O*-glucoside accumulates to higher levels in GT or why pelargonin and pelargonidin 3-*O*-beta-D-glucoside, pelargonidin *O*-acetylhexoside, cyanidin *O*-acetylhexoside are only detected in PT, will require additional investigations.

The activities of the structural genes in the flavonoid-anthocyanin biosynthetic pathways are regulated by other genes, predominantly transcription factors (TF) from the families of MYB, bHLH and WD40, which form ternary complexes called MBW [28,29,31,33]. Accordingly, in this study we also uncovered several members of MYB and bHLH families as the main drivers of transcriptional changes between the two turnips (Figure 3). However, we did not find any annotated gene corresponding to the WD40 within the DEGs. An et al. [32] have shown that the ternary complexes MBW is not indispensable for the regulation of anthocyanin genes in apple, therefore, with pending in-depth investigation, we are tempted to speculate that MYB and bHLH may be sufficient for the regulation of anthocyanin structural genes in turnips. Many WRKY genes were also differentially expressed between the two turnips (Figure 3), suggesting that they may play key roles. Lei et al. [60] demonstrated that *WRKY2* and *WRKY34* negatively regulate the expression of certain MYBs during plant male gametogenesis. Similarly, *AtWRKY40* binds to the W-box in promoters *AtMYB2* to inhibit its expression [61]. Later on, Verweij et al. [38] also showed in two different species how a WRKY gene negatively regulates the complex MYB-bHLH-WD40. In our study, the MYB and bHLH genes were mostly up-regulated while WRKY genes were mainly down-regulated in TP, implying that WRKYs may repress the expression levels of MYB and bHLH members. Among others, we propose four candidate TFs, including *c42189.graph_c1* (WRKY) *c33188.graph_c0* (MYB), *c44079.graph_c0* (MYB) and *c37493.graph_c0* (bHLH) for future thorough functional characterizations in turnip skin pigmentation.

4. Materials and Methods

4.1. Plant Materials

Two *Brassica rapa* ssp. rapa varieties with different root skin colors (purple turnip 'PT' and green turnip 'GT') were used as plant materials. We collected healthy and consistent purple and green turnips during five consecutive developmental stages, including seedling stage (S1, 15 days after sowing (DAS)), early stage of fleshy root expansion (S2, 30 DAS), full expansion stage of fleshy root (S3, 45 DAS), maturity stage of fleshy root (S4, 55 DAS) and harvest stage of fleshy root (S5, 65 DAS), at the Anningqu experimental site, Xinjiang, China. Plants were grown in natural environment conditions in July 2017 and skin samples were collected from three different plants of each variety (three biological replicates). In total, 30 samples were quickly frozen in liquid nitrogen and stored at −80 °C in a refrigerator until further use.

4.2. Metabolic Profiling

The sample preparation, extract analysis, metabolite identification and quantification were performed at Wuhan MetWare Biotechnology Co., Ltd. (www.metware.cn) following their standard procedures and previously described by Zhang et al. [48]

4.3. Metabolite Data Analysis

Before the data analysis, quality control (QC) analysis was conducted to confirm the reliability of the data. The QC sample was prepared by the mixture of sample extracts and inserted into every five samples to monitor the changes in repeated analyses. Data matrices with the intensity of the metabolite features from the 30 samples were uploaded to the Analyst 1.6.1 software (AB SCIEX, Ontario, Canada) for statistical analyses. The supervised multivariate method, partial least squares-discriminant analysis (PLS-DA), was used to maximize the metabolome differences between the pair of samples. The relative importance of each metabolite to the PLS-DA model was checked using the parameter called variable importance in projection (VIP). Metabolites with VIP ≥ 1 and fold change ≥ 2 or fold change ≤ 0.5 were considered as differential metabolites for group discrimination [48].

4.4. RNA Extraction, cDNA Library Construction, and Transcriptome Sequencing

Total RNAs were extracted using Spin Column Plant total RNA Purification Kit following the manufacturer's protocol (Sangon Biotech, Shanghai, China). Purity of the extracted RNAs was assessed on 1% agarose gels followed by NanoPhotometer spectrophotometer (IMPLEN, Los Angeles, CA, USA). RNA quantification was performed using Qubit RNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA, USA). Next, RNA integrity was checked by the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA).

Sequencing libraries was created using NEB Next Ultra RNA Library Prep Kit following manufacturer's instructions. The index codes were added to each sample. Briefly, the mRNA was purified from 3 μg total RNA of each of three replicate using poly-T oligo-attached magnetic beads and then broken into short fragments to synthesize first strand cDNA. The second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. PCR was carried out with Phusion High Fidelity DNA polymerase using universal PCR primers and index (×) primer. Finally, six paired-end cDNA libraries with an insert size of 300 bp were constructed for transcriptome sequencing and sequenced on Illumina HiSeq 2500 platform (Illumina Inc., San Diego, USA) by Biomarker Technology Corporation (www.biomarker.com.cn). The raw RNAseq data are submitted at: www.ncbi.nlm.nih.gov/bioproject/PRJNA558197.

4.5. De novo Assembly, Functional Annotation, Classification and Metabolic Pathway Analysis

The clean reads were retrieved after trimming adapter sequences, removal of low quality (containing > 50% bases with a Phred quality score < 15) and reads with unknown nucleotides (more than 1% ambiguous residues N) using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The high quality reads from all the 30 libraries were de novo assembled into transcripts using Trinity (Version r20140717, [62]) by employing paired-end method [63]. Next, the transcripts were realigned to construct unigenes. The assembled unigenes were then annotated by searching against various databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [64], Gene Ontology (GO) [65], Clusters of Orthologous Groups (COG) [66], PfAM, Swissprot [67], egNOG [68], NR [69], euKaryotic Orthologous Groups (KOG) [70] using BLAST [71] with a threshold of E-value $< 1.0 \times 10^{-5}$.

The software KOBAS2.0 [72] was employed to get the unigene KEGG orthology. The analogs of the unigene amino acid sequences were searched against the Pfam database [73] using HMMER tool [74] with a threshold of E-value $< 1.0 \times 10^{-10}$. The sequenced reads were compared with the unigene library using Bowtie [75], and the level of expression was estimated in combination with RSEM [76]. The gene expression level was determined according to the fragments per kilobase of exon per million fragments mapped (FPKM).

4.6. Differential Expression and Enrichment Analysis

The read count was normalized and EdgeR Bioconductor package [77] was used to determine the differential expression genes (DEGs) between the two varieties at each developmental stage with

the fold change > 2 [44] and false discovery rate (FDR) correction set at $p < 0.01$. GO enrichment analysis was performed using the topGO method based on the wallenius noncentral hypergeometric distribution with $p < 0.05$ [78]. KEGG pathway enrichment analysis of the DEGs was done using KOBAS2.0 [72]. The FDR correction was employed ($p < 0.05$) to reduce false positive prediction of enriched GO terms and KEGG pathways.

4.7. SNP Analysis

The reads and unigene sequences of each sample were compared using the software STAR [79] and the single nucleotide polymorphism (SNP) was identified through the pipeline (SNP Calling) for RNA-Seq by GATK2 [80]. Raw vcf files were filtered with GATK standard filter method and other parameters (clusterWindowSize: 35; MQ0 ≥ 4 and (MQ0/(1.0*DP)) > 0.1; QUAL < 10; QUAL < 30.0 or QD < 5.0 or HRun > 5), and only SNPs with distance > 5 were retained.

4.8. Gene Expression Using Quantitative Real Time-PCR

The qRT-PCR was performed on RNA extracted from root samples of both varieties at the five developmental stages as described by Dossa et al. [81] using the *Actin* gene as the internal control. Specific primer pairs of 15 selected genes were designed using the Primer Premier 5.0 [82] (Table S4). The qRT-PCR was conducted on a Roche Lightcyler® 480 instrument using the SYBR Green Master Mix (Vazyme), according to the manufacturer's protocol. Each reaction was performed using a 20 µL mixture containing 10 µL of 2 × ChamQ SYBR qPCR Master Mix, 6 µL of nuclease-free water, 1 µL of each primer (10 mM), and 2 µL of 4-fold diluted cDNA. All of the reactions were run in 96-well plates and each cDNA was analyzed in triplicate. The following cycling profile was used: 95 °C for 30 s, followed by 40 cycles of 95 °C/10 s, 60 °C/30 s. Data are presented as relative transcript level based on the $2^{-\Delta\Delta Ct}$ method [83].

5. Conclusions

In summary, this study generated tremendous genomic and metabolic resources and elucidated the mechanisms of the differential anthocyanin accumulation in purple and green turnips. It provides an important theoretical basis for further in-depth analysis of the candidate structural genes along with the key transcription factors predicted to modulate anthocyanins in turnip towards developing new turnip varieties with improved nutritional quality.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/18/4387/s1. Table S1. Statistics of the read mapping results. GP (Green Turnip) and PT (Purple Turnip); Table S2. Quantification of the detected anthocyanins in two turnips. GP (Green Turnip) and PT (Purple Turnip) and S1–S5 represent the different developmental stages. Data from three biological replicates are shown; Table S3. List of the shared and stage-specific differentially accumulated anthocyanins between Green Turnip and Purple Turnip at different developmental stages S1–S5; Table S4. The primer sequences of genes used for quantitative real time PCR.

Author Contributions: H.Z. and H.W. designed the research and drafted the manuscript. H.Z., Q.L. and Z.T. performed the research. Q.W., H.H. and H.L. collected samples in the field. H.Z., Y.M. prepared plant materials and carried out qRT-PCR analysis. All authors read and approved the final manuscript.

Funding: This research was funded by the National natural Science Foundation of China (grant number: 31960600) and the Project of Renovation Capacity Building for the Young Sci-Tech Talents Sponsored by Xinjiang Academy of Agricultural Sciences (grant number: xjnkq-2019003).

Acknowledgments: We thank Biomarker Technologies Corporation (Beijing, China) for assistance with the PacBio RSII platform and Metware Biotechnology Co., Ltd. (Wuhan, China) for the self-compiled metabolite database MWDB.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fahey, J.W.; Talalay, P.; Gustine, D.I.; Flores, H.E. The role of crucifers in cancer chemoprotection. In *Phytochemicals and Health*; American Society of Plant Physiologist: Rochville, MD, USA, 1995; pp. 87–93.
2. Zhang, H.; Schonhof, I.; Krumbein, A.; Gutezeit, B.; Li, L.; Stützel, H.; Schreiner, M. Water supply and growing season influence glucosinolate concentration and composition in turnip root (*Brassica rapa* ssp. *rapifera* L.). *J. Plant Nutr. Soil Sci.* **2008**, *171*, 255–265. [CrossRef]
3. Chung, I.M.; Rekha, K.; Rajakumar, G.; Thiruvengadam, M. Production of glucosinolates, phenolic compounds and associated gene expression profiles of hairy root cultures in turnip (*Brassica rapa* ssp. *rapa*). *3 Biotech* **2016**, *6*, 175. [CrossRef]
4. Rochfort, S.J.; Imsic, M.; Jones, R.; Trenerry, V.C.; Tomkins, B. Characterization of flavonol conjugates in immature leaves of pak choi (*Brassica rapa* L. ssp. *chinensis* L. (Hanelt.)) by HPLC-DAD and LC-MS/MS. *J. Agric. Food Chem.* **2006**, *54*, 4855–4860. [CrossRef]
5. Thiruvengadam, M.; Chung, I.M. Selenium, putrescine, and cadmium influence health-promoting phytochemicals and molecular-level effects on turnip (*Brassica rapa* ssp. *rapa*). *Food Chem.* **2015**, *173*, 185–193. [CrossRef]
6. Podsędek, A. Natural antioxidants and antioxidant capacity of Brassica vegetables: A review. *LWT-Food Sci. Technol.* **2007**, *40*, 1–11. [CrossRef]
7. Ellinger, S.; Gordon, A.; Kürten, M.; Jungfer, E.; Zimmermann, B.F.; Zur, B.; Ellinger, J.; Marx, F.; Stehle, P. Bolus consumption of a specifically designed fruit juice rich in anthocyanins and ascorbic acid did not influence markers of antioxidative defense in healthy humans. *J. Agric. Food Chem.* **2012**, *60*, 11292–11300. [CrossRef]
8. Poulose, S.M.; Fisher, D.R.; Larson, J.; Bielinski, D.F.; Rimando, A.M.; Carey, A.N.; Schauss, A.G.; Shukitt-Hale, B. Anthocyanin-rich Açai (*Euterpe oleracea* Mart.) fruit pulp fractions attenuate inflammatory stress signaling in mouse brain BV-2 microglial cells. *J. Agric. Food Chem.* **2012**, *60*, 1084–1093. [CrossRef]
9. Pojer, E.; Mattivi, F.; Johnson, D.; Stockley, C.S. The case for anthocyanin consumption to promote human health: A review. *Compr. Rev. Food Sci. Food Saf.* **2013**, *12*, 483–508. [CrossRef]
10. Kruger, M.J.; Davies, N.; Myburgh, K.H.; Lecour, S. Proanthocyanidins, anthocyanins and cardiovascular diseases. *Food Res. Int.* **2014**, *59*, 41–52. [CrossRef]
11. Newsome, A.G.; Culver, C.A.; Van Breemen, R.B. Nature's palette: The search for natural blue colorants. *J. Agric. Food Chem.* **2014**, *62*, 6498–6511. [CrossRef] [PubMed]
12. Fragoso, M.F.; Romualdo, G.R.; Vanderveer, L.A.; Franco-Barraza, J.; Cukierman, E.; Clapper, M.L.; Carvalho, R.F.; Barbisan, L.F. Lyophilized açaí pulp (*Euterpe oleracea* Mart) attenuates colitis-associated colon carcinogenesis while its main anthocyanin has the potential to affect the motility of colon cancer cells. *Food Chem. Toxicol.* **2018**. [CrossRef]
13. Shirley, B.W.; Kubasek, W.L.; Storz, G.; Bruggemann, E.; Koornneef, M.; Ausubel, F.M.; Goodman, H.M. Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J.* **1995**, *8*, 659–671. [CrossRef]
14. Moyano, E.; Martı'nez-Garcia, J.F.; Martin, C. Apparent redundancy in myb gene function provides gearing for the control of flavonoid biosynthesis in antirrhinum flowers. *Plant Cell* **1996**, *8*, 1519–1532.
15. Weisshaar, B.; Jenkins, G.I. Phenylpropanoid biosynthesis and its regulation. *Curr. Opin. Plant Biol.* **1998**, *1*, 251–257. [CrossRef]
16. Harborne, J.B.; Williams, C.A. Advances in flavonoid research since 1992. *Phytochemistry* **2000**, *55*, 481–504. [CrossRef]
17. Pietta, P.-G. Flavonoids as antioxidants. *J. Nat. Prod.* **2000**, *63*, 1035–1042. [CrossRef]
18. Shirley, B.W. Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **2002**, *5*, 218–223. [CrossRef]
19. Cominelli, E.; Galbiati, M.; Vavasseur, A.; Conti, L.; Sala, T.; Vuylsteke, M.; Leonhard, N.; Dellaporta, S.L.; Tonelli, C. A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr. Biol.* **2005**, *15*, 1196–1200. [CrossRef]
20. Zhang, L.; Wang, Y.; Sun, M.; Wang, J.; Kawabata, S.; Li, Y. *BrMYB4*, a Suppressor of Genes for Phenylpropanoid and Anthocyanin Biosynthesis, is Down-Regulated by UV-B but not by Pigment-Inducing Sunlight in Turnip cv. Tsuda. *Plant Cell Physiol.* **2014**, *55*, 2092–2101. [CrossRef]

21. Liu, Y.; Tikunov, Y.; Schouten, R.E.; Marcelis, L.F.M.; Visser, R.G.F.; Bovy, A. Anthocyanin Biosynthesis and Degradation Mechanisms in *Solanaceous* Vegetables: A Review. *Front. Chem.* **2018**, *6*. [CrossRef]
22. Jaakola, L. New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* **2013**, *18*, 477–483. [CrossRef]
23. Tanaka, Y.; Ohmiya, A. Seeing is believing: Engineering anthocyanin and carotenoid biosynthetic pathways. *Curr. Opin. Biotechnol.* **2008**, *19*, 190–197. [CrossRef]
24. Broun, P. Transcriptional control of flavonoid biosynthesis: A complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. *Curr. Opin. Plant Biol.* **2005**, *8*, 272–279. [CrossRef]
25. Guo, N.; Cheng, F.; Wu, J.; Liu, B.; Zheng, S.; Liang, J.; Wang, X. Anthocyanin biosynthetic genes in *Brassica rapa*. *BMC Genomics* **2014**, *15*, 426. [CrossRef]
26. Li, Z.; Zhao, M.; Jin, J.; Zhao, L.; Xu, Z. Anthocyanins and their biosynthetic genes in three novel-colored Rosa rugosa cultivars and their parents. *Plant Physiol. Biochem.* **2018**, *129*, 421–428. [CrossRef]
27. Guo, N.; Han, S.; Zong, M.; Wang, G.; Zheng, S.; Liu, F. Identification and differential expression analysis of anthocyanin biosynthetic genes in leaf color variants of ornamental kale. *BMC Genom.* **2019**, *20*, 564. [CrossRef]
28. To, K.Y.; Wang, C.K. Molecular breeding of flower color. In *Floriculture Ornamental and Plant Biotechnology: Advances and Topical Issues Volume I*; Teixeira da Silva, J.A., Ed.; Advances and Topical Issues: Isleworth, UK, 2006; pp. 300–310.
29. Gonzalez, A.; Zhao, M.; Leavitt, J.M.; Lloyd, A.M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* **2008**, *53*, 814–827. [CrossRef]
30. Sapir, M.; Oren-Shamir, M.; Ovadia, R.; Reuveni, M.; Evenor, D.; Tadmor, Y.; Nahon, S.; Shlomo, H.; Chen, L.; Meir, A.; et al. Molecular aspects of Anthocyanin fruit tomato in relation to high pigment-1. *J. Hered.* **2008**, *99*, 292–303. [CrossRef]
31. Shi, M.Z.; Xie, D.Y. Engineering of red cells of Arabidopsis thaliana and comparative genome-wide gene expression analysis of red cells versus wild-type cells. *Planta* **2011**, *233*, 787–805. [CrossRef]
32. An, X.; Tian, Y.; Chen, K.; Wang, X.; Hao, Y. The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. *J. Plant Physiol.* **2012**, *169*, 710–717. [CrossRef]
33. Zhou, L.L.; Shi, M.Z.; Xie, D.Y. Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of Arabidopsis thaliana. *Planta* **2012**, *236*, 825–837. [CrossRef]
34. Chen, Y.; Mao, Y.; Liu, H.; Yu, F.; Li, S.; Yin, T. Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. *PLoS ONE* **2014**, *9*, e90842.
35. Jiao, Y.; Ma, R.J.; Shen, Z.J.; Yan, J.; Yu, M.L. Gene regulation of anthocyanin biosynthesis in two blood-flesh peach (*Prunus persica* (L) Batsch) cultivars during fruit development. *J. Zhejiang Univ. Sci. B (Biomed. Biotechnol.)* **2014**, *15*, 809–819. [CrossRef]
36. Dixon, R.A.; Steele, C.L. Flavonoids and isoflavonoids—A gold mine for metabolic engineering. *Trends Plant. Sci.* **1999**, *4*, 394–400. [CrossRef]
37. Stommel, J.R.; Dumm, J.M. Coordinated regulation of biosynthetic and regulatory genes coincides with Anthocyanin Accumulation in developing eggplant fruit. *J. Am. Soc. Hortic. Sci.* **2015**, *140*, 129–135. [CrossRef]
38. Verweij, W.; Spelt, C.E.; Bliek, M.; de Vries, M.; Wit, N.; Faraco, M.; Koes, R.; Quattrocchio, F.M. Functionally similar WRKY proteins regulate vacuolar acidification in Petunia and hair development in Arabidopsis. *Plant Cell* **2016**, *28*, 786–803. [CrossRef]
39. Yuan, Y.; Rebocho, A.B.; Sagawa, J.M.; Stanley, L.E.; Bradshaw, H.D., Jr. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between Mimulus species. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2448–2453. [CrossRef]
40. Onik, J.C.; Hu, X.; Lin, Q.; Wang, Z. Comparative Transcriptomic Profiling to Understand Pre- and Post-Ripening Hormonal Regulations and Anthocyanin Biosynthesis in Early Ripening Apple Fruit. *Molecules* **2018**, *23*, 1908. [CrossRef]
41. He, L.; Tang, R.; Shi, X.; Wang, W.; Cao, Q.; Liu, X.; Wang, T.; Sun, Y.; Zhang, H.; Li, R.; et al. Uncovering anthocyanin biosynthesis related microRNAs and their target genes by small RNA and degradome sequencing in tuberous roots of sweet potato. *BMC Plant Biol.* **2019**, *19*, 232. [CrossRef]

42. Lin, K.; Zhang, N.; Severing, E.I.; Nijveen, H.; Cheng, F.; Visser, R.G.F.; Wang, X.; de Ridder, D.; Bonnema, G. Beyond genomic variation—Comparison and functional annotation of three Brassica rapa genomes: A turnip, a rapid cycling and a Chinese cabbage. *BMC Genom.* **2014**, *15*, 250. [CrossRef]
43. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; et al. The genome of the mesopolyploid crop species *Brassica Rapa*. *Nat. Genet.* **2011**, *43*, 1035–1039. [CrossRef]
44. Anders, S.; McCarthy, D.J.; Chen, Y.; Okoniewski, M.; Smyth, G.K.; Huber, W.; Robinson, M.D. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **2013**, *8*, 1765–1786. [CrossRef]
45. Banerjee, N.; Zhang, M.Q. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Res.* **2003**, *31*, 7024–7031. [CrossRef]
46. Stintzing, F.C.; Carle, R. Functional properties of anthocyanins and betalains in plants, food and in human nutrition. *Trends Food Sci. Technol.* **2004**, *15*, 19–38. [CrossRef]
47. Zhang, S.; Ying, H.; Pingcuo, G.; Wang, S.; Zhao, F.; Cui, Y.; Shi, J.; Zeng, H.; Zeng, X. Identification of Potential Metabolites Mediating Bird's Selective Feeding on *Prunus mira* Flowers. *BioMed Res. Int.* **2019**, *2019*, 1395480. [CrossRef]
48. Takos, A.M.; Jaffé, F.W.; Jacob, S.R.; Bogs, J.; Robinson, S.P.; Walker, A.R. Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* **2006**, *142*, 1216–1232. [CrossRef]
49. Rahim, M.A.; Busatto, N.; Trainotti, L. Regulation of anthocyanin biosynthesis in peach fruits. *Planta* **2014**, *240*, 913. [CrossRef]
50. Wang, G.; Guo, X.; Floros, J. Human SP-A 3'-UTR variants mediate differential gene expression in basal levels and in response to dexamethasone. *Am. J. Physiol. Lung Cell Mol. Physiol.* **2003**, *284*, L738–L748. [CrossRef]
51. Buchweitz, M. Natural Solutions for Blue Colors. In *Schweiggert, Food in Handbook on Natural Pigments in Food and Beverages*; Carle, R., Ralf, M., Eds.; Woodhead Publishing: Sawston, UK, 2016; pp. 355–384.
52. He, J.; Giusti, M.M. Anthocyanins: Natural Colorants with Health-Promoting Properties. *Annu. Rev. Food Sci. Technol.* **2010**, *1*, 163–187. [CrossRef]
53. Lo Scalzo, R.; Genna, A.; Branca, F.; Chedin, M.; Chassaigne, H. Anthocyanin composition of cauliflower (*Brassica oleracea* L. var. botrytis) and cabbage (*B. oleracea* L. var. capitata) and its stability in relation to thermal treatments. *Food Chem.* **2008**, *107*, 136–144. [CrossRef]
54. Li, C.X. The Study of Physicochemical Property of Anthocyanin in Purple Pakchoi. Master's Thesis, Northwest A & F University, Yangling, China, 2011.
55. Guo, N.; Zheng, S.N.; Wu, J.; Chen, F.; Liang, J.L.; Wang, X.W. The anthocyanin metabolic profiling analysis of three *Brassica rapa* vegetables. *Acta Hortic. Sin.* **2014**, *41*, 1707–1715.
56. Park, C.H.; Yeo, H.J.; Kim, N.S.; Eun, P.Y.; Kim, S.-J.; Arasu, M.V.; Al-Dhabi, N.A.; Park, S.-Y.; Kim, J.K.; Park, S.U. Metabolic profiling of pale green and purple kohlrabi (*Brassica oleracea* var. gongylodes). *Appl. Biol. Chem.* **2017**, *60*, 249–257. [CrossRef]
57. Lloyd, A.; Brockman, A.; Aguirre, L.; Campbell, A.; Bean, A.; Cantero, A.; Gonzalez, A. Advances in the MYB–bHLH–WD repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. *Plant. Cell Physiol.* **2017**, *58*, 1431–1441. [CrossRef]
58. Naing, A.H.; Kim, C.K. Roles of R2R3-MYB transcription factors in transcriptional regulation of anthocyanin biosynthesis in horticultural plants. *Plant Mol. Biol.* **2018**, *98*, 1–18. [CrossRef]
59. Fosket, D.E. The Genetic Basis of Plant Development. In *Fosket, Plant Growth and Development*; Donald, E., Ed.; Academic Press: Cambridge, MA, USA, 1994; pp. 41–78.
60. Lei, R.; Ma, Z.; Yu, D. WRKY2/34–VQ20 Modules in Arabidopsis thaliana Negatively Regulate Expression of a Trio of Related MYB Transcription Factors during Pollen Development. *Front. Plant. Sci.* **2018**, *9*, 331. [CrossRef]
61. Shang, Y.; Yan, L.; Liu, Z.Q.; Cao, Z.; Mei, C.; Xin, Q.; Wu, F.Q.; Wang, X.F.; Du, S.Y.; Jiang, T.; et al. The Mg-chelatase H subunit of *Arabidopsis* antagonizes a group of WRKY transcription repressors to relieve ABA—Responsive genes of inhibition. *Plant. Cell* **2010**, *22*, 1909–1935. [CrossRef]
62. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full length transcriptome assembly from RNA Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]

63. Yu, R.; Liang, X.; Wei, Z.; Wang, Y.; Luo, X.; Wang, R.; Zhu, X.; Xie, Y.; Karanja, B.; Liu, L. De novo taproot transcriptome sequencing and analysis of major genes involved in sucrose metabolism in Radish (*Raphanus sativus* L.). *Front. Plant Sci.* **2016**, *7*, 585. [CrossRef]
64. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [CrossRef]
65. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
66. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A. The COG database: A tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [CrossRef]
67. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119. [CrossRef]
68. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2015**, *44*, 286–293. [CrossRef]
69. Deng, Y.Y.; Li, J.Q.; Wu, S.F.; Zhu, Y.P.; Chen, Y.W.; He, F.C. Integrated nr Database in Protein Annotation System and Its Localization. *Comput. Eng.* **2006**, *32*, 71–74.
70. Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **2004**, *5*, R7. [CrossRef]
71. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
72. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [CrossRef]
73. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, 222–230. [CrossRef]
74. Eddy, S.R. Profile hidden Markov models. *Bioinform. Ital.* **1998**, *14*, 755–763. [CrossRef]
75. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef]
76. Li, B.; Colin, N.D. RSEM: Accurate transcript quantification from RNA Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef]
77. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
78. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [CrossRef]
79. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
80. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
81. Dossa, K.; Mmadi, M.A.; Zhou, R.; Zhou, Q.; Yang, M.; Cisse, N.; Diouf, D.; Wang, L.; Zhang, X. The contrasting response to drought and waterlogging is underpinned by divergent DNA methylation programs associated with transcript accumulation in sesame. *Plant. Sci.* **2018**, *277*, 207–217.
82. Lalitha, S. Primer premier 5. *Biotechnol. Softw. Int. Rep.* **2000**, *1*, 270–272. [CrossRef]
83. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Transcriptome Profiling Provides Insight into the Genes in Carotenoid Biosynthesis during the Mesocarp and Seed Developmental Stages of Avocado (*Persea americana*)

Yu Ge [1], Zhihao Cheng [1], Xiongyuan Si [2], Weihong Ma [1], Lin Tan [1], Xiaoping Zang [1], Bin Wu [1], Zining Xu [1], Nan Wang [1], Zhaoxi Zhou [1], Xinge Lin [1], Xiangshu Dong [3,*] and Rulin Zhan [1,*]

1 Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou 570102, China
2 Biotechnology Center, Anhui Agricultural University, Hefei 230036, China
3 College of Agriculture, Yunnan University, Kunming 650091, China
* Correspondence: dongxiangshu@ynu.edu.cn (X.D.); zhanrulin@catas.cn (R.Z.); Tel.: +86-871-6503-1539 (X.D.); +86-898-6677-3067 (R.Z.)

Received: 12 August 2019; Accepted: 21 August 2019; Published: 23 August 2019

Abstract: Avocado (*Persea americana* Mill.) is an economically important crop because of its high nutritional value. However, the absence of a sequenced avocado reference genome has hindered investigations of secondary metabolism. For next-generation high-throughput transcriptome sequencing, we obtained 365,615,152 and 348,623,402 clean reads as well as 109.13 and 104.10 Gb of sequencing data for avocado mesocarp and seed, respectively, during five developmental stages. High-quality reads were assembled into 100,837 unigenes with an average length of 847.40 bp (N50 = 1725 bp). Additionally, 16,903 differentially expressed genes (DEGs) were detected, 17 of which were related to carotenoid biosynthesis. The expression levels of most of these 17 DEGs were higher in the mesocarp than in the seed during five developmental stages. In this study, the avocado mesocarp and seed transcriptome were also sequenced using single-molecule long-read sequencing to acquired 25.79 and 17.67 Gb clean data, respectively. We identified 233,014 and 238,219 consensus isoforms in avocado mesocarp and seed, respectively. Furthermore, 104 and 59 isoforms were found to correspond to the putative 11 carotenoid biosynthetic-related genes in the avocado mesocarp and seed, respectively. The isoform numbers of 10 out of the putative 11 genes involved in the carotenoid biosynthetic pathway were higher in the mesocarp than those in the seed. Besides, alpha- and beta-carotene contents in the avocado mesocarp and seed during five developmental stages were also measured, and they were higher in the mesocarp than in the seed, which validated the results of transcriptome profiling. Gene expression changes and the associated variations in gene dosage could influence carotenoid biosynthesis. These results will help to further elucidate carotenoid biosynthesis in avocado.

Keywords: avocado; carotenoid biosynthesis; mesocarp; seed; de novo assembly from short read sequencing; full-length transcript sequencing; differentially expressed genes; gene dosage

1. Introduction

Avocado (*Persea americana* Mill.) is a member of the family Lauraceae of the order Laurales, and widely grown in countries and regions with a tropical-to-cool climate [1–3]. Avocado is among the most economically important subtropical/tropical fruit crops worldwide, with considerable increases in yield reported in several countries, including Mexico, the USA, Indonesia, Chile, Spain, Israel, Colombia, South Africa, and Australia [4]. Certain avocado constituents, such as carotenoids, lipids,

sugars, proteins, minerals, vitamins, and other nutrients and active ingredients, provide nutritional and health benefits [5–7].

Carotenoids in fruits have been extensively studied because of their nutritional benefits for humans [5,8]. Moreover, some carotenoids serve as precursors of vitamin A, strigolactones, and abscisic acid, and some apocarotenoids are also potent antioxidants and colorants [9]. Additionally, carotenoid derivatives substantially affect the aromatic flavor of fruits, thereby making the fruits more desirable to consumers and seed dispersers [10,11]. Carotenoids are natural yellow-to-red pigments that are mainly a type of C40 terpenoid distributed in most photosynthetic organisms as well as in some non-photosynthetic fungi and bacteria [12]. In plants, carotenoids primarily contribute to photosynthesis, photomorphogenesis, photoprotection, light-harvesting processes, and growth and development [12–14]. The plant carotenoid metabolic pathway has been well elucidated. This pathway consists of a series of chemical reactions, including condensations, dehydrogenations, cyclizations, hydroxylations, and epoxidations [12].

Next-generation high-throughput sequencing technology (NGST) have recently become popular options for transcriptome sequencing experiments because they enable high-throughput, efficient, accurate, and reproducible analyses [15–20]. Previous studies suggested that changes in carotenoid biosynthesis and accumulation are correlated with changes to the expression of genes encoding carotenoid metabolic pathway enzymes [8,12]. Transcriptome sequencing methods have been used to investigate the carotenoid biosynthetic mechanism in many species, including *Momordica cochinchinensis* [21], *Brassica campestris* [22], green alga [23], celery [9], carrot [24,25], and *Euscaphi skonishii* [26]. However, few reports have focused on transcriptome sequencing to investigate the carotenoid biosynthetic mechanism in avocado.

Among the third-generation sequencing platforms, PacBio RS II, which is regarded as the first commercialized third-generation sequencer, is based on single-molecule real-time (SMRT) technology [27]. The PacBio RS II system can produce much longer reads than second-generation sequencing platforms, and has been applied to effectively capture full-length transcript sequences [28]. Single-molecule real-time technology has the following three main advantages over second-generation sequencing options: it generates longer reads, it has a higher consensus accuracy, and it is less biased [29]. A previous study revealed that SMRT technology can precisely ascertain alternative polyadenylation sites and full-length splice isoforms, and also detect a higher isoform density than that for the reference genome [30]. The application of SMRT technology for nearly 3 years has helped to elucidate the complexity of the transcriptome and molecular mechanism underlying metabolite synthesis in safflower [27], *Zanthoxylum bungeanum* [28], *Trifolium pretense* [30], sugarcane [31], switchgrass [32], *Medicago sativa* [33], *Zanthoxylum planispinum* [34], bermudagrass [35], *Camellia sinensis* [36], and *Cassia obtusifolia* [37]. So far, no report has been found about the application of SMRT technology in a plant species from the family Lauraceae.

In this study, Illumina HiSeq 2000 next-generation sequencing technology and PacBio RS II third-generation sequencing platform were integrated to generate transcriptome data for exploring the carotenoid biosynthetic pathways in the avocado mesocarp and seed. We integrated the gene dosage variation and the associated changes in gene expression to identify genes that are likely important for carotenoid accumulation. Additionally, metabolite profiling (alpha- and beta-carotene contents) via high-performance liquid chromatography (HPLC) was used to auxiliarily validate the transcriptomic analyses. The data obtained in this comprehensive study involving the full-length transcript sequences and de novo transcriptome assembly from short read sequencing will be useful for investigating the main physiological and biochemical molecular metabolic mechanisms in the avocado mesocarp and seed.

2. Results

2.1. Overview of the Morphology and NGST Profiling

Morphological measurements, including mesocarp weight, seed weight, fruit length, and fruit width, gradually increased from 75 to 215 days after full bloom (DAFB), peaking at 114.74 g FW, 11.81 g FW, 83.07 mm, and 62.42 mm, respectively (Table S1). The photos of the tested avocado samples per fruit developmental stage are also presented in Figure S1. The RNA extracted from 15 mesocarp and seed samples were analyzed by RNA sequencing (RNA-seq), with three replicates per avocado fruit developmental stage. The sequencing of 30 cDNA libraries resulted in 20–26 million clean reads and 6.03–8.28 Gb of sequence data (Table S2). The generated avocado transcriptome data were deposited in the GenBank database (accession number PRJNA541745). The default parameters of the Trinity program were used to assemble the high-quality reads into 205,415 transcripts with a mean length of 1199.57 bp (N50 = 2063 bp) as well as 100,837 unigenes with a mean length of 847.40 bp (N50 = 1725 bp). Of these 100,837 unigenes, 59,969 (59.47%) were short (i.e., up to 500 bp), 16,511 (16.37%) were 501–1000 bp long, and 24,357 (24.16%) were longer than 1000 bp. The length distributions of all transcripts and unigenes are presented in Figure S2. These results demonstrated that the sequencing quality was sufficient for subsequent analyses.

2.2. Annotation and Identification of Unigenes

Regarding the gene annotations, the BLASTX program revealed that 14,565 (37.05%), 20,712 (52.69%), 12,638 (32.15%), 19,065 (48.50%), 23,009 (58.54%), 22,403 (56.99%), 34,394 (87.50%), and 35,021 (89.09%) of the 39,309 avocado unigenes had significant matches with sequences in the Clusters of Orthologous Groups (COG), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Eukaryotic Orthologous Groups (KOG), Pfam, Swiss-Prot, eggNOG, and the NCBI non-redundant protein sequence (Nr) databases, respectively. To further predict and classify functions, the annotated unigenes were analyzed according to GO assignments, COG classifications, and KEGG pathway assignments. A total of 20,712 unigenes were assigned to 49 sub-categories of the three main GO functional categories (cellular component, biological process, and molecular function) (Figure 1; Table S3). The molecular function category comprised the most unigenes, followed by the biological process and cellular component categories. The most common molecular function GO terms were catalytic activity (10,855 unigenes, GO: 0003824) and binding (10,148 unigenes, GO: 0005488). The top three biological process GO terms were metabolic process (10,654 unigenes, GO: 0008152), cellular process (9833 unigenes, GO: 0009987), and single-organism process (6516 unigenes, GO: 0044699). The most frequently observed cellular component GO terms were cell (8781 unigenes, GO: 0005623) and cell part (8769 unigenes, GO: 0044464) (Figure 1).

2.3. Screening of Differentially Expressed Genes during Avocado Mesocarp and Seed Development

An analysis of differentially expressed genes (DEGs) in the avocado mesocarp and seed during five fruit developmental stages revealed 16,903 DEGs (Table S5). There were 4013–4828 DEGs between the mesocarp and seed at five time-points, with some minor variability in the number of DEGs among the five fruit developmental stages. The number of DEGs increased considerably during mesocarp development. The largest number of DEGs (1516) was detected between 75 and 215 DAFB. Similarly, the number of DEGs sharply increased during the whole seed development stage, with a 4.40-fold increase from 75 vs. 110 DAFB to 75 vs. 215 DAFB. These results indicated that the fifth mesocarp and seed developmental stages may be associated with the most dramatic changes in enzyme contents and multiple metabolic pathways.

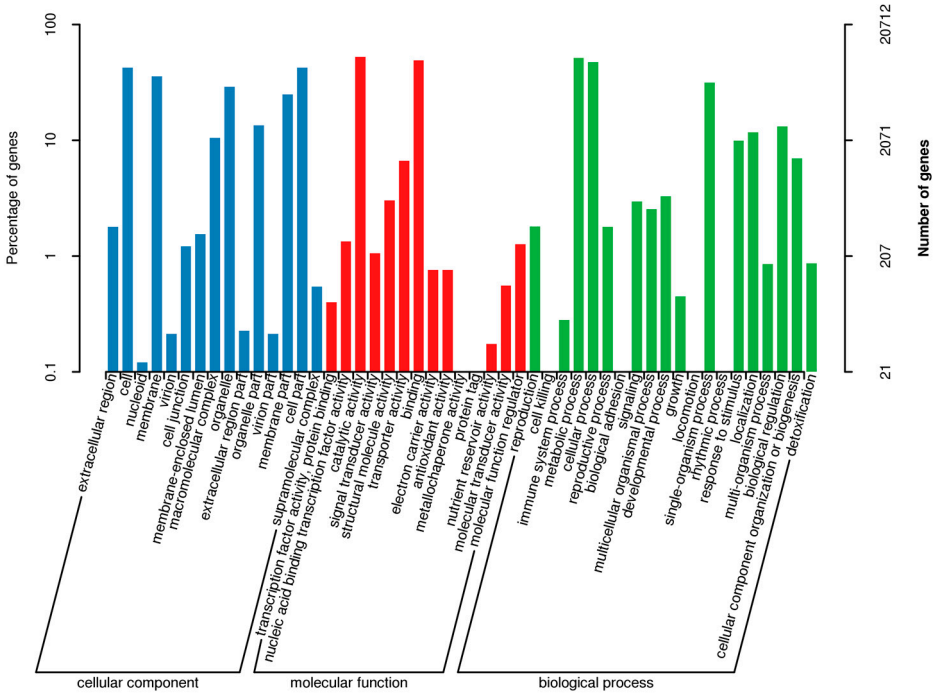


Figure 1. Gene Ontology functional annotations of avocado unigenes. The unigenes were divided into three categories (biological process, cellular component, and molecular function). A total of 19,065 unigenes were functionally characterized based on 26 Eukaryotic Orthologous Groups (KOG) categories (Figure 2). Among the 26 categories, 'General function prediction only' represented the largest group (5557 unigenes, 26.38%), followed by 'Posttranslational modification, protein turnover, chaperones' (1959 unigenes, 9.30%), and 'Signal transduction mechanisms' (1729 unigenes, 8.19%). 'Extracellular structures' and 'Cell motility' were the two smallest groups. Additionally, 12,638 unigenes were assigned to 130 pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Table S4), with carbon metabolism (571 unigenes), biosynthesis of amino acids (516 unigenes), and ribosome (465 unigenes) representing the three most common pathways.

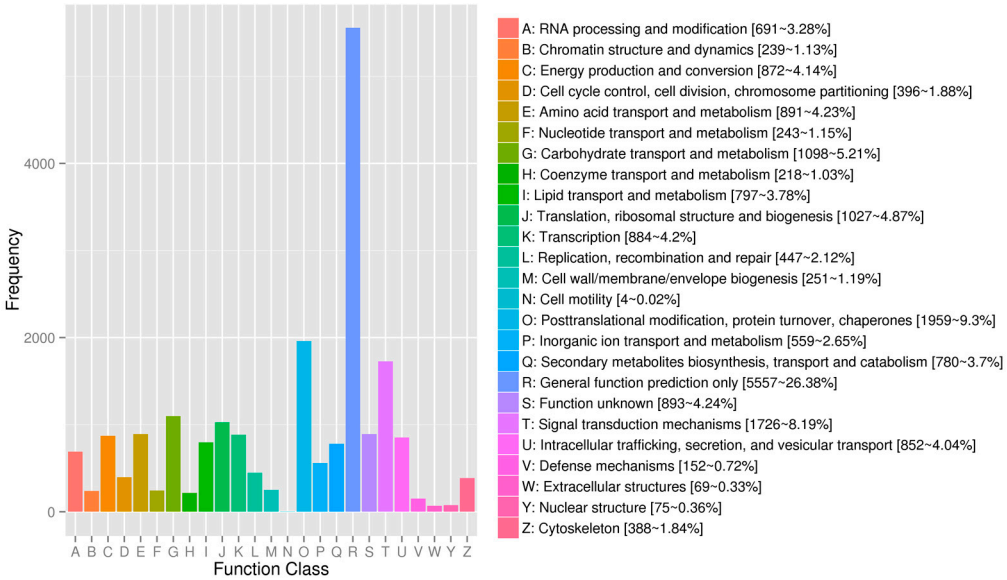
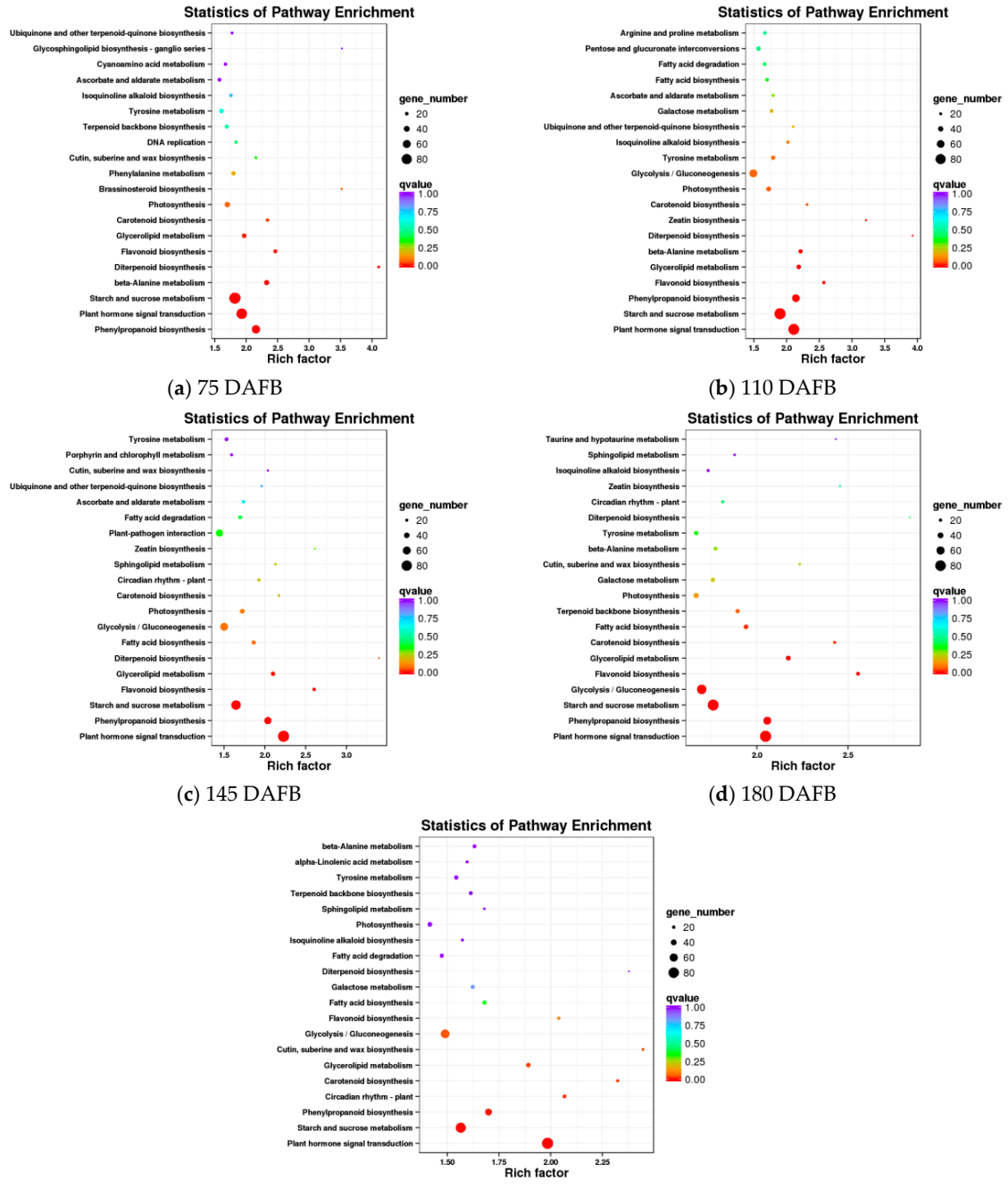


Figure 2. Eukaryotic Orthologous Groups classification of the assembled unigenes.

2.4. Differentially Expressed Carotenoid Biosynthetic Genes between the Avocado Mesocarp and Seed

A comparison of the avocado mesocarp and seed at five developmental stages based on the KEGG pathway enrichment among all DEGs resulted in the identification of the carotenoid biosynthetic pathway in four of the five developmental stages (Figure 3). The DEGs detected in the avocado mesocarp and seed transcriptomes included 17 unigenes that putatively encode 11 enzymes in the carotenoid biosynthetic pathway (Table 1).



(**a**) 75 DAFB (**b**) 110 DAFB

(**c**) 145 DAFB (**d**) 180 DAFB

(**e**) 215 DAFB

Figure 3. Results of the KEGG enrichment analysis of differentially expressed genes (DEGs) between the avocado mesocarp and seed at five developmental stages.

Table 1. Unigenes related to carotenoid biosynthesis.

Gene Name	Functional Protein Name	Enzyme Commission Number	Unigene ID
PaPSY	15-cis-phytoene synthase	2.5.1.32	c103350.graph_c0, c113873.graph_c5
PaPDS	Phytoene desaturase	1.3.5.5	c103201.graph_c0, c104826.graph_c4
PaZ-ISO	15-cis-ζ-carotene isomerase	5.2.1.12	c109620.graph_c1
PaZDS	ζ-carotene desaturase	1.3.5.6	c108741.graph_c1, c115069.graph_c3
PaCRTISO	Carotenoid isomerase	5.2.1.13	c108133.graph_c1
PaLCY-E	Lycopene ε-cyclase	5.5.1.18	c117627.graph_c3
PaLCY-B	Lycopene β-cyclase	5.5.1.19	c92930.graph_c0, c110018.graph_c0
PaCYP97A	P450 β-ring carotene hydroxylase	1.14.13.129	c106779.graph_c0
PaCYP97C	P450 ε-ring carotene hydroxylase	1.14.99.45	c110544.graph_c0
PaZEP	Zeaxanthin epoxidase	1.14.15.21	c109893.graph_c0, c116714.graph_c5
PaNSY	Neoxanthin synthase	5.3.99.9	c106233.graph_c1, c92501.graph_c0

An analysis of the unigenes related to carotenoid biosynthesis that were differentially expressed during five mesocarp and seed developmental stages (Figure 4) revealed that the following 15 unigenes were more highly expressed in the mesocarp than in the seed at each of the five examined time-points: *PaPSY* (c103350.graph_c0 and c113873.graph_c5), *PaPDS* (c103201.graph_c0 and c104826.graph_c4), *PaZ-ISO* (c109620.graph_c1), *PaZDS* (c108741.graph_c1 and c115069.graph_c3), *PaCRTISO* (c108133.graph_c1), *PaLCY-E* (c117627.graph_c3), *PaLCY-B* (c92930.graph_c0 and c110018.graph_c0), *PaCYP97C* (c110544.graph_c0), *PaZEP* (c109893.graph_c0 and c116714.graph_c5), and *PaNSY* (c92501.graph_c0). In contrast, the *PaNSY* (c106233.graph_c1) expression level was considerably lower in the mesocarp than in the seed at each of the five time-points (Figure 4; Table S6). Additionally, *PaCYP97A* (c106779.graph_c0) was expressed at lower levels in the mesocarp than in the seed from 75 to 180 DAFB, but the opposite pattern was observed at 215 DAFB (Figure 4; Table S6). The *PaPSY*, *PaPDS*, *PaZ-ISO*, *PaZDS*, *PaCRTISO*, *PaLCY-E*, and *PaLCY-B* expression levels were higher in the mesocarp than in the seed at each of the five time-points, and increased by 1.09 to 22.41 fold (Table S6). To confirm the accuracy of the high-throughput sequencing results, the expression levels of ten unigenes involved in the carotenoid biosynthetic pathway (i.e., *PaPSY*, *PaPDS*, *PaZ-ISO*, *PaLCY-E*, *PaCYP97C*, *PaZEP*, and *PaNSY*) were analyzed by a quantitative real-time polymerase chain reaction (qRT-PCR) assay (Figure 5). The resulting expression patterns of these genes during the five mesocarp and seed developmental stages were consistent with the RNA-seq data.

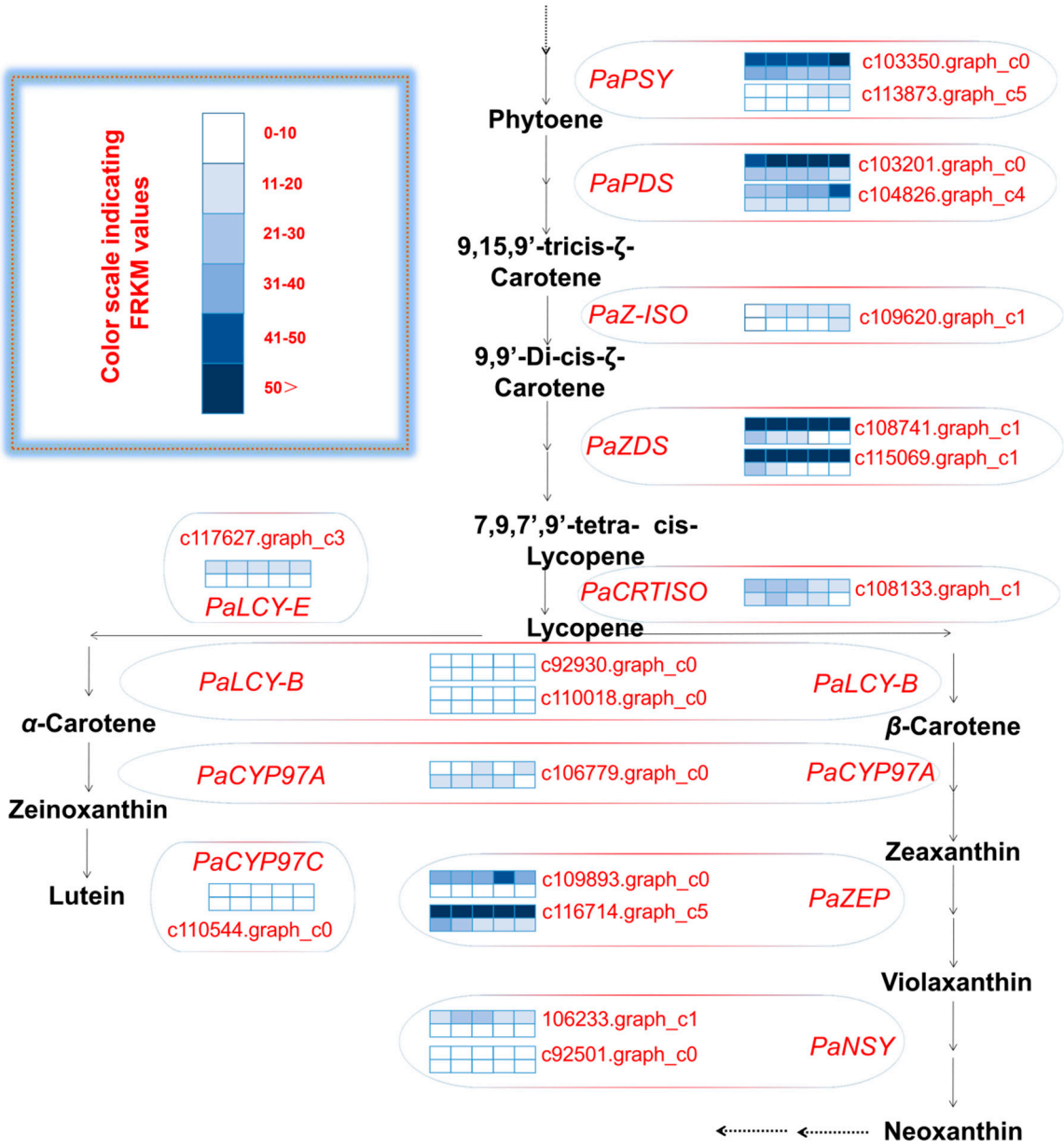
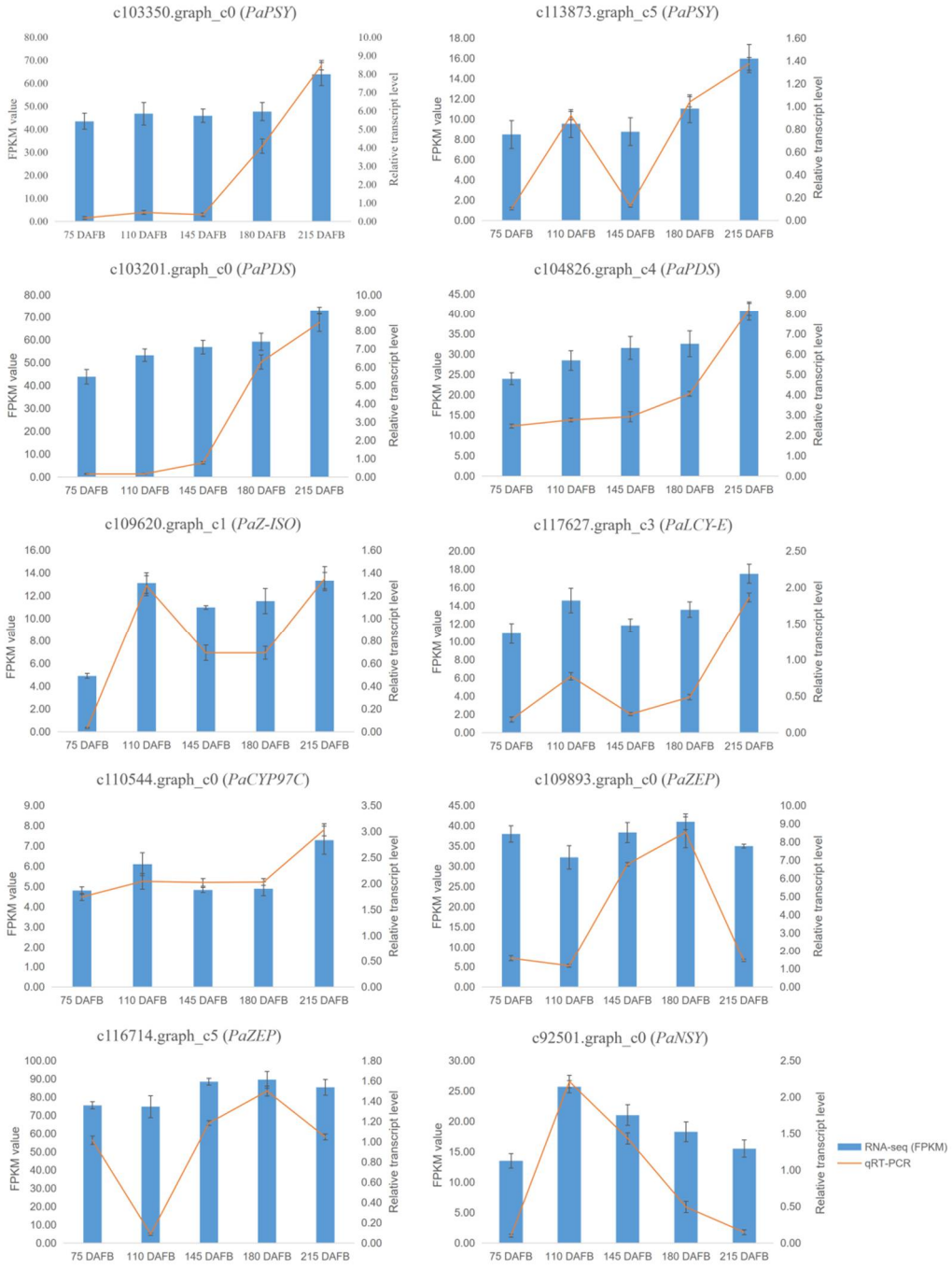
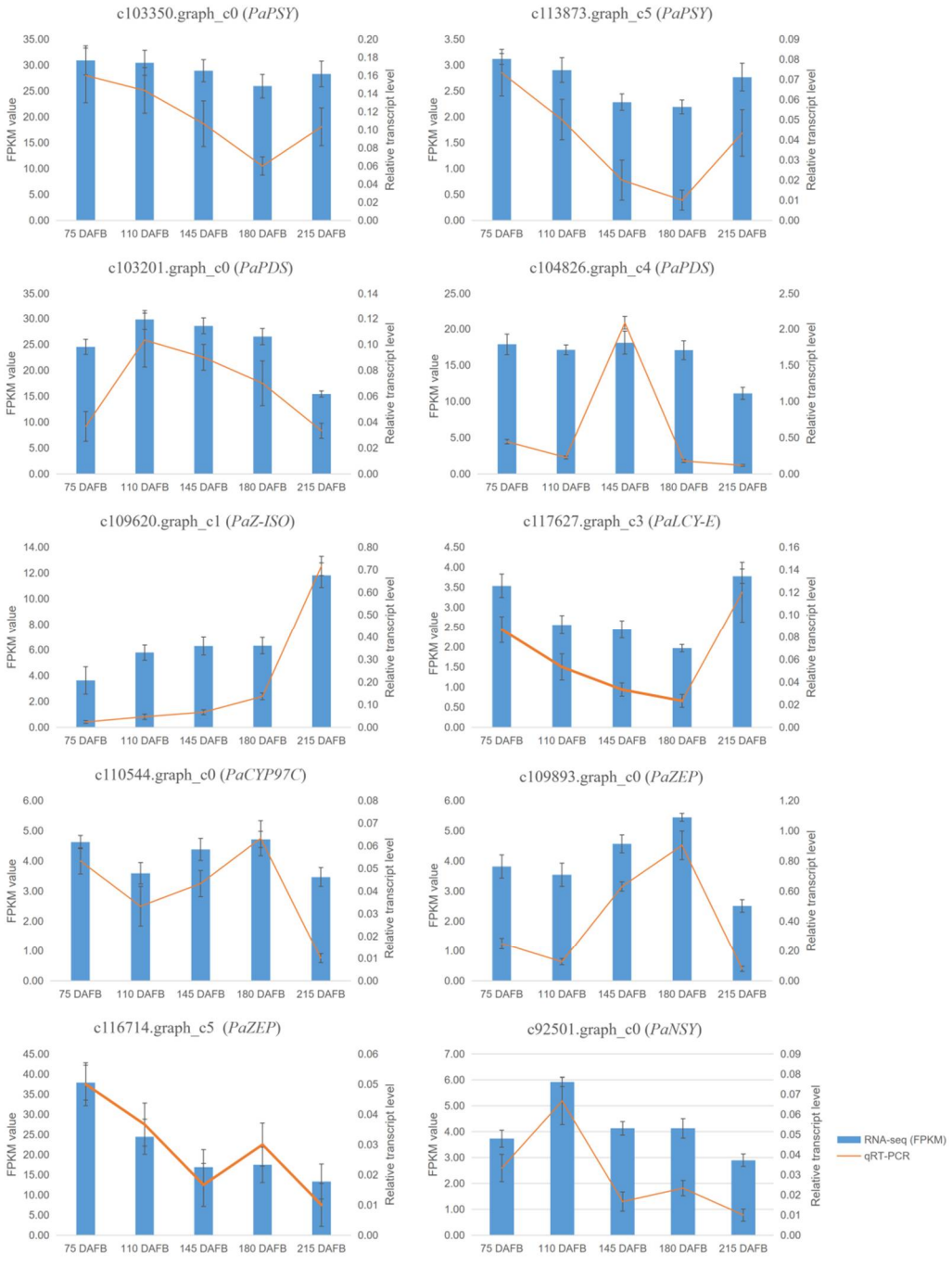


Figure 4. Carotenoid biosynthetic pathway and transcript levels during avocado mesocarp and seed developmental stages. FPKM: fragments per kilobase of transcript sequence per million base pairs sequenced; gene expression levels at 75, 110, 145, 180, and 215 days after full bloom (DAFB) are indicated with colored bars; the top and bottom barfor eachunigene demonstrates avocado mosocarp and seed, respectively; PSY: 15-cis-phytoene synthase; PDS: phytoene desaturase; Z-ISO: 15-cis-ζ-carotene isomerase; ZDS: ζ-carotene desaturase; CRTISO: carotenoid isomerase; LCY-E: lycopene ε-cyclase; LCY-B: lycopene β-cyclase; CYP97A: P450 β-ring carotene hydroxylase; CYP97C: P450 ε-ring carotene hydroxylase; ZEP: zeaxanthin epoxidase; NSY: neoxanthin synthase.



(**a**)

Figure 5. *Cont.*



(**b**)

Figure 5. The FPKM values and relative expression levels of 10 carotenoid biosynthesis unigenes in avocado mesocarp (**a**) and seed (**b**). Error bars are standard errors of the mean from three biological replicates and two technical replicates.

2.5. General Properties of Single-Molecule Long-Reads

Full-length cDNA sequences derived from poly-A-tailed RNA samples were normalized and subjected to SMRT sequencing with the PacBio RS II platform. A total of 25.79 and 17.67 Gb clean data were generated for the library in avocado mesocarp and seed, respectively. Each SMRT cell produced 651,260 and 586,430 reads of inserts (ROIs) from the library (1–6 kb) in avocado mesocarp and seed, respectively. These ROIs were successfully extracted in avocado mesocarp and seed, respectively, with a mean length of 2200 and 2239 bp, a quality score of 0.96 and 0.94. All ROIs were further

classified into 495,245 and 403,108 full-length nonchimeric in avocado mesocarp and seed, respectively. On the basis of the iterative isoform-clustering algorithm, 233,014 and 238,219consensus isoforms were acquired in avocado mesocarp and seed, respectively, with a mean length of 2170 and 2027 bp (Table S7). After removing the redundant sequences for all high-quality transcripts and corrected low-quality transcripts with CD-HIT (c = 0.90), 76,345 and 68,618 nonredundant transcripts remained. The SMRT and Illumina HiSeq 2000 sequencing data were deposited in the GenBank database (accession numbers PRJNA551932 and PRJNA559779).

2.6. Isoforms in Carotenoid Biosynthetic Pathway between the Avocado Mesocarp and Seed

KEGG analysis in the avocado mesocarp and seed indicated that a total of 104 and 59 isoforms were found to correspond to the putative 11 genes in the carotenoid biosynthetic pathway, respectively (Figure 6). Two to 23 isoforms were found in the putative 11 genes in avocado mesocarp, and one to 15 isoforms were generated from the putative 11 genes in avocado seed. *PaPSY* possessed the most isoform number in avocado mesocarp and seed, respectively. The number of isoforms correspond to the putative 10 genes in the carotenoid biosynthetic pathway were higher in the mesocarp than those in the seed, and increased by 1.33–5.50 fold. However, the number of isoforms corresponding to *PaCYP97A* was lower in the mesocarp than those in the seed.

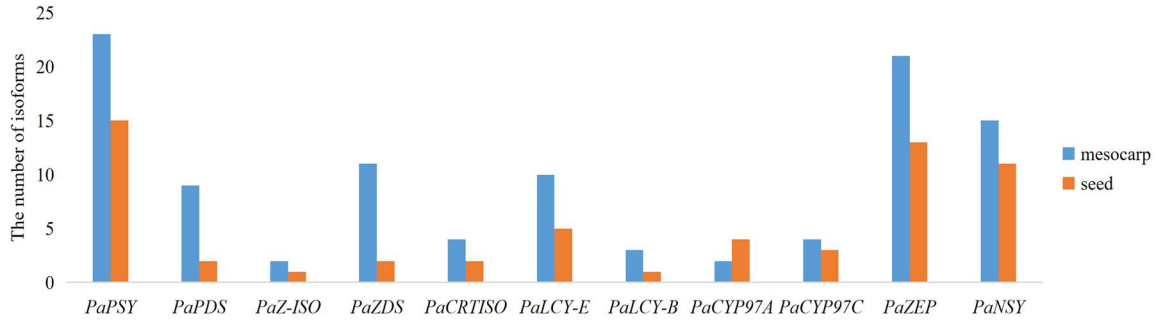


Figure 6. Number of single-molecule real-time (SMRT) isoforms corresponding to the putative 11 genes in the carotenoid biosynthetic pathway.

2.7. Verification of Transcriptome Profiling in Carotenoid Biosynthetic Pathway between the Avocado Mesocarp and Seed by Metabolite Profiling via HPLC

At last, in order to validate transcriptome profiling via NGST and SMRT sequencing in carotenoid biosynthetic pathway between the avocado mesocarp and seed, alpha- and beta-carotene were selected to measure contents during five avocado mesocarp and seed developmental stages by HPLC (Figure S3). The mesocarp alpha- and beta-carotene contents increased slightly from 75 days after full bloom (DAFB) (0.21 and 0.13 μg/g fresh weight (FW), respectively) to 110 DAFB (0.24 and 0.19 μg/g FW, respectively). They then decreased to their lowest levels (0.18 and 0.12 μg/g FW, respectively) at 145 DAFB, but then increased again up to 210 DAFB, peaking at 0.27 and 0.28 μg/g FW, respectively (Figure 7). Trace amounts of alpha- and beta-carotenes were detected in developing seeds, with the contents fluctuating between 0.01 and 0.02 μg/g FW from 75 to 215 DAFB (Figure 7).

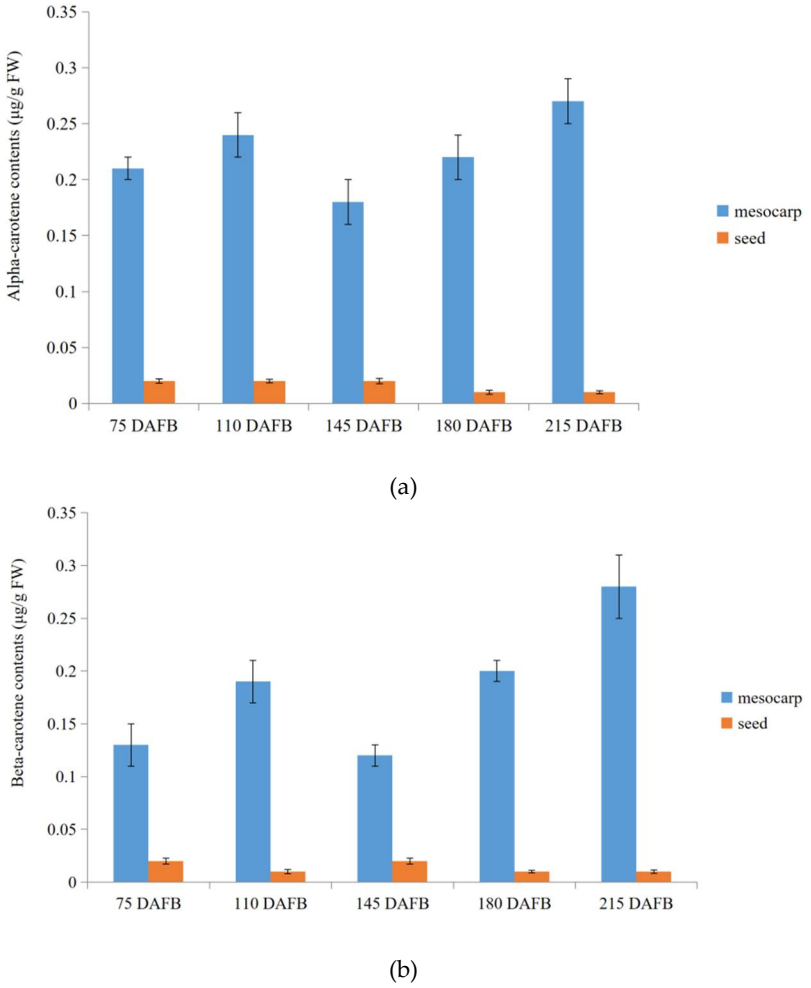


Figure 7. Alpha-carotene and beta-carotene contents during five avocado mesocarp (**a**) and seed (**b**) developmental stages.

3. Discussion

As it is inexpensive and can be completed rapidly, the transcriptome sequencing technique is useful for obtaining a large number of unigene sequences for an organism that lacks an available reference sequence [38]. To the best of our knowledge, for avocado, NGST transcriptome sequencing has been used to investigate fatty acid biosynthesis [39–41], but not any other metabolic biosynthetic pathway. Within our transcriptome assembly, 109.13 and 104.10 Gb of sequence data were respectively generated for the avocado mesocarp and seed during five developmental stages. Additionally, the 100,837 identified unigenes may be useful for subsequent analyses of metabolic biosynthetic pathways in avocado or related species. The N50 and mean lengths of avocado unigenes in our study were 1725 and 847.40 bp, respectively, which implies that our sequence assembly was accurate and effective. The N50 value in this study was higher than those obtained for avocado samples generated from mesocarp during four developmental stages (1050 bp) [41] and our previous avocado samples from five mixed organs sampled (1283 bp) [42], while the mean length in this study was lower than those obtained for both studies (987 and 922 bp) [41,42]. Recently, one of the advances in transcriptome sequencing technology has been the development of the long-read SMRT sequencing technique, which enables researchers to obtain a substantial number of full-length sequences from a cDNA library [28]. In the current study, PacBio SMRT system was applied to generate the full-length transcriptome of avocado mesocarp and seed. The 25.79 and 17.67 Gb SMRT data produced in this study provide the comprehensive insights into the avocado mesocarp and seed, respectively, and might

serve as the genetic basis for future research on avocado. Interestingly, the full-length transcriptome sequence described herein is also the first such sequence for a plant species from the family Lauraceae.

Carotenoids are widely distributed isoprenoid pigments with very diverse biological functions in plants [12]. Carotenoids accumulate as secondary metabolites in leaves [9,22], fruits [21,26,43], and roots [24,25]. The carotenoid biosynthetic pathway has been extensively studied in many photosynthetic and non-photosynthetic organisms, and some researchers confirmed that in most plant species, carotenoid accumulation is mainly controlled by regulating the transcription of genes related to carotenoid biosynthesis [12]. However, the transcript profiles of genes related to carotenoid biosynthesis in avocado fruit remained unclear. In our avocado NGST transcriptome database, we identified 17 unigenes encoding 11 putative enzymes involved in the carotenoid biosynthetic pathway in avocado fruit. The 15 out of 17 unigenes were more highly expressed in the mesocarp than in the seed at each of the five examined time-points. Meanwhile, SMRT transcriptome database in our study indicated that the number of isoforms correspond to the putative 10 genes in the carotenoid biosynthetic pathway were higher in the mesocarp than those in the seed. Furthermore, the metabolite (alpha- and beta-carotene) profiling via HPLC in the avocado mesocarp and seed during five developmental stages in this study validated the results of our NGST and SMRT transcriptome profiling. These results clearly showed that the upregulated expression levels of most unigenes encoding 11 putative enzymes involved in the carotenoid biosynthetic pathway might contribute to the higher carotenoid pathway flux in the avocado mesocarp than in the seed. Besides, gene dosage (isoform number) increase of most carotenoid biosynthetic-related genes could also accelerate the carotenoid accumulation. Previous studies revealed gene dosage balance impacts on agronomic traits in plants, and defined the linkage between quantitative trait and gene dosage variation [44–46]. Consequently, we might imply that the gene dosage variation and the associated changes in gene expression of these unigenes might be important for controlling the carotenoid contents in avocado during the mesocarp and seed developmental stage.

An earlier investigation proved that upregulated *PSY* and *PDS* expression levels are correlated with the total carotenoid content during the tomato fruit maturation stage [47]. Similarly, *PSY*, *ZDS*, *CRTISO*, and *LCY-E* might be key genes for controlling carotenoid contents in *M. cochinchinensis* ripening fruits [21]. Additionally, *LCY-B* expression contributes to the accumulation of carotenoids in papaya [48], kiwifruit [49], and citrus [50] fruits. Another study indicated that *PSY* expression is also related to the alpha- and beta-carotene as well as total carotenoid contents in red pepper fruits [51]. In *B. campestris* L. subsp. *chinensis* var. *rosularis* Tsen and Lee leaves, *LCY-E* and *ZDS* expression may be vital for carotenoid biosynthesis [22]. In celery, *PSY* and *LCY-E* expression may be important for promoting beta-carotene biosynthesis. In the potato tuber, *PSY* expression is considered to increase the beta-carotene content [52]. Welsch [53] also suggested that *PSY* expression mediates the beta-carotene accumulation in cassava roots. Thus, analyses of the differences in gene expression profiles may yield new insights into carotenoid biosynthetic mechanisms and identify diverse carotenogenic genes expressed in various developmental stages, tissues, and species as well as in response to specific treatments.

The identification of genes encoding enzymes related to the carotenoid biosynthetic pathway not only facilitates the characterization of physiological functions in higher plants, it also provides useful information relevant for metabolic engineering. On the basis of NGST and SMRT transcriptome sequencing in this study, we investigated the differences in carotenoid biosynthesis between the avocado mesocarp and seed. However, carotenoid biosynthesis involves complex biological processes regulated by many biological pathways (i.e., the MVA, MEP, and carotenoid biosynthetic pathways) and genes. The NGST and SMRT transcriptome database described herein may represent a useful resource for clarifying carotenoid biosynthesis in various avocado tissues. Additionally, to the best of our knowledge, this study is the first to integrate Illumina with PacBio SMRT sequencing platforms for investigating avocado mesocarp and seed developmental stages via transcriptome sequencing and assembly without a reference genome. We believe that the transcriptome dataset will provide a

solid foundation for future functional and genomics-based analyses of avocado, and will be useful for elucidating metabolic biosynthetic mechanisms.

4. Materials and Methods

4.1. Plant Materials

Avocado fruits (cultivar 'Hass') were harvested from six 10-year-old trees grafted onto Zutano clonal rootstock (two trees were used as a unit for each biological replicate) from April 2018 to September 2018 at the Chinese Academy of Tropical Agricultural Sciences (Danzhou, Hainan, China: 19.52°N, 109.57°E; altitude = 200 m above sea level). In these trees, fruits that developed during the main flowering season (i.e., February 2018) were marked, after which samples were collected at five time-points (75, 110, 145, 180, and 215 DAFB) until the fruits reached physiological maturity (defined as the ability to ripen after harvest). Two sets of fruits were randomly collected for each biological replicate during each developmental stage. The first set of nine fruits was used to measure fruit, mesocarp, and seed phenotypic traits. The second set of nine fruits was used for transcript and carotenoid analyses. Fruits were quickly brought to the laboratory, after which their phenotypic traits were measured as previously described [7] or they were immediately frozen at −80 °C for transcript and carotenoid analyses.

4.2. NGST Sequencing

Total RNA was extracted using a Plant RNA Kit (OMEGA Bio-Tek, Norcross, GA, USA). RNA concentration was measured using NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA).mRNA was purified from total RNA with poly-T oligo-attached magnetic beads. Samples underwent an RNA-seq analysis involving three biological replicates per sample. The fragmentation step was completed with divalent cations in the NEBNext First Strand Synthesis Reaction Buffer (5×) at an elevated temperature. First-strand cDNA was synthesized with a series of random hexamer primers and reverse transcriptase, and second-strand cDNA was subsequently produced with DNA Polymerase I and RNase H. The cDNA libraries were constructed by ligating cDNA fragments to sequencing adapters and amplifying fragments by PCR. The libraries were then sequenced with the Illumina HiSeq 2000 platform (Nanxin Bioinformatics Technology Co., Ltd., Guangzhou, China).

4.3. Transcriptome Assembly, Annotation, and Coding Sequence Prediction

Clean data (clean reads) were obtained by discarding reads with adapters, reads with ambiguous poly-N sequences, and low-quality reads in which more than 50% of the bases had a Q-value ≤ 20. The two read files that were independently established for the libraries/samples were used for assembling the transcriptome with the Trinity program (version 2.5.1) [54]; the min_kmer_cov was set to 2 and all other parameters were set to default values. The assembled transcripts were hierarchically clustered to unigenes through shared reads and expression by the Corset program [55].

Unigenes were functionally annotated with a BLASTX alignment algorithm (E-value threshold of 10^{-5}) and the following databases: KOG/COG, Swiss-Prot (manually annotated and reviewed protein sequence database), Pfam (along with the HMMER3.0 package), Nr, and Nt (comprising non-redundant nucleotide sequences). The KEGG Automatic Annotation Server [56] was used to map these genes according to the KEGG metabolic pathway database. Rich factor = (the number of DEGs/the number of all DEGs)/(the number of all unigenes in pathways/the number of all unigenes in KEGG). Additionally, Blast2GO (version 2.5) [57] was used for the annotation of unigenes with GO terms based on the BLASTX hits against the Pfam and Nr databases, with a cut-off E-value of 10^{-6}. To predict the coding sequences, the unigenes were first used to screen the Nr and Swiss-Prot databases with a BLAST algorithm, after which the open reading frame data for sequence matches

were acquired directly. The coding sequences for the remaining unigenes were predicted with ESTScan (version 3.0.3) (https://sourceforge.net/projects/estscan/).

4.4. Identification of Differentially Expressed Genes

To identify DEGs between two samples, the gene expression levels were quantified with the FPKM method. The read counts were adjusted with the edgeR program package, with one scaling-normalized factor for each sequenced library. The DEGs between two samples were analyzed with the DEGSeq R package (version 1.20.0). The *p*-values were adjusted according to the Benjamini and Hochberg method. A corrected *p*-value of 0.005 and a $\log_2$ (fold-change) of 1 were set as the threshold for identifying significant DEGs. Significantly enriched GO terms and KEGG pathways were determined based on a corrected *p*-value ≤ 0.05. The GO functional enrichment and KEGG pathway enrichment analyses of the DEGs were completed with GOseq R packages and KOBAS (version 2.0) (http://kobas.cbi.pku.edu.cn/home.do), respectively.

4.5. Validation of Transcripts by Quantitative Real-Time PCR

The expression levels of 10 unigenes related to carotenoid biosynthesis in the avocado mesocarp and seed were validated by a qRT-PCR assay, which was completed with a 96-well plate and the QuantStudio 7 Flex Real Time PCR System (Applied Biosystems, Foster City, CA, USA). Details regarding the qRT-PCR primers are presented in Table S8. Total RNA was extracted from the mesocarp and seed at the five developmental stages using RNAiso Plus Reagent (TaKaRa Bio Inc., Kusatsu, Japan) based on the manufacturer's protocol, then treated with RNase-free DNase I (New England Biolabs, Ipswich, MA, USA) to eliminate all contaminating DNA. The resulting RNA was applied for first strand synthesis by the PrimeScriptRT reagent Kit with gDNA Eraser (TaKaRa Bio Inc.). The concentration of cDNA was determined and diluted to 12.5 ng/μL. PCR was performed using QuantStudio7 Flex Real Time PCR System (Applied Biosystems, Foster City, CA, USA).The 20-μL reaction volumes comprised 2 μL cDNA, 10 μL SYBR *Premix Ex* Taq™ II (TliRNaseH Plus) (TaKaRa Bio Inc.), 1.0 μL each 10 μM primer, and 6 μL distilled water. The PCR program was as follows: 95 °C for 30 s; 40 cycles of 95 °C for 5 s, melting temperature of each primer for 30 s. The *PaActin7* gene was used as an endogenous control for normalizing data and $2^{-\Delta CT}$ method was used for PCR data analysis. For each sample, the qRT-PCR analysis involved three biological replicates and two technical replicates.

4.6. SMRT Sequencing

Poly-T oligo-attached magnetic beads were used to purify mRNA from the total RNA extracted from mesocarp and seed samples collected at each analyzed developmental stage. The mRNA from all five developmental stages was combined to serve as the template to synthesize cDNA with the SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). After a PCR amplification, quality control check, and purification, full-length cDNA fragments were acquired according to the BluePippin Size Selection System protocol, ultimately resulting in the construction of a cDNA library (1–6 kb). Selected full-length cDNA sequences were ligated to the SMRT bell hairpin loop. The concentration of the cDNA library was then determined with the Qubit 2.0 fluorometer, whereas the quality of the cDNA library was assessed with the 2100 Bioanalyzer (Agilent). Finally, one SMRT cell each was sequenced respectively with the PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA) for avocado mesocarp and seed.

4.7. Quality Filtering and Correction of PacBio Long-Reads

Raw reads were processed into error-corrected reads of insert (ROIs) using an isoform sequencing pipeline, with minimum full pass = 0.00 and minimum predicted accuracy = 0.80. Next, full-length, non-chimeric transcripts were detected by searching for the poly-A tail signal and the 5′ and 3′ cDNA primer sequences in the ROIs. Iterative clustering for error correction was used to obtain high-quality consensus isoforms, which were then polished with Quiver. The low-quality full-length transcript

isoforms were corrected based on Illumina short-reads with the default setting of the Proovread program. High-quality and corrected low-quality transcript isoforms were confirmed as nonredundant with the CD-HIT (version 1) (http://weizhongli-lab.org/cd-hit/).

4.8. Analysis of Alpha- and Beta-Carotenes by HPLC

Avocado mesocarp and seed extracts were prepared as previously described [58], with minor modifications. Briefly, fresh avocado mesocarp and seed samples were separately ground in a mortar containing liquid nitrogen. Samples (approximately 2 g) were added to centrifuge tubes, after which they were treated with 4 mL acetone and homogenized (intermediate speed) for 1.5 min at 4 °C. The supernatant was then transferred to a new centrifuge tube, and the extraction of the residue was repeated twice. The extracts were mixed with 5 mL methanolic potassium hydroxide (15%, *w*/*v*), and then saponified for 2 h in the presence of nitrogen. A 3-mL aliquot of the mixture was diluted with 1 mL 10% sodium chloride and then added to a 2-mL solution comprising methylene chloride and water. The supernatant was washed three times with water, evaporated to dryness in the presence of nitrogen, and reconstituted in methanol/methyl tert-butyl ether (85:15). The subsequent HPLC analysis of carotenoids was completed with the 1290 HPLC system (Agilent, Santa Clara, CA, USA) and a YMC carotenoid C30 column (250 × 4.6 mm, 5 µm; Waters, Santa Clara, CA, USA) analyzed at 445 nm. The HPLC mobile phase consisted of methanol/water (96:4, *v*/*v*) and tert-butyl ether at a flow rate of 1.0 mL/min and the column temperature was maintained at 30 °C. The alpha- and beta-carotenes were identified by comparing the retention times of the peaks with those of commercial standards purchased from Sigma-Aldrich (Shanghai, China). Carotenoid contents were quantified based on external calibration curves ($R^2 \geq 0.999$). Alpha- and beta-carotene contents were expressed as microgram per gram of fresh weight (µg/g FW). Samples were analyzed with three biological replicates and two technical replicates.

5. Conclusions

This study provides a comprehensive overview of the NGST transcriptomes of the avocado mesocarp and seed at five developmental stages. NGST and SMRT transcriptomes results implied that the gene dosage variation and the associated changes in gene expression of most carotenoid biosynthetic-related genes might contribute to the higher carotenoid pathway flux in the avocado mesocarp than in the seed, and accelerate the carotenoid accumulation. The metabolite (alpha- and beta-carotene) profiling via HPLC in the avocado mesocarp and seed during five developmental stages in this study validated the results of our NGST and SMRT transcriptome profiling. Our study results provide new insights into the carotenoid contents and the molecular mechanisms underlying carotenoid accumulation in avocado.

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/17/4117/s1. Table S1. Phenotypes of avocado cultivar 'Hass' at 75, 110, 145, 180, and 215 days after full bloom during the fruit developmental stage; Table S2. Transcriptomic data for 30 avocado samples; Table S3. Gene ontology annotations for the assembled avocado unigenes; Table S4. Enriched Kyoto Encyclopedia of Genes and Genomes pathways among the assembled avocado unigenes; Table S5. Data for the differentially expressed genes in various cDNA libraries for developing mesocarps and seeds; Table S6. Analysis of differential expression and annotation of unigenes related to the carotenoid biosynthetic pathway; Table S7. PacBio library and sequencing results in avocado mesocarp and seed; Table S8. Details regarding qRT-PCR primers; Figure S1. The photos of the tested avcoado samples per fruit developmental stage. Figure S2. Transcript and unigene length distributions; Figure S3. Chromatogram of α-carotene and β-carotene extracted from 215 DAFB mesocarp and seed of avocado 'Hass'.

Author Contributions: Y.G., X.D. and R.Z. conceived and designed the experiments; X.S., L.T. performed the experiments; Z.C., B.W. and N.W. analyzed the data; Z.X., Z.Z. and X.L. helped complete the experiments; W.M. and X.Z. contributed materials; and Y.G. wrote the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 31701883) and the Central Public-interest Scientific Institution Basal Research Fund for the Chinese Academy of Tropical Agricultural Sciences (grant number 1630092018003).

Acknowledgments: We gratefully acknowledge Pingzhen Lin from the Haikou Experimental Station of the Chinese Academy of Tropical Agricultural Sciences for their valuable support related to the avocado resource collection. We gratefully acknowledge three anonymous reviewers for their constructive comments. We thank Yajima for editing the English text of a draft of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gross-German, E.; Viruel, M.A. Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: Genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. *Tree Genet. Genomes* **2013**, *9*, 539–555. [CrossRef]
2. Galindo-Tovar, M.E.; Ogata-Aguilar, N.; Arzate-Fernandez, A.M. Some aspects of avocado (Persea americana Mill.) diversity and domestication in Mesoamerica. *Genet. Resour. Crop Evol.* **2008**, *55*, 441–450. [CrossRef]
3. Ge, Y.; Zhang, T.; Wu, B.; Tan, L.; Ma, F.; Zou, M.; Chen, H.; Pei, J.; Liu, Y.; Chen, Z.; et al. Genome-Wide Assessment of Avocado Germplasm Determined from Specific Length Amplified Fragment Sequencing and Transcriptomes: Population Structure, Genetic Diversity, Identification, and Application of Race-Specific Markers. *Genes* **2019**, *10*, 215. [CrossRef] [PubMed]
4. Schaffer, B.; Wolstenholme, B.N.; Whiley, A.W. *The Avocado: Botany, Production and Uses*, 2nd ed.; CPI Group Ltd.: Croydon, UK, 2012.
5. Dreher, M.L.; Davenport, A.J. Hass Avocado Composition and Potential Health Effects. *Crit. Rev. Food Sci. Nutr.* **2013**, *53*, 738–750. [CrossRef] [PubMed]
6. Galvão, M.D.S.; Narain, N.; Nigam, N. Influence of different cultivars on oil quality and chemical characteristics of avocado fruit. *Food Sci. Technol.* **2014**, *34*, 539–546. [CrossRef]
7. Ge, Y.; Si, X.; Cao, J.; Zhou, Z.; Wang, W.; Ma, W. Morphological Characteristics, Nutritional Quality, and Bioactive Constituents in Fruits of Two Avocado (Persea americana) Varieties from Hainan Province, China. *J. Agric. Sci.* **2017**, *9*, 8–17. [CrossRef]
8. Joung, J.G.; Chung, M.Y.; Tieman, D.; Lee, J.M.; Joung, J.; McQuinn, R.; Chung, M.; Fei, Z.; Klee, H.; Giovannoni, J. Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. *Plant J.* **2012**, *70*, 191–204.
9. Li, J.W.; Ma, J.; Feng, K.; Liu, J.X.; Que, F.; Xiong, A.S. Carotenoid Accumulation and Distinct Transcript Profiling of Structural Genes Involved in Carotenoid Biosynthesis in Celery. *Plant Mol. Biol. Rep.* **2018**, *36*, 663–674. [CrossRef]
10. Goff, S.A.; Klee, H.J. Plant Volatile Compounds: Sensory Cues for Health and Nutritional Value? *Scinece* **2006**, *311*, 815–819. [CrossRef]
11. Walter, M.H.; Floss, D.S.; Strack, D. Apocarotenoids: Hormones, mycorrhizal metabolites and aroma volatiles. *Planta* **2010**, *232*, 1–17. [CrossRef]
12. Nisar, N.; Li, L.; Lu, S.; Khin, N.C.; Pogson, B.J. Carotenoid Metabolism in Plants. *Mol. Plant* **2015**, *8*, 68–82. [CrossRef] [PubMed]
13. Britton, G. Structure and properties of carotenoids in relation to function. *FASEB J.* **1995**, *9*, 1551–1558. [CrossRef] [PubMed]
14. Yuan, H.; Zhang, J.; Nageswaran, D.; Li, L. Carotenoid metabolism and regulation in horticultural crops. *Hortic. Res.* **2015**, *2*, 15036. [CrossRef] [PubMed]
15. Li, J.; Guo, H.; Wang, Y.; Zong, J.; Chen, J.; Li, D.D.; Li, L.; Wang, J.J.; Liu, J.X. High-throughput SSR marker development and its application in a centipedegrass (Eremochloa ophiuroides (Munro) Hack.) genetic diversity analysis. *PLoS ONE* **2018**, *13*, e0202605. [CrossRef] [PubMed]
16. Li, W.; Zhang, C.; Jiang, X.; Liu, Q.; Liu, Q.; Wang, K. De Novo Transcriptomic Analysis and Development of EST–SSRs for Styrax japonicus. *Forests* **2018**, *9*, 748. [CrossRef]
17. Liu, F.M.; Hong, Z.; Yang, Z.J.; Zhang, N.N.; Liu, X.J.; Xu, D.P. De Novo transcriptome analysis of Dalbergia odorifera T. Chen (Fabaceae) and transferability of SSR markers developed from the transcriptome. *Forests* **2019**, *10*, 98. [CrossRef]
18. Ma, S.; Dong, W.; Lyu, T.; Lyu, Y. An RNA Sequencing Transcriptome Analysis and Development of EST-SSR Markers in Chinese Hawthorn through Illumina Sequencing. *Forests* **2019**, *10*, 82. [CrossRef]

19. Penin, A.A.; Klepikova, A.V.; Kasianov, A.S.; Gerasimov, E.S.; Logacheva, M.D. Comparative Analysis of Developmental Transcriptome Maps of Arabidopsis thaliana and Solanum lycopersicum. *Genes* **2019**, *10*, 50. [CrossRef]
20. Weisberg, A.J.; Kim, G.; Westwood, J.H.; Jelesko, J.G. Sequencing and De Novo Assembly of the Toxicodendron radicans (Poison Ivy) Transcriptome. *Genes* **2017**, *8*, 317. [CrossRef]
21. Hyun, T.K.; Rim, Y.; Jang, H.J.; Kim, C.H.; Park, J.; Kumar, R.; Lee, S.; Kim, B.C.; Bhak, J.; Nguyen-Quoc, B.; et al. De novo transcriptome sequencing of Momordica cochinchinensis to identify genes involved in the carotenoid biosynthesis. *Plant Mol. Biol.* **2012**, *79*, 413–427. [CrossRef]
22. Tian, H.M.; Fang, L.; Zhang, Q.A.; Wang, M.X.; Wang, Y.; Jia, L. Transcriptome analysis of carotenoid biosynthesis in the Brassica campestris L. subsp. chinensis var. rosularis Tsen. *Sci. Hortic.* **2018**, *235*, 116–123. [CrossRef]
23. He, Y.; Ma, Y.; Du, Y.; Shen, S. Differential gene expression for carotenoid biosynthesis in a green alga Ulva prolifera based on transcriptome analysis. *BMC Genom.* **2018**, *19*, 916. [CrossRef] [PubMed]
24. Ma, J.; Li, J.; Xu, Z.; Wang, F.; Xiong, A. Transcriptome profiling of genes involving in carotenoid biosynthesis and accumulation between leaf and root of carrot (*Daucus carota* L.). *Acta Biochim. Biophys. Sin.* **2018**, *50*, 481–490. [CrossRef] [PubMed]
25. Machaj, G.; Bostan, H.; Macko-Podgórni, A.; Iorizzo, M.; Grzebelus, D. Comparative Transcriptomics of Root Development in Wild and Cultivated Carrots. *Genes* **2018**, *9*, 431. [CrossRef] [PubMed]
26. Yuan, X.; Sun, W.; Zou, X.; Liu, B.; Huang, W.; Chen, Z.; Li, Y.; Qiu, M.-Y.; Liu, Z.-J.; Mao, Y.; et al. Sequencing of Euscaphis konishii Endocarp Transcriptome Points to Molecular Mechanisms of Endocarp Coloration. *Int. J. Mol. Sci.* **2018**, *19*, 3209. [CrossRef] [PubMed]
27. Chen, J.; Tang, X.; Ren, C.; Wei, B.; Wu, Y.; Wu, Q.; Pei, J. Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genom.* **2018**, *19*, 548. [CrossRef] [PubMed]
28. Tian, J.; Feng, S.; Liu, Y.; Zhao, L.; Tian, L.; Hu, Y.; Yang, T.; Wei, A. Single-Molecule Long-Read Sequencing of Zanthoxylum bungeanum Maxim. Transcriptome: Identification of Aroma-Related Genes. *Forest* **2018**, *9*, 765. [CrossRef]
29. Roberts, R.J.; Carneiro, M.O.; Schatz, M.C. The advantages of SMRT sequencing. *Genome Biol.* **2013**, *14*, 405. [CrossRef] [PubMed]
30. Chao, Y.H.; Yuan, J.B.; Li, S.F.; Jia, S.Q.; Han, L.B.; Xu, L.X. Analysis of transcripts and splice isoforms in red clover (Trifolium pratense L.) by single-molecule long-read sequencing. *BMC Plant Biol.* **2018**, *18*, 300. [CrossRef] [PubMed]
31. Hoang, N.V.; Furtado, A.; Mason, P.J.; Marquardt, A.; Kasirajan, L.; Thirugnanasambandam, P.P.; Botha, F.C.; Henry, R.J. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genom.* **2017**, *18*, 395. [CrossRef] [PubMed]
32. Zuo, C.; Blow, M.; Sreedasyam, A.; Kuo, R.C.; Ramamoorthy, G.K.; Torres-Jerez, I.; Li, G.; Wang, M.; Dilworth, D.; Barry, K.; et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Biotechnol. Biofuels* **2018**, *11*, 170. [CrossRef] [PubMed]
33. Chao, Y.; Yuan, J.; Guo, T.; Xu, L.; Mu, Z.; Han, L. Analysis of transcripts and splice isoforms in Medicago sativa L. by single-molecule long-read sequencing. *Plant Mol. Biol.* **2019**, *99*, 219–235. [CrossRef] [PubMed]
34. Kim, J.A.; Roy, N.S.; Lee, I.H.; Choi, A.Y.; Choi, B.S.; Yu, Y.S.; Park, N.I.; Park, K.C.; Kim, S.; Yang, H.S.; et al. Genome-wide transcriptome profiling of the medicinal plant Zanthoxylum planispinum using a single-molecule direct RNA sequencing approach. *Genomics* **2019**, *111*, 973–979. [CrossRef] [PubMed]
35. Zhang, B.; Liu, J.; Wang, X.; Wei, Z. Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol. Biochem.* **2018**, *132*, 95–103. [CrossRef] [PubMed]
36. Xu, Q.; Zhu, J.; Zhao, S.; Hou, Y.; Li, F.; Tai, Y.; Wan, X.; Wei, C. Transcriptome Profiling Using Single-Molecule Direct RNA Sequencing Approach for In-depth Understanding of Genes in Secondary Metabolism Pathways of Camellia sinensis. *Front. Plant Sci.* **2017**, *8*, 1205. [CrossRef] [PubMed]
37. Deng, Y.; Zheng, H.; Yan, Z.; Liao, D.; Li, C.; Zhou, J.; Liao, H. Full-Length Transcriptome Survey and Expression Analysis of Cassia obtusifolia to Discover Putative Genes Related to Aurantio-Obtusin Biosynthesis, Seed Formation and Development, and Stress Response. *Int. J. Mol. Sci.* **2018**, *19*, 2476. [CrossRef] [PubMed]

38. Du, M.; Li, N.; Niu, B.; Liu, Y.; You, D.; Jiang, D.; Ruan, C.; Qin, Z.; Song, T.; Wang, W. De novo transcriptome analysis of Bagarius yarrelli (*Siluriformes: Sisoridae*) and the search for potential SSR markers using RNA-Seq. *PLoS ONE* **2018**, *13*, e0190343. [CrossRef]
39. Ibarra-Laclette, E.; Méndez-Bravo, A.; Pérez-Torres, C.A.; Albert, V.A.; Mockaitis, K.; Kilaru, A.; López-Gómez, R.; Cervantes-Luevano, J.I.; Herrera-Estrella, L. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genom.* **2015**, *16*, 599. [CrossRef]
40. Kilaru, A.; Cao, X.; Dabbs, P.B.; Sung, H.-J.; Rahman, M.M.; Thrower, N.; Zynda, G.; Podicheti, R.; Ibarra-Laclette, E.; Herrera-Estrella, L.; et al. Oil biosynthesis in a basal angiosperm: Transcriptome analysis of Persea Americana mesocarp. *BMC Plant Biol.* **2015**, *15*, 203. [CrossRef]
41. Vergara-Pulgar, C.; Rothkegel, K.; González-Agüero, M.; Pedreschi, R.; Campos-Vargas, R.; Defilippi, B.G.; Meneses, C. De novo assembly of Persea americana cv. 'Hass' transcriptome during fruit development. *BMC Genom.* **2019**, *20*, 108. [CrossRef]
42. Ge, Y.; Tan, L.; Wu, B.; Wang, T.; Zhang, T.; Chen, H.; Zou, M.; Ma, F.; Xu, Z.; Zhan, R. Transcriptome Sequencing of Different Avocado Ecotypes: De novo Transcriptome Assembly, Annotation, Identification and Validation of EST-SSR Markers. *Forests* **2019**, *10*, 411. [CrossRef]
43. Wisutiamonkul, A.; Ampomah-Dwamena, C.; Allan, A.C.; Ketsa, S. Carotenoid accumulation in durian (Durio zibethinus) fruit is affected by ethylene via modulation of carotenoid pathway gene expression. *Plant Physiol. Biochem.* **2017**, *115*, 308–319. [CrossRef] [PubMed]
44. Bastiaanse, H.; Zinkgraf, M.; Canning, C.; Tsai, H.; Lieberman, M.; Comai, L.; Henry, I.; Groover, A. A comprehensive genomic scan reveals gene dosage balance impacts on quantitative traits in Populus trees. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 13690–13699. [CrossRef] [PubMed]
45. Birchler, J.A.; Bhadra, U.; Bhadra, M.P.; Auger, D.L. Dosage-Dependent Gene Regulation in Multicellular Eukaryotes: Implications for Dosage Compensation, Aneuploid Syndromes, and Quantitative Traits. *Dev. Biol.* **2001**, *234*, 275–288. [CrossRef] [PubMed]
46. Förster, S.; Schumann, E.; Baumann, M.; Weber, W.E.; Pillen, K. Copy number variation of chromosome 5A and its association with Q gene expression, morphological aberrations, and agronomic performance of winter wheat cultivars. *Theor. Appl. Genet.* **2013**, *126*, 3049–3063. [CrossRef] [PubMed]
47. Bramley, P.M. Regulation of carotenoid formation during tomato fruit ripening and development. *J. Exp. Bot.* **2002**, *53*, 2107–2113. [CrossRef] [PubMed]
48. Devitt, L.C.; Fanning, K.; Dietzgen, R.G.; Holton, T.A. Isolation and functional characterization of a lycopene β-cyclase gene that controls fruit colour of papaya (*Carica papaya* L.). *J. Exp. Bot.* **2010**, *61*, 33–39. [CrossRef] [PubMed]
49. Ampomah-Dwamena, C.; McGhie, T.; Wibisono, R.; Montefiori, M.; Hellens, R.P.; Allan, A.C. The kiwifruit lycopene beta-cyclase plays a significant role in carotenoid accumulation in fruit. *J. Exp. Bot.* **2009**, *60*, 3765–3779. [CrossRef]
50. Kato, M.; Ikoma, Y.; Matsumoto, H.; Sugiura, M.; Hyodo, H.; Yano, M. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol.* **2004**, *134*, 824–837. [CrossRef]
51. Hornero-Méndez, D.; Gómez-Ladrón, D.G.R.; Mínguez-Mosquera, M.I. Carotenoid biosynthesis changes in five red pepper (*Capsicum annuum* L.) cultivars during ripening. Cultivar selection for breeding. *J. Agric. Food Chem.* **2000**, *48*, 3857–3864. [CrossRef]
52. Ducreux, L.J.; Morris, W.L.; Hedley, P.E.; Shepherd, T.; Davies, H.V.; Millam, S.; Taylor, M.A. Metabolic engineering of high carotenoid potato tubers containing enhanced levels of beta-carotene and lutein. *J. Exp. Bot.* **2005**, *56*, 81–89. [CrossRef] [PubMed]
53. Welsch, R.; Arango, J.; Bär, C.; Salazar, B.; Al-Babili, S.; Beltrán, J.; Chavarriaga, P.; Ceballos, H.; Tohme, J.; Beyer, P. Provitamin A Accumulation in Cassava (*Manihot esculenta*) Roots Driven by a Single Nucleotide Polymorphism in a Phytoene Synthase Gene. *Plant Cell* **2010**, *22*, 3348–3356. [CrossRef] [PubMed]
54. Lu, Q.Y.; Zhang, Y.; Wang, Y.; Wang, D.; Lee, R.P.; Gao, K.; Byrns, R.; Heber, D. California Hass Avocado: Profiling of Carotenoids, tocopherol, fatty acid, and fat content during maturation and from different growing areas. *J. Agric. Food Chem.* **2009**, *57*, 10408–10413. [CrossRef] [PubMed]

55. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
56. Davidson, N.M.; Oshlack, A. Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* **2014**, *15*, 410. [PubMed]
57. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, 480–484. [CrossRef]
58. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Review

Modern Trends in Plant Genome Editing: An Inclusive Review of the CRISPR/Cas9 Toolbox

Ali Razzaq [1], Fozia Saleem [1], Mehak Kanwal [2], Ghulam Mustafa [1], Sumaira Yousaf [2], Hafiz Muhammad Imran Arshad [2], Muhammad Khalid Hameed [3], Muhammad Sarwar Khan [1] and Faiz Ahmad Joyia [1,*]

1 Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad 38040, Pakistan
2 Nuclear Institute for Agriculture and Biology (NIAB), P.O. Box 128, Faisalabad 38000, Pakistan
3 School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: faizahmad1980@uaf.edu.pk

Received: 14 June 2019; Accepted: 15 August 2019; Published: 19 August 2019

Abstract: Increasing agricultural productivity via modern breeding strategies is of prime interest to attain global food security. An array of biotic and abiotic stressors affect productivity as well as the quality of crop plants, and it is a primary need to develop crops with improved adaptability, high productivity, and resilience against these biotic/abiotic stressors. Conventional approaches to genetic engineering involve tedious procedures. State-of-the-art OMICS approaches reinforced with next-generation sequencing and the latest developments in genome editing tools have paved the way for targeted mutagenesis, opening new horizons for precise genome engineering. Various genome editing tools such as transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs), and meganucleases (MNs) have enabled plant scientists to manipulate desired genes in crop plants. However, these approaches are expensive and laborious involving complex procedures for successful editing. Conversely, CRISPR/Cas9 is an entrancing, easy-to-design, cost-effective, and versatile tool for precise and efficient plant genome editing. In recent years, the CRISPR/Cas9 system has emerged as a powerful tool for targeted mutagenesis, including single base substitution, multiplex gene editing, gene knockouts, and regulation of gene transcription in plants. Thus, CRISPR/Cas9-based genome editing has demonstrated great potential for crop improvement but regulation of genome-edited crops is still in its infancy. Here, we extensively reviewed the availability of CRISPR/Cas9 genome editing tools for plant biotechnologists to target desired genes and its vast applications in crop breeding research.

Keywords: CRISPR/Cas9; genome editing; plant breeding; multiplex genome editing; crop improvement; TALEN; ZFN; biotic stress; abiotic stress

1. Introduction

Food security is the most crucial challenge in the current scenario of a rapidly growing global population. According to cautious estimates, the global population will escalate to ten billion by the end of 2050 and a 60–100% rise in global food production will be necessary [1]. Besides extreme weather, increasing biotic and abiotic stressors, a growing population, and shrinking availability of agricultural land and water resources are important constraints for food production and farming. Over the past few decades, improvements in crop plants have contributed by deciphering numerous biological mechanisms and elucidating the role of genetic and epigenetics factors [2]. Crop breeders and plant scientist are striving hard to understand mainly the genetic mechanism underlying unique plant responses towards environmental stressors. Recently, numerous novel genes and their regulatory pathways have been identified in plants [3,4]. For crop improvement and development of elite cultivars

with increased productivity, a breeding strategy of "cross the elite with the elite and wait for the best" has been applied, concentrating on the genes linked with vital agronomic traits [5]. Classical plant breeding strategies for crop improvement are more challenging which take a long time for germplasm selection. On the other hand, modern tools for genome editing (GE) exhibit the capability of integrating a foreign gene into a predetermined site of the genome precisely, allowing for accurate substitution of an existing allele with an alternative one [6,7]. Genome editing has emerged as a tremendous strategy for efficient and targeted genome manipulations, especially for crops which have complex genomes and which are difficult to improve through conventional breeding approaches [6].

For basic as well as applied plant biology, the unstable and non-specific transgene incorporation in the host genome has been a matter of concern for edible crop species [8]. The discovery of programmed sequence-specific nucleases (SSNs) has facilitated precise gene editing. In both plant and animal systems, application of SSNs for accurate GE has been recognized as a breakthrough in genome engineering. The SSNs can be applied to produce several kinds of mutations, such as insertions, deletions, replacement, substitutions, integration of specific sequence of DNA at a desired locus, and site-directed substitutions across many organisms and cell types. Though all types of SSNs have unique features, the mechanism for producing double-strand breaks (DSBs) in the target DNA is similar for all. The DSBs created by SSNs are reconstructed via non-homologous end joining (NHEJ) or homology-directed recombination (HDR). Non-homologous end joining is an error-prone DNA repair mechanism that facilitates direct end-joining of DSBs without involving a homologous template and can generate insertions or deletions at target sites to develop gene knockouts. Additionally, NHEJ can also be applied to introduce insertions at the point of the DSB during operation of the repair mechanism. On the other hand, the HDR repair pathway is a highly accurate mechanism that needs a homologous template to mediate repair and can be used to attain precise changes like gene insertion and gene replacement [6,9,10]. As compared to transgenic strategies, which result in inadvertent gene insertions and sometimes random phenotypical characters, GE approaches produce well-defined mutants, proving GE as a powerful technique for plant breeding and functional genomics. In contrast to transgenic plants, genome-edited plants have the added benefit of site specificity [11]. In breeding programs, these improved plants can be proven useful and subsequent species can be employed reliably with less concerns and comparatively minor monitoring methods are needed in contrast to traditional genetically engineered plants [12].

2. Modern Trends in Plant Genome Editing

In recent years, many fascinating GE approaches have been established because of the advancements in molecular biology, which have permitted site-specific and accurate editing in many genomes [8]. In GE, engineered nucleases are composed of a sequence-specific DNA binding domain merged with a non-specific nucleases domain. Targeted genes can be precisely cleaved by such fused nucleases and nicks can be repaired with the help of HDR or NHEJ [13,14]. A vital strategy to execute targeted GE through SSNs is to generate DSBs at targeted sites; these nicks prompt the activation of the DNA repair mechanism through the HDR or NHEJ pathway [15]. The DNA repair system of the HDR pathways requires a homologous template to repair the DSB, whereas the two ends of DSBs are directly ligated in the NHEJ pathway [15]. Though NHEJ is more common, there are some flaws which make it undesirable in many studies. The major disadvantage of this process is that it produces insertions or deletions of different sizes during the repair mechanism, which may produce off-targets. In contrast to NHEJ, the repair mechanism via HDR is more accurate and reliable, which depends on homologous DNA to repair the DSB [16]. Thus, SSNs can be applied to manipulate the genomic sequences by targeted addition or deletion of specific nucleotides in the targeted locus [10].

Recently, great achievements have been made in the era of genome engineering with the development of meganulceases (MNs), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein 9 (CRISPR/Cas9). Experimental proofs gradually showed that these SSNs were not only used

for gene insertion or inactivation, but also significantly enhanced the effectiveness of homologous recombination and, thus, allowed more precise gene replacement events. In 1993, Puchta and co-workers [17] provided the first evidence of homologous recombination in plant cells by using SSNs. After the discovery of ZFNs in 1996 by Kim and colleagues [18], extensive efforts have been made for progressive advancement with this tool, which offered a significant breakthrough in plant GE. In 2003, scientists were able to inactivate genes using ZFNs for the first time [19]. In 2005, the first SSN-based mutagenesis via ZFNs was carried out successfully in plants [20]. Therefore, ZFNs have the ability to produce site-specific DSBs and have many applications in genome engineering [21]. Later TALENs were included to the toolbox of SSNs for programmed genome engineering [22]. Genome manipulation through engineered nucleases had gained much importance by the end of 2011, and *Nature Methods* crowned it as the "Method of the Year". In 2012, *Science* chose it as the "Breakthrough of the Year" due to the significant progress achieved in GE using TALENs. Recently, an emerging GE nuclease, "CRISPR/Cas9", was added to the toolbox for editing nucleases. In 2013, the first CRISPR/Cas9-based GE event was reported in eukaryotes [23]. In 2015, Ma's group [24] developed the multiplex genome editing mechanism in monocots and dicots. In 2013 and 2015, it was selected by *Science* as the "Breakthrough of the Year". Furthermore, advancements in the CRISPR/Cas system introduced a more precise technique of base editing, which was heralded again by *Science* in 2017 as the "Breakthrough of the Year". CRISPR/Cas9-based GE has tremendously revolutionized genome engineering since the initial few research articles were published in *Nature Biotechnology* [25,26]. All the above mentioned approaches have been extensively employed for GE and caused mutations via site-directed substitutions, replacement, deletions, and insertions at specific sites in the genome [10].

TALENs, ZFNs, and MNs are the first-generation editing tools for genome manipulation as, illustrated in Figure 1. However, they are time consuming and require lengthy protocols to attain target specificity. As compared to first-generation GE approaches, second-generation GE tools such as the CRISPR-Cas9 technique are easier to design, cost effective, and robust [26–28]. The CRISPR/Cas9 toolkit is very simple to design, as it involves only single-guided RNA (sgRNA) and the Cas9 protein in contrast to TALENs and ZFNs. Additionally, the procedure involved in TALENs and ZFNs are complex because they require protein engineering for their construction. Due to the presence of these constraints, applications of TALENs and ZFNs in plants have been limited [9]. Continuous innovation for efficient GE has expanded the applications of the CRISPR/Cas9 system in several fields of plant science and is quickly becoming a highly promising GE tool [6,11,23–26]. The plant GE tools and their corresponding applications are depicted in Figure 2, while successive steps involving the GE strategies are shown in Figure 3.

In the present review, we discuss fascinating GE tools for crop improvement. We briefly describe first-generation genome editing tools such as TALENs, ZFNs, and MNs and comprehensively elaborate on second-generation genome editing strategies with special focus on the applications of the CRISPR/Cas9 system in plant breeding for crop improvement. We briefly outline historical background, structural organization, and mode of action of the CRISPR/Cas9 toolbox. We describe the workflow of CRISPR/Cas9 from vector design to mutant screening. We also highlight recent breakthrough events in technology improvement in the CRISPR/Cas9 system. Furthermore, we discuss the recent role of CRISPR/Cas9 technology in crop breeding to develop the best performing cultivars with biotic and abiotic stress resilience, improving yield-related traits and production of high-quality crops. Finally, we outline the future outlook of CRISPR/Cas9 and pinpoint the current challenges with respect to the regulation of edited crops and their safe use.

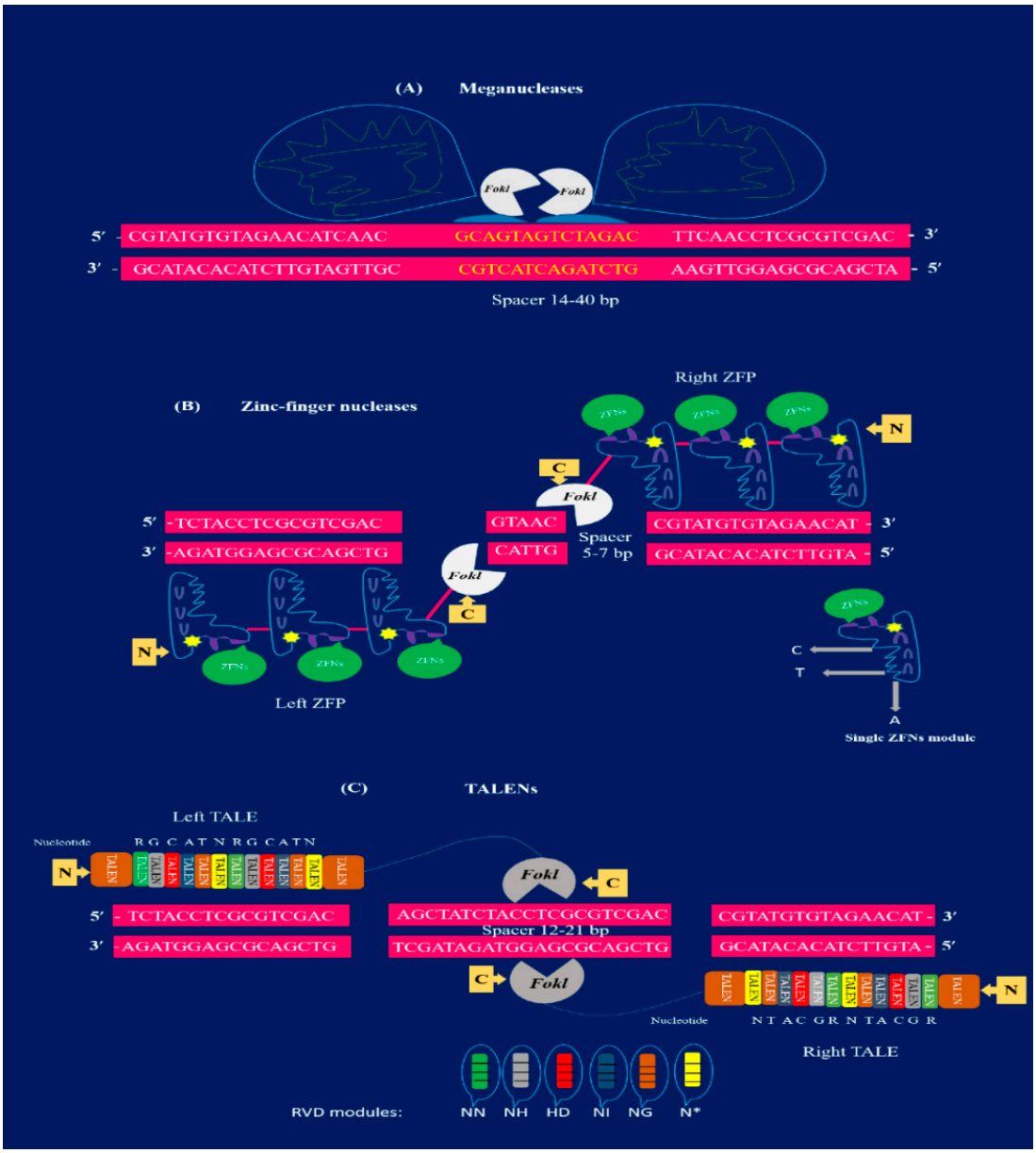


Figure 1. Structural illustration of first-generation genome editing tools: (**A**) meganucleases (MNs) have multifunctional domains with the ability to bind double-stranded target DNA and generate DSBs. The meganuclease is demonstrated to bind a target spacer sequence of 14–40 bp (yellow). The FokI nuclease cuts the target sequence (color). (**B**) Representation of ZFN bound to the target sequence of 18–36 bp long. Each monomer of ZFN (blue) is made by ZFP. There are two basic domains: the DNA binding domain at N-terminus and the catalytic domain with FokI nuclease (white) present at the C-terminus. The connection among these domains is indicated with a pink line. The ZFN modules are merged with Fok1 (white) and dimerized to cut the target sequence at a spacer 5–7 bp (pink) to produce DSBs. (**C**) Two TALEN dimers bound to the target sequence (pink) site. Each module of TALENs are composed of TALE that contain 33–35 amino acid repeats. The pair of TALENs are separated by a spacer region of 12–21 bp (pink). There are specific RVD modules (green NN, grey NH, red HD, dark blue NI, orange NG, and yellow N) that can recognize only one single nucleotide. TALE modules are dimerized to fuse with FokI (at C-terminus) to produce DSBs in the spacer region.

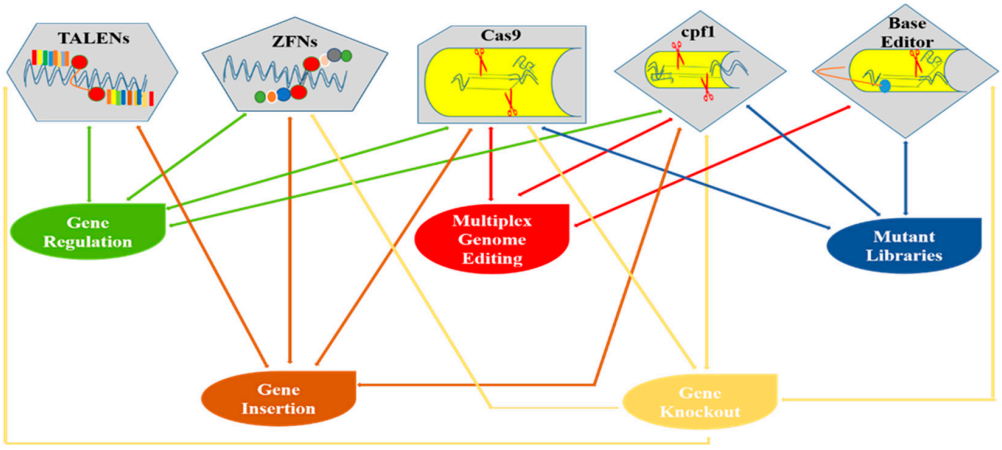


Figure 2. Diagrammatical representation of various genome editing tools and their applications in plants.

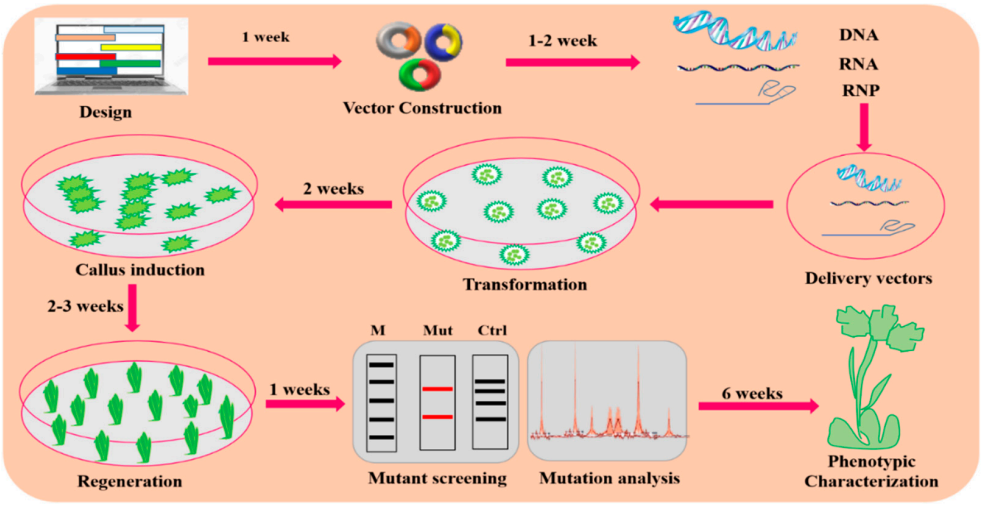


Figure 3. A general description of the GE mechanism in plants. Plant GE typically consists of the following steps: designing and construction of vectors, targeted delivery of vectors via *Agrobacterium*-mediated transformation or biolistic for transformation, callus induction and regeneration, mutation screening and analysis, and phenotypic characterization for the desired trait.

2.1. Meganulceases

In SSNs, MNs were the pioneering class of nucleases (Figure 1A), extensively applied for plant GE [29,30]. Meganulceases were also termed as homing endonucleases. Later, they were utilized for generating DSBs in several genomes [31]. Meganulceases have the ability to recognize target DNA sequences of about ~12–40 bp, that make MNs the most efficient delivery approach for all vectors including plant RNA viruses [32]. As compared to different SSNs, MNs are difficult to re-design for target sequences different than their natural ones. Non-modular properties of the specific proteins are the main reason for hindrance in re-designing MNs. So, for plants, the applications of MNs have been restricted to only naturally existing MNs such as I-SceI and I-CreI nucleases [29].

2.2. Zinc-Finger Nucleases

In plants, ZFNs are extensively being applied for plant GE [33]. Zinc-finger nucleases are one of the main techniques for genome manipulations which are very beneficial in various GE applications. Zinc-finger nucleases have been widely used for target specific mutagenesis to disrupt the gene function and produce several gene knockouts [34]. GE with ZFNs has demonstrated the production

of herbicide-resistant plants, and various kinds of targeted and specific gene insertion have also been unveiled [35]. In plant biotechnology, zinc-finger proteins (ZFPs) can be exploited in two ways: ZFNs and ZFN-TFs. Due to the flexible nature of ZFPs, it provides a striking basis for modeling ZFNs with desired sequence-specific domains to produce DSBs and facilitate GE [36,37]. In 1996, ZFNs were reported for the first time and named as chimeric restriction enzymes. According to this research, chimeric restriction enzymes were developed by associating the non-specific FokI with the DNA binding domain of two dissimilar ZFPs. The ZFNs were constructed by the fusion of chimeric proteins that were composed of DNA cleavage and a DNA binding domain. A set of 3–6 Cys2His 2 ZFs constructed the DNA binding domain, while a Fok1 restriction enzyme was generated by the DNA cleavage domain [18]. FokI is homodimeric in nature and belongs to the type IIS class of restriction enzymes isolated from *Flavobacterium okeanokoites* [18]. The domain of FokI nuclease needs to dimerize in order to cut DNA [38]. Two ZFN monomers are required having FokI dimerization and C-terminal fusion for active cleavage when binding to DNA. Target DNA with 9–18 bp has been recognized by each monomer containing 3–6 ZFs (Figure 1B). Consequently, each monomer of ZFN targets the spacer region of 5–7 bp located in the adjacent half-site and dimerize to perform the cleavage activity for targeted DNA [10]. ZFNs, as compared to MNs, are small in size (about 300 aa in one monomer and 600 aa in a pair of nucleases), enabling them responsive to many delivery procedures. In the last two decades, ZFNs have been applied for site-specific mutations in plants such as *Arabidopsis thaliana*, soybean, maize, tobacco, and petunia [36,39,40]. However, modular association of ZFs has gained partial achievement [41]. Currently, ZFNs are not recommended in several cases due to the lower target specificity, limited amount of specific target domains, and large number of non-targeted editing [42].

2.3. TALENs

Another interesting tool for GE is termed as transcription activator-like effector nucleases (TALENs). These are modern inclusions to the SSNs resources and have been extensively applied for GE in plants [43]. In 1989, TALENs were first discovered when a pathogenic bacterium called *Xanthomonas* was studied for many plant varieties [44]. *Xanthomonas* is responsible for uncontrolled growth of plant cells due to the synthesis of a novel protein termed as transcription activator-like effectors (TALEs) that target specific DNA sequences and greatly influence gene expression [45]. For targeted GE, TALENs are manipulated by changing the TALE repeated domains required for specific target identification and are successively linked to Fok1 nuclease to obtain suitable TALEN. The TALENs that recognize 12–21 bp extend, likewise to ZFNs, and require a spacer region of 14–20 bp for Fok1 dimerization with a pair of TALENs (Figure 1C) [10].

TALENs have the advantage over other SSNs such as MNs and ZFNs because of their modular domain. The domain for TALENs contains 33–35 aa in direct sequence repeats and two amino acids are called repeat variable di-residues (RVDs) in these repeats. The RVDs are responsible for recognition of specific nucleotides which includes thymine; NI, HD, cytosine; NG and adenine; and NN. A single RVD associated with every single nucleotide in combined mechanisms was identified that has the ability to design specific DNA binding motifs and remove the remodeling issues faced in the case of ZFNs and MNs [22,46–48]. Target specificity is another advantage of TALENs over other nucleases. Typically, 15–20 RVDs are used in order to design TALEN monomers with more than a 30 bp target site. As compared to ZFNs, TALENs reduce the toxicity and are more specific having large target sites [49]. A large size of about ~950 aa to ~1900 aa is the only drawback of TALENs for use as an accurate tool for GE. TALENs are usually carried to cells via direct integration of DNA or by integration of a construct-harboring TALEN-encoding unit into the genome. TALENs have been successfully applied in plants such as rice [43], *Arabidopsis* [22], tobacco [42], and *Brachypodium* [26] for GE. TALENs are more extensively exploited for targeted GE as compared to ZFNs, but they still need an efficient way to assemble tandem repeats for binding to the targeted DNA region. Furthermore, the repetitive nature and large size pose as big hurdles for the successful delivery of TALENs [45]. Some reported events of MN-, TALEN-, and ZFN-mediated mutagenesis in plants are described in Table 1.

Table 1. Applications of first-generation genome editing tools in crop plants.

Gene Editor Nucleases	Species	Modification Type	Delivery Technique	Repair Pathway	Target Gene	Desired Trait	Reference
Meganuclease	*Zea mays*	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*MS26*	Male-sterile plants	[50]
	Gossypium hirsutum	Trait stacking	Particle bombardment	HR	*EPSPS*	Herbicide tolerance	[30]
	Zea mays	Gene knockout	*Agrobacterium*-mediated transformation	HR	*LG1*	Heritable targeted mutagenesis	[51]
Zinc-finger nucleases	*Oryza sativa*	Trait stacking	*Agrobacterium*-mediated transformation	HR	*OsQQR*	Detection of safe harbor loci	[37]
	Zea mays	Trait stacking	Microparticle bombardment	HR	*ZmTLP*	Herbicide tolerance	[52]
	Brassica napus	Gene expression	*Agrobacterium*-mediated transformation	Transcriptional activation	*KasII*	Improved seed oil composition	[21]
	Glycine max	Gene knockout	*Agrobacterium rhizogenes*	HR	*DCL*	Heritable transmission	[40]
	Zea mays	Gene replacement	Whiskers	NHEJ	*IPK1*	Herbicide tolerance	[39]
TALENs	*Saccharum officinarum*	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*COMT*	Improved saccharification efficiency	[53]
	Zea mays	Frame-shift mutation	*Agrobacterium*-mediated transformation	NHEJ	*ZmMTL*	Induction of haploid plants	[54]
	Glycine max	Gene knockout	*Agrobacterium rhizogenes*	NHEJ	*GmPDS11, GmPDS18*	Albino and dwarf phenotype	[55]
	Glycine max	Gene stacking	*Agrobacterium rhizogenes*	NHEJ	*FAD2-1A, FAD2-1B, FAD3A*	High oleic, low linoleic contents	[56]
	Solanum tuberosum	Gene knockout	Particle bombardment	NHEJ	*VInv*	Minimizing reducing sugars	[57]
	Saccharum officinarum	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*COMT*	Improved cell wall composition	[58]
	Solanum tuberosum	Gene knockout	Particle bombardment	NHEJ	*ALS*	Transient expression in protoplasts	[59]
	Zea mays	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*ZmGL2*	Reduced epicuticular wax in leaves	[60]
	Oryza sativa	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*OsBADH2*	Fragrant rice	[61]
	Triticum aestivum	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*MLO*	Powdery mildew resistance	[62]
	Hordeum vulgare	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	Transgene	*GFP*	[63]
	Glycine max	Gene knockout	*Agrobacterium rhizogenes*	NHEJ	*FAD2-1A/B*	Improved oil quality	[64]
	Nicotiana tabacum	Gene knockout	Peg-mediated protoplast transformation	NHEJ	*Sur A, Sur B*	Targeted mutation	[42]
	Oryza sativa	Gene knockout	*Agrobacterium*-mediated transformation	NHEJ	*Os11N3*	Bacterial blight resistance	[43]

Description: Non-homologous end joining (NHEJ), Homology repair (HR), Male sterile 26 (*MS26*), 3-phosphoshikimate 1-carboxyvinyltransferase 2 (*EPSPS*), Liguleless 1 (*LG1*), Trait landing pads (*ZmTLP*), 3-oxoacyl-[acyl-carrier-protein] synthase II, chloroplastic (*KasII*), Protein DCL homolog, chloroplastic (*DCL*), Inositol pentakisphosphate 2-kinase (*IPK1*), Catechol-O-methyltransferase (*COMT*), MATRILINEAL (*MTL*), Phytoene desaturase (*PDS*), Fatty acid desaturase (*FAD*), Vacuolar invertase gene (*VInv*), Acetolactate synthase gene (*ALS*), Maize glossy2 (*GL2*), Betaine aldehyde dehydrogenase (*BADH2*), MILDEW-RESISTANCE LOCU (*MLO*), Green fluorescent protein (*GFP*), Acetolactate synthase genes (*Sur A, Sur B*), Rice bacterial blight susceptibility gene (*Os11N3).*

3. CRISPR/Cas9 System

Recently, a fascinating GE tool CRISPR/Cas9 was identified for targeted genome manipulations and to express desired genes in numerous organisms [6,65]. The CRISPR/Cas9 system has emerged as the most powerful tool for GE in many species including plants [26]. The latest ground-breaking technology of CRISPR/Cas9 is basically present as an adaptive immune system of type II prokaryotes and protects them against invading organisms during phage infection by spacer acquisition, biogenesis, and target degradation [9]. The toolbox of CRISPR/Cas9 was adapted from bacteria as well as Archaea and in included in the toolbox of engineered nucleases [23,66]. There are two main components of the CRISPR/Cas9 system: a single guide RNA (sgRNA) that identifies a specific DNA sequence and the Cas9 protein which produces DSBs at a targeted site [9]. Therefore, when changing the design of sgRNA, numerous desired sites can be targeted, which makes it simpler to handle than TALENs and ZFNs [10].

3.1. Discovery of CRISPR/Cas9 Wonder

The discovery of the CRISPR/Cas9 system dates back to 1987 when Ishino and his colleagues first identified CRISPR while studying the *iap* gene in the genome of *E. coli*. During the cloning of the *iap* gene, they unexpectedly cloned a specific portion of CRISPR, and at the conclusion of their experiment, revealed that the bacterial genome consisted of a successive array of repeats [67]. After this discovery, an Archaea (*Haloferax mediteranii*) was also found to contain the CRISPR sequences [68]. Mojica et al. (2000) reported a similar type of regularly spaced repeats in *Haloferax mediterranei* and *Haloferax volcanii*, having interrelated functions [69]. Only prokaryotes were considered to have such repetitive sequences, which were named as CRISPR but were not present in eukaryotes and viruses [70]. These short repeats have an average length of 32 bp but are of different sizes from 21 to 47 bp in different organisms. Every repeat has a unique sequence of nucleotides that are extremely conserved in specific species [71]. It was unveiled that the short regular repeats are transcribed into small RNAs [72].

Four *Cas* genes (*Cas1–4*) were discovered in prokaryotes having CRISPR DNA sequences during that period [70]. From then on, many CRISPR/Cas sequences and multiple Cas proteins were identified [72]. In 2005, CRISPR spacers were discovered in plasmids and phages by three independent research groups by applying computational and sequencing technologies [73–75]. The function of CRISPR/Cas was still ambiguous before Barrangou et al. (2007) successfully demonstrated for the first time that CRISPR protected *Streptococcus thermophilus* from viral attack [76]. It was revealed that the CRISPR defense mechanism prevents the horizontal gene flow in *Staphylococci* [77]. In another study, it was observed that CRISPR RNAs regulate the CRISPR interference [78]. The presence of the CRISPR/Cas system in the bacterial genome was identified to cut specific sites in plasmid DNA and bacteriophages [79]. In 2011, the CRISPR/Cas machinery of *S. thermophilus* was exploited to confer immunity in *E. coli* [80]. Some of these important events are highlighted in Figure 4.

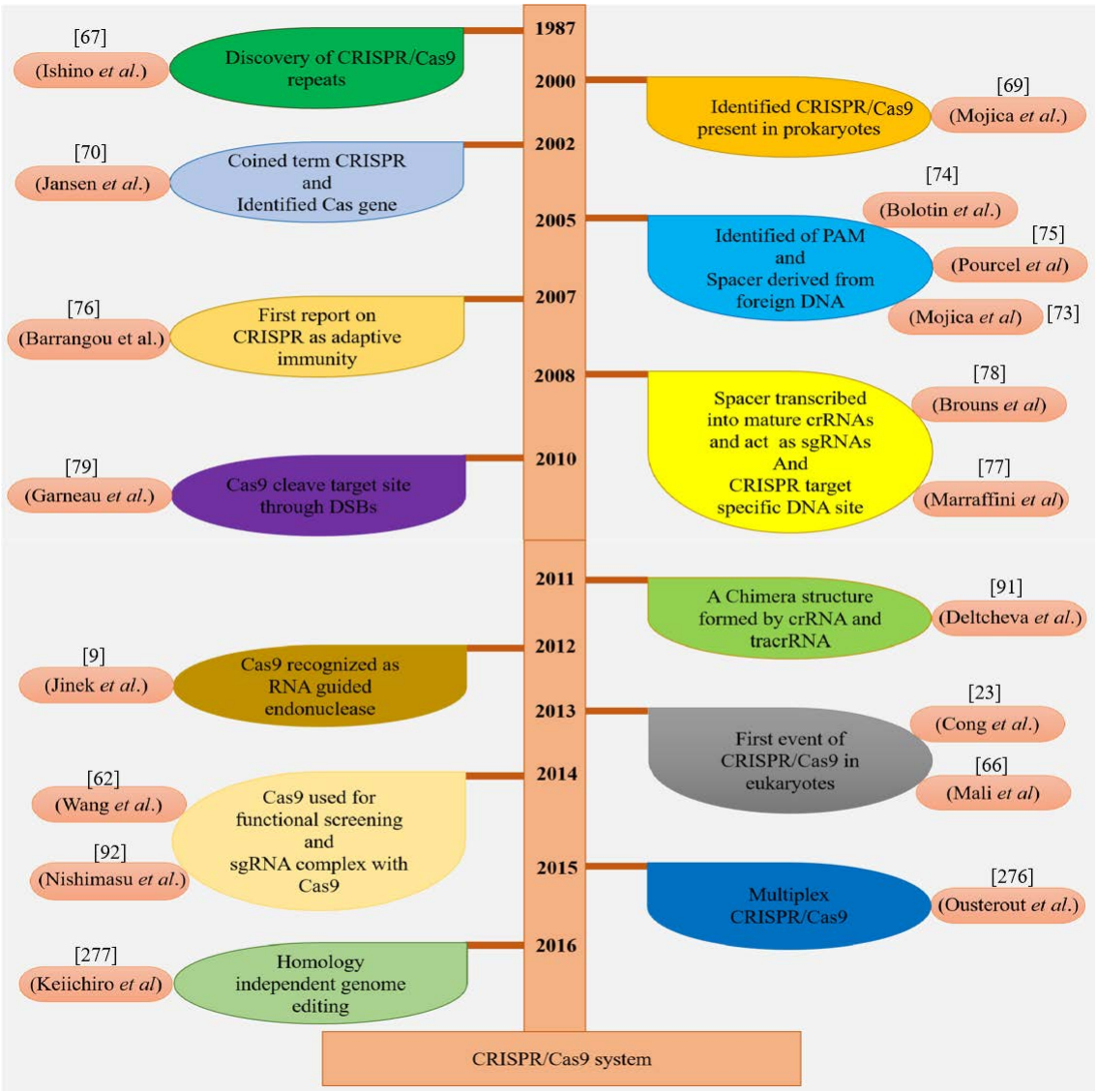


Figure 4. Historical chart illuminating key developments in the CRISPR/Cas9 system.

3.2. Architectural Organization of CRISPR/Cas9 System and Its Functions

Prokaryotic organisms such as bacteria and Archaea have a special type of defense machinery called CRISPR/Cas [81] in their adaptive immune network to protect them against the attack of viruses and phages [82].

All the natural CRSIPR/Cas networks are composed of many *Cas* genes, which are encoded by homologous palindromic repeated units, RNA-mediated endonucleases, and novel short RNAs, termed as "spacers" produced by the introduction of short mobile sequences called protospacers. The protospacers are derived from the unique spacers which move among the homologous palindromic sequences repeats when the cell is attacked by invaders. These spacers work as identifying units for the invaded cell and allow the CRISPR-Cas system to cut foreign DNA sequences. There are three steps to the CRISPR/Cas-mediated immune system, including adaptation, expression, and interference. The first step in this mechanism is adaptation, which is associated with the sequential layout of the CRISPR-array via a new spacer's procurement. Precursor CRISPR-RNA (pre-crRNA) and *Cas* genes are expressed in the expression stage. Mature cr-RNAs are produced from precursor CRISPR-RNA using RNase III and Cas proteins in the interference event-specific targeted portion memorized by the combinative properties of both Cas proteins and cr-RNA [83]. The CRISPR motif, termed as protospacer

adjacent motif (PAM), is connected with each protospacer and closely situated in the target portion of the sequence. The PAM was found to be a highly specific part of the foreign phage or virus genome but is not present on the CRISPR locus in bacterial genome [78].

The PAM sequence consisting of conserved dinucleotides is required upstream of the binding sites of crRNA for Cas proteins to recognize the target sequence [9]. The Cas proteins are unable to detect target DNA for effective cleavage during PAM site recognition. The PAM is also exceptionally crucial to distinguish between the bacteria's own DNA and invader DNA sequences. Such features enable bacteria to defend their own DNA from nucleases [84]. In several kinds of CRISPR networks, PAM sequences are essential for Cas proteins functions, such as PAM sequence 5′-NNNNGATT which is targeted by Cas proteins in *Neissseria meningiditis* [85]. Similarly, Cas9 proteins target the PAM sequence 5′-NGGNG or 5′-NNAGAA in *S. thermophiles* [79,86] and 5′-NGG in *S. pyogenes* [9].

Two independent groups of scientists discovered the CRISPR/Cas9 machinery with three major kinds (type I–II–III) and two classes in host cells [87]. In 2015, Markarova and coworkers executed the comparative genomic analysis of existing data and found two further reputed types and five subtypes [88]. During the defense mechanism of CRISPR/Cas9 against invader DNA, these two classes behave differently, such as class 1 which consists of subtypes (I, III, IV) and uses many Cas proteins, while only a large Cas protein is used by the class 2 system, which has the subtypes (II, V) [89]. In the adaptation phase of the CRISPR/Cas mechanism, spacers are added by Cas1 and Cas2 proteins and pre-crRNA develop involving Cas5 or Cas6 in the type I system. The Cas6 protein is also used in the type III system for a similar process but stimulation of 3′ end is accomplished by an uncertain element. For crRNA maturation, trans-activating crRNA (tracrRNA) and RNase are utilized in the type II mechanism [90], as shown in (Figure 5A). Currently, the immune system of *S. pyogenes* operates as a type II system, which is a well-established GE technique known as CRISPR. This CRISPR/Cas9 system is modified by two major units: a non-coding chimeric RNA and Cas9 endonucleases for double-stranded (dsDNA) breaks in DNA (Figure 5B) [9]. The Cas9 protein is directed by the guide RNA (gRNA) and Cas9 proteins recognize targeted DNA in the presence of the "seed" sequence, which is produced by spacers derived from crRNA and the *S. pyogenses* Cas9 (SpCas9) 5′ NGG ′3 sequence lying closely to the target region [9]. The crRNA and tracrRNA are complementary to each other and it directs pre-crRNA to mature crRNA by means of RNase III. After the maturation of crRNA, it guides Cas9 proteins to break specific DNA sequences [91]. In 2014, Nishimasu et al. (2014) demonstrated that the SpCas9 and gRNA DNA endonuclease has unique lobes, such as an assembly composed of a target detection lobe which is attached to the heteroduplex of sgRNA: a DNA molecule and a nuclease lobe which nicked the target DNA sequence [92], as illustrated in (Figure 5C).

3.3. Genome Editing Mechanism of CRISPR/Cas9 System

The mode of GE is established by the healing process of the genome. After the identification of the target site, Cas9 allows sgRNA to pair with the target DNA sequence. The Cas9 endonuclease is composed of the HNH and RucV-like domain, which cuts the target DNA strands three to four bases upstream of the PAM site. The HNH domain cuts the complementary DNA strands while the RuvC domain cleaves the non-complementary to gRNA. The blunt-ended DSBs can be repaired by the HDR and NHEJ repair pathways (Figure 5D). The NHEJ is error prone and causes DNA insertion or deletion at the target sequence [23]. The expression of sgRNA as pair, NHEJ mechanism came up with large deletions. The large deletions in chromosomes were attained by the NHEJ mechanism utilizing co-expressed sgRNAs. The HDR repair mechanism is only operational when a specific homologous target site is available with respect to the DSB site. In plants, through GE, many outstanding repairs were achieved via HDR, such as gene replacement, DNA correction, and targeted knock-in [93,94].

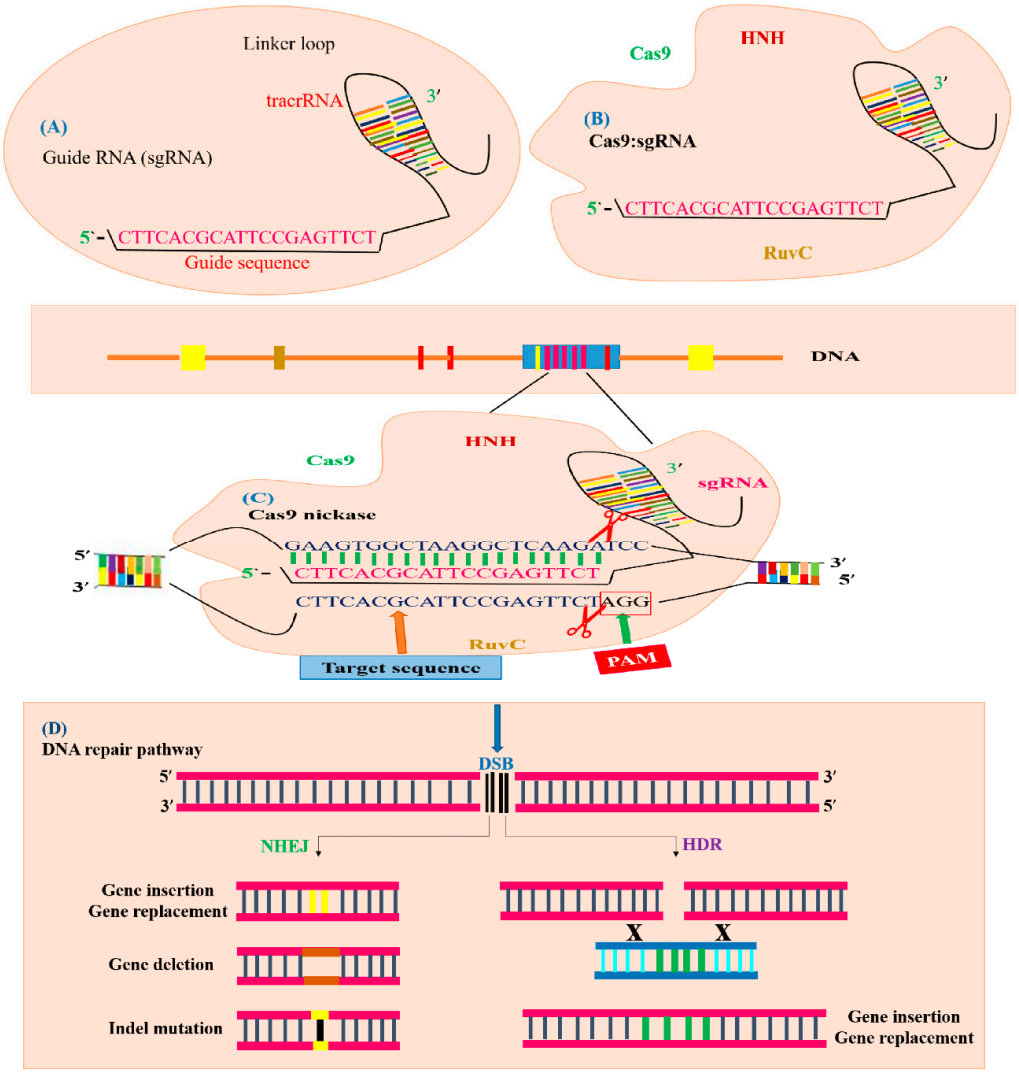


Figure 5. Illustration of CRISPR/Cas9-mediated GE. The CRISPR/Cas9 system is composed of sgRNA and Cas9. (**A**) sgRNA with a guide sequence (colored pink) is developed by the combination of protospacer with crRNA and tracrRNA. (**B**) Cas9 machinery combines with sgRNA to form a complex to trigger CRISPR/Cas9 editing. The Cas9 nuclease consists of two parts, depending on its function and structure. The recognition site identifies the target DNA and interacts with sgRNA. The nuclease site contains two domains RuvC-like and HNH which cleave the target DNA site non-complementary by the RuvC domain and complementary by the HNH domain to the gRNA. (**C**) The Cas9 nuclease detects the genomic target site (indicated with blue color) having a 20 bp target sequence that is homologous to seed or guide sequence (indicated with pink color), which is crucial for Cas9 activity and specificity. The specific PAM sequence (indicated with red color) is detected by Cas9: sgRNA complex and DSBs created by the Cas9 endonuclease three base pairs upstream of the PAM sequence. (**D**) Targeted mutagenesis of a desired gene is achieved by filling the DSB (indicated with black color) by means of the HDR or NHEJ mechanism. The NHEJ repair mechanism generally produces insertion (indicated with yellow color), deletion (indicated with brown color) or indels (indicated with black line) at the break point, generating targeted mutants. The HDR repair mechanism uses a template DNA sequence for homologous recombination to produce gene replacement or gene insertion (indicated with green color).

Biolistic and *Agrobacterium*-mediated transformation can be applied to transfer the sgRNA and Cas9 protein into desired cells [95]. GE by CRISPR/Cas9 is heavily dependent on the choice of sgRNA promoters and ubiquitous expression of the Cas9 enzyme. Universal *CaMV35S* RNA polymerase II promoters have been extensively used for Cas9 expression in plants. Similarly, for sgRNA expression,

U3 or U6 RNA pol III promoters are applied [96]. The expression level of sgRNAs is significantly greater in endogenous promoters as compared to exogenous promoters [97]. Moreover, sgRNA expression is guided by U6 promoters which were derived from monocotyledonous or dicotyledonous varieties and can only be used in monocot or dicot plants [98]. For successful integration of CRISPR/Cas9 machinery in plant nuclei, Cas9 proteins must join with nuclear localization signals [96].

To bind the target DNA by synthetically developed short gRNA sequences of approximately 20 nucleotides, the mechanism of CRISPR/Cas editing demands the PAM 5′ NGG motif for the Cas9 enzyme to cleave 3–4 bases in the target DNA sequence after the generation of the protospacer [9]. There are two domains of Cas nucleases which have the ability to cut one strand of DNA like the HNH domain and RuvC-like domain. Simple steps involving the execution of the CRISPR mechanism are recognition of the PAM sequence; sgRNA development; cloning of sgRNA; transformation into the host cell; selection of transformed individual organisms; and edited lines confirmation, as described in Figure 6.

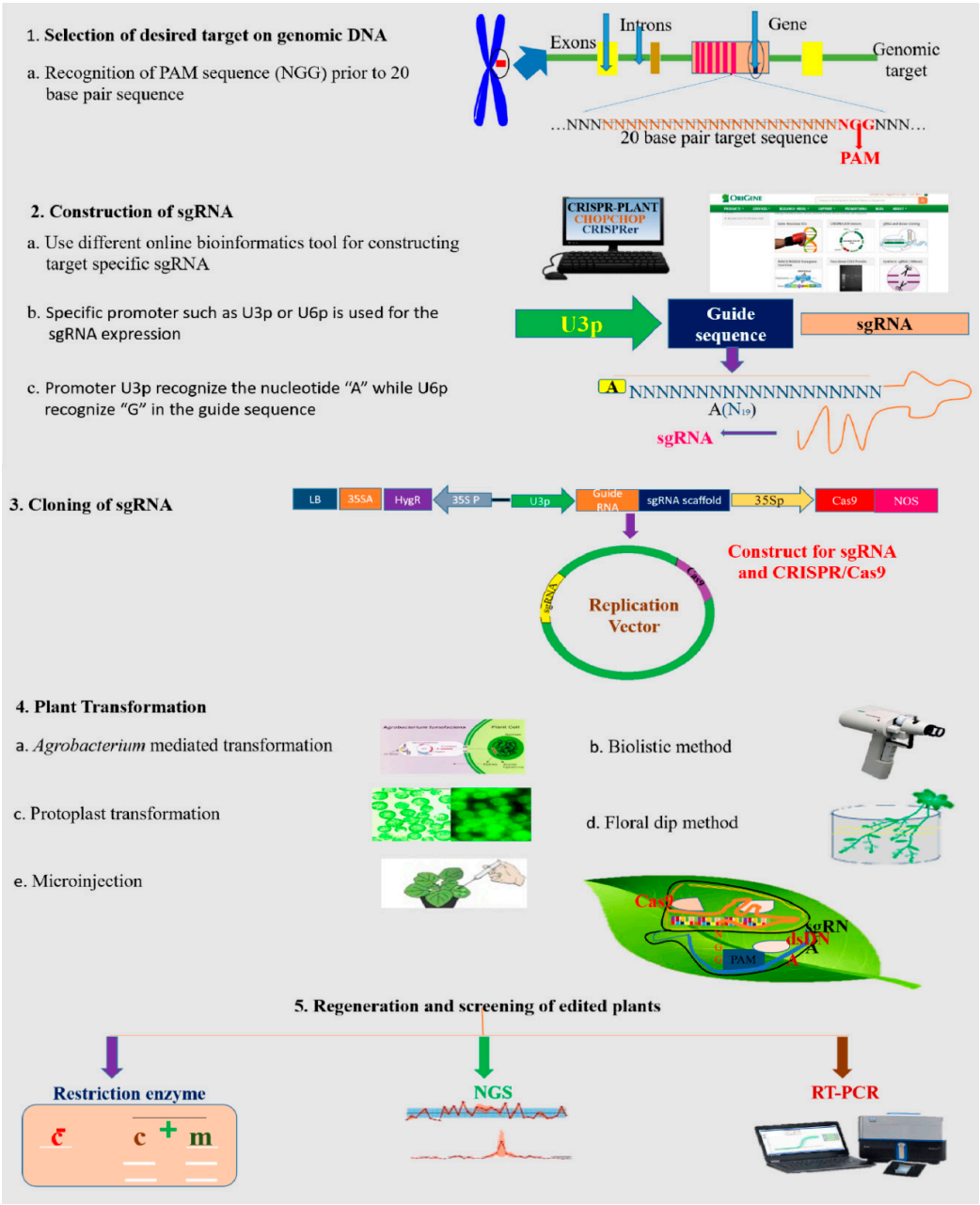


Figure 6. Basic steps in the workflow of CRISPR/Cas9-based genome editing.

4. CRISPR/Cas9-Mediated GE in Plants

To date, numerous efforts have been successfully employed for targeted gene editing in model as well as in major crop plants via the CRISPR/Cas9-based GE toolbox. Many factors have been reported that affect the editing ability of the CRISPR/Cas9 system including targeted DNA, GC contents, Cas9 codons, sgRNA structure, and expression of cas9 and sgRNA. All these factors must be highly optimized to achieve greater efficiency of CRISPR/Cas9 system [24].

4.1. Designing the CRISPR/Cas9 Delivery System

In the past, numerous attempts have been made at gene editing in plants, but the efficiency of CRIPSR/Cas9 was low [25,26]. With the improvements in technology, many highly efficient vector delivery systems for CRISPR/Cas9 have been designed for plant GE, such as supersession of viral infection, gene disruption of cis-elements, genomic deletion, gene knockout, and multiplex genome editing. Continuous progress in this editing toolkit has allowed more precise, accurate, and targeted delivery of the Cas9 system into plant cells, which include discovery of new Cas9 variants, efficient screening methods for knockout mutants, vector selection, and construction and employment of the most appropriate delivery system for the Cas9 expression cassette. In this section, we describe the construction, screening, and delivery of the CRISPR/Cas9 system into plant cells.

4.2. Cargo-Vectors for the CRISPR/Cas9 System

Two kinds of vector systems are used in CRISPR/Cas9-mediated GE, such as a single-vector system and a binary-vector system. A binary-vector system has been utilized for many years because of its ability for fast primary testing. Any specific vector having several gRNAs and a Cas9 protein expression cassette already constructed in it can be applied for plant transformation. Different structural construct of gRNAs can be utilized for numerous Cas9 proteins to design a unique gRNA: Cas9 nuclease, which allows more accuracy and easiness in experimental design. A single vector harboring both expression cassettes of gRNA and Cas9 protein is becoming more promising. Generally, in a single-vector system, RNA polymerase III-driven promoters (U6/U3) are designed for gRNA expression, whereas ubiquitin and *CaMV35S* promoters based on RNA polymerase II are exploited for *Cas9* gene expression. Continuous advancements in CRISPR/Cas9 permit researchers to develop smarter vector systems to regulate expression of gRNA and the *Cas9* gene. Recently, some new adjustments were made in the single-vector system, such as single polymerase II and dual polymerase II promoters. Single polymerase II vectors are applied to govern the expression of gRNA and the *Cas9* gene at the same time, while the dual polymerase II vectors exploit two different promoters to drive the expression of gRNA and the *Cas9* gene. The addition of all these latest technologies in the CRISPR/Cas9 delivery system assists to decrease the vector length, which eventually, improves the transformation efficiencies [99].

4.3. Bioinformatics Tools for Designing the CRISPR/Cas9 Construct

One of the most crucial steps for highly precise GE is to design an sgRNA construct for CRISPR/Cas9. To date, numerous bioinformatics tools have been developed and are available online for sgRNA designing. There are many accessible online tools with plant databases and which permit the design of sgRNA for identification of new target sites [100], as shown in Table 2. For example CRISPR Design (http://www.genome-engineering.org) was developed by Zhang and colleagues for designing sgRNA and it also assists in assessing off-target mutation [101]. In 2014, Xie and coworkers successfully developed a web tool named CRISPR-PLANT (http://www.genome.arizona.edu/CRISPR) in order to design efficient sgRNA constructs for CRISPR/Cas9-based GE [102]. For example, a novel web tool was developed by Michano and colleagues for rapid detection of target loci in soybean for CRISPR/Cas9-mediated GE [103]. Similarly, CRISPR-P (http://cbi.hzau.edu.cn/crispr/) eases the designing of sgRNA for every plant having an available sequenced genome and also helps to evaluate off-targets [104].

Table 2. List of various single guide RNA (sgRNA) designing bioinformatics tools for the CRISPR/Cas9 system.

Tool Name	Description & Function	Year	Web Link	Reference
CRISPRlnc	Design sgRNA for lncRNAs, works for all species	2019	(http://www.crisprlnc.org)	[105]
CRISPR-Local	Design sgRNA for non-reference cultivars, predict sgRNA that can target multiple genes	2018	(http://crispr.hzau.edu.cn/CRISPR-Local/)	[106]
sgRNA Scorer 2.0	Design sgRNA for several PAM sites	2017	(http://crispr.med.harvard.edu/sgRNAScorerV2)	[107]
CRISPR-P 2.0	Predict on-target scores, analyze and detect guide sequence	2017	(http://cbi.hzau.edu.cn/CRISPR2/)	[108]
CRISPRpred	Efficient designing of sgRNA based on target in silico prediction	2017	(https://github.com/khaled-buet/CRISPRpred)	[109]
CRISPR-DO	Specific for both coding and non-coding targets, provides information regarding off-targeted sites and its functional conservation	2016	(http://cistrome.org/crispr/)	[110]
phytoCRISP-Ex	UNIX-based standalone, Cas9 target prediction	2016	(http://www.phytocrispex.biologie.ens.fr/CRISP-Ex/)	[111]
CRISPy	Target prediction for sgRNA, graphical representation of results	2016	(http://crispy.secondarymetabolites.org/)	[112]
Cas-Designer	RNA-guided endonucleases, provides all information about off-targets and out-of frame scores	2015	(http://rgenome.net/cas-designer/)	[113]
CCTop	Predict target sgRNA sequence based on possible off-targets	2015	(https://crispr.cos.uni-heidelberg.de/)	[100]
Azimuth	Design sgRNA for both on-target and off-target models	2015	(https://research.microsoft.com/en-us/projects/azimuth/)	[114]
CRISPRdirect	Design sgRNA with minimal off-targets	2014	(https://crispr.dbcls.jp/)	[115]
CRISPR-PLANT	Construct specific sgRNAs for particular plant species	2014	(https://www.genome.arizona.edu/crispr/)	[102]
CRISPRseek	Screen sgRNA for targeted sequences, produce cleavage scores for predicted off-targets	2014	(https://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html)	[116]
Cas-OFFinder	Based on RNA-guided endonucleases, robust for detecting off-target sites	2014	(http://www.rgenome.net/cas-oinder/)	[117]
E-CRISP	Potential target site evaluation	2014	(https://www.e-crisp.org/E-CRISP/designcrispr.html)	[118]
SSFinder	High-throughput detection of target sites	2014	(https://code.google.com/p/ssinder/)	[119]
GPP Web Portal	Produce potential sgRNA scores	2014	(https://www.broadinstitute.org/rnai/public/analysis-tools/sgrnadesign)	[120]
CRISPR-P	Generate synthetic sgRNA, predict potential sites for enzyme cut	2014	(https://cbi.hzau.edu.cn/crispr)	[104]
CHOPCHOP	Detect optimal target sites for sgRNA, produce potential scores for target sites	2014	(https://chopchop.cbu.uib.no/)	[121]
sgRNAcas9	Rapid design of sgRNA with less off-targets	2014	(https://www.biootools.com/col.jsp?id=103/)	[122]
CRISPR Design	Precise sgRNA construction for target sites, assess off-target sites	2013	(http://www.genome-engineering.org)	[101]

4.4. Construction of the sgRNA Expression Cassette

Construction of a unique sgRNA expression cassette is the most important step in CRISPR/Cas9-mediated gene editing, and it works as a guide system for the Cas9/sgRNA complex consisting of 98 nucleotides with a 20 nucleotide target sequence [92]. In plants, RNA polymerase III is used to transcribe sgRNA and its expression is mostly governed by U3 or U6 promoters [93]. As the expression cassettes of sgRNA:U3/U6 promoters are very small in length, approximately ~300–600 bp, overlapping PCR or adaptor ligation can be applied to construct these expression cassettes [123]. In 2015, Ma and colleagues developed a robust cloning-free approach for sgRNA expression cassette development based on the PCR technique. Gibson assembly or the Golden Gate cloning strategy was used for direct cloning of sgRNA expression cassette into binary vectors for the CRISPR/Cas9 system [24]. In another approach, Gao and Zhao utilized a ribozyme mechanism to generate sgRNA by transcription of pre-RNA through RNA polymerase II, whereby inducible or constitutive promoters can be ligated to obtain the desired function of sgRNA [124].

4.5. Construction of Cas9 Expression Cassettes

Cas9 is composed of 4107 bp of coding sequence. For Cas9 nuclear localization in eukaryotes, the Cas9 coding sequence must be fused with the nuclear localization signal. Plant usage–bias codons have been used to design highly efficient and optimized Cas9 expression cassettes for improved GE in plants [125]. For example, utilization of codon-optimized Cas9p in rice and Gramineae family has been improved by enhancing GC contents [24], which imitate the genes from the Gramineae family [126]. Commonly, constitutive promoters like 35S *Cauliflower mosaic virus* (CaMV) and ubiquitin from *A. thaliana*, rice, and maize can govern the expression of Cas9 in dicots and monocots for highly targeted gene editing using callus-based transformation approaches.

4.6. Transformation Approaches for CRISPR/Cas9-Based Vector Delivery into Plants

For CRISPR/Cas9-mediated GE, cargo–vector harboring the expression cassettes of both sgRNA and the *Cas9* gene must be carried to targeted sites in plant cells. For cargo–vector transformation, floral dip and biolistic approaches are generally executed. Nowadays, advanced strategies like ribonucleo-protein complex, plasmid delivery, and virus-mediated delivery systems are applied for plant transformation. There are certain limitations in using the virus-mediated delivery system, but several studies have been carried out in plants using the virus delivery system [127,128]. A transient expression system is commonly used by researchers to transfer the vector into protoplast for analyzing the efficiency and feasibility of the CRISPR/Cas9 toolkit [129]. Biolistic and PEG-mediated transformation techniques can be used for direct delivery of *Cas9* gene expression cassettes [130]. However, it is difficult to regenerate plants from protoplast due to the heritable targeted mutations, which poses a major drawback associated with this approach in many plant species. *Agrobacterium*-mediated transformation is a highly efficient approach for stable transformation of the CRISPR/Cas9 system in dicot and monocots [131,132].

4.7. Strategies for Mutant Screening

The CRISPR/Cas9 system is a groundbreaking innovation in GE technology for developing desired mutants and numerous mutant libraries have been created by the CRISPR/Cas9 system so far, such as the genomic-scale mutant library for tomato [133] and rice [134,135]. As the applications for GE approaches are increasing day by day, scientists are required to screen huge numbers of mutants, using a strategy that includes the detection of off-target and on-target edits and which later removes the transgenes in edited plant off-springs.

To overcome the limitations associated with mutant screening, different techniques have been developed, including annealing at critical temperature polymerase chain reaction (ACT-PCR) [136], high-resolution melting analysis (HRMA) [137], polyacrylamide gel electrophoresis (PAGE)-mediated genotyping [138], T7 endonuclease I (T7EI) approach [139], and restriction enzyme site loss technique [140].

There are certain pros and cons for each technique and they are centered on genotyping differences. A mutant can be detected rapidly when it has a clear evident phenotype. For example, a visible albino phenotype was observed when a gene phytoen desaturase mutated via the CRISPR/Cas9 system. It was applied as a phenotypic marker to detect rice- and tobacco-edited plants [141,142]. Additionally, transgenic plants can also be screened using some herbicide/antibiotic selectable markers [141,143]. But making a connection among visual phenotypes and targeted genes is the only challenge associated with phenotyping-mediated screening [135]. In other approaches, high-throughput sequencing is highly efficient and precise strategy to screen all the mutants generated by the CRISPR/Cas9 system [144]. For the detection of DNA-free plants edited by CRISPR/Cas9, whole genome sequencing is quite beneficial and helpful [145].

5. Recent Breakthroughs in CRISPR/Cas9-Mediated Genome Editing in Plants

The GE tool of the CRISPR/Cas9 system has gained remarkable importance in agriculture; however, there are some drawbacks which limit its application in plant GE. The major concerns related to this technology are non-specific off-targets; HR inefficiency; PAM sequences constriction; cargo-vector inefficiency; and many others. With the advancement of plant biotechnology many unique innovative steps are regularly incorporated into the GE network to tackle these challenges.

5.1. CRISPR/Cas DNA as Cargo-Delivery Vector

The conventional cargo-vectors for GE have certain limitations which hinder the accurate editing mechanism in plants. The CRISPR/Cas DNA-based cargo-vector is now extensively used to overcome the drawbacks in plant GE.

5.1.1. Stable Expression

The abovementioned approaches of vector delivery into plant cells, such as biolistic and *Agrobacterium*-mediated transformation, are used to deliver the CRISPR/Cas DNA, and by screening the mutants, the DNA is inserted into the target location in the plant genome and carry out the editing phenomena. Currently, there are numerous applications of this system in plant GE. In 2016, Gao and colleagues incorporated a fluorescent gene in the expression cassette of CRISPR/Cas9 [146]. Recently, an interesting strategy was developed in which *BARNASE* and *CMS* were used as suicide genes to kill embryos and pollens containing the transgene of T0 plant progeny [147].

5.1.2. Transient Expression

Another delivery strategy that can be employed to obtain transgene-free edited plant is the CRISPR transient expression system. In this technique, the screening of mutants using selectable markers (antibiotic/herbicide) is omitted to regenerate the edited progeny without the incorporation of foreign genes into the plant genome. In 2016, Zhang and co-workers successfully executed this strategy for the first time in wheat to obtain transgene-free plants [148]. Furthermore, DNA base editors such as adenine and cytidine base editors have also been transferred by the transient expression approach for attaining DNA-free base substitutions [149,150]. *Agrobacterium*-mediated transformation-based transient expression has also been successfully developed in tobacco [142], and protoplast transformation-based transgenes-free expression can also be achieved in tobacco and potato protoplasts [151,152].

5.2. DNA-Free Genome Editing Through Ribonucleoproteins (RNPs)

Foreign DNA-free GE is another approach in the CRSIPR/Cas9 toolkit to produced transgene-free plants [153]. In conventional approaches to genetic engineering, the foreign DNA with editing components is incorporated into the host organism. Due to the random incorporation, the genetic modifications can be unpredictable. Even if the expression cassettes are abolished, the fragments of

foreign DNA can still be integrated into the host genome and cause mutations [154]. Furthermore, the introduction of genetically modified organisms has increased globally [155]. So, there is an increased demand to produce transgene-free plants. To develop DNA-free genome-edited plants, particle bombardment and protoplast transformation techniques have been employed.

Transgene integration may be reduced by using the CRISPR/Cas DNA transient expression system, but it does not entirely eliminate it; additionally, the discarded DNA portion may still get incorporated into different non-targeted locations within the plant genome. To escape the shortcomings of mRNA and plasmid-based expression system of sgRNA/Cas9, a most competent transgene-free editing strategy was established by designing the RNPs sgRNA/Cas9 system in plants [156–158]. Hence, sgRNA/Cas9 RNPs have a greater ability to produce DNA-free edited plants with low off-target frequency and is more efficient than a plasmid-mediated editing system. The RNP-based system does not need transcriptional and translational apparatuses for creating nicks in the target sites and, after cleaving, it is then disintegrates itself. In 2015, Woo and colleagues performed DNA-free GE for the first time in rice, tobacco, lettuce, and *Arabidopsis* using the RNPs system [156]. Potato, apple, and grape explants were subjected to targeted mutations carrying CRISPR/Cas9-mediated RNPs [159–161]. In addition, DNA-free GE in maize and wheat has been developed by particle bombardment-mediated transformation of RNPs and Cas9 proteins into cells [58,148,162]. Recently, a new CRISPR/Cas variant Cpf1 has been added to the RNP-based GE toolkit, carrying AsCpf1/crRNA and LbCpf1/crRNA RNPs into tobacco and soybean [163]. However, in several cereal crops protoplast regeneration is a bigger task, hence, the biolistic-mediated RNP editing system is the most appropriate technique for GE in plants. The delivery of RNP-mediated CRISPR/Cas9 machinery has been demonstrated by two different groups in wheat and maize [145,146]. The discovery of a DNA-free editing system will surely simplify the GE of plants and helps to commercialize the edited plants in the future.

5.3. CRISPR/Cas9 Toolbox: Ways Toward Precise Editing

5.3.1. Base Editing

As compared to DSB-governed GE, single-nucleotide modification at a specific site of the genome is called base editing, and is not based on donor DNA or an HDR mechanism and also does not require DSB generation, which provides a simple, highly accurate, and universal mechanism for editing a single base at a target site. Thus, base editing with the CRISPR/Cas9 tool is gaining interest for precise targeted gene editing in plants [164]. Currently, the use of the HDR repair mechanism with donor DNA for DSBs has been found to be less effective in contrast to NHEJ repair with a template-free system, posing a great hurdle in plants for base substitution. Genome-wide association studies (GWAS) have demonstrated that crop plants having a single nucleotide insertion/deletion are more significant for screening the elite germplasm [16]. Therefore, powerful tools are required immediately for generating accurate base editing in crop plants [165]. The CRISPR/Cas9 is an exceptional technique for precise substitution of a single base in target DNA [166]. The CRISPR/Cas9-directed base editing strategy has used the gRNA system, which is homologous to the natural CRISPR system. But in case of a cytosine base-editor (CBE) system, modified Cas9 endonucleases called nickase (nCas9) are used as compared to the natural CRISPR system. These nCas9 proteins, in addition to dead Cas9 proteins fused with an enzyme having base cleaving activity such as cytidine, are converted to uridine by the cytidine deaminase [149,167]. Recently, an effective base-editor 3 (BE3) platform was developed, which involves the merger of APOBECI known as rat cytidine deaminase and which has been extensively employed for GE in many organisms including plants [168]. In addition, several improvements in the BE3 system has allowed modifications in PAM sites to enhance its editing specificity and accuracy [168]. Likewise, three cytidine deaminase orthologs such as human APOBEC3A [150,169], human AID [170], and lamprey PmCDA1 [167] have been fused with nCas9 to attain highly precise C-to-T substitution. For example, a plant CBE system based on APOBEC3A has been widely applied for C-to-T substitution in potato, rice, and wheat [150,171]. In rice and *Arabidopsis*, CBE has been used to create point mutations.

In addition, CBE can also be applied to generate non-specific mutations that manipulate the desired gene and disrupt its function. CBE was found to be precise and more accurate than SSN-mediated editors, producing rare if any indels [172].

Similarly, adenine converts to inosine by adenine deaminase [168]. In wheat and watermelon, this strategy has been adopted to develop herbicide-resistance plants [50,173]. Yan and colleagues identified a fluorescence-tracking mechanism in rice which converts the adenine to guanine by a single-base editing system [174]. An adenine base editor (ABE) was developed for multiplex base substitutions in rice [175]. Similarly, ABE was applied to study the germline transmission and preferred phenotypic changes in *Arabidopsis* [176]. Recently, Li et al. (2018) upgraded the ABE for generating base editing in wheat and rice plants. They have also successfully developed an herbicide-resistant rice by producing the point mutation [177]. So, GE has been provided novel dimensions by base-editing tools, widening its prospective applications by manipulating desired nucleotides in the plant genome.

5.3.2. Multiplex Genome Editing

In plants, cellular processes are fine-tuned by several redundant genes. Sometimes, mutating a single gene may not confer a desired phenotype because of the compensation effect produced by other genes in same gene family. Hence, an upgraded editing system with improved efficiency is needed for multiplex gene editing in plants. In CRISPR/Cas9-mediated multiplex GE, many sgRNA cassettes can be designed by using single or multiple promoters into a single-vector system [108,178]. In 2013, Mao and coworkers designed two sgRNAs for two homologous of *magnesium-chelatase subunit I* (*CHLI*) having function in the photosynthesis mechanism, and it successfully transformed the vector in *Arabidopsis thaliana*. The result showed the albino phenotype in plants in which both genes were disrupted [179]. In another study, four subunits of katanin p80 were mutated in *A. thaliana* using multiplex genome editing. For this, three sgRNA expression cassettes were designed for simultaneous gene editing and the results demonstrated the dwarf phenotype in quadruple-mutant plants [180].

The group of Xie reported the editing of eight genes simultaneously by designing multiple sgRNA expression cassettes. An endogenous t-RNA-processing platform was used for the expression of multiple sgRNAs. All the sgRNAs were released after the nick produced by endogenous t-RNA-processing-based RNase [181]. Similarly, this t-RNA-based strategy has also been efficiently demonstrated in *Zea mays* [182]. A multiplexing system was developed by Tang and colleagues in which hammerhead self-cleaving ribozyme was applied. Additionally, the same promoter Po1II was used for the expression of multiple sgRNAs that govern Cas9 activity. Ribozyme cleavages separated sgRNAs and Cas9 after transcription and released functional sgRNAs and Cas9 [183]. Furthermore, the ability of the CRISPR-Cpf1 system was harnessed for multiplex GE in rice. A single promoter was applied to produce a construct composed of numerous repeated units of crRNA attached with a target sequence. A target repeat sequence was recognized by Cpf1 and produced cleavage, which resulted in releasing of crRNAs [184]. Hence, CRISPR/Cas9-mediated multiplex GE is a convenient approach for knocking out multiple genes at once and helping to decipher the function of a desired gene family that regulates multiple biological networks. Moreover, it is also beneficial in finding out the epistatic association among genes in numerous genetic processes.

5.4. Beyond Cas9: New Cas Variants Broadening the CRISPR Toolbox

The CRISPR/Cas9 system which originated from *Streptococcus pyogenes* has some drawbacks which hinder its editing activity like multiple incompatible off-targets due to the gRNA mismatches. Thus, several changes have been made to enhance the editing efficiency and to minimize the off-target nicking of Cas9 enzymes including SpCas9n (Cas9n) [23], Dead cas9 (dcas9) [185], and *FokI* Cas9 (fCas9) [186,187]. Various bacterial species have been used for the extraction of Cas9 proteins having novel- and stretched-PAM sequences that can help in enhancing the non-target cleavages. *Neisseria meningitides* have unique a CRISPR/Cas machinery named Nmecas9 which is specific for 8-mer (5′-NNNNGATT) PAM sequence targets that can minimize the chance of off-target cleavage and

enhance specificity [188]. Besides, of the other identified orthologs of Cas9, SpCas9 has been most commonly used for GE. SpCas9, derived from *Staphylcoccus aureus*, detects the 6-mer PAM sequence (5′-NNGRRT) [189]. The modification of SpCas9 has been carried out which targets the PAM sequence (5′ NGA) and edits the target gene efficiently [190]. Besides SpCas9, the shorter length of SaCas9 permits it to overcome the delivery challenges faced by SpCas9 in utilizing the multi-dimensional adeno-virus cargo vectors [189]. The CRISPR/SaCas9 system has been efficiently used to edit many plant genomes such as citrus, rice, tobacco, and *A. thaliana* [191]. Furthermore, *Streptococcus thermophilus*-derived St1Cas9 and St3Cas9 have also been employed for CRISPR-mediated GE [191]. Different types of tracrRNA and crRNA are used by these orthologs to identify different PAM sites [192].

CRISPR/Cpf1 System

Recently, *Francisella novicida* was studied to discover the Class II type CRISPR-Cpfl system [193], recently named Cas13 [194]. In comparison to Cas9 for cleavage and production of cohesive ends, Cpfl needs a single RNA-guided complex having 4–5 nucleotides 5′-overhangs. The CRISPR-Cpfl system has been used successfully with none or fewer off-targets in both animals and plants. In 2016, the CRISPR/Cpf1 tool was effectively applied for GE in plants [195]. Due to the exceptional properties of Cpf1, type V CRISPR/Cpf1 has been considered as another powerful technique for plant GE [196]. The CRISPR/Cpf1 machinery like the conventional CRISPR/Cas9 system is formed by the two major elements: one Cpfl nuclease for target specificity and the other one for target sequence identification called crRNA. Although, in contrast to the Cas9 network, which recognizes PAM sequences with G-rich contents (5′-NGG-3′), the Cpf1 recognizes a PAM sequence (5′-TTN-3′) having T-rich contents [194]. Furthermore, CrRNA and tracrRNA interaction is not needed in the Cpfl system, although it is necessary for the Cas9 technique. A size of about 42 to 44 crRNA is required in the CpfI system, which is smaller than that of gRNA [194]. In rice and tobacco, targeted mutagenesis has been carried out through the CRISPR/Cpf1 mechanism derived from *Francisella novicida* (FnCpf1) [195]. In rice, *Lachnospiraceae*-derived Cpf1 (LbCpf1)-mediated targeted mutations have been reported [197,198]. Similarly, LbCpf1 and FnCpf1 nucleases have great potential for precise GE for specific gene addition via HR mechanism [199].

6. Applications of CRISPR/Cas9 in Plant Breeding

Climate change and rapid increases in the world's population are two major concerns that threaten agriculture production and food security globally [200]. Several biotic stressors (bacteria, viruses, fungi, insects, nematodes, etc.) and abiotic stresses (drought, salinity, heat, cold, waterlogging, etc.) hamper crop production and compromise food security around the world. Crop breeders are striving hard to develop climate-resilient, stress-tolerant crops with better quality and increased production [201]. Thus, the CRISPR/Cas9 system has numerous applications for the functional genomic research of plant genes that play a crucial role in genetic improvement of many significant agronomic traits. Especially, the knockout of some genes can encourage superior traits including disease resistance, adaptation to various abiotic stressors, nutrient usage, and yield improvements. Thus, CRISPR/Cas9-mediated GE has great potential in plant breeding for crop improvement.

6.1. CRISPR/Cas9 System for Plant Disease Resistance

Virus, bacteria, fungi, nematodes, and insects are the major causal agents inducing biotic stressors and crop yield reduction. Moreover, the persistent upsurge in several new strains of lethal pests make the battle very challenging against these pathogens [202]. Thus, to protect agriculture from the devastating impact of biotic stressors, it is very crucial to understand the plant–pathogen interaction [203]. GE strategies have been successfully applied to explore plant–pathogen interactions and mechanisms underlying plant responses against pathogen attack.

CRISPR/Cas9-mediated GE can be employed directly to disrupt disease-causing genes, known as "S-genes" and develop disease-resistant crops. For example, targeted knockout plants for the

ethylene-responsive gene *OsERF922* were generated via the CRISPR/Cas9 tool, which showed reduced blast lesions and increased resistance against rice blast caused by *Magnaporthe oryzae* [204]. Likewise, bacterial blight-resistant plants were produced by targeted mutagenesis of the *SWEET13* gene [205]. CRISPR/Cas9-based mutagenesis was applied to the promoter region, and transcription factor (TF) of canker *CsLOB1* in *Citrus paradise* was identified. Due to the presence of such mutations, two mutant lines $D_{LOB}10$ and $D_{LOB}9$ with high mutation rates have been produced. The frame-shift mutation and disruption of the *CsLOB1* gene improved resistance against *Xanthomonas citri* [206]. Peng and colleagues reported the editing of effector binding elements (EBEs) by the CRISPR/Cas9 system in the *CsLOB1* gene promoter region to increase disease resistance in *Citrus sinensis* against *Xanthomonas citri* [207].

In wheat protoplasts, the CRISPR/Cas9 technique was applied by Shan et al. to edit the *TaMLO* gene [141] and produce wheat lines resistant to powdery mildew caused by *Blumeria graminis* f. sp. *Tritici* [62]. In another study, CRISPR/Cas9-mediated multiplex GE was performed to mutate three homologs of the *EDR1* gene to develop resistance against powdery mildew in wheat [208]. Similarly, CRISPR/Cas9-based mutants of *MLO* were produced in tomatoes which conferred resistance against powdery mildew [209].

It has been estimated that about half of plant diseases are caused by virulent viruses, which result in heavy crop losses globally [201]. Gene-targeting efficiencies were improved many folds by DNA virus amplicons. Geminiviral-based DNA replicons of wheat was utilized for transient expression of the CRISPR/Cas9 system against *wheat dwarf virus* (*WDV*), in hexaploid wheat and 12 fold upregulation was observed in ubiquitin gene expression [210]. Stable over-expression of sgRNAs and Cas9 that particularly target the genome of the Gemini-virus to prevent its growth has been applied for virus-resistant crop breeding programs [211–213]. Furthermore, the CRISPR/Cas9 system can also be used to mutate viral genomes in addition to tackling diseases caused by them [201]. The efficiency of CRISPR/Cas9-mediated viral GE can be increased by using virus promoters to govern sgRNA/Cas9 expression cassettes [211]. Recently, a new ortholog of Cas9 has been discovered in *Francisella novicida* (FnCas9) to edit RNA virus genomes. FnCas9 has successfully inhibited the replication of the tobacco mosaic virus as well as the cucumber mosaic virus and provides immunity against them [214]. Therefore, CRISPR/Cas9-mediated GE is an exceptional tool to improve genetic make-up and enables them to combat various pathogens. A list of recent studies indicating the significant success of the CRISPR/Cas9 system against various plant diseases is compiled in Table 3.

Table 3. Summary of disease-resistant crops developed via CRISPR/Cas9.

Crop	Target Gene	Pathogen	Gene Function	Trait Improvement	Editing Result	Repair Pathway	Delivery Technique	Reference
Oryza sativa	*eIF4G*	*Rice tungro spherical virus*	Translation initiation factor	Resistance against *Rice tungro spherical virus*	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[215]
Vitis vinifera	*VvWRKY52*	*Botrytis cinerea*	Transcription factor	Increased resistance against *Botrytis cinerea*	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[216]
Gossypium hirsutum	*Gh14-3-3d*	*Verticillium dahliae*	Negative regulator of disease resistance	Resistance to Cotton verticillium wilt	Knock-in	NHEJ	*Agrobacterium*-mediated transformation	[217]
Solanum lycopersicum	*SlJAZ2*	*Pseudomonas syringae*	co-receptor of coronatine	Bacterial speck resistant	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[218]
Solanum lycopersicum	CP and Rep sequences	*Tomato yellow leaf curl virus*	Negative regulator of viral resistance	Improved resistance against *Tomato yellow leaf curl virus*	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[219]
Solanum lycopersicum	*SlMlo1*	*Oidium neolycopersici*	Encoding powdery mildew resistance	Improved resistant against powdery mildew	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[209]
Triticum aestivum	*EDR1*	*Erysiphe cichoracearum*	Encoding powdery mildew resistance	Improved resistant against powdery mildew	Knock-out	NHEJ	Particle bombardment	[208]
Citrus paradise	*CsLOB1*	*Xanthomonas citri* subsp. *citri*	Increase susceptibility against citrus canker	Citrus canker resistant	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[207]
Citrus sinensis	CsLOB1	*Xanthomonas citri* subsp. *citri*	Increase susceptibility against citrus canker	Citrus canker resistant	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[206]
Oryza sativa	*OsERF922*	*Magnaporthe oryzae*	*ERF* transcription factor	Resistance against blast fungus	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[204]
Cucumis sativus	*eIF4E*	Multiple viruses	Translation initiation factor	Broad virus resistance	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[220]
Oryza sativa	*OsSWEET13*	*X. oryzae* pv. *oryzae*	Sucrose transporter gene	Resistance against bacterial blight	Knock-out	NHEJ	*Agrobacterium*-mediated transformation	[205]

Description: Translation initiation factor 4 gamma gene (*eIF4G*), WRKY transcription factor 52 (*WRKY52*), 14-3-3 protein 6-like (*Gh14-3-3d*), Jasmonate ZIM-domain protein 2 (*JAZ2*), Powdery mildew resistance protein (*Mlo1*), Enhanced disease resistance1 (*EDR1*), LATERAL ORGAN BOUNDARIES 1 (*CsLOB1*), Ethylene-responsive gene (*OsERF922*), Eukaryotic translation initiation factor 4E-1 (*elF4E*), Bidirectional sugar transporter SWEET13-like (*SWEET13*).

6.2. CRISPR/Cas9 for the Production of Climate Smart Crops

The CRISPR/Cas9 technology has been extensively applied in major crop plants such as wheat, rice, maize, cotton, soybean, tomato, and potato to cope with various abiotic stressors. Development of climate smart abiotic stress-tolerant crops via the CRISPR/Cas9 tool has modernized plant breeding programs. Major events for crop improvement via CRISPR/Cas9 are described in Table 4.

For example, in wheat protoplast, two genes related to abiotic stress, *TaDREB3* and *TaDREB2*, have been studied using the CRISPR/Cas9 technique. With a T7 endonuclease assay, the expression of mutated genes has been confirmed in approximately 70% of transfected protoplasts. The mutated plants showed increased tolerance against drought as compared to wild cultivars [221]. Three rice genes named mitogen-activated protein kinase (*OsMPK2*), phytoene desaturase (*OsPDS*), and betaine aldehyde dehydrogenase (*OsBADH2*) have been edited using the CRISPR/Cas9 technique. For transformation of CRISPR/Cas9 machinery, particle bombardment and protoplast transformation methods were used and revealed that these genes are responsible for regulating many abiotic stressors [26].

To protect plants from abiotic stressors, plant annexins play a major role. The annexin *OsAnn3* gene in rice has been studied under cold stress and its function was determined in edited knockouts developed by the CRISPR/Cas9 system [222]. Similarly, the gene *SAPK2* was mutated to study the stress tolerance mechanism in rice. The results revealed that the expression level of *SAPK2* was enhanced under drought and salinity stress conditions [223]. Drought tolerance in transgenic maize was enhanced by the overexpression of *AGROS* genes and they are of great prominence for maize breeding. To identify new allelic variants, CRISPR/Cas9 was applied to mutate the *ARGOS8* gene [224]. Curtin and colleagues carried out CRISPR/Cas9-based knockout mutagenesis of two genes *Drb2a* and *Drb2b* and found that these genes regulate salt and drought tolerance in soybean [225]. In tomato, important signaling molecules, i.e., mitogen-activated protein kinases (*MAPKs*) that respond against drought stress by protecting the membrane of cells from oxidative destruction and regulating genes transcription to tackle drought stress. The association of the *SlMAPK3* gene in controlling the drought tolerance mechanism has been reported in tomato by creating knockout mutants of the *SlMAPK3* gene under drought stress through the CRISPR/Cas9 system [226].

Many important traits like stress tolerance and crop yield are controlled by multiple genes. Many studies have been carried out to locate the quantitative trait loci (QTLs) that are controlling important traits in crop improvement programs. For the development of better performing varieties, such QTLs have been transferred into the elite lines. But this introgression is laborious for closely associated QTLs and if non-target regions are introduced into best performing varieties, it may produce many deleterious effects. Conversely, CRISPR/Cas technology can be a fascinating approach to generate and examine targeted mutagenesis. Using a CRISPR/Cas9-mediated QTL editing approach, the functions of grain number QTLs (*Gn1a*) and grain size (*GS3*) in rice varieties were examined [227]. Hence, the above studies revealed that CRISPR/Cas9-based GE has massive potential for the development of climate-resilient crops.

Table 4. Summary of CRISPR/Cas9 applications in major crops for abiotic stress tolerance.

Crop	Target Gene	Trait Study	Editing Result	Repair Mechanism	Delivery Technique	Reference
Oryza sativa	*OsNAC041*	Salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[228]
Oryza sativa	*OsOTS1*	Salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[229]
Oryza sativa	*OsRR22*	Salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[230]
Solanum lycopersicum	*SlNPR1*	Drought tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[231]
Glycine max	*Drb2a, Drb2b*	Drought and salt tolerance	Knockout	NHEJ	*Agrobacterium rhizogenes*	[225]
Oryza sativa	*OsNAC14*	Drought tolerance	Knock-in	HDR	*Agrobacterium*-mediated transformation	[232]
Oryza sativa	*SAPK1* and *SAPK2*	Salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[233]
Zea mays	*ZmHKT1*	Salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[234]
Solanum lycopersicum	*SlCBF1*	Cold tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[235]
Triticum aestivum	*TaDREB2, TaDREB3*	Drought tolerance	Knockout	NHEJ	PEG-mediated transformation	[221]
Oryza sativa	*OsAnn3*	Cold tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[222]
Oryza sativa	*SAPK2*	Drought and salinity tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[223]
Zea mays	*ARGOS8*	Drought tolerance	Knockout	HDR	Particle bombardment	[224]
Solanum lycopersicum	*SlMAPK3*	Drought tolerance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[226]
Oryza sativa	*OsMPK2, OsPDS, OsBADH2*	Multiple stress tolerance	Knockout	HDR	Particle bombardment	[26]

Description: NAC transcription factor coding gene (*OsNAC041*), Small Ubiquitin-like Modifier (*OsOTS1*), Two-component response regulator (*OsRR22*), Regulatory protein NPR1 (*SlNPR1*), dsRNA-binding protein (*Drb*), NAC domain-containing protein 2 (*OsNAC14*), ABA-activated protein kinase 1 (*SAPK1*), Sodium transporter HKT1 (*ZmHKT1*), C-repeat-binding factor-1 (*SlCBF1*), Wheat dehydration responsive element binding protein 2 (*TaDREB2*), Wheat ethylene responsive factor 3 (*TaDREB3*), Aluminum-induced protein superfamily pseudogene (*ARGOS8*), Mitogen-activated protein kinase 3 (*SlMAPK3*), Mitogen-activated protein kinase 2 (*OsMPK2*).

6.3. Crop Yield and Quality Improvements via CRISPR/Cas9

Two other important agricultural traits are crop yield and quality that need to be improved through the CRISPR/Cas9 system to ensure food security worldwide (Table 5). Crop yield is a complex, multi-genic, and quantitative trait that is influenced by several features. The CRISPR/Cas9 technology has demonstrated its worth for quick yield improvement in crops.

In many studies, the CRISPR/Cas9 technique has been used to knockout the genes that negatively regulate yield related-traits including tiller number (*OsAAP3*), panicle size (*OsDEP1*, *TaDEP1*), grain weight (*TaGW2*, *TaGASR7*), grain size (*OsGS3, OsGRF4*), and grain number (*OsGn1a*). The results demonstrated that CRISPR/Cas9 is an efficient technology for improving crop yield [236–240]. A multiplexing GE strategy has been employed to mutate three genes simultaneously including *GS3*, *GW2*, and *GW5*, and *TGW6* which headed towards trait pyramiding and enhancing grain size and weight in rice [241]. Similarly, Li and coworkers applied the CRISPR/Cas9 system to knockout three yield-related genes, *Hd2*, *Hd 4* and *Hd5*, which resulted in early heading in rice [242]. It was reported that the *OsSWEET11* gene has a crucial role in grain filling and sucrose transportation. So, the CRISPR/Cas9 system was applied to disrupt the *OsSWEET11* gene, which led to decreased sucrose concentration and reduced grain weight. This study suggested that the overexpression of this gene may be beneficial for maximizing rice yield [243]. In wheat, CRISPR/Cas9-mediated GE knockout of the *GASR7* gene increased kernel weight [148]. Recently, a new approach has been established for gene identification on a large-scale that assists in examining the complex quantitative traits, including yield, by integrating the CRISPR/Cas9 tool, whole genome sequencing, and pedigree analysis. In a study, 30 varieties of "Green Revolution miracle rice" were subjected to genome sequencing and 57 genes controlling yield-related traits were screened. Knockout mutants of those 57 genes were created using the CRISPR/Cas9 technique. Phenotyping indicated that several genes are crucial for regulating yield-related traits in rice [244].

There are also many applications of CRISPR/Cas9 technology for quality improvement in crops such as storage quality, nutritional value, fragrance, and starch contents. For example, the cooking and eating quality of rice has been improved by mutating the *Waxy* gene using CRISPR/Cas9 [245]. The nutritional value of rice has also been improved by knocking out the *SBEIIb* gene which resulted in more amylose synthesis [246]. Similarly, the starch synthase gene *GBSS* was mutated via CRISPR/Cas9 in potato. The mutated lines showed decrease levels of amylose and enhanced the concentration of the amylose/amylopectin ratio [152]. In 2018, Sanchez and colleagues, carried out CRISPR/Cas9-mediated GE of the gluten-encoding gene family *α–gliadin* to produce low-gluten wheat [247]. To improve oil composition of soybean, the CRISPR/Cpf1 system was employed to disrupt the *FAD2-1B* and *FAD2-1A*. The results revealed high-yielding soybean plants with improved levels of oleic acid [163]. Moreover, it was reported that the zein protein has been reduced by 12.5% in kernels by disrupting the *PPR* and *RPL* genes in maize. The mutated plants indicate increased production levels of healthy tryptophan and lysine in maize [182]. Sorghum nourishment quality has been improved by targeting *k1C* genes which were responsible for poor digestibility and hindered production of important amino acids [248]. Recently, some other studies for quality improvement have been carried out via the CRISPR/Cas9 system, such as *Brassica napus* with high oleic acid concentration [249], long shelf life of tomatoes [250], and increased lycopene levels in tomato [251].

In summary, the above described studies reveal that CRISPR/Cas9-mediated modern breeding techniques can be utilized to attain valuable mutations for improving crop yield and quality.

Table 5. Summary of CRISPR/Cas9 applications in major crops for yield and quality improvement.

Crop	Target Gene	Trait Improvement	Editing	Repair Mechanism	Delivery Technique	Reference
Triticum aestivum	*TaGW2*	Grain weight	Knockout	HR	Particle bombardment	[236]
Oryza sativa	*OsAAP3*	Grain yield	Knock-in	NHEJ	*Agrobacterium*-mediated transformation	[237]
Oryza sativa	*OsCCD7*	High-tillering	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[252]
Glycine max	*GmFT2a*	Delayed flowering	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[253]
Oryza sativa	*GW5*	Grain weight	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[238]
Oryza sativa	*Hd2, Hd 4, Hd5*	Early heading	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[242]
Oryza sativa	*OsSWEET11*	Grain weight	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[243]
Solanum lycopersicum	*SP5G*	Early yielding	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[254]
Oryza sativa	*OsGRF4*	Grain size	Knock-in	NHEJ	*Agrobacterium*-mediated transformation	[239]
Oryza sativa	*IPA, GS3, DEP1, Gn1a*	Improved yield	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[240]
Oryza sativa	*GS3, GW2, GW5, TGW6*	Grain weight	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[241]
Triticum aestivum	*GASR7*	Kernel weight	Knockout	HDR	Particle bombardment	[148]
Triticum aestivum	*α–gliadin*	Low gluten	Knockout	HDR	Particle bombardment	[247]
Oryza sativa	*Waxy*	Enhanced glutinosity	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[245]
Solanum lycopersicum	*lncRNA1459*	long shelf life	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[250]
Solanum lycopersicum	*SGR1, LCY-E, Blc, LCY-B1*	Increased lycopene	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[251]
Oryza sativa	*SBEIIb*	Amylose, starch resistance	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[246]
Glycine max	*FAD2-1A, FAD2-1B*	Improved oil quality	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[163]
Solanum tuberosum	*GBSS*	Increase amylopectin/amylose	Knockout	NHEJ	PEG-mediated transfection	[152]
Solanum lycopersicum	*SlGAD2, SlGAD3*	Enhance ϒ-Aminobutyric acid	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[255]
Zea mays	*PPR, RPL*	Reduced zein protein	Knockout	NHEJ	*Agrobacterium*-mediated transformation	[182]

Description: Carotenoid cleavage dioxygenase 7, chloroplastic-like (*CCD7*), Phosphatidylethanolamine-binding protein FT2a (*FT2a*), Protein IQ-DOMAIN 14 (*GW5*), Protein SELF PRUNING 5G (*SP5G*), Growth-regulating factor 4-like (*GRF4*), Glutathione synthetase (GS3), Protein STRICTOSIDINE SYNTHASE-LIKE 10 (*TGW6*), Cytokinin dehydrogenase 2-like (*Gn1a*), Keratin-associated protein 5-5 (*DEP1*), Ubiquitin-protein ligase (*GW2*), GA-induced protein (*GASR7*), Senescence-inducible chloroplast stay-green protein 1 (*SGR1*), Lycopene epsilon-cyclase (*LCY-E*), Glycoside hydrolase family 13 protein (*SBEIIb*), Granule-bound starch synthase (*GBSS*), 2-oxoglutarate-dependent dioxygenase 2 (*GAD2*), Pentatricopeptide repeat (*PPR*), Ribosomal protein lateral (*RPL*).

7. Regulatory Affairs of Genome-Edited Crops

Since the development of the first genetically modified organism (GMO) in 1995, firm rules and sanctions have been imposed to regulate GM crops worldwide. In most European countries, GM crops are still banned for commercial production and release of GM crops in the field and consumer market is prohibited. Similarly, GE non-transgenic crops may also be banned if regulatory bodies consider them as GMOs. GM crops continue to provoke extensive public misunderstanding and mistrust despite 22 years of commercialization and cultivation on 189.8 million hectares in 2017 with approximately US$18.2 billion economic gains in 2016. Globally, 82% of the total crop area for soybeans, 68% for cotton, 30% for maize, and 25% for oilseed rape were planted with GM varieties in 2014. Despite high adoption rates by farmers, the cumbersome regulatory processes and delayed cultivation approval procedures have reduced the value of innovation "the GM crops".

Quick action may be needed for strict legislation to distinguish between GE and transgenic crop. Modern breeding technologies, particularly GE, are highly feasible alternative options to GM crops and involve a reduced degree of regulatory oversight. New discoveries in GE technology and continuous progress in delivery systems that do not need to insert any specific foreign DNA in host cells for crop improvement may strongly challenge the legislative laws regulating the transgenic crops [130].

Traditional plant breeding approaches mainly depend on chromosomal modification via homologous recombination. Selection and crossbreeding have been applied for many years to screen the best performing varieties. Conventional plant breeding techniques have been used for several decades to detect novel traits and introduce them into individual plants for desired results. Traditional breeding has successfully developed many new cultivars, but these approaches need rigorous and continuous selection for many generations [256]. Genetic variability has significantly decreased due to the progressive evolution of many major crops through traditional breeding [257]. Therefore, modern plant breeding approaches have become essential to overcome the certain limitations of traditional breeding such as self-incompatibility, long generation time, heterozygosity, polyploidy, and time consuming.

Mutation breeding and engineering of transgenic plants are other crucial strategies used for crop improvement [258]. Conventional mutagenesis has helped to produced genetic variations that ultimately help in improving food quality and crop yield. For genetic analysis, natural or artificial mutagenesis has been induced using chemical and physical mutagens like ethyl methanesulfonate (EMS) and gamma rays [259]. Screening of large numbers of mutants is the biggest challenge. Such laborious, time-consuming, and untargeted breeding platforms cannot maintain pace with global food demands [258]. Foreign genes have been transferred into the elite crop lines to obtain desired traits through transgenic breeding. As compared to traditional breeding, transgenic techniques eliminate all crossing barriers and have increased genetic variability. Transgenic technology has provided enormous opportunities for crop improvement but, at the same time, provoked public concerns about its potential effect on human health and environment. Hence, commercialization of genetically modified crops (GMOs) is under strict control and limited by lengthy and expensive regulatory assessment procedures [260].

The advanced GE technology assists to produce precise and targeted mutations without the integration of any DNA sequence in plant genomes. This permits the development of non-transgenic crops with increased yield and improved quality and stress tolerance [261]. These approaches are speedy in contrast to traditional breeding techniques and allow development of transgene-free plants [156]. Plants produced via GE approaches are very similar to plants developed by conventional breeding. Additionally, GE technology takes less time to incorporate desired traits into the plant genome as compared to transgenic breeding, mutation breeding, and traditional breeding. It will take approximately 4–6 years to develop a GE plant with desired trait as compared to the transgenic breeding, which require 8–12 years for the creation of transgenic plant. On the other hand, mutation breeding and conventional breeding need 8–10 years approximately to obtain a desired phenotype [258]. The emergence of advanced GE tools not only revolutionized the world of science, but its economics

are also very spectacular as compared to genetically engineered plants. Generally, the total expenditure required to execute a single transformation event would have been approximately a quarter of a million US dollars [262], while the budget needed for GE could be $30 [263], which is astonishingly economical and reliable. Consequently, large amounts of money can be saved on developing and approving genome-edited crops, avoiding laborious and time-consuming field experiments which normally demand many years to regulate a GM crop. In addition, it will eliminate the uncertainty and fear regarding the use of GM crops [212]

There is a primary need to review the current rules and regulations regarding GMOs. In addition, genome manipulation done via GE tools are quite different from a transgenic approach. For example, mutations produced by CIRSPR/Cas9 are small indels as compared to large gene sequence insertion or deletion [264]. Such small indels are most often produced in plants under normal growth environments and can also be generated through conventional mutagens. Additionally, in contrast to GMOs which need stable integration of foreign DNA in the genome, CRISPR/Cas9-mediated GE can be used to develop DNA-free non-transgenic plants with improved traits. As far as the regulatory affairs of these gene-edited plants are concerned, there is no international regulatory framework present at the moment. Two major stakeholders, the USA and European Union (EU), have opposite policies for the regulation of genome-edited plants. The United States Department of Agriculture (USDA) has exempted GE crops from its strict rules and regulations [265], while the EU holds the position to treat genome-edited plants as GMOs. Recently, the European Court of Justice has ordered the verdict that GE plants should be subjected to similar regulatory procedures as in the case of GMOs [266]. This judicial ruling may impede investment in GE techniques and limit their use in modern plant breeding platforms in European countries. In Germany, CIBUSTM canola cases are undecided and no legitimate information has been issued by the European Commission (EC) [267]. In 2011, independent legal experts at the EU suggested some legal categorization of modern plant breeding approaches, including GE technology. A committee was formed by the EC, called the "New Techniques Working Group", to evaluate plants developed using different breeding techniques that fall under the category of GMOs legislation. By the end of 2011, the assessment was completed and the report finalized but it was never published [268]. Although, some important regulatory entities of the EU (including the French High Council for Biotechnology, the European Plant Science Organization, German Academy of Sciences, and the British Biotechnology and Biological Sciences Research Council) have already proposed that the assessment of GE plants should be based on the specific trait improvement instead of technology executed to develop them. However, a study was carried out by the German Federal Agency for Nature Conservation about GE organisms and they decided that GE organisms must be treated under the same regulations as GMOs, arguing that since GE is a strategy to manipulate the genome to produce targeted modification linked with unfamiliar risks, regardless of genome alterations that happen in nature [269].

In several countries, emerging crop editing tools like meganulcease, TALENs, ZFNs, and CRISPR/Cas9 have been applied for the past decade and they do not come under the category of GMO regulatory laws. The USA and Canada regard gene editing as equivalent to traditional breeding. The United States Department of Agriculture (USDA) granted permission to regularize and develop the CRISPR/Cas9-mediated genome-edited crops. Besides the USA, many other countries such as Brazil, Chile, and Argentina have established advanced regulatory principles for genome-edited crops. Every new discovery in biotechnology has been flawlessly approved due to the trait-based scheme in Canada. A strong and authentic regulatory policy is required to distinguish between GMOs and GE plants. Unfortunately, many countries have not established a clear regulatory policy for GE plants. The extensive use of GE strategies brings many challenges for regulatory bodies, as it requires great technical expertise and reliable evaluating procedures for the regulation of GE crops. Evidently, science-based guidelines that judge genome edited plants in a similar way as plants developed by conventional breeding programs are required to boost the applications of GE for crop improvement. For this, many countries such as the United States, Argentina, Australia, Brazil, Canada, and Chile

have issued legal interpretations of various omissions in regulatory rules and exempted GE crops from the strict regulations of GMOs. However, this exemption may be dependent on some strict requirements like absence of foreign gene (Australia), no signs of pest characteristics (USA), and type of trait modification (Canada and other countries).

The USA is the main stakeholder in the world and several regulatory authorities govern the regulation of GE crops including the Food and Drug Administration (FDA), Environmental Protection Agency (EPA), Animal and Plant Health Inspection Service (APHIS), and the and United States Department of Agriculture (USDA). The current policy of the USA regarding GE crops was developed by the USDA and depends on the "Plant Protection Act". Any GE plant that poses pest characteristics and food safety issues is closely assessed and monitored by the regulatory bodies (USDA, EPA, and FDA). The USDA does not treat GE plants under GMOs regulations (https://www.aphis.usda.gov/aphis/ourfocus/biotechnology/brs-news-and-information/pbi-details). The APHIS has proposed many verdicts regarding the risk assessment of DNA-free GE crops and suggested modifications to the rules in order to eradicate the legislation application about pest and GE crops in 2017 [270]. The United States Department of Agriculture (USDA) acknowledged gene editing as a much faster form of traditional breeding. The USDA has allowed more than ten case-by-case studies of genome editing for cultivation without regulatory permits. These included the development of a high level of amylopectin producing Wax corn by applying the CRISPR/Cas9 tool which has been mutated for the *Wx1* gene. Similarly, a CRISPR/Cas9 strategy was applied to produce browning resistance to white button mushrooms by mutating the *polyphenol oxidase* gene at the Pennsylvania State University [265]. Herbicide-resistant rape seed was produced by the RTDS mechanism. In soybean, drought-resistant genes such as *Drb2b* and *Drb2a* were knocked out using the CRIPSR/Cas9 system. *Setaria viridis* was subjected to the CRISPR/Cas9 technique for delayed flowering by disrupting the *ID1* gene and the *Camelina* genome was edited by Yield10 Bioscience for enhanced oil production. In addition, low phytate level corn has been established using Dow's ZFN, and resistant wheat against powdery mildew, soybean with a mutated *FAD3* gene, and potato with *PPO* knockout using a TALENs approach were also approved.

Similarly, there is no difference in Canada's approach towards the GE techniques from the techniques that have foreshadowed it. Canadian plants with novel traits (PNTs) regulations are activated only if the technique produces any specific trait, causing toxicity, allergenicity, and effects on any other organism. All the plant cultivars with specific traits are subjected to PNT regulations, irrespective of how they were produced, suggesting that the plant cultivar could be produced via conventional breeding, conventional mutagenesis, genetic engineering or gene editing. It is anticipated that some of GE techniques may produce novel cultivars that are PNTs, while many of them may not be treated under the PNTs regulations. Thus, in Canada, plant cultivars that are carried through the PNT regulations need open release approval from Health Canada and the Canadian Food Inspection Agency (CFIA) in order to register as approved cultivars for commercial use by industry [271].

Argentina established a functional regulatory framework for regularization of modern plant breeding products (Whelan 2015) [272]. Policy-makers and regulatory bodies have made flexible assessment protocols that depend on case-by-case evaluations. Fundamentally, the regulatory framework of Argentina determines the overall process of developing a GE plant. The plant developed without any transgene integration has been designated as non-GMO. Moreover, if any transgene strategy was applied but the final product is DNA-free, then this is also treated as non-GMO. A regulating body in Argentina, CONABIA, assesses the genome-edited material before giving approval. Under the rule No. 763/11, a simple deletion in genome is not regarded as a GM crop.

Most of the GE tools are introduced by industries, but the CRISPR/Cas technology was discovered by academic research groups. These academic institutions and different companies are contesting to establish intellectual property (IP) sets for speedy commercialization of CRISPR/Cas-based products. Since 2005, a 15 fold increase has been reported in the number of patent applications and 42 patent applications were registered in the USA in 2014. Over the last couple of years, investment in GE bio-enterprise has increased fivefold [273]. It was estimated that the market value of GE technology

was about $1.84 billion at the end of 2014, and it is predicted to grow with a 13.75% compound annual growth rate of about $3.51 billion by 2019 [274]. Additionally, the private companies that use CRISPR/Cas in the sector of agriculture, health, and industry have equally played a significant part in the current growth of the GE market. Over $600 million has been received by major companies which use CRISPR/Cas technology over the last decade [275].

To conclude, regulatory authorities need to develop comprehensive regulatory frameworks which direct the utilization of GE tools without constraining research. Furthermore, issues of IP rights and licensing policies need to be scrutinized for GE plants which can be used for commercial purposes.

8. Conclusion and Outlook

The production of safe, low-cost, and nutritive food by adopting sustainable agricultural practices will be a huge task. In this regard, the availability of modern technologies to improve cultivars will be a vital aspect. GE is a powerful tool which is expected to play a crucial role in meeting the increasing demands of crop production to fulfill the needs of an exploding population under a climate change scenario. As compared to conventional breeding methodologies, the molecular breeding strategies aided by GE tools allow scientists to precisely target and edit for desired traits. GE can be used to enhance crop productivity, nutritional value, and develop resistance against biotic as well as abiotic stressors by improving the crop genome. The advanced tools in plant GE have been extensively employed to edit crops for a specific agronomic trait and have been utilized in several breeding platforms for carrying the desired trait for the development of an elite local variety. Thus, modern plant breeding approaches will increase the performance of plant breeding, and gene-edited elite cultivars can be approved for cultivation in specified locations without strict regulatory laws. For the last two decades SSNs such as meganucleases, ZFNs, and TALENs have revolutionized plant GE. These SSNs have many applications in plant GE and can be used for gene insertion, gene deletion, and increasing the efficiency of homologous recombination which allows for more precise and accurate events of gene replacement.

Beside other GE techniques, CRISPR/Cas9 is the most powerful tool for crop improvement. In many plant systems it has been vigorously applied over the last five years for combating abiotic and biotic stressors and to improve other agronomic traits. CRISPR/Cas9 as a GE technology for site-direct mutagenesis has many excellent characteristics including great target specificity, easiness to execute, and low cost, which are unachievable through conventional mutagenic strategies. Further, CRISPR/Cas9 is superior to other first-generation SSNs because RNAs guide the Cas9 nuclease instead of proteins. Several Cas9-mediated techniques are being employed in different plant varieties, and these techniques will offer exceptional knowledge about plant biology and facilitate us to develop improved cultivars with great accuracy and speed via modern plant breeding. Recently some striking developments have been achieved in the CRISPR/Cas9 toolbox to increase the targeted mutagenesis with increased efficiency via base editing, multiplex GE [276], and generation of DNA-free plants. The CRSIPR/Cas9 is a versatile tool for plant GE, due to the fact of its sophisticated toolbox of Cas9 variants such as the CRISPR/Cpf1 system and online accessible bioinformatics tools for designing highly precise delivery systems. The CRISPR/Cas9-based precise GE produces gene replacement, gene insertion, and knockout mutations that are rapidly being used to increase yield, improve quality, and enhance tolerance in crops to boost crop domestication and hybrid breeding. Moreover, CRISPR/Cas9 technology is gaining interest day by day and will be a fundamental GE approach to developing improved plants with desired traits that will aid in accomplishing the goal of zero hunger in the world.

Although CRISPR/Cas9-mediated GE has gained remarkable achievements in crop improvement, there are certain challenges that need to be addressed to develop a more efficient system for plant GE. This includes assembling pangenomes for crop improvement, programmed identification of candidate sites for gene editing via functional genomics, designing of highly efficient delivery systems for GE, and reducing the frequency of off-target editing, deciphering novel pathways for this reduction, and optimization of the Cas9 function. The major pitfalls of CRISPR/Cas9 is the inefficient delivery

system for plant transformation because the current protocols are limited to certain tissues [277], genotypes, and crop varieties. The packaging of Cas proteins into delivery vectors poses large barriers for efficient delivery of CRISPR/Cas machinery. Recently, some novel cargo-vector systems have been introduced which show promising potential for efficient delivery systems. For example, carbon nanotubes have been utilized to transfer CRISPR/Cas9 editing constructs into plant leaves. Some other nano-products such as mesoporous silica nanoparticles and layered double hydroxides also have great potential to broaden the accessibility of delivery systems, as they have high transformation efficiencies and little toxicity and cellular damage. Developing improved delivery systems will be vital for efficient targeted and more precise GE for crop improvement. On the other hand, the frequency of off-target effects needs to be addressed more comprehensively, as there are many safety issues linked with CRISRP/Cas9-based bio-products. Luckily, off-target mutations are mostly bearable in plants and mutants, and off-target effects can be detected and eliminated through segregation over successive crosses. Selection of Cas9 requiring long PAMs and designing of sgRNA with close affinity for target sequence, may help to reduce the off-target effects in the future. Therefore, continuous efforts are required to overcome these hurdles in order to increase the experimental versatility and applied skills of the CRISPR/Cas9 toolbox in the future.

The advancement in modern breeding approaches has been greatly acknowledged as an innovation in our capacity to manipulate genomes and has subsequently challenged our understanding and assessment of current regulatory policies. As GE tools are extensively employed in plants, the safety of GE plants is the matter of debate around the globe. Development of regulatory policies for novel crop innovations should be multidimensional, transparent, and be able to distinguish between GMOs and GE events. Hence, to explore the large prospective of modern plant breeding approaches for improved yield and food security, it is necessary to illuminate the clear status of these approaches, including GE, and to fix current regulatory uncertainties. Harnessing the innovative ideas of system biology, synthetic biology, next-generation sequencing, and the latest developments in functional genomic approaches integrated with the advanced tools of CRISPR/Cas9 will permit the development of smart crops with higher yields and improved qualities. In the near future, CRISPR/Cas9 technology can be integrated with speed breeding programs to revolutionize the global agriculture and promise of food security.

Author Contributions: A.R. conceived the idea. A.R. and M.K. wrote the manuscript. F.S. helped in original draft preparation. F.A.J. and A.R. developed figures and organized tables. M.K.H. helped in the literature review. F.A.J., S.Y., F.S., G.M., and H.M.I.A. read, reviewed, and edited the manuscript. M.S.K. provided expert opinion, technical assistance, and finalized the manuscript. All authors listed have made substantial, direct, and intellectual contributions to the work and approved the manuscript.

Funding: This research received no external funding.

Acknowledgments: We are grateful to all the researchers whose contributions have been cited in this review paper, which have helped us to prepare this review paper. Furthermore, we apologize to those authors whose excellent work could not be cited due to space limitations.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
Cas9	CRISPR-associated protein 9
GE	Genome Editing
SSNs	Site-specific nucleases
DSB	Double-stranded breaks
NHEJ	Non-homologous end joining
HDR	Homology-directed recombination
MNs	Meganulceases
ZFNs	Zinc-finger nucleases
TALENs	Transcription activator-like effector nucleases

TALEs	Transcription activator like effectors
RVD	Repeat variable di-residues
sgRNA	single guide RNA
Pre-crRNA	Precursor CRISPR-RNA
PAM	Protospacer adjacent motif
RNPs	Ribonucleoproteins
GMO	Genetically modified organism
CBE	Cytosine base editor
ABE	Adenine base editor

References

1. Dhankher, O.P.; Foyer, C.H. Climate resilient crops for improving global food security and safety. *Plant Cell Environ.* **2018**, *41*, 877–884. [CrossRef] [PubMed]
2. Ma, X.; Zhu, Q.; Chen, Y.; Liu, Y.G. CRISPR/Cas9 platforms for genome editing in plants: Developments and applications. *Mol. Plant* **2016**, *9*, 961–974. [CrossRef] [PubMed]
3. Cao, D.; Li, Y.; Liu, B.; Kong, F.; Tran, L.S.P. Adaptive Mechanisms of Soybean Grown on Salt-Affected Soils. *Land Degrad. Dev.* **2018**, *29*, 1054–1064. [CrossRef]
4. Esfahani, M.N.; Inoue, K.; Chu, H.D.; Nguyen, K.H.; Van Ha, C.; Watanabe, Y.; Burritt, D.J.; Herrera-Estrella, L.; Mochida, K.; Phan Tran, L.S. Comparative transcriptome analysis of nodules of two *Mesorhizobium*-chickpea associations with differential symbiotic efficiency under phosphate deficiency. *Plant J.* **2017**, *91*, 911–926. [CrossRef] [PubMed]
5. Abdelrahman, M.; Sawada, Y.; Nakabayashi, R.; Sato, S.; Hirakawa, H.; El-Sayed, M.; Hirai, M.Y.; Saito, K.; Yamauchi, N.; Shigyo, M. Integrating transcriptome and target metabolome variability in doubled haploids of *Allium cepa* for abiotic stress protection. *Mol. Breed.* **2015**, *35*, 195. [CrossRef]
6. Feng, Z.; Zhang, B.; Ding, W.; Liu, X.; Yang, D.L.; Wei, P.; Cao, F.; Zhu, S.; Zhang, F.; Mao, Y.; et al. Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* **2013**, *23*, 1229–1232. [CrossRef]
7. Abdelrahman, M.; Jogaiah, S.; Burritt, D.J.; Tran, L.S.P. Legume genetic resources and transcriptome dynamics under abiotic stress conditions. *Plant Cell Environ.* **2018**, *41*, 1972–1983. [CrossRef]
8. Brooks, C.; Nekrasov, V.; Lippman, Z.B.; Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* **2014**, *166*, 1292–1297. [CrossRef]
9. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA —Guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816–821. [CrossRef]
10. Kim, H.; Kim, J.S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **2014**, *15*, 321–334. [CrossRef]
11. Zhang, Z.; Hua, L.; Gupta, A.; Tricoli, D.; Edwards, K.J.; Yang, B.; Li, W. Development of an *Agrobacterium*-delivered CRISPR/Cas9 system for wheat genome editing. *Plant Biotechnol. J.* **2019**, *17*, 1623–1635. [CrossRef]
12. Waltz, E. With a free pass, CRISPR-edited plants reach market in record time. *Nat. Biotechnol.* **2018**, *36*, 6–7. [CrossRef]
13. Shan, Q.; Wang, Y.; Chen, K.; Liang, Z.; Li, J.; Zhang, Y.; Zhang, K.; Liu, J.; Voytas, D.F.; Zheng, X.; et al. Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Mol. Plant* **2013**, *6*, 1365–1368. [CrossRef]
14. Puchta, H.; Dujon, B.; Hohn, B. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5055–5060. [CrossRef]
15. Puchta, H. The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. *J. Exp. Bot.* **2005**, *56*, 1–14. [CrossRef]
16. Li, J.; Meng, X.; Zong, Y.; Chen, K.; Zhang, H.; Liu, J.; Li, J.; Gao, C. Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat. Plants* **2016**, *2*, 1–6. [CrossRef]
17. Puchta, H.; Dujon, B.; Hohn, B. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.* **1993**, *21*, 5034–5040. [CrossRef]

18. Kim, Y.G.; Cha, J.; Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to *FokI* cleavage domain. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1156–1160. [CrossRef]
19. Bibikova, M.; Beumer, K.; Trautman, J.K.; Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **2003**, *300*, 764. [CrossRef]
20. Lloyd, A.; Plaisier, C.L.; Carroll, D.; Drews, G.N. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2232–2237. [CrossRef]
21. Gupta, M.; Dekelver, R.C.; Palta, A.; Clifford, C.; Gopalan, S.; Miller, J.C.; Novak, S.; Desloover, D.; Gachotte, D.; Connell, J.; et al. Transcriptional activation of *Brassica napus* β-ketoacyl-ACP synthase II with an engineered zinc finger protein transcription factor. *Plant Biotechnol. J.* **2012**, *10*, 783–791. [CrossRef]
22. Cermak, T.; Doyle, E.L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J.A.; Somia, N.V.; Bogdanove, A.J.; Voytas, D.F. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **2011**, *39*, 7879. [CrossRef]
23. Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex genome engineering using CRISPR/Cas Systems. *Science* **2013**, *339*, 819–823. [CrossRef]
24. Ma, X.; Zhang, Q.; Zhu, Q.; Liu, W.; Chen, Y.; Qiu, R.; Wang, B.; Yang, Z.; Li, H.; Lin, Y.; et al. A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol. Plant* **2015**, *8*, 1274–1284. [CrossRef]
25. Nekrasov, V.; Staskawicz, B.; Weigel, D.; Jones, J.D.G.; Kamoun, S. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **2013**, *31*, 691–693. [CrossRef]
26. Shan, Q.; Wang, Y.; Li, J.; Zhang, Y.; Chen, K.; Liang, Z.; Zhang, K.; Liu, J.; Xi, J.J.; Qiu, J.L.; et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **2013**, *31*, 686–688. [CrossRef]
27. Moscou, M.J.; Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **2009**, *326*, 1501. [CrossRef]
28. Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **2009**, *326*, 1509–1512. [CrossRef]
29. Prieto, J.; Redondo, P.; Padró, D.; Arnould, S.; Epinat, J.C.; Pâques, F.; Blanco, F.J.; Montoya, G. The Cterminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Res.* **2007**, *35*, 3262–3271. [CrossRef]
30. D'Halluin, K.; Vanderstraeten, C.; Van Hulle, J.; Rosolowska, J.; Van Den Brande, I.; Pennewaert, A.; D'Hont, K.; Bossut, M.; Jantz, D.; Ruiter, R.; et al. Targeted molecular trait stacking in cotton through targeted double-strand break induction. *Plant Biotechnol. J.* **2013**, *11*, 933–941. [CrossRef]
31. Stoddard, B.L. Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure* **2011**, *19*, 7–15. [CrossRef]
32. Marton, I.; Zuker, A.; Shklarman, E.; Zeevi, V.; Tovkach, A.; Roffe, S.; Ovadis, M.; Tzfira, T.; Vainstein, A. Nontransgenic genome modification in plant cells. *Plant Physiol.* **2010**, *154*, 1079–1087. [CrossRef]
33. Ran, Y.; Patron, N.; Kay, P.; Wong, D.; Buchanan, M.; Cao, Y.Y.; Sawbridge, T.; Davies, J.P.; Mason, J.; Webb, S.R.; et al. Zinc finger nuclease-mediated precision genome editing of an endogenous gene in hexaploid bread wheat (*Triticum aestivum*) using a DNA repair template. *Plant Biotechnol. J.* **2018**, *16*, 2088–2101. [CrossRef]
34. Bonawitz, N.D.; Ainley, W.M.; Itaya, A.; Chennareddy, S.R.; Cicak, T.; Effinger, K.; Pareddy, D.R. Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. *Plant Biotechnol. J.* **2018**, *17*, 750–761. [CrossRef]
35. Cai, C.Q.; Doyon, Y.; Ainley, W.M.; Miller, J.C.; DeKelver, R.C.; Moehle, E.A.; Rock, J.M.; Lee, Y.L.; Garrison, R.; Schulenberg, L.; et al. Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol. Biol.* **2009**, *69*, 699–709. [CrossRef]
36. Zhang, F.; Maeder, M.L.; Unger-Wallace, E.; Hoshaw, J.P.; Reyon, D.; Christian, M.; Li, X.; Pierick, C.J.; Dobbs, D.; Peterson, T.; et al. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12028–12033. [CrossRef]
37. Cantos, C.; Francisco, P.; Trijatmiko, K.R.; Slamet-Loedin, I.; Chadha-Mohanty, P.K. Identification of "safe harbor" loci in indica rice genome by harnessing the property of zinc-finger nucleases to induce DNA damage and repair. *Front. Plant Sci.* **2014**, *5*, 302. [CrossRef]
38. Vanamee, É.S.; Santagata, S.; Aggarwal, A.K. *FokI* requires two specific DNA sites for cleavage. *J. Mol. Biol.* **2001**, *309*, 69–78. [CrossRef]

39. Shukla, V.K.; Doyon, Y.; Miller, J.C.; Dekelver, R.C.; Moehle, E.A.; Worden, S.E.; Mitchell, J.C.; Arnold, N.L.; Gopalan, S.; Meng, X.; et al. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **2009**, *459*, 437–441. [CrossRef]
40. Curtin, S.J.; Zhang, F.; Sander, J.D.; Haun, W.J.; Starker, C.; Baltes, N.J.; Reyon, D.; Dahlborg, E.J.; Goodwin, M.J.; Coffman, A.P.; et al. Targeted mutagenesis of duplicated genes in soybean with Zinc-Finger Nucleases. *Plant Physiol.* **2011**, *156*, 466–473. [CrossRef]
41. Ramirez, C.L.; Foley, J.E.; Wright, D.A.; Müller-Lerch, F.; Rahman, S.H.; Cornu, T.I.; Winfrey, R.J.; Sander, J.D.; Fu, F.; Townsend, J.A.; et al. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* **2008**, *5*, 575. [CrossRef]
42. Zhang, Y.; Zhang, F.; Li, X.; Baller, J.A.; Qi, Y.; Starker, C.G.; Bogdanove, A.J.; Voytas, D.F. Transcription Activator-Like Effector Nucleases enable efficient plant genome engineering. *Plant Physiol.* **2013**, *161*, 20–27. [CrossRef]
43. Li, T.; Liu, B.; Spalding, M.H.; Weeks, D.P.; Yang, B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* **2012**, *30*, 390–392. [CrossRef]
44. Bonas, U.; Stall, R.E.; Staskawicz, B. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris pv. vesicatoria*. *MGG Mol. Gen. Genet.* **1989**, *218*, 127–136. [CrossRef]
45. Pennisi, E. The tale of the TALEs. *Science* **2012**, *338*, 1408–1411. [CrossRef]
46. Reyon, D.; Tsai, S.Q.; Khgayter, C.; Foden, J.A.; Sander, J.D.; Joung, J.K. FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **2012**, *30*, 460–465. [CrossRef]
47. Briggs, A.W.; Rios, X.; Chari, R.; Yang, L.; Zhang, F.; Mali, P.; Church, G.M. Iterative capped assembly: Rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.* **2012**, *40*, e117. [CrossRef]
48. Schmid-Burgk, J.L.; Schmidt, T.; Kaiser, V.; Höning, K.; Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.* **2013**, *31*, 76–81. [CrossRef]
49. Mussolino, C.; Morbitzer, R.; Lütge, F.; Dannemann, N.; Lahaye, T.; Cathomen, T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* **2011**, *39*, 9283–9293. [CrossRef]
50. Djukanovic, V.; Smith, J.; Lowe, K.; Yang, M.; Gao, H.; Jones, S.; Nicholson, M.G.; West, A.; Lape, J.; Bidney, D.; et al. Male-sterile maize plants produced by targeted mutagenesis of the cytochrome P450-like gene (*MS26*) using a re-designed I-CreI homing endonuclease. *Plant J.* **2013**, *76*, 888–899. [CrossRef]
51. Gao, H.; Smith, J.; Yang, M.; Jones, S.; Djukanovic, V.; Nicholson, M.G.; West, A.; Bidney, D.; Falco, S.C.; Jantz, D.; et al. Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J.* **2010**, *61*, 176–187. [CrossRef]
52. Ainley, W.M.; Sastry-Dent, L.; Welter, M.E.; Murray, M.G.; Zeitler, B.; Amora, R.; Corbin, D.R.; Miles, R.R.; Arnold, N.L.; Strange, T.L.; et al. Trait stacking via targeted genome editing. *Plant Biotechnol. J.* **2013**, *11*, 1126–1134. [CrossRef]
53. Kannan, B.; Jung, J.H.; Moxley, G.W.; Lee, S.; Altpeter, F. TALEN-mediated targeted mutagenesis of more than 100 *COMT* copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnol. J.* **2018**, *16*, 856–866. [CrossRef]
54. Kelliher, T.; Starr, D.; Richbourg, L.; Chintamanani, S.; Delzer, B.; Nuccio, M.L.; Green, J.; Chen, Z.; McCuiston, J.; Wang, W.; et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **2017**, *542*, 105–109. [CrossRef]
55. Du, H.; Zeng, X.; Zhao, M.; Cui, X.; Wang, Q.; Yang, H.; Cheng, H.; Yu, D. Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* **2016**, *217*, 90–97. [CrossRef]
56. Demorest, Z.L.; Coffman, A.; Baltes, N.J.; Stoddard, T.J.; Clasen, B.M.; Luo, S.; Retterath, A.; Yabandith, A.; Gamo, M.E.; Bissen, J.; et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol.* **2016**, *16*, 225. [CrossRef]
57. Clasen, B.M.; Stoddard, T.J.; Luo, S.; Demorest, Z.L.; Li, J.; Cedrone, F.; Tibebu, R.; Davison, S.; Ray, E.E.; Daulhac, A.; et al. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol. J.* **2016**, *14*, 169–176. [CrossRef]

58. Jung, J.H.; Altpeter, F. TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. *Plant Mol. Biol.* **2016**, *92*, 131–142. [CrossRef]
59. Nicolia, A.; Proux-Wéra, E.; Åhman, I.; Onkokesung, N.; Andersson, M.; Andreasson, E.; Zhu, L.H. Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. *J. Biotechnol.* **2015**, *204*, 17–24. [CrossRef]
60. Char, S.N.; Unger-Wallace, E.; Frame, B.; Briggs, S.A.; Main, M.; Spalding, M.H.; Vollbrecht, E.; Wang, K.; Yang, B. Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnol. J.* **2015**, *13*, 1002–1010. [CrossRef]
61. Shan, Q.; Zhang, Y.; Chen, K.; Zhang, K.; Gao, C. Creation of fragrant rice by targeted knockout of the *OsBADH2* gene using TALEN technology. *Plant Biotechnol. J.* **2015**, *13*, 791–800. [CrossRef]
62. Wang, Y.; Cheng, X.; Shan, Q.; Zhang, Y.; Liu, J.; Gao, C.; Qiu, J. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **2014**, *32*, 947–952. [CrossRef]
63. Gurushidze, M.; Hensel, G.; Hiekel, S.; Schedel, S.; Valkov, V.; Kumlehn, J. True-breeding targeted gene knock-out in barley using designer TALE-nuclease in haploid cells. *PLoS ONE* **2014**, *9*, e92046. [CrossRef]
64. Haun, W.; Coffman, A.; Clasen, B.M.; Demorest, Z.L.; Lowy, A.; Ray, E.; Retterath, A.; Stoddard, T.; Juillerat, A.; Cedrone, F.; et al. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* **2014**, *12*, 934–940. [CrossRef]
65. Alagoz, Y.; Gurkok, T.; Zhang, B.; Unver, T. Manipulating the biosynthesis of bioactive compound alkaloids for next-generation metabolic engineering in opium poppy using CRISPR-Cas 9 genome editing technology. *Sci. Rep.* **2016**, *6*, 1–9. [CrossRef]
66. Mali, P.; Yang, L.; Esvelt, K.M.; Aach, J.; Guell, M.; DiCarlo, J.E.; Norville, J.E.; Church, G.M. RNA-guided human genome engineering via Cas9. *Science* **2013**, *339*, 823–826. [CrossRef]
67. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakatura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **1987**, *169*, 5429–5433. [CrossRef]
68. Mojica, F.J.M.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F. Long stretches of short tandem repeats are present in the largest replicons of the *Archaea Haloferax* mediterranei and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **1995**, *17*, 85–93. [CrossRef]
69. Mojica, F.J.M.; Díez-Villaseñor, C.; Soria, E.; Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **2000**, *36*, 244–246. [CrossRef]
70. Jansen, R.; Van Embden, J.D.A.; Gaastra, W.; Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **2002**, *43*, 1565–1575. [CrossRef]
71. Karginov, F.V.; Hannon, G.J. The CRISPR system: Small RNA-guided defense in bacteria and archaea. *Mol. Cell* **2010**, *37*, 7–19. [CrossRef]
72. Haft, D.H.; Selengut, J.; Mongodin, E.F.; Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **2005**, *1*, e60. [CrossRef]
73. Mojica, F.J.M.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **2005**, *60*, 174–182. [CrossRef]
74. Bolotin, A.; Quinquis, B.; Sorokin, A.; Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **2005**, *151*, 2551–2561. [CrossRef]
75. Pourcel, C.; Salvignol, G.; Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **2005**, *151*, 653–663. [CrossRef]
76. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D.A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–1712. [CrossRef]
77. Marraffini, L.A.; Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **2008**, *322*, 1843–1845. [CrossRef]
78. Brouns, S.J.J.; Jore, M.M.; Lundgren, M.; Westra, E.R.; Slijkhuis, R.J.H.; Snijders, A.P.L.; Dickman, M.J.; Makarova, K.S.; Koonin, E.V.; Oost, J. Antiviral defense in prokaryotes. *Science* **2008**, *321*, 960–964. [CrossRef]

79. Garneau, J.E.; Dupuis, M.È.; Villion, M.; Romero, D.A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadán, A.H.; Moineau, S. He CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNAT. *Nature* **2010**, *468*, 67–71. [CrossRef]
80. Sapranauskas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **2011**, *39*, 9275–9282. [CrossRef]
81. Pawluk, A.; Davidson, A.R.; Maxwell, K.L. Anti-CRISPR: Discovery, mechanism and function. *Nat. Rev. Microbiol.* **2018**, *16*, 12–17. [CrossRef]
82. Bhaya, D.; Davison, M.; Barrangou, R. CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **2011**, *45*, 273–297. [CrossRef]
83. Bortesi, L.; Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **2015**, *33*, 41–52. [CrossRef]
84. Mojica, F.J.M.; Díez-Villaseñor, C.; García-Martínez, J.; Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **2009**, *155*, 733–740. [CrossRef]
85. Zhang, Y.; Heidrich, N.; Ampattu, B.J.; Gunderson, C.W.; Seifert, H.S.; Schoen, C.; Vogel, J.; Sontheimer, E.J. Processing-Independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell* **2013**, *50*, 488–503. [CrossRef]
86. Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2579–E2586. [CrossRef]
87. Makarova, K.S.; Aravind, L.; Wolf, Y.I.; Koonin, E.V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **2011**, *6*, 38. [CrossRef]
88. Makarova, K.S.; Wolf, Y.I.; Alkhnbashi, O.S.; Costa, F.; Shah, S.A.; Saunders, S.J.; Barrangou, R.; Brouns, S.J.J.; Charpentier, E.; Haft, D.H.; et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2015**, *13*, 722–736. [CrossRef]
89. Makarova, K.S.; Haft, D.H.; Barrangou, R.; Brouns, S.J.J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F.J.M.; Wolf, Y.I.; Yakunin, A.F.; et al. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2011**, *9*, 467–477. [CrossRef]
90. Rath, D.; Amlinger, L.; Rath, A.; Lundgren, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* **2015**, *117*, 119–128. [CrossRef]
91. Deltcheva, E.; Chylinski, K.; Sharma, C.M.; Gonzales, K.; Chao, Y.; Pirzada, Z.A.; Eckert, M.R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **2011**, *471*, 602–607. [CrossRef]
92. Nishimasu, H.; Ran, F.A.; Hsu, P.D.; Konermann, S.; Shehata, S.I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014**, *156*, 935–949. [CrossRef]
93. Li, J.F.; Norville, J.E.; Aach, J.; McCormack, M.; Zhang, D.; Bush, J.; Church, G.M.; Sheen, J. Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **2013**, *31*, 688–691. [CrossRef]
94. Schiml, S.; Fauser, F.; Puchta, H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant J.* **2014**, *80*, 1139–1150. [CrossRef]
95. Miao, J.; Guo, D.; Zhang, J.; Huang, Q.; Qin, G.; Zhang, X.; Wan, J.; Gu, H.; Qu, L.J. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* **2013**, *23*, 1233–1236. [CrossRef]
96. Belhaj, K.; Chaparro-Garcia, A.; Kamoun, S.; Nekrasov, V. Plant genome editing made easy: Targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* **2013**, *9*, 39. [CrossRef]
97. Sun, X.; Hu, Z.; Chen, R.; Jiang, Q.; Song, G.; Zhang, H.; Xi, Y. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* **2015**, *5*, 10342. [CrossRef]
98. Mao, Y.; Botella, J.R.; Zhu, J.K. Heritability of targeted gene modifications induced by plant-optimized CRISPR systems. *Cell. Mol. Life Sci.* **2017**, *74*, 1075–1093. [CrossRef]
99. Lowder, L.G.; Zhang, D.; Baltes, N.J.; Paul, J.W.; Tang, X.; Zheng, X.; Voytas, D.F.; Hsieh, T.F.; Zhang, Y.; Qi, Y. A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiol.* **2015**, *169*, 971–985. [CrossRef]

100. Stemmer, M.; Thumberger, T.; Del Sol Keyer, M.; Wittbrodt, J.; Mateo, J.L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* **2015**, *10*, e0124633. [CrossRef]
101. Hsu, P.D.; Scott, D.A.; Weinstein, J.A.; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **2013**, *31*, 827–832. [CrossRef]
102. Xie, K.; Zhang, J.; Yang, Y. Genome-wide prediction of highly specific guide RNA spacers for CRISPR–Cas9-Mediated genome editing in model plants and major crops. *Mol. Plant* **2014**, *7*, 923–926. [CrossRef]
103. Michno, J.M.; Wang, X.; Liu, J.; Curtin, S.J.; Kono, T.J.; Stupar, R.M. CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. *GM Crop. Food* **2015**, *6*, 243–252. [CrossRef]
104. Lei, Y.; Lu, L.; Liu, H.Y.; Li, S.; Xing, F.; Chen, L.L. CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in Plants. *Mol. Plant* **2014**, *7*, 1494–1496. [CrossRef]
105. Chen, W.; Zhang, G.; Li, J.; Zhang, X.; Huang, S.; Xiang, S.; Hu, X.; Liu, C. CRISPRlnc: A manually curated database of validated sgRNAs for lncRNAs. *Nucleic Acids Res.* **2019**, *47*, D63–D68. [CrossRef]
106. Sun, J.; Liu, H.; Liu, J.; Cheng, S.; Peng, Y.; Zhang, Q.; Yan, J.; Liu, H.J.; Chen, L.L. CRISPR-Local: A local single-guide RNA (sgRNA) design tool for non-reference plant genomes. *Bioinformatics* **2018**. [CrossRef]
107. Chari, R.; Yeo, N.C.; Chavez, A.; Church, G. Sgrna Scorer 2.0—A species independent model to predict CRISPR/Cas9 activity. *ACS Synth. Biol.* **2017**, *6*, 902–904. [CrossRef]
108. Liu, H.; Ding, Y.; Zhou, Y.; Jin, W.; Xie, K.; Chen, L.L. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Mol. Plant* **2017**, *10*, 530–532. [CrossRef]
109. Rahman, M.K.; Rahman, M.S. CRISPRpred: A flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. *PLoS ONE* **2017**, *12*, e0181943. [CrossRef]
110. Ma, J.; Köster, J.; Qin, Q.; Hu, S.; Li, W.; Chen, C.; Cao, Q.; Wang, J.; Mei, S.; Liu, Q.; et al. CRISPR-DO for genome-wide CRISPR design and optimization. *Bioinformatics* **2016**, *32*, 3336–3338. [CrossRef]
111. Rastogi, A.; Murik, O.; Bowler, C.; Tirichine, L. PhytoCRISP-Ex: A web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinform.* **2016**, *17*, 261. [CrossRef]
112. Blin, K.; Pedersen, L.E.; Weber, T.; Lee, S.Y. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **2016**, *1*, 118–121. [CrossRef]
113. Park, J.; Bae, S.; Kim, J.S. Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* **2015**, *31*, 4014–4016. [CrossRef]
114. Fusi, N.; Smith, I.; Doench, J.; Listgarten, J. In Silico Predictive Modeling of CRISPR/Cas9 guide efficiency. *BioRxiv* **2015**, 021568. [CrossRef]
115. Naito, Y.; Hino, K.; Bono, H.; Ui-Tei, K. CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **2015**, *31*, 1120–1123. [CrossRef]
116. Zhu, L.J.; Holmes, B.R.; Aronin, N.; Brodsky, M.H. CRISPRseek: A Bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS ONE* **2014**, *9*, e108424. [CrossRef]
117. Bae, S.; Park, J.; Kim, J.S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **2014**, *30*, 1473–1475. [CrossRef]
118. Heigwer, F.; Kerr, G.; Boutros, M. E-CRISP: Fast CRISPR target site identification. *Nat. Methods* **2014**, *11*, 122–123. [CrossRef]
119. Upadhyay, S.K.; Sharma, S. SSFinder: High Throughput CRISPR-Cas Target Sites Prediction Tool. *Biomed Res. Int.* **2014**, *2014*, 4. [CrossRef]
120. Doench, J.G.; Hartenian, E.; Graham, D.B.; Tothova, Z.; Hegde, M.; Smith, I.; Sullender, M.; Ebert, B.L.; Xavier, R.J.; Root, D.E. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **2014**, *32*, 1262–1267. [CrossRef]
121. Montague, T.G.; Cruz, J.M.; Gagnon, J.A.; Church, G.M.; Valen, E. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **2014**, *42*, 401–407. [CrossRef]
122. Xie, S.; Shen, B.; Zhang, C.; Huang, X.; Zhang, Y. SgRNAcas9: A software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS ONE* **2014**, *9*, e0100448. [CrossRef]
123. Xie, K.; Yang, Y. RNA-Guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **2013**, *6*, 1975–1983. [CrossRef]

124. Gao, Y.; Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **2014**, *56*, 343–349. [CrossRef]
125. Fauser, F.; Schiml, S.; Puchta, H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* **2014**, *79*, 348–359. [CrossRef]
126. Wong, G.K.S.; Wang, J.; Tao, L.; Tan, J.; Zhang, J.; Passey, D.A.; Yu, J. Compositional gradients in *Gramineae* genes. *Genome Res.* **2002**, *12*, 851–856. [CrossRef]
127. Baltes, N.J.; Gil-Humanes, J.; Cermak, T.; Atkins, P.A.; Voytas, D.F. DNA Replicons for plant genome engineering. *Plant Cell* **2014**, *26*, 151–163. [CrossRef]
128. Ali, Z.; Eid, A.; Ali, S.; Mahfouz, M.M. Pea early-browning virus-mediated genome editing via the CRISPR/Cas9 system in *Nicotiana benthamiana* and *Arabidopsis*. *Virus Res.* **2018**, *244*, 333–337. [CrossRef]
129. Jiang, W.; Zhou, H.; Bi, H.; Fromm, M.; Yang, B.; Weeks, D.P. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res.* **2013**, *41*, e188. [CrossRef]
130. Wolter, F.; Puchta, H. Knocking out consumer concerns and regulator's rules: Efficient use of CRISPR/Cas ribonucleoprotein complexes for genome editing in cereals. *Genome Biol.* **2017**, *18*, 682–698. [CrossRef]
131. Char, S.N.; Neelakandan, A.K.; Nahampun, H.; Frame, B.; Main, M.; Spalding, M.H.; Becraft, P.W.; Meyers, B.C.; Walbot, V.; Wang, K.; et al. An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* **2017**, *15*, 257–268. [CrossRef]
132. Zhang, S.; Zhang, R.; Song, G.; Gao, J.; Li, W.; Han, X.; Chen, M.; Li, Y.; Li, G. Targeted mutagenesis using the Agrobacterium tumefaciens-mediated CRISPR-Cas9 system in common wheat. *BMC Plant Biol.* **2018**, *18*, 302. [CrossRef]
133. Jacobs, T.B.; Zhang, N.; Patel, D.; Martin, G.B. Generation of a collection of mutant tomato lines using pooled CRISPR libraries. *Plant Physiol.* **2017**, *174*, 2023–2037. [CrossRef]
134. Meng, X.; Yu, H.; Zhang, Y.; Zhuang, F.; Song, X.; Gao, S.; Gao, C.; Li, J. Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant* **2017**, *10*, 1238–1241. [CrossRef]
135. Lu, Y.; Ye, X.; Guo, R.; Huang, J.; Wang, W.; Tang, J.; Tan, L.; Zhu, J.; Chu, C.; Qian, Y. genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. *Mol. Plant* **2017**, *10*, 1242–1245. [CrossRef]
136. Hua, Y.; Wang, C.; Huang, J.; Wang, K. A simple and efficient method for CRISPR/Cas9-induced mutant screening. *J. Genet. Genomics* **2017**, *44*, 207–213. [CrossRef]
137. Thomas, H.R.; Percival, S.M.; Yoder, B.K.; Parant, J.M. High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis. *PLoS ONE* **2014**, *9*, e114632. [CrossRef]
138. Zhu, X.; Xu, Y.; Yu, S.; Lu, L.; Ding, M.; Cheng, J.; Song, G.; Gao, X.; Yao, L.; Fan, D.; et al. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.* **2014**, *4*, 6420. [CrossRef]
139. Vouillot, L.; Thélie, A.; Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 Genes Genomes Genet.* **2015**, *5*, 407–415. [CrossRef]
140. Shan, Q.; Wang, Y.; Li, J.; Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **2014**, *9*, 2395–2410. [CrossRef]
141. Zhang, H.; Zhang, J.; Wei, P.; Zhang, B.; Gou, F.; Feng, Z.; Mao, Y.; Yang, L.; Zhang, H.; Xu, N.; et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* **2014**, *12*, 797–807. [CrossRef]
142. Chen, L.; Li, W.; Katin-Grazzini, L.; Ding, J.; Gu, X.; Li, Y.; Gu, T.; Wang, R.; Lin, X.; Deng, Z.; et al. A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants. *Hortic. Res.* **2018**, *5*, 13. [CrossRef]
143. Lu, H.P.; Liu, S.M.; Xu, S.L.; Chen, W.Y.; Zhou, X.; Tan, Y.Y.; Huang, J.Z.; Shu, Q.Y. CRISPR-S: An active interference element for a rapid and inexpensive selection of genome-edited, transgene-free rice plants. *Plant Biotechnol. J.* **2017**, *15*, 1371–1373. [CrossRef]
144. Bell, C.C.; Magor, G.W.; Gillinder, K.R.; Perkins, A.C. A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. *BMC Genomics* **2014**, *15*, 1002. [CrossRef]
145. Braatz, J.; Harloff, H.J.; Mascher, M.; Stein, N.; Himmelbach, A.; Jung, C. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* **2017**, *174*, 935–942. [CrossRef]

146. Gao, X.; Chen, J.; Dai, X.; Zhang, D.; Zhao, Y. An effective strategy for reliably isolating heritable and *Cas9* -free *Arabidopsis* mutants generated by CRISPR/Cas9-mediated genome editing. *Plant Physiol.* **2016**, *171*, 1794–1800. [CrossRef]
147. He, Y.; Zhu, M.; Wang, L.; Wu, J.; Wang, Q.; Wang, R.; Zhao, Y. Programmed self-elimination of the CRISPR/Cas9 construct greatly accelerates the isolation of edited and transgene-free rice plants. *Mol. Plant* **2018**, *11*, 1210–1213. [CrossRef]
148. Zhang, Y.; Liang, Z.; Zong, Y.; Wang, Y.; Liu, J.; Chen, K.; Qiu, J.L.; Gao, C. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **2016**, *7*, 12617. [CrossRef]
149. Zong, Y.; Wang, Y.; Li, C.; Zhang, R.; Chen, K.; Ran, Y.; Qiu, J.L.; Wang, D.; Gao, C. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **2017**, *35*, 438–440. [CrossRef]
150. Zong, Y.; Song, Q.; Li, C.; Jin, S.; Zhang, D.; Wang, Y.; Qiu, J.L.; Gao, C. Efficient c-to-t base editing in plants using a fusion of ncas9 and human apobec3a. *Nat. Biotechnol.* **2018**, *36*, 950. [CrossRef]
151. Lin, C.S.; Hsu, C.T.; Yang, L.H.; Lee, L.Y.; Fu, J.Y.; Cheng, Q.W.; Wu, F.H.; Hsiao, H.C.W.; Zhang, Y.; Zhang, R.; et al. Application of protoplast technology to CRISPR/Cas9 mutagenesis: From single-cell mutation detection to mutant plant regeneration. *Plant Biotechnol. J.* **2018**, *16*, 1295–1310. [CrossRef]
152. Andersson, M.; Turesson, H.; Nicolia, A.; Fält, A.S.; Samuelsson, M.; Hofvander, P. Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* **2017**, *36*, 117–128. [CrossRef]
153. Veillet, F.; Perrot, L.; Chauvin, L.; Kermarrec, M.P.; Guyon-Debast, A.; Chauvin, J.E.; Nogué, F.; Mazier, M. Transgene-free genome editing in tomato and potato plants using *Agrobacterium*-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.* **2019**, *20*, 402. [CrossRef]
154. Kim, S.; Kim, D.; Cho, S.W.; Kim, J.; Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **2014**, *24*, 1012–1019. [CrossRef]
155. Jones, H.D. Regulatory uncertainty over genome editing. *Nat. Plants* **2015**, *1*, 10–38. [CrossRef]
156. Woo, J.W.; Kim, J.; Kwon, S.; Corvalán, C.; Cho, S.W.; Kim, H.; Kim, S.G.; Kim, S.T.; Choe, S.; Kim, J.S. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* **2015**, *33*, 1162–1164. [CrossRef]
157. Liang, Z.; Chen, K.; Zhang, Y.; Liu, J.; Yin, K.; Qiu, J.L.; Gao, C. Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or ribonucleoproteins. *Nat. Protoc.* **2018**, *13*, 413–430. [CrossRef]
158. Liang, Z.; Chen, K.; Li, T.; Zhang, Y.; Wang, Y.; Zhao, Q.; Liu, J.; Zhang, H.; Liu, C.; Ran, Y.; et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* **2017**, *8*, 14261. [CrossRef]
159. Andersson, M.; Turesson, H.; Olsson, N.; Fält, A.S.; Ohlsson, P.; Gonzalez, M.N.; Samuelsson, M.; Hofvander, P. Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. *Physiol. Plant.* **2018**, *164*, 378–384. [CrossRef]
160. Mcvey, M.; Khodaverdian, V.Y.; Meyer, D.; Heyer, W.; Genetics, M.; Biology, C.; Cordova, R. Eukaryotic DNA polymerases in homologous recombination. *Annu. Rev. Genet.* **2017**, 393–421. [CrossRef]
161. Malnoy, M.; Viola, R.; Jung, M.H.; Koo, O.J.; Kim, S.; Kim, J.S.; Velasco, R.; Nagamangala Kanchiswamy, C. DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Front. Plant Sci.* **2016**, *7*, 1904. [CrossRef]
162. Svitashev, S.; Schwartz, C.; Lenderts, B.; Young, J.K.; Mark Cigan, A. Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nat. Commun.* **2016**, *7*, 13274. [CrossRef]
163. Kim, H.; Kim, S.T.; Ryu, J.; Kang, B.C.; Kim, J.S.; Kim, S.G. CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* **2017**, *8*, 14406. [CrossRef]
164. Shimatani, Z.; Kashojiya, S.; Takayama, M.; Terada, R.; Arazoe, T.; Ishii, H.; Teramura, H.; Yamamoto, T.; Komatsu, H.; Miura, K.; et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **2017**, *35*, 441–443. [CrossRef]
165. Shimatani, Z.; Ariizumi, T.; Fujikura, U.; Kondo, A.; Ezura, H.; Nishida, K. Targeted base editing with CRISPR-deaminase in tomato. *Methods Mol. Biol.* **2019**, *1917*, 297–307. [CrossRef]
166. Komor, A.C.; Kim, Y.B.; Packer, M.S.; Zuris, J.A.; Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**, *533*, 420–424. [CrossRef]

167. Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K.Y.; et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **2016**, *353*. [CrossRef]
168. Hess, G.T.; Tycko, J.; Yao, D.; Bassik, M.C. Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **2017**, *68*, 26–43. [CrossRef]
169. Gehrke, J.M.; Cervantes, O.; Clement, M.K.; Wu, Y.; Zeng, J.; Bauer, D.E.; Pinello, L.; Joung, J.K. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **2019**, *36*, 977–982. [CrossRef]
170. Ma, Y.; Zhang, J.; Yin, W.; Zhang, Z.; Song, Y.; Chang, X. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **2016**, *13*, 1029–1035. [CrossRef]
171. Li, Z.; Xiong, X.; Li, J.F. New cytosine base editor for plant genome editing. *Sci. China Life Sci.* **2018**, *61*, 1602–1603. [CrossRef]
172. Billon, P.; Bryant, E.E.; Joseph, S.A.; Nambiar, T.S.; Hayward, S.B.; Rothstein, R.; Ciccia, A. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of stop codons. *Mol. Cell* **2017**, *67*, 1068–1079. [CrossRef]
173. Tian, S.; Jiang, L.; Cui, X.; Zhang, J.; Guo, S.; Li, M.; Zhang, H.; Ren, Y.; Gong, G.; Zong, M.; et al. Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* **2018**, *37*, 1353–1356. [CrossRef]
174. Yan, F.; Kuang, Y.; Ren, B.; Wang, J.; Zhang, D.; Lin, H.; Yang, B.; Zhou, X.; Zhou, H. Highly Efficient A·T to G·C Base Editing by Cas9n-Guided tRNA Adenosine Deaminase in Rice. *Mol. Plant* **2018**, *11*, 631–634. [CrossRef]
175. Hua, K.; Tao, X.; Yuan, F.; Wang, D.; Zhu, J.K. Precise A·T to G·C base editing in the rice genome. *Mol. Plant* **2018**, *11*, 627–630. [CrossRef]
176. Kang, B.C.; Yun, J.Y.; Kim, S.T.; Shin, Y.J.; Ryu, J.; Choi, M.; Woo, J.W.; Kim, J.S. Precision genome engineering through adenine base editing in plants. *Nat. Plants* **2018**, *4*, 427–431. [CrossRef]
177. Li, C.; Zong, Y.; Wang, Y.; Jin, S.; Zhang, D.; Song, Q.; Zhang, R.; Gao, C. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **2018**, *19*, 59. [CrossRef]
178. Xing, H.L.; Wang, Z.P.; Zhang, H.Y.; Han, C.Y.; Liu, B.; Wang, X.C.; Chen, Q.J.; Dong, L. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **2014**, *14*, 327. [CrossRef]
179. Mao, Y.; Zhang, H.; Xu, N.; Zhang, B.; Gou, F.; Zhu, J.K. Application of the CRISPR–Cas system for efficient genome engineering in plants. *Mol. Plant* **2013**, *6*, 2008–2011. [CrossRef]
180. Wang, C.; Liu, W.; Wang, G.; Li, J.; Dong, L.; Han, L.; Wang, Q.; Tian, J.; Yu, Y.; Gao, C.; et al. KTN80 confers precision to microtubule severing by specific targeting of katanin complexes in plant cells. *EMBO J.* **2017**, *36*, 3435–3447. [CrossRef]
181. Xie, K.; Minkenberg, B.; Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3570–3575. [CrossRef]
182. Qi, W.; Zhu, T.; Tian, Z.; Li, C.; Zhang, W.; Song, R. High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. *BMC Biotechnol.* **2016**, *16*, 58. [CrossRef]
183. Tang, X.; Zheng, X.; Qi, Y.; Zhang, D.; Cheng, Y.; Tang, A.; Voytas, D.F.; Zhang, Y. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol. Plant* **2016**, *9*, 1088–1091. [CrossRef]
184. Wang, M.; Mao, Y.; Lu, Y.; Tao, X.; Zhu, J. Multiplex gene editing in rice using the CRISPR-Cpf1 System. *Mol. Plant* **2017**, *10*, 1011–1013. [CrossRef]
185. Mali, P.; Aach, J.; Stranges, P.B.; Esvelt, K.M.; Moosburner, M.; Kosuri, S.; Yang, L.; Church, G.M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **2013**, *31*, 833–838. [CrossRef]
186. Guilinger, J.P.; Thompson, D.B.; Liu, D.R. Fusion of catalytically inactive Cas9 to *FokI* nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **2014**, *32*, 577–582. [CrossRef]
187. Chylinski, K.; Le Rhun, A.; Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **2013**, *10*, 726–737. [CrossRef]
188. Lee, C.M.; Cradick, T.J.; Bao, G. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* **2016**, *24*, 645–654. [CrossRef]

189. Ran, F.A.; Cong, L.; Yan, W.X.; Scott, D.A.; Gootenberg, J.S.; Kriz, A.J.; Zetsche, B.; Shalem, O.; Wu, X.; Makarova, K.S.; et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **2015**, *520*, 186–191. [CrossRef]
190. Hu, X.; Meng, X.; Liu, Q.; Li, J.; Wang, K. Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. *Plant Biotechnol. J.* **2018**, *16*, 292–297. [CrossRef]
191. Jia, H.; Xu, J.; Orbović, V.; Zhang, Y.; Wang, N. Editing citrus genome via SaCas9/sgRNA System. *Front. Plant Sci.* **2017**, *8*, 2135. [CrossRef]
192. Steinert, J.; Schiml, S.; Fauser, F.; Puchta, H. Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *Plant J.* **2015**, *84*, 1295–1305. [CrossRef]
193. Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; Degennaro, E.M.; Winblad, N.; Choudhury, S.R.; Abudayyeh, O.O.; Gootenberg, J.S.; et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **2017**, *35*, 31–34. [CrossRef]
194. Zetsche, B.; Gootenberg, J.S.; Abudayyeh, O.O.; Slaymaker, I.M.; Makarova, K.S.; Essletzbichler, P.; Volz, S.E.; Joung, J.; Van Der Oost, J.; Regev, A.; et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015**, *163*, 759–771. [CrossRef]
195. Endo, A.; Masafumi, M.; Kaya, H.; Toki, S. Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Sci. Rep.* **2016**, *6*, 38169. [CrossRef]
196. Zhang, H.; Zhang, J.; Lang, Z.; Botella, J.R.; Zhu, J.K. Genome Editing—Principles and Applications for Functional Genomics Research and Crop Improvement. *Crit. Rev. Plant Sci.* **2017**, *34*, 291–309. [CrossRef]
197. Yin, X.; Biswal, A.K.; Dionora, J.; Perdigon, K.M.; Balahadia, C.P.; Mazumdar, S.; Chater, C.; Lin, H.C.; Coe, R.A.; Kretzschmar, T.; et al. CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene *EPFL9* in rice. *Plant Cell Rep.* **2017**, *36*, 745–757. [CrossRef]
198. Xu, R.; Qin, R.; Li, H.; Li, D.; Li, L.; Wei, P.; Yang, J. Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnol. J.* **2017**, *15*, 713–717. [CrossRef]
199. Begemann, M.B.; Gray, B.N.; January, E.; Gordon, G.C.; He, Y.; Liu, H.; Wu, X.; Brutnell, T.P.; Mockler, T.C.; Oufattole, M. Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Sci. Rep.* **2017**, *7*, 11606. [CrossRef]
200. Scheben, A.; Edwards, D. Genome editors take on crops. *Science* **2017**, *355*, 1122–1123. [CrossRef]
201. Zaidi, S.S.A.; Tashkandi, M.; Mansoor, S.; Mahfouz, M.M. Engineering plant immunity: Using crispr/cas9 to generate virus resistance. *Front. Plant Sci.* **2016**, *7*, 1673. [CrossRef]
202. Al-Sadi, A.M.; Al-Moqbali, H.S.; Al-Yahyai, R.A.; Al-Said, F.A. AFLP data suggest a potential role for the low genetic diversity of acid lime (*Citrus aurantifolia Swingle*) in Oman in the outbreak of witches' broom disease of lime. *Euphytica* **2012**, *188*, 285–297. [CrossRef]
203. Kettles, G.J.; Kanyuka, K. Dissecting the molecular interactions between wheat and the fungal pathogen Zymoseptoria tritici. *Front. Plant Sci.* **2016**, *7*, 508. [CrossRef]
204. Wang, F.; Wang, C.; Liu, P.; Lei, C.; Hao, W.; Gao, Y.; Liu, Y.G.; Zhao, K. Enhanced rice blast resistance by CRISPR/ Cas9-Targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PLoS ONE* **2016**, *11*, e0154027. [CrossRef]
205. Zhou, J.; Peng, Z.; Long, J.; Sosso, D.; Liu, B.; Eom, J.S.; Huang, S.; Liu, S.; Vera Cruz, C.; Frommer, W.B.; et al. Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* **2015**, *82*, 632–643. [CrossRef]
206. Jia, H.; Zhang, Y.; Orbović, V.; Xu, J.; White, F.F.; Jones, J.B.; Wang, N. Genome editing of the disease susceptibility gene CsLOB1 in citrus confers resistance to citrus canker. *Plant Biotechnol. J.* **2017**, *15*, 817–823. [CrossRef]
207. Peng, A.; Chen, S.; Lei, T.; Xu, L.; He, Y.; Wu, L.; Yao, L.; Zou, X. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene *CsLOB1* promoter in citrus. *Plant Biotechnol. J.* **2017**, *15*, 1509–1519. [CrossRef]
208. Zhang, Y.; Bai, Y.; Wu, G.; Zou, S.; Chen, Y.; Gao, C.; Tang, D. Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant J.* **2017**, *91*, 714–724. [CrossRef]
209. Nekrasov, V.; Wang, C.; Win, J.; Lanz, C.; Weigel, D.; Kamoun, S. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* **2017**, *7*, 482. [CrossRef]

210. Gil-Humanes, J.; Wang, Y.; Liang, Z.; Shan, Q.; Ozuna, C.V.; Sánchez-León, S.; Baltes, N.J.; Starker, C.; Barro, F.; Gao, C.; et al. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* **2017**, *89*, 1251–1262. [CrossRef]
211. Ji, X.; Si, X.; Zhang, Y.; Zhang, H.; Zhang, F.; Gao, C. Conferring DNA virus resistance with high specificity in plants using virus-inducible genome-editing system. *Genome Biol.* **2018**, *19*, 197. [CrossRef]
212. Baltes, N.J.; Hummel, A.W.; Konecna, E.; Cegan, R.; Bruns, A.N.; Bisaro, D.M.; Voytas, D.F. Conferring resistance to geminiviruses with the CRISPR-Cas prokaryotic immune system. *Nat. Plants* **2015**, *1*, 15145. [CrossRef]
213. Ali, Z.; Abulfaraj, A.; Idris, A.; Ali, S.; Tashkandi, M.; Mahfouz, M.M. CRISPR/Cas9-mediated viral interference in plants. *Genome Biol.* **2015**, *16*, 238. [CrossRef]
214. Zhang, T.; Zheng, Q.; Yi, X.; An, H.; Zhao, Y.; Ma, S.; Zhou, G. Establishing RNA virus resistance in plants by harnessing CRISPR immune system. *Plant Biotechnol. J.* **2018**, *16*, 1415–1423. [CrossRef]
215. Macovei, A.; Sevilla, N.R.; Cantos, C.; Jonson, G.B.; Slamet-Loedin, I.; Čermák, T.; Voytas, D.F.; Choi, I.R.; Chadha-Mohanty, P. Novel alleles of rice eIF4G generated by CRISPR/Cas9-targeted mutagenesis confer resistance to Rice tungro spherical virus. *Plant Biotechnol. J.* **2018**, *16*, 1918–1927. [CrossRef]
216. Wang, X.; Tu, M.; Wang, D.; Liu, J.; Li, Y.; Li, Z.; Wang, Y.; Wang, X. CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnol. J.* **2018**, *16*, 844–855. [CrossRef]
217. Zhang, Z.; Ge, X.; Luo, X.; Wang, P.; Fan, Q.; Hu, G.; Xiao, J.; Li, F.; Wu, J. Simultaneous Editing of Two Copies of Gh14-3-3d Confers Enhanced Transgene-Clean Plant Defense Against Verticillium dahliae in Allotetraploid Upland Cotton. *Front. Plant Sci.* **2018**, *9*, 842. [CrossRef]
218. Ortigosa, A.; Gimenez-Ibanez, S.; Leonhardt, N.; Solano, R. Design of a bacterial speck resistant tomato by CRISPR/Cas9-mediated editing of SlJAZ2. *Plant Biotechnol. J.* **2018**, *17*, 665–673. [CrossRef]
219. Tashkandi, M.; Ali, Z.; Aljedaani, F.; Shami, A.; Mahfouz, M.M. Engineering resistance against Tomato yellow leaf curl virus via the CRISPR/Cas9 system in tomato. *Plant Signal. Behav.* **2018**, *13*. [CrossRef]
220. Chandrasekaran, J.; Brumin, M.; Wolf, D.; Leibman, D.; Klap, C.; Pearlsman, M.; Sherman, A.; Arazi, T.; Gal-On, A. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* **2016**, *7*, 1140–1153. [CrossRef]
221. Kim, D.; Kim, D.; Alptekin, B.; Budak, H. CRISPR/Cas9 genome editing in wheat. *Funct. Integr. Genomics* **2017**, *18*, 31–41. [CrossRef]
222. Shen, C.; Que, Z.; Xia, Y.; Tang, N.; Li, D.; He, R.; Cao, M. Knock out of the annexin gene OsAnn3 via CRISPR/Cas9-mediated genome editing decreased cold tolerance in rice. *J. Plant Biol.* **2017**, *60*, 539–547. [CrossRef]
223. Lou, D.; Wang, H.; Liang, G.; Yu, D. *OsSAPK2* confers abscisic acid sensitivity and tolerance to drought stress in rice. *Front. Plant Sci.* **2017**, *8*, 993. [CrossRef]
224. Shi, J.; Gao, H.; Wang, H.; Lafitte, H.R.; Archibald, R.L.; Yang, M.; Hakimi, S.M.; Mo, H.; Habben, J.E. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* **2017**, *15*, 207–216. [CrossRef]
225. Curtin, S.J.; Xiong, Y.; Michno, J.M.; Campbell, B.W.; Stec, A.O.; Čermák, T.; Starker, C.; Voytas, D.F.; Eamens, A.L.; Stupar, R.M. CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. *Plant Biotechnol. J.* **2018**, *16*, 1125–1137. [CrossRef]
226. Wang, L.; Chen, L.; Li, R.; Zhao, R.; Yang, M.; Sheng, J.; Shen, L. Reduced drought tolerance by CRISPR/Cas9-mediated SlMAPK3 mutagenesis in tomato plants. *J. Agric. Food Chem.* **2017**, *65*, 8674–8682. [CrossRef]
227. Shen, L.; Wang, C.; Fu, Y.; Wang, J.; Liu, Q.; Zhang, X.; Yan, C.; Qian, Q.; Wang, K. QTL editing confers opposing yield performance in different rice varieties. *J. Integr. Plant Biol.* **2018**, *60*, 89–93. [CrossRef]
228. Bo, W.; Zhaohui, Z.; Huanhuan, Z.; Xia, W.; Binglin, L.; Lijia, Y.; Xiangyan, H.; Deshui, Y.; Xuelian, Z.; Chunguo, W.; et al. Targeted Mutagenesis of NAC Transcription Factor Gene, OsNAC041, Leading to Salt Sensitivity in Rice. *Rice Sci.* **2019**, *26*, 98–108. [CrossRef]
229. Sadanandom, A.; Srivastava, A.K.; Zhang, C. Targeted mutagenesis of the SUMO protease, Overly Tolerant to Salt1 in rice through CRISPR/Cas9-mediated genome editing reveals a major role of this SUMO protease in salt tolerance. *BioRxiv* **2019**, 555706. [CrossRef]

230. Zhang, A.; Liu, Y.; Wang, F.; Li, T.; Chen, Z.; Kong, D.; Bi, J.; Zhang, F.; Luo, X.; Wang, J.; et al. Enhanced rice salinity tolerance via CRISPR/Cas9-targeted mutagenesis of the OsRR22 gene. *Mol. Breed.* **2019**, *39*, 47. [CrossRef]
231. Li, R.; Liu, C.; Zhao, R.; Wang, L.; Chen, L.; Yu, W.; Zhang, S.; Sheng, J.; Shen, L. CRISPR/Cas9-Mediated SlNPR1 mutagenesis reduces tomato plant drought tolerance. *BMC Plant Biol.* **2019**, *19*, 38. [CrossRef]
232. Shim, J.S.; Oh, N.; Chung, P.J.; Kim, Y.S.; Choi, Y.D.; Kim, J.K. Overexpression of OsNAC14 Improves Drought Tolerance in Rice. *Front. Plant Sci.* **2018**, *9*, 310. [CrossRef]
233. Lou, D.; Wang, H.; Yu, D. The sucrose non-fermenting-1-related protein kinases SAPK1 and SAPK2 function collaboratively as positive regulators of salt stress tolerance in rice. *BMC Plant Biol.* **2018**, *18*, 203. [CrossRef]
234. Zhang, M.; Cao, Y.; Wang, Z.; Wang, Z.Q.; Shi, J.; Liang, X.; Song, W.; Chen, Q.; Lai, J.; Jiang, C. A retrotransposon in an HKT1 family sodium transporter causes variation of leaf Na+exclusion and salt tolerance in maize. *New Phytol.* **2018**, *217*, 1161–1176. [CrossRef]
235. Li, R.; Zhang, L.; Wang, L.; Chen, L.; Zhao, R.; Sheng, J.; Shen, L. Reduction of Tomato-Plant Chilling Tolerance by CRISPR-Cas9-Mediated SlCBF1 Mutagenesis. *J. Agric. Food Chem.* **2018**, *66*, 9042–9051. [CrossRef]
236. Zhang, Y.; Li, D.; Zhang, D.; Zhao, X.; Cao, X.; Dong, L.; Liu, J.; Chen, K.; Zhang, H.; Gao, C.; et al. Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *Plant J.* **2018**, *94*, 857–866. [CrossRef]
237. Lu, K.; Wu, B.; Wang, J.; Zhu, W.; Nie, H.; Qian, J.; Huang, W.; Fang, Z. Blocking amino acid transporter *OsAAP3* improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol. J.* **2018**, *16*, 1710–1722. [CrossRef]
238. Liu, J.; Chen, J.; Zheng, X.; Wu, F.; Lin, Q.; Heng, Y.; Tian, P.; Cheng, Z.J.; Yu, X.; Zhou, K.; et al. *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* **2017**, *3*, 17043. [CrossRef]
239. Li, S.; Gao, F.; Xie, K.; Zeng, X.; Cao, Y.; Zeng, J.; He, Z.; Ren, Y.; Li, W.; Deng, Q.; et al. The *OsmiR396c-OsGRF4-OsGIF1* regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* **2016**, *14*, 2134–2146. [CrossRef]
240. Li, M.; Li, X.; Zhou, Z.; Wu, P.; Fang, M.; Pan, X.; Lin, Q.; Luo, W.; Wu, G.; Li, H. Reassessment of the four yield-related genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* **2016**, *7*, 377. [CrossRef]
241. Xu, R.; Yang, Y.; Qin, R.; Li, H.; Qiu, C.; Li, L.; Wei, P.; Yang, J. Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice. *J. Genet. Genomics* **2016**, *43*, 529–532. [CrossRef]
242. Li, X.; Zhou, W.; Ren, Y.; Tian, X.; Lv, T.; Wang, Z.; Fang, J.; Chu, C.; Yang, J.; Bu, Q. High-efficiency breeding of early-maturing rice cultivars via CRISPR/Cas9-mediated genome editing. *J. Genet. Genomics* **2017**, *44*, 175–178. [CrossRef]
243. Ma, L.; Zhang, D.; Miao, Q.; Yang, J.; Xuan, Y.; Hu, Y. Essential role of sugar transporter *OsSWEET11* during the early stage of rice grain filling. *Plant Cell Physiol.* **2017**, *58*, 863–873. [CrossRef]
244. Huang, J.; Li, J.; Zhou, J.; Wang, L.; Yang, S.; Hurst, L.D.; Li, W.H.; Tian, D. Identifying a large number of high-yield genes in rice by pedigree analysis, whole-genome sequencing, and CRISPR-Cas9 gene knockout. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E7559–E7567. [CrossRef]
245. Zhang, J.; Zhang, H.; Botella, J.R.; Zhu, J. Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the *Waxy* gene in elite rice varieties. *J. Integr. Plant Biol.* **2018**, *60*, 369–375. [CrossRef]
246. Sun, Y.; Jiao, G.; Liu, Z.; Zhang, X.; Li, J.; Guo, X.; Du, W.; Du, J.; Francis, F.; Zhao, Y.; et al. Generation of High-Amylose Rice through CRISPR/Cas9-Mediated Targeted Mutagenesis of Starch Branching Enzymes. *Front. Plant Sci.* **2017**, *8*, 298. [CrossRef]
247. Sánchez-León, S.; Gil-Humanes, J.; Ozuna, C.V.; Giménez, M.J.; Sousa, C.; Voytas, D.F.; Barro, F. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* **2018**, *16*, 902–910. [CrossRef]
248. Li, A.; Jia, S.; Yobi, A.; Ge, Z.; Sato, S.; Zhang, C.; Angelovici, R.; Clemente, T.E.; Holding, D.R. Editing of an alpha-kafirin gene family increases digestibility and protein quality in sorghum. *Plant Physiol.* **2018**, *177*, 1425–1438. [CrossRef]
249. Okuzaki, A.; Ogawa, T.; Koizuka, C.; Kaneko, K.; Inaba, M.; Imamura, J.; Koizuka, N. CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in Brassica napus. *Plant Physiol. Biochem.* **2018**, *131*, 63–69. [CrossRef]

250. Li, R.; Fu, D.; Zhu, B.; Luo, Y.; Zhu, H. CRISPR/Cas9-mediated mutagenesis of lncRNA1459 alters tomato fruit ripening. *Plant J.* **2018**, *94*, 513–524. [CrossRef]
251. Li, X.; Wang, Y.; Chen, S.; Tian, H.; Fu, D.; Zhu, B.; Luo, Y.; Zhu, H. Lycopene Is Enriched in Tomato Fruit by CRISPR/Cas9-Mediated Multiplex Genome Editing. *Front. Plant Sci.* **2018**, *9*, 559. [CrossRef]
252. Butt, H.; Jamil, M.; Wang, J.Y.; Al-Babili, S.; Mahfouz, M. Engineering plant architecture via CRISPR/Cas9-mediated alteration of strigolactone biosynthesis. *BMC Plant Biol.* **2018**, *18*, 174. [CrossRef]
253. Cai, Y.; Chen, L.; Liu, X.; Guo, C.; Sun, S.; Wu, C.; Jiang, B.; Han, T.; Hou, W. CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. *Plant Biotechnol. J.* **2018**, *16*, 176–185. [CrossRef]
254. Soyk, S.; Müller, N.A.; Park, S.J.; Schmalenbach, I.; Jiang, K.; Hayama, R.; Zhang, L.; Van Eck, J.; Jiménez-Gómez, J.M.; Lippman, Z.B. Variation in the flowering gene *SELF PRUNING 5G* promotes day-neutrality and early yield in tomato. *Nat. Genet.* **2017**, *49*, 162–168. [CrossRef]
255. Nonaka, S.; Arai, C.; Takayama, M.; Matsukura, C.; Ezura, H. Efficient increase of Γ-aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci. Rep.* **2017**, *7*, 7057. [CrossRef]
256. Chahal, G.S.; Gosal, S.S. *Principles and Procedures of Plant Breeding: Biotechnological and Conventional Approaches*; Alpha Science Int'l Ltd.: Oxford, UK, 2002.
257. Pacher, M.; Puchta, H. From classical mutagenesis to nuclease-based breeding–directing natural DNA repair for a natural end-product. *Plant J.* **2017**, *90*, 819–833. [CrossRef]
258. Scheben, A.; Wolter, F.; Batley, J.; Puchta, H.; Edwards, D. Towards CRISPR/Cas crops–bringing together and genome editing. *New Phytol.* **2017**, *216*, 682–698. [CrossRef]
259. Sikora, P.; Chawade, A.; Larsson, M.; Olsson, J.; Olsson, O. Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Int. J. Plant Genomics* **2011**, *2011*, 1–3. [CrossRef]
260. Hartung, F.; Schiemann, J. Precise plant breeding using new genome editing techniques: Opportunities, safety and regulation in the EU. *Plant J.* **2014**, *78*, 742–752. [CrossRef]
261. Xu, R.F.; Li, H.; Qin, R.Y.; Li, J.; Qiu, C.H.; Yang, Y.C.; Ma, H.; Li, L.; Wei, P.C.; Yang, J.B. Generation of and "transgene clean" targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* **2015**, *5*, 11491. [CrossRef]
262. Visser, B.; Eaton, D.; Louwaars, N.; van der Meer, I.M.; Beekwilder, J.; van Tongeren, F. *Potential of Genetic Use Restriction Technologies (GURTs) on Agrobiodiversity and Agricultural Production Systems*; FAO: Rome, Italy, 2001.
263. Ledford, H. CRISPR, the disruptor. *Nat. News* **2015**, *522*, 20–24. [CrossRef]
264. Feng, Z.; Mao, Y.; Xu, N.; Zhang, B.; Wei, P.; Yang, D.L.; Wang, Z.; Zhang, Z.; Zheng, R.; Yang, L.; et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4632–4637. [CrossRef]
265. Waltz, E. Gene-edited CRISPR mushroom escapes US regulation. *Nat. News* **2016**, *532*, 293. [CrossRef]
266. Callaway, E. CRISPR plants now subject to tough GM laws in European Union. *Nature* **2018**, *560*, 16. [CrossRef]
267. Sprink, T.; Eriksson, D.; Schiemann, J.; Hartung, F. Regulatory hurdles for genome editing: Process-vs. Product-based approaches in different regulatory contexts. *Plant Cell Rep.* **2016**, *35*, 1493–1506. [CrossRef]
268. Kahrmann, J.; Bömeke, O.; Leggewie, G. Aged GMO legislation meets new genome editing techniques. *Z. Eur. Umw. Plan.* **2017**, *15*, 176–182. [CrossRef]
269. Fears, R.; Ter Meulen, V. Point of View: How should the applications of genome editing be assessed and regulated? *Elife* **2017**, *6*, e26295. [CrossRef]
270. Ishii, T.; Araki, M. A future scenario of the global regulatory landscape regarding genome-edited crops. *GM Crop. Food* **2017**, *8*, 44–56. [CrossRef]
271. Smyth, S.J. Canadian regulatory perspectives on genome engineered crops. *GM Crop. Food* **2017**, *8*, 35–43. [CrossRef]
272. Whelan, A.I.; Lema, M.A. Regulatory framework for gene editing and other new breeding techniques (NBTs) in Argentina. *GM Crop. Food* **2015**, *6*, 253–265. [CrossRef]
273. Brinegar, K.K.; Yetisen, A.; Choi, S.; Vallillo, E.; Ruiz-Esparza, G.U.; Prabhakar, A.M.; Khademhosseini, A.; Yun, S.H. The commercialization of genome-editing technologies. *Crit. Rev. Biotechnol.* **2017**, *37*, 924–932. [CrossRef]

274. Genome Editing/Genome Engineering Market by Application (Cell Line Engineering, Animal & Plant Genetic Engineering), Technology (CRISPR, Antisense, TALEN, Zinc Finger Nuclease) & End User (Biotechnology & Pharmaceutical, Crop)-Global Forecast to 2019. Available online: http://www.marketsandmarkets.com/Market-Reports/genome-editing-engineerinmarket231037000.html (accessed on 12 April 2016).
275. Van Erp, P.B.; Bloomer, G.; Wilkinson, R.; Wiedenheft, B. The history and market impact of CRISPR RNA-guided nucleases. *Curr. Opin. Virol.* **2015**, *12*, 85–90. [CrossRef]
276. Ousterout, D.G.; Kabadi, A.M.; Thakore, P.I.; Majoros, W.H.; Reddy, T.E.; Gersbach, C.A. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* **2015**, *6*, 6244. [CrossRef]
277. Suzuki, K.; Tsunekawa, Y.; Hernandez-Benitez, R.; Wu, J.; Zhu, J.; Kim, E.J.; Hatanaka, F.; Yamamoto, M.; Araoka, T.; Li, Z.; et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **2016**, *540*, 144. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Transgenerational Perpetuation of CHS Gene Expression and DNA Methylation Status Induced by Short Oligodeoxynucleotides in Flax (*Linum usitatissimum*)

Magdalena Dzialo [1], Jan Szopa [1,2,3], Agata Hnitecka [1] and Magdalena Zuk [1,2,*]

[1] Department of Genetic Biochemistry, Faculty of Biotechnology, Wroclaw University, Przybyszewskiego 63, 51-148 Wroclaw, Poland
[2] Linum Foundation, pl. Grunwaldzki 24a, 50-363 Wroclaw, Poland
[3] Department of Genetics, Plant Breeding and Seed Production, Wroclaw University of Environmental and Life Sciences, pl. Grunwaldzki 24a, 50-363 Wroclaw, Poland
* Correspondence: mzuk@ibmb.uni.wroc.pl

Received: 18 June 2019; Accepted: 14 August 2019; Published: 16 August 2019

Abstract: Over two decades ago, short oligodeoxynucleotides (ODNs) were proven to be an effective and rapid technique for analysis of gene function without interference in the plant genome. Our previous research has shown the successful regulation of chalcone synthase (CHS) gene expression in flax by ODN technology. The CHS gene encodes a pivotal enzyme in flavonoid biosynthesis. The manipulation of its transcript level was the result of the specific methylation status developed after treatment with ODNs. In further analysis of the application of oligodeoxynucleotides in plants, we will focus on maintaining the methylation status induced originally by ODNs homologous to the regulatory regions of the CHS gene in flax. This article reports the latest investigation applied to stabilization and inheritance of the epigenetic marks induced by plants' treatment with ODNs. The methylation status was analyzed in the particular CCGG motifs located in the CHS gene sequence. Individual plants were able to maintain alterations induced by ODNs. In order to confirm the impact of methylation marks on the nucleosome rearrangement, chromatin accessibility assay was performed. The perpetuation of targeted plant modulation induced by ODNs exhibits strong potential for improving crops and intensified application for medicine, nutrition and industry.

Keywords: chalcone synthase; chromatin accessibility; epigenetic inheritance; genetic engineering; methylation; ODNs

1. Introduction

A considerable number of studies have confirmed that epigenetic mechanisms play a crucial role in the ability of plants to adapt to the environment during stress conditions [1–3]. Heritable epigenetic variability can lead to physiological, morphological and ecological changes. Alterations of the epigenome can improve the plant features valuable for agronomical reasons [4]. Precise manipulation of the DNA methylation status enables one to control chromatin condensation, and thus gene expression. The assessment of detailed mechanisms that regulate the epigenetic changes under particular stress factors could provide information about the characteristic epigenetic marks that lead to adaptation in encountered stress conditions [5].

Epigenetic modifications induced under stress factors can be reversed and epigenomic status can return to the initial state, when the stress conditions are no longer present. However, the latest findings have shown that some of these changes can be preserved as stress memory leading to inheritance of the epigenetic status and thus enhancement of the adaptable ability of the progeny to the

environment [5]. In comparison to mammalian cells, in plant cells, massive erosion of the epigenetic marks during gametogenesis occurs [6]. Particular short RNA molecules are assumed to crucially contribute to the inheritance of the epigenetic changes. Due to the RNA-dependent DNA methylation, the methylation implemented to the genome of parental cells can be maintained after cell division. The maintenance of the DNA methylation after replication is the effect of the feed-forward maintenance of de novo methylation in plants. The essence of this process is the establishment of two processes. The first mechanism involves methylation in the CpG and CpHpG contexts (H indicates cytosine, adenine or thymine) through methyltransferase 1 (MET1) and chromomethylase 3 (CMT3) in the newly synthesized DNA strand. However, cytosine methylation is also maintained through RNA-dependent DNA methylation by short RNA molecules stimulated by DNA methylation. The occurrence of these two processes leads to independence of the DNA methylation mechanism in the plant cell [7].

The latest research has shown that apart from environmental stress, epigenetic changes may occur as an effect of short oligodeoxynucleotides' (ODNs/OLIGOs) activity. The mechanism of ODNs is based on the development of particular short non-coding RNAs, which may lead to the elevated accumulation (RNA activation—RNAa) or degradation (RNA interference—RNAi) of the gene transcript and the RNA-directed DNA methylation (RdDM) [8,9].

In our previous work [9], we demonstrated the effectiveness of the short oligodeoxynucleotides in the induction of epigenetic changes in flax. The plant infiltration with ODNs complementary to the regulatory regions (ODN1 to 5'UTR, ODN6 to intron, ODN11 to 3'UTR) of the chalcone synthase (CHS) genes *LuCHS6* and *LuCHS7* led to changes in the CHS genes' expression. The alteration of the *CHS* transcript levels was assumed to be triggered by the changes in the DNA methylation. The identification of the 5-methylcytosine (5-mC) in the CCGG motifs located in the CHS genes made it possible to distinguish the stably unmethylated motifs, independently of the analyzed plant (*LuCHS6*/*LuCHS7*: −232/−, +217/−, +1606/+1156) and variable methylated motifs (*LuCHS6*/*LuCHS7*: +996/+552, +1219/+775, +1273/+829). The in silico analysis of nucleosome formation energy also indicates the alteration of nucleosome position upon the changes in the methylation status of *LuCHS6*. The preliminary examination of this theoretical presumption suggests that the increased level of methylation in the +1273 CCGG motif of the transgenic plants led to the visible changes in the accessibility of DNA for the restriction enzymes. Thus, the obtained results suggested the variable susceptibility of DNA upon methylation to the interaction with the particular effector proteins (e.g., transcription machinery) [9].

The subject of our research was the linseed variety of common flax *Linum usitatissimum* L.–Linola. Linseed is widely known for its biomedical applications due to an appropriate ratio of unsaturated fatty acids in the linseed-derived oil (the ω-6:ω-3 ratio is 1:4) [10], as well as high content of phenylpropanoid compounds (flavonoids, lignans and phenolic acids) [11,12] and fiber [13]. Flavonoids present in the flaxseed are popular in the context of health-promoting properties, including antioxidative [14] and antimicrobial activity [15]. Recent findings indicate a preventive or therapeutic role of flavonoids from flax in metabolic disorders [16] and carcinogenesis [17,18].

Chalcone synthase (CHS) catalyzes the first step of flavonoid biosynthesis [19]. In many plant species, there occur more than one gene encoding CHS (multigenomic family). It was demonstrated that there are 5 CHS genes (Phytozome ID)—*LuCHS3* (Lus10041508), *LuCHS4* (Lus10023670), *LuCHS5* (Lus10011746), *LuCHS6* (Lus10033717), *LuCHS7* (Lus10031622)—and 4 CHS-like genes (Phytozome ID): *LuCHS1* (Lus10039904), *LuCHS2* (Lus10042388), *LuCHS8* (Lus10026286), *LuCHS9* (Lus10002187) in the genome of *Linum usitatissimum* L. [20]. The transcripts derived from the 5 mentioned CHS genes were also confirmed by our group [9,21]. Moreover, our findings indicated the presence of 2 additional CHS genes, named in order: *LuCHS10* and *LuCHS11* (recognized fragment of *LuCHS11* sequence is presented in Supplemental Figure S3). Initially, they were only suspected to be transcriptional variants of *LuCHS3* and *LuCHS4,* respectively. However, the in silico analysis showed the diverse localization of these genes on the chromosomes. The systematized names, numbers in databases and genomic localization are presented in Table 1. Apart from the fact that CHS genes encode the key enzyme in

flavonoid formation, they display a susceptibility to the response to stress factors and can be specifically regulated under particular environmental conditions. The expression of the chalcone synthase genes may also depend on the plant developmental stage and tissue specificity. In our research, we focused on two CHS genes: *LuCHS6* and *LuCHS7*. The genomic sequences of these genes are similar to the CHS-A encoding gene from *Petunia* x *hybrida*, used for the generation of genetically modified (GM-CHS) flax, which served as the model for changes in CHS gene expression in our initial research.

Table 1. Genes encoding CHS in flax.

Gene	Gene ID (Phytozome)	Gene Accession No (GenBank)	Chromosomal Localization
LuCHS3 [a]	Lus10041508 [a]	AFSQ01010505 [b]	CP027628.1 (Lu4) Range 1: 14557172 to 14557366 Range 2: 14558491 to 14559582
LuCHS4 [a]	Lus10023670 [a]	AFSQ01007984 [b]	CP027629.1 (Lu5) Range 1: 8911370 to 8911554 Range 2: 8911645 to 8912061 Range 3: 8912194 to 8912636
LuCHS5 [a]	Lus10011746 [a]	AFSQ01007986 [b]	CP027629.1 (Lu5) Range 1: 8892102 to 8892257 Range 2: 8893137 to 8894143
LuCHS6 [a]	Lus10033717 [a]	AFSQ01005163 [b]	CP027630.1 (Lu6) Range 1: 1398135 to 1398320 Range 2: 1396599 to 1397611
LuCHS7 [a]	Lus10031622 [a]	AFSQ01023955 [b]	CP027622.1 (Lu12) Range 1: 19609154 to 19609339 Range 2: 19609419 to 19610428
LuCHS10	-	AFSQ01010603 [b]	CP027619.1 (Lu1) Range 1: 5753813 to 5754010 Range 2: 5753070 to 5752694
LuCHS11	-	AFSQ01012744 [c]	CP027632.1 (Lu8) Range 1: 14778517 to 14778701 Range 2: 14777435 to 14778426

[a]—indicates the CHS genes (sequences and names) published by [20]; [b]—indicates the CHS genes (sequences and names) published by [21]; [c]—recently identified CHS gene.

The main purpose of our research was the monitoring of changes induced by the treatment with ODNs after a time longer than 48 h since the beginning of incubation. The plants from subsequent generations derived from ODN-treated flax were analyzed in the context of the *LuCHS6* and *LuCHS7* gene expression and methylation of CCGG motifs, in order to verify the perpetuation of the induced changes. With a view to the relative susceptibility to ODNs' degradation by nucleases, modification of the oligodeoxynucleotide molecules was applied—by methylation of cytosines present in the ODN sequence (met) and by the replacement of oxygen by a sulfur at the phosphodiester bond (thiophosphate–pto). Stability of the epigenetic changes induced by the modified ODNs was also analyzed in the next generation of plants. Moreover, the chromatin accessibility assay has shown the differences between the control (Linola) and plants with stable epigenetic modification (W92.40). Performed experiments contributed to the understanding of the processes involved in the introduction of exogenous oligonucleotides into plant cells.

2. Results

2.1. Heredity of CHS Gene Expression in the F1, F2 and F3 Generations of ODN-Treated Flax.

The nomenclature of the CHS gene isoforms *CHS1* (Phytozome database, Lus10031622) and *CHS2* (Phytozome database, Lus10033717) published in our previous work [9] will be followed in the present paper by the numbers used by [20], sc. *LuCHS6* (Phytozome database, Lus10033717) and *LuCHS7* (Phytozome database, Lus10031622).

The level of the mRNA derived from both studied isoforms, *LuCHS6* and *LuCHS7*, was analyzed in the F1 and F2 generation of plants treated with ODNs. In our previous study, the series of ODN sequences homologous to *CHS* was investigated. The oligodeoxynucleotides directed to the regulatory regions, 5'UTR-ODN1, intron-ODN6 and 3'UTR-ODN11 have shown significant modulation in the CHS gene expression. For ODN1 and ODN11, overexpression of the endogenous CHS gene was observed. Alternatively, for ODN6, the gene repression was noted for the studied gene. These three ODNs were selected for further analyses. Table 2 presents the maintenance of the modulation of CHS gene expression induced by ODNs in the next generations (F1–F3).

Table 2. Maintenance of gene expression modulation of CHS gene in F1, F2 and F3 generations of plants treated with short oligodeoxynucleotides (ODNs).

Generation	Type of Plant	Percentage of Modulation Maintenance (%)
F1	ODN1	71
	ODN6	5
	ODN11	86
F2	ODN1	33
	ODN6	5
	ODN11	75
F3	ODN1	38
	ODN6	61
	ODN11	6

A part of seeds obtained from the F0 plants treated with ODNs were introduced into the in vitro culture in order to select the lines that maintained the initially induced changes in the *CHS* expression. The percentage proportion of the plants that maintained the tendency of the initial modulation to all analyzed plants was as follows: ODN1-71%, ODN6-5%, ODN11-86%, where the studied groups varied between 7 to 20 individuals in the F1 plant generation. The most plants with the stable modification in the *CHS* gene were noted for the ODN11, whereas the least for ODN6. Seeds of selected plants that maintained the generated characteristics served to obtain the next generation (F1) under field conditions.

Seeds collected from the F1 generation of plants were also partially introduced to the in vitro conditions. The total gene expression of *LuCHS6* and *LuCHS7* (the sum) was determined in the F2 flax. Among the series of plants, individuals that presented a tendency of *CHS* expression similar to the parental generation were selected. The most plants that perpetuated the alteration in the CHS gene expression was observed according to the ODN11 treatment (75%—6 of 8 ODN11 treated plants). A minor percentage (33%) of the ODN1 treated plants maintained in the F2 generation induced primarily changes. The repression noted after ODN6 in the F0 and F1 plants was not observed in the F2 generation and the massive overexpression in those plants was investigated.

The F3 field generation of plants treated with ODNs was also determined in the context of the maintenance of the induced changes in the CHS transcript level. A decreased percentage of modulation

maintenance was observed for ODN1 (38%) and ODN11 (6%) in relation to the previous generations. In contrast, the number of plants that perpetuated repression of the CHS gene was the highest among the F3 generation of ODN-treated plants and the maintenance of the induced changes reached 61%.

2.2. Heredity of Epigenetic Changes in the F1 Generation of ODN-Treated Flax Cultivated in the Field

It has been repeatedly demonstrated that CHS gene expression may be induced in plants under stress conditions such as UV light, and bacterial or fungal infection; therefore it was decided to check during plant growth, as the environmental stress present in the field will affect the presence of induced epigenetic changes. The seeds from F0 generation plants treated with ODNs were cultivated in the experimental field. During growth of F1 plants in natural conditions, individual young plants (6-week-old plants in the middle of the vegetative growth period) were selected (and harvested) for testing from a pool of plants in order to determine the maintenance of the induced features.

2.2.1. CHS Gene Expression in the Selected Plants

Among plants cultivated in the field, the highest percentage of the flax treated ODN1 presented the primarily induced modulation of *CHS* expression (75%) Only 33% of the analyzed F1 generation of plants incubated with ODN11 possessed overexpression of the CHS gene. For ODN6, none of the analyzed plants presented expected repression of the studied gene. In Figure 1, the CHS gene expression for selected individuals is presented.

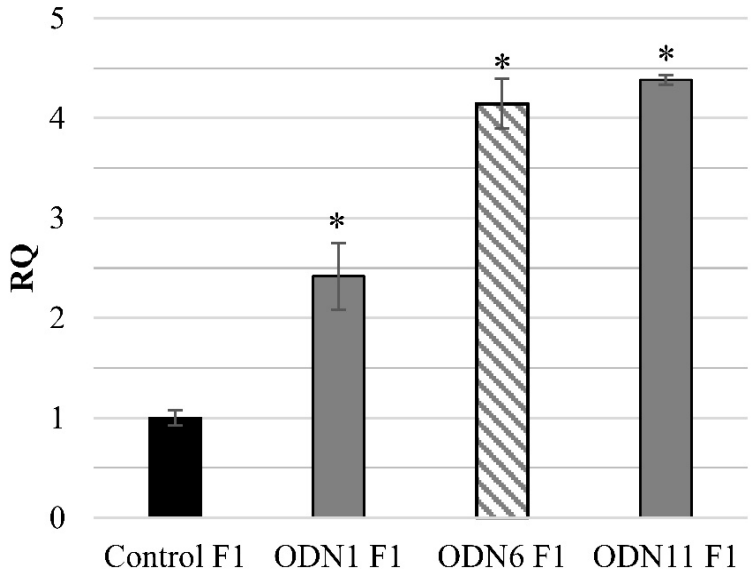


Figure 1. CHS gene expression in the F1 generation of plants that maintained modifications induced by short oligodeoxynucleotides (ODNs), cultivated in the experimental field. The seeds obtained from F0 plants were sown and plants were cultivated in the field. The total expression of both CHS genes, *LuCHS6* and *LuCHS7*, was determined by the real-time PCR reaction. The values are referred to the reference gene expression actin. The relative quantity (RQ) presents the transcript level in comparison to the control (set as 1, black). Plants treated with ODN6 met did not maintain in the F1 generation the initially observed level of CHS gene expression (bar presented in diagonal stripes). Data represent the mean value ± SD from at least three repeats of the experiment. The significance of the differences between each mean and control was determined by Student's *t*-test. Asterisk indicates * $p < 0.05$.

2.2.2. Pattern of DNA Methylation in Variable CCGG Motifs of the CHS Gene

For the plants presented in Figure 1, the methylation status of the variable CCGG sites was evaluated (Figure 2A–D). In particular F1 individuals, methylation of CCGG motifs was mostly similar to the profiles observed in our previous study [9]. The methylation status of F1 control plants resembled

the results for F0 control plants. However, at the +1273 site, a reduced percentage of demethylation (22.5%) for the F1 generation was noted (Figure 2A), while for the F0 generation, the relevant status measured 33.9%. For ODN1, the percentage of CCmGG methylation at the +996 site was similar to the methylation level observed in the plants after treatment. The methylation profile at the +1273 site for plants modified by ODN1 was corresponding between generations; hence, in the F1 generation an increased methylation status (by 15.8%) was observed (Figure 2B). According to ODN6, despite the observed reversal of the tendency of change in CHS gene expression, the percentage of CCmGG at the +996 motif was stable, as observed after 48 h after incubation. Also at the +1273 site, the tendency of change was maintained and a minor increase in demethylation was observed (Figure 2C). For ODN11, significant stable methylation of internal cytosine CCmGG was noted in the +996 and +1273 motifs, which correlates with primarily induced ODN modifications (Figure 2D).

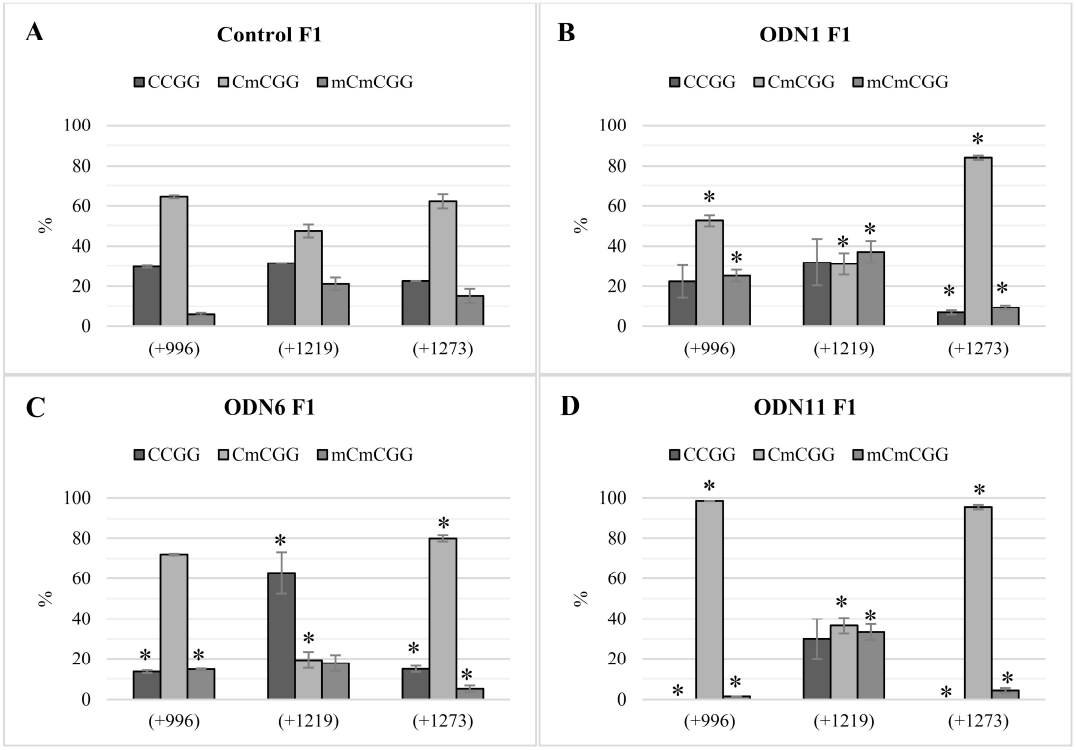


Figure 2. Methylation in the variable CCGG motifs identified in the CHS gene sequence in the F1 generation of control plants (**A**) and individuals that maintained modifications induced by ODNs (**B**–**D**), cultivated in the experimental field. The seeds obtained from F0 plants were sown and plants were cultivated in the field. The figure presents the percentage of cytosine methylation in the variable CCGG sites (+996, +1219, +1273) located in the coding region of the CHS gene sequences. The genomic DNA was digested by restriction enzymes HpaII and MspI. The amount of non-digested DNA was determined by real-time PCR. The percentage of a particular modification is presented for studied plants and control: CCGG—lack of methylation (dark grey), CCmGG—methylation of internal cytosine (light grey) and CmCmGG—methylation of both cytosines (medium grey). The site positions were presented according to the *LuCHS6* sequence, due to the presence of all CCGG sites. Data represent the mean value ± SD from at least three independent + experimental repeats. The significance of the differences between each mean and control was determined by Student's *t*-test. Asterisk indicates * $p < 0.05$.

2.3. Activity of ODNs Modified by Methylation (Met) and Thiophosphate (Pto)

Susceptibility of the oligodeoxynucleotide sequences for the nucleolytic digestion limits the lifetime of ODNs. In order to extend the time of action of short oligonucleotides, two modifications of the nucleotides included in the sequences were performed: 1. methylation of all cytosines in the

particular ODN (met) 2. substitution with a sulfur atom instead of one of the nonbridging oxygen atoms in the phosphate backbone (pto).

2.3.1. CHS Gene Expression Studies 10 Days after ODN Incubation

The impact of the unmodified, and modified 1. met and 2. pto ODNs on the gene expression of both studied isoforms is presented in Figure 3. In order to determine whether the effect obtained by modified ODNs lasts longer, the gene expression was investigated in the material harvested after at least 10 days after the incubation with oligodeoxynucleotides. The value of *CHS* expression after ODN1 and ODN11 treatments was at a similar level (RQ ~2), regardless of the sequence modification, in comparison to the control. However, the ODN6 modified by methylation and thiophosphate triggered repression more prominently than unmodified ODN6, respectively by 41% and 63% in comparison to the control.

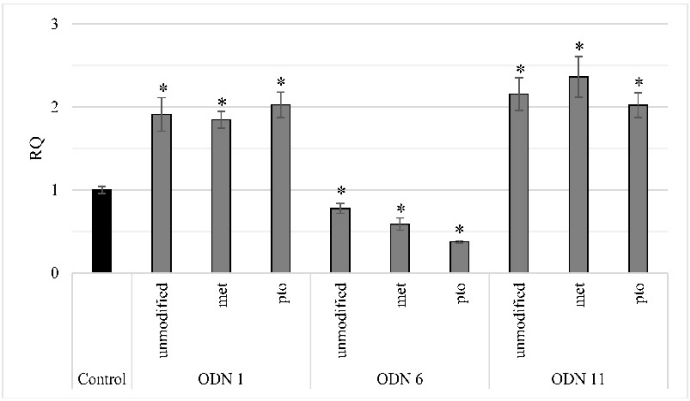


Figure 3. CHS gene expression in in vitro plants treated with modified ODNs after 10 days of incubation. The 4-week-old plants cultured in vitro were incubated with particular sequences of previously analyzed ODN 1, 6 and 11: unmodified ODN, ODN met (sequence with methylated cytosines) and ODN pto (sequence with thiophosphate bonds). The total expression of the two CHS genes *LuCHS6* and *LuCHS7* was determined by the real-time PCR reaction in plants harvested 10 days after incubation. The values are referred to the reference gene expression actin. The relative quantity (RQ) presents the transcript level in comparison to the control (set as 1, black). Data represent the mean value ± SD from at least three independent experiments. The significance of the differences between each mean and control was determined by Student's *t*-test. Asterisk indicates * $p < 0.05$.

2.3.2. Pattern of DNA Methylation in Variable CCGG Motifs of the CHS Gene after Treatment by Unmodified and Modified Oligos

In order to determine a specific pattern of the CHS gene methylation, crucial -CCGG- motifs in the CHS gene sequence were analyzed. As presented in the previous work, among the studied sites, the stable demethylated and highly variable in methylation -CCGG- were observed. Stable lack of cytosine methylation was observed in the following regions: 5'UTR (−232), non-coding (+217) and exon 2 (+1606), which were basically unchanged in comparison to the control. Variable sites were noted only in the coding region (+996, +1219, +1273). The figures show the percentage of lack of methylation, methylation of single inner cytosine and both cytosines. Since the stable sites do not differ in the cytosine methylation profile between studied plants, only results for variable CCGG sites are presented.

The analysis of the CCGG sites for unmodified and modified sequences of ODN11 is presented in Figure 3. Similarly as in Figure 4, the experiment was performed in the material harvested 10 days after the moment of incubation with oligodeoxynucleotides. After 10 days, the original sequence ODN11 did not lead to maintaining the changes in the methylation profile. The data presented by [9], shown in Figure 4D, indicate the potent increase in CmCGG methylation after 48 h in unmodified ODN11,

in comparison to the control. However, at 10 days after the flax incubation with the ODN11, the methylation profile became similar to that observed in the control. In case of ODN11 met and ODN11 pto, the perpetuation of initially induced cytosine methylation was observed (Figure 4). In ODN11 met, a potent increase of CmCGG was observed in all three variable sites +996, +1219, +1273, respectively by 19.3%, 23.0% and 13.8%, whereas for ODN11 pto in comparison to the control a higher percentage was noted in the site +1219, by 15.0% of single cytosine methylation and 2.01% of double cytosines methylation. An increase in CmCmGG methylation after ODN11 pto was also observed at the +996 site (by 8.1 %).

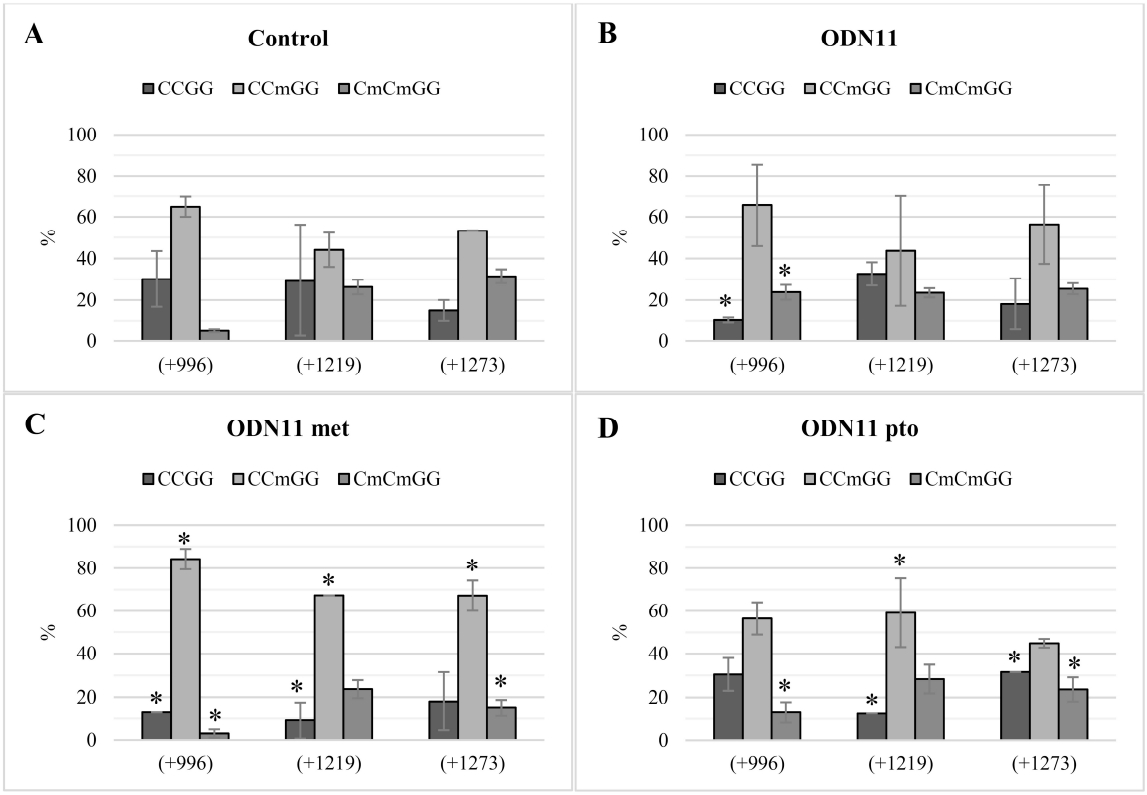


Figure 4. Methylation profiles for variable CCGG motifs in control plants (**A**). Plants treated with unmodified ODN11 (**B**) and modified ODN11 met (**C**) and ODN11 pto (**D**), 10 days after the treatment. The figure presents the percentage of the cytosine methylation in the variable CCGG sites (+996, +1219, +1273) located in the coding region of the CHS gene sequence. The genomic DNA was digested by restriction enzymes HpaII and MspI. The amount of non-digested DNA was determined by real-time PCR. The percentage of particular modification was presented for studied plants and control: CCGG—lack of methylation (dark grey), CCmGG—methylation of internal cytosine (light grey) and CmCmGG—methylation of both cytosines (medium grey). The site positions were presented according to the *LuCHS6* sequence, due to the presence of all CCGG sites. Data represent the mean value ± SD from at least three independent experiments. The significance of the differences between each mean and control was determined by Student's *t*-test. Asterisk indicates * $p < 0.05$.

2.3.3. Heredity of CHS Gene Expression in the F1 and F2 Generations of Modified ODN-Treated Flax Cultivated In Vitro

The flax plants treated with modified ODNs were cultivated in the experimental field. A part of the seeds obtained were sterilized and introduced to the in vitro culture. The stabilization of the modulation of CHS gene expression by modified oligodeoxynucleotides in the F1 generation of plants was investigated and the results are presented in Table 3. For ODN1 met, 4 out of 5 (80%) studied plants maintained the overexpression of *CHS* in comparison to the control (set as 1).

Table 3. Maintenance of gene expression modulation of CHS gene in F1 and F2 generations of plants treated with modified ODNs (met and pto).

Generation	Type of Plant	Percentage of Modulation Maintenance (%)
F1	ODN1 met	80
	ODN6 met	50
	ODN11 met	80
	ODN11 pto	100
F2	ODN6 met	50
	ODN11 met	38
	ODN11 pto	47

Regarding ODN6, the modification of oligonucleotide sequences via methylation made it possible to select one individual (out of five analyzed plants) with a repressed level of *CHS* expression (other four plants demonstrate similar too control level of *CHS* expression). For ODN11, both met and pto modifications were investigated in the next generation. Out of 7 analyzed plants, 6 presented transmission of the primarily induced modification in the chalcone synthase gene expression (data not presented).

The F2 generation plants obtained after treatment with modified ODNs were analyzed in order to assess maintenance of the induced changes in chalcone synthase gene expression. In comparison to the F1 generation, the number of analyzed plants was higher (25–30 individual plants for each ODN, except ODN6 with 6 plants). Despite the percentage of maintaining induced changes being lower, the results of F2 analysis were more significant.

2.4. Chromatin/DNA Accessibility Assay

For the Linola and genetically modified GM-CHS plants (W92.40), the chromatin/DNA accessibility assay was conducted. The chromatin isolated from in vitro cultured plants was digested with particular restriction enzymes: AatII and PvuI (particular sites for digestion were localized in the CHS gene, around the +1273 CCGG motif). Subsequently, DNA derived from the isolated chromatin was also incubated with restriction enzymes. The nucleic acid was purified and the effect of DNA digestion was assessed by real-time PCR at the methylation CCGG site (+1273). Figure 5 presents the relative quantity of the product obtained for the motif +1273 for undigested control, chromatin and DNA digested with restriction enzymes for Linola and GM-CHS plants. The results of this study have shown that between Linola and genetically modified plants, significant differences in the quantity of the product were obtained. For the GM-CHS flax with the stable methylation in the +1273 CCGG motif an increased level of the product was noted; thus, for transgenic plants, less accessibility of chromatin for restriction enzymes in comparison to the non-transgenic Linola was assumed. No relevant changes were observed in the +1273 motif for the DNA accessibility between the Linola and GM-CHS plants.

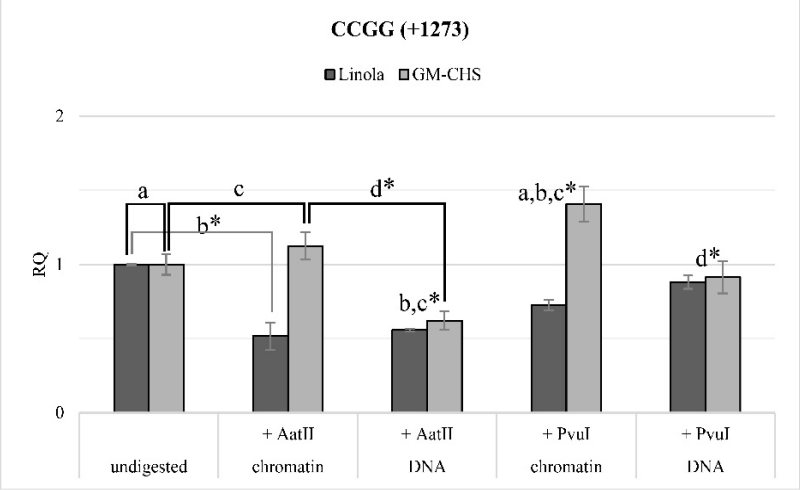


Figure 5. Chromatin/DNA accessibility assay. Chromatin from non-transgenic (Linola) and GM-CHS flax (W92.40) was isolated using a dedicated technical kit. The chromatin and chromatin-derived DNA (primarily treated with proteinase) were incubated with restriction enzymes AatII and PvuI (particular sites for digestion were localized in the CHS gene). After being cut, the DNA was purified and the effect of DNA digestion was assessed by real-time PCR at the methylation CCGG site (+1273). The values are referred to the reference gene expression actin. The relative quantity (RQ) presents the transcript level in comparison to the samples not incubated with the restriction enzymes (undigested, set as 1 for each line). Data represent the mean value ± SD from at least three independent experiments. The significance of the differences between the samples was determined by Student's *t*-test: a—the reference of the methylated motif to unmethylated, b—in reference to the undigested sample for unmethylated motif, c—in reference to the undigested sample for methylated motif, d—the reference of the DNA to chromatin, for the appropriate restriction enzyme. Asterisk indicates * $p < 0.05$.

3. Discussion

Genetic engineering methods allow precise modification of the phenotype by introducing changes into the sequence of the plant genome. A well-known and commonly used method of plant genetic modification is *Agrobacterium*-mediated plant transformation. An important aspect of this method is the possibility to acquire stable transformants. Unfortunately, the practical application of transgenic plants for industrial purposes is limited in most European countries, due to restrictive legal regulations. The main reason is the lack of trust of potential GM plant consumers in relation to the harmless impact on people and animals. Another issue would be the risk of the disturbance to the natural environment by transgenic crops. The lack of public acceptance of GM crops encourages scientists to search for new methods that enable induction of organism variability without the need to change the genome [22].

Recently, a lot of attention has been devoted to methods based on the use of site-specific nucleases (SSNs), including TALENs, ZFNs and CRISPR/Cas9 [23]. Although the principle of SSN technology is based only on the editing of endogenous genes, the legal regulations regarding this method are not entirely resolved. Some genome editions can be considered as mutations and thus might be controlled by legal restrictions that apply to GMO [24]. Methods that will allow the use of the natural cell genetic repertoire are eagerly in demand. Epigenetic changes are modifications that take place in nature, mainly during the plant adjustment to constantly changing environmental conditions. In addition, the epigenetic DNA pattern can be maintained during cell division and be inherited [7,25]. Therefore, epigenetic modifications have a large potential for application in the phenotype variability of plants since they are naturally occurring changes.

Plants are considered as the best model for epigenetic regulation research. All the epigenetic mechanisms described so far occur in plants. In the plant genome, methylation in the 5th position of the cytosine in a DNA chain is a significant epigenetic modification. In plants, 5-methylcytosine can apply

to three nucleotide contexts: CpG, CpHpG and CpHpH. The equilibrium of the methylation pattern is obtained through the interaction of de novo methylation processes, maintenance of methylation and demethylation. Recently, the methylation of adenine in the N6 position (N6-mA) has also been demonstrated in the genome of *Arabidopsis thaliana*. It has been shown that this modification is essential for plant development and occurs in actively expressed genes [26]. DNA methylation plays an important role in many processes in plants, such as vegetative development, reproduction, fertilization or gametogenesis. Additionally, DNA methylation specifically interacts with other epigenetic mechanisms: posttranslational histone modifications and processes related to non-coding RNA molecules. Due to the constant need to adapt to changing environmental conditions, the "plasticity" of the plant epigenome enables phenotypic traits to be adjusted by specifically regulating gene expression [6].

The use of short oligodeoxynucleotides (ODNs) proved to be a precise method of inducing changes at the epigenomic level. The ODN technology is based on the introduction of short oligodeoxynucleotides of size 12 to 25 nucleotides into cells, which are complementary to the corresponding regions of the target gene. The technology has proven to be an effective method of studying gene function and transformation in animal cells. The method has gained popularity due to the fact that it may impact the modulation of the gene expression of a particular gene and does not lead to modification of the genomic sequence. Currently, ODN technology is optimized to be used in plant research. The favorable property of this method for plant research is the possibility of direct introduction of ODNs to plant cells and limitation of the pleiotropic effects. Although the mechanism of ODN action has not been fully explained yet, it was suggested that the following processes are involved: RNA interference, RNA activation and RNA-dependent DNA methylation [8,9,27]. The majority of reports indicate that the silencing of gene expression is probably the result of the degradation of RNA:DNA duplexes by RNase H [28]. The formation of a triple complex consisting of a DNA double helix and ODN sequence is also not excluded; hence the ability to form a transcript may be limited [29].

The main purpose of our research was to induce significant changes in the transcript level of the CHS gene, which would be maintained after cell divisions. Previously [9], we indicated that the ODN technology is effective in modulation of CHS gene expression and DNA methylation. The changes triggered by the new technology were similar to those observed initially in genetically modified flax (GM-CHS) [9].

Modified ODNs proved to be effective in prolonging the time of maintaining induced changes in methylation of cytosines in the flax genome. Among the two used modifications, methylation of cytosines present in the ODN sequence proved to be more effective than the unmodified version. Analysis of subsequent generations of plants treated with ODNs confirmed the possibility of maintaining induced epigenetic changes during cell division [5]. Due to the declining number of F2 generation plants that maintained the changes induced by unmodified ODNs, F1 generation plants treated with modified oligonucleotides were obtained. The changes induced through ODN6 were maintained in the F1 and F2 generations with the lowest percentage of maintenance. In plants growing in vitro after treatment with ODN6, the tendency of primarily induced changes remains despite subcultures (Figure S1), which confirms the transmission of epigenetic changes during vegetative reproduction [30]. However, at the same time, we cannot exclude that the elevated level of genomic methylation, and hence the reduced expression of the CHS gene, might be triggered by repeated in vitro sub-culturing of plants [31]. More importantly, the modification of ODN 6 through cytosines' methylation led to improvement of the perpetuation of changes in subsequent generations.

Analysis of methylation of CCGG motifs in the F1 generation of plants treated with ODNs showed the maintenance of the methylation pattern in two of the three analyzed motifs. Instability of methylation status observed in motif +1219/+775 concerned not only plants treated with ODNs, but also control plants. Thus, we presume that particular CCGG motifs may be more susceptible to methylation/demethylation driven by environmental factors, just as the cytosines localized in the transcription factor binding region can be more susceptible to methylation [32]. Despite the fact that in

the F1 plants treated with ODN6, the repression of CHS gene was not observed, the methylation pattern of these plants remained after cell divisions. The possible reason for not keeping the CHS expression decreased after the action of ODN6 might be the compensation for effects driven by environmental conditions, and hence increased expression of the CHS gene was observed.

In order to confirm the specific nature of the ODN action, we also determined whether modulation of CHS gene expression displays any changes to the transcript level of the other genes encoding enzymes related to the phenylpropanoid pathway (Figure S2). In plants treated with ODN1, there were no significant changes in the expression of analyzed genes other than *CHS* involved in the synthesis of phenylpropanoids. Only *PAL* and *HCT* showed a minor reduction in gene expression. In the *Arabidopsis thaliana* model plant, HCT gene repression led to a decrease in lignin levels, which is associated with increased CHS gene activity and redirection of metabolic changes to the flavonoid compound synthesis pathway [33]. However, in the flax with CHS gene repression, decreased levels of lignin and HCT gene repression were observed, which did not affect the flavonoid content [19]. Regarding the flax treated with oligodeoxynucleotides modified by methylation, a positive relation was also observed between expression of genes CHS and HCT, where an elevated transcript level of CHS gene was accompanied by increased HCT gene expression. Despite the prolonged maintenance of changes in DNA methylation induced by ODN with thiophosphate bonding, pleiotropic effects concerning the genes related to the synthesis of flavonoids and lignin were observed. Analyzed genes displayed elevated transcript levels. According to a research by [19], enzymes involved in the biosynthesis of flavonoid compounds and lignin occur in the form of multienzyme complexes in which the main regulating component is possibly CHS [19]. Therefore, modulation of CHS gene expression may lead to control of the expression of other genes encoding the components of these complexes. Moreover, in potato transgenic plants overexpressing the CHS gene, effects on other metabolic pathways were also observed, which may be caused by competition for substrates or the effect of the synthesized flavonoids on plant metabolism [34]. However, in plants treated with ODN1 met, no changes in the expression of other genes encoding enzymes involved in the synthesis of phenylpropanoids (Supplemental Figure S2) were observed, which may suggest that this modification not only stabilizes the modulation of target gene expression, but also increases the specificity of ODN. The group of Khorova have proven that methylation of oligonucleotides increases nucleic acid resistance to nuclease degradation and increases binding strength when interacting with the complementary sequence [35].

DNA methylation is known to cooperate with the positioning of nucleosomes to regulate the structure of chromatin, which leads to an appropriate modulation of gene expression [36]. We have already presented the analysis of the nucleosome energy diagram, which showed that methylation of variable motifs in the CpG island region results in the formation of less stable nucleosomes with higher energy and their shift towards the 3′ end of the DNA [9]. In the vicinity of one of the methylated motifs, sites recognized by AatII and PvuI endonucleases were identified; hence experimental verification of the is silico analysis was performed. In the chromatin structure, DNA regions bound to the histone proteins are less available for endonucleases [37]. Thus, DNA present in nucleosomes is protected from degradation by restriction enzymes. Analysis of chromatin digestion and DNA showed that there is probably a change in the positioning of nucleosomes as a result of the increase in the level of methylation in CCGG +1273 sites.

The present results confirmed that ODN technology induces changes in the plant genome that can be inherited. These data contribute to the statement that the use of natural genetic repertoire through epigenetic variability can be a valuable alternative to improve crop quality.

4. Materials and Methods

4.1. Designing Short Oligodeoxynucleotide Sequences (ODN) Homologous to CHS Gene Regions

On the basis of our previous work [9], for this study, three ODNs homologues to the particular CHS gene regions were selected: ODN1 (5′UTR), ODN6 (intron) and ODN11 (3′UTR). The ODN

sequences were designed using Mfold software (version 3.2, Genetics Computer Group, Madison, WI, USA) in the antisense orientation and correspond simultaneously to two CHS gene isoforms: LuCHS6 (Phytozome data base: Lus10033717) and LuCHS7 (Phytozome database: Lus10031622). Detailed information about used ODN sequences were reported by [9] (Table 2, Supplementary Table S1). In order to enhance the resistance of the sequences to the degradation by nucleases, the sequences of the three mentioned oligonucleotides were modified by methylation (all cytosines occurred in the sequences) and thiophosphate modification (at each phosphodiester bond). The oligodeoxynucleotides were synthesized by Genomed S.A. (Warszawa, Poland).

4.2. Plant Material

Flax seeds (*Linum usitatissimum* L., cv. Linola) were obtained from the Flax and Hemp Collection of the Institute of Natural Fibers (Poznań, Poland).

Flax Growing Conditions for ODN Technology

To investigate whether modified ODNs are more effective in the stability of the induced epigenetic modulation, flax plants were cultured in in vitro conditions in the phytotron at 16 h light (22 °C), 8 h darkness (16 °C). The seeds were first sterilized for 15 min with 50% PPM—a broad-spectrum biocide/fungicide for plant tissue culture (Plant Preservative Mixture; Plant Cell Technology, Washington, DC, USA) and then germinated on Murashige and Skoog (MS) basal medium (Sigma-Aldrich, St. Louis, MO, USA), supplemented with 2% sucrose, pH = 5.8, solidified with 0.8% agar on Petri dishes. Mature flax plants were cultured on the MS basal medium, supplemented with 2% sucrose, pH = 5.8, solidified with 0.8% agar in sterile glass jars. In order to minimize plant infection by pathogens, the medium was complemented with 0.1% PPM™ (Plant Preservative Mixture) Plant Cell Technology (Washington, USA).

In order to obtain seeds for raising the next generation of plants in the field, *Linola* flax seeds were germinated in the soil in the phytotron conditions in 16 h light (22 °C), 8 h darkness (16 °C), before they were treated with ODNs.

4.3. Identification of cDNA Sequences

The isolated mRNA from Linola flax was submitted to sequencing. Necessary sample preparations, sequencing and data processing were performed by Genomed S.A. (Warszawa, Poland). Data from three biological replicates were analyzed.

4.4. Flax Treatment with ODNs by Infiltration

Four-week-old Linola flax cultured in in vitro conditions were used for the treatment with modified ODNs. The plants were cut above their roots and submerged in water solution of the particular ODN in the 10 μM concentration. The incubation was carried out in the vacuum chamber for 20 min. After the treatment, the plants were put into an MS medium. The material for the analysis was harvested 10 days after infiltration. The harvesting time was indicated by previous experiments [8] and enabled determination of the ODNs' activity in comparison to the unmodified sequences.

For the field cultivation, whole two-week-old seedlings germinated in the soil were treated with the ODNs by the forced osmosis method. It was described in [38] that the sugar starvation method of introducing ODNs is effective in the induction of changes in DNA methylation. First, the plants were transferred from the soil into the water and kept in the darkness in the phytotron chamber in order to deplete endogenous sucrose. Then seedlings were put in the water solution of 100 mM sucrose in the presence of 10 μM antisense ODNs. After 24 h of incubation in the darkness with particular oligodeoxynucleotides, plants were put into the soil and cultivated in the experimentation field. The matured seed capsules were harvested approximately four months later. Collected seeds were stored in dry, cool conditions. In order to determine the stability of induced modifications in the subsequent generations, some of the seeds were sterilized and cultured in in vitro conditions.

4.5. Flax cultivation in the Field

In general, the growing season for flax in a moderate climate is between April and July. Seeds collected from plants treated with ODNs were sown in the subsequent season to obtain an F1 generation of plants. For an F2 generation, plants were cultivated in the analogous manner.

4.6. Gene Expression Analysis

The expression of investigated genes was analyzed via real-time PCR. The total RNA was isolated from freeze-ground green tissue of the transgenic plants. Isolation was performed using the Trizol method (Invitrogen, Carlsbad, CA, USA), following the protocol of the manufacturer. The isolated RNA was deprived of DNA contamination by DNase I (Invitrogen, Carlsbad, USA). The purified RNA was used as a template for cDNA synthesis using reverse transcriptase—High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA).

Real-time PCR reactions were performed using a DyNAmo SYBR Green qPCR kit (Thermo Scientific, Waltham, MA, USA). Primers used for the reaction are presented in the Supplementary data Table S1. The used system was the Applied Biosystems StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, USA). The reaction was conducted according to the protocol provided by the manufacturer. Each reaction was performed in three repeats. As the reference, the actin gene was used. The results were presented as the relative quantification (RQ) to the reference gene.

4.7. DNA Methylation Patterns in Specific Regions of Chalcone Synthase Gene

The methylation patterns of the chalcone synthase gene sequence were established in the control and ODN-treated plants. The DNA was incubated with restriction enzymes MspI and HpaII for at least 3 h (New England Biolabs, Ipswich, MA, USA); it differs in sensitivity to cytosine methylation. The genomic DNA digested by the restriction enzymes and undigested DNA were used as a templates for the real-time PCR reaction. The reaction was performed similar to the gene expression analysis. The primers for the reaction (Supplementary data Table S1) were designed for specific sites of methylation predicted by the NEB cutter V2.0. In the chalcone synthase gene, six CCGG islands were analyzed. The sites of analyzed -CCGG- motifs were indicated by their positions towards the ATG site (+1) as follows: 5'UTR (site −232), non-coding region (+217) and coding region (+996, +1219, +1273, +1606).

The quantity of DNA measured by the real-time PCR was calculated in order to estimate the methylation of cytosines. The "CCGG" non-methylated DNA was calculated as the difference between undigested DNA and DNA incubated with HpaII. The single methylated cytosine "CCmGG" was estimated as the difference between samples digested by HpaII and digested by MspI. The level of the "CmCmGG" was equal to the DNA incubated with MspI. The values were presented as a percentage referring to the undigested DNA, set as 100%.

4.8. Relative Accessibility of Chromatin and DNA

The chromatin was isolated from 1 g of freeze-ground green tissue by ChromaFlash Plant Chromatin Extraction Kit (EpiGentek, Farmingdale, NY, USA) according to a protocol provided by the manufacturer. The sonication stage was set for the QSonica 700 (QSonica, Newtown, CT, USA) instrument and performed with the following parameters: Amplitude–25%, Process time–3 min 20 s, Pulse ON–20 s, Pulse OFF–30 s. Isolated chromatin was divided into portions (approximately 300–400 ng per aliquot), snap-frozen in liquid nitrogen, and stored at −80 °C.

Digestion of chromatin and DNA was performed. First, the chromatin was digested with the following FastDigest enzymes: AatII and PvuI (Thermo Scientific, Waltham, MA, USA). The restriction enzymes were defined by in silico analysis. Sites of digestion for AatII and PvuI are located in the nucleosomes, where DNA wraps around nucleosomes. After digestion with restriction enzymes (15 min, 37 °C) and enzymes' inactivation (AatII, PvuI –80 °C, 5 min), the samples were incubated with Proteinase K (Qiagen, Hilden, Germany) and RNase A (Qiagen, Hilden, Germany) at 37 °C for 1 h.

Simultaneously, chromatin samples were incubated first with Proteinase K and RNase A, in order to release the DNA from the nucleosome. The DNA was digested with restriction enzymes in an analogous manner as chromatin was cut. Undigested control samples were also prepared. Total DNA was purified from each sample with phenol/chloroform/isoamyl alcohol (PCI-25:24:1, *v*/*v*), mixed vigorously and centrifuged (10,000× *g*, for 5 min, at room temperature (RT)). The upper phase was transferred into a new tube and DNA was precipitated by adding 0.5 mL of cold isopropanol. In order to facilitate precipitation, 2 µL (20 µg) of linear polyacrylamide was added (5 mg of acrylamide was dissolved in 200 µL of water, 1 µL of 10% APS and 1 µL of TEMED were added, left to polymerase overnight at RT, 2.5 volumes of ethanol were added, centrifuged at 12,000× *g*, 5 min, dried and re-suspended in 500 µL of sterile water). After 10 min of incubation on ice, the samples were centrifuged 12,000× *g*, for 15 min, at 4°C. The DNA pellet was washed with 0.5 mL of cold 70% (*v*/*v*) ethanol and centrifuged at 12,000× *g* for 10 min at 4 °C. The dry pellet was re-suspended in 40 µL of sterile water.

The relative accessibility of chromatin/DNA was assessed by the real-time PCR method. During the reaction, the region surrounding motif CCGG located at +1273 was amplified. The selection of this region was preferred due to the methylation status of the particular motif and the location of the sites of cutting by AatII and PvuI restriction enzymes. The real-time PCR reaction was performed similarly to the gene expression analysis. The quantity of measured DNA was calculated in order to estimate the relative accessibility of chromatin/DNA to the restriction cut. The actin gene was used as the reference. The results were presented as the relative chromatin/DNA accessibility in comparison to the undigested control, set as 1.

4.9. In Silico Analysis

Chromosome Localization of CHS Genes from Flax

The genomic sequence of each CHS gene was aligned in BLASTn (Available online: https://blast.ncbi.nlm.nih.gov/Blast.cgi) to whole genome sequencing of *Linum usitatissimum* [39]. The localization (range) of exons in the chromosome was identified for the particular gene.

4.10. Statistical Analysis

Presented experiments were performed at least three times, independently (three technical repeats of each of three biological samples). Data represent the mean value ± standard deviation (SD) from at least three independent experiments. The significance of the differences between each mean and control was determined by Student's t-test. An asterisk above each bar indicates $p < 0.05$. The statistical analysis was performed using STATISTICA (ver. 12, StatSoft Inc.—Dell, Round Rock, TX, USA).

5. Conclusions

In this article, we presented the latest research regarding ODN technology in flax, considering two genes: *LuCHS6* and *LuCHS7*. The main purpose was to demonstrate the potential maintenance of induced epigenetic changes in CHS gene expression and methylation via short oligodeoxynucleotides complementary to the gene of interest. Induced changes in the methylation status (5-methylcytosine) and CHS gene expression can be perpetuated in the next generation. Modified ODNs (especially, by methylation) were assumed to be successful in prolonging the ODN-triggered effect. Although the mechanism of ODNs' action and heredity of ODN-induced changes are not fully explained yet, the processes involving the sRNAs (small non-coding RNAs) and RdDM (RNA-directed DNA methylation) along with the rearrangement of nucleosome positioning are suspected to play a substantial role in defining the epigenome [8,9]. Thus, we suggested the great potential of ODN technology in plant research as a significant tool not only for genetic and epigenetic studies, but also crop improvement.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/16/3983/s1.

Author Contributions: M.D. and M.Z. conceived and designed the experiments and analyzed the data; M.D. performed the experiments and wrote the manuscript; J.S. and M.Z. supervised the work and revised the manuscript; A.H. performed the bioinformatic in silico analysis. All authors reviewed and approved the manuscript.

Funding: The research was supported by grants no. 2016/23/N/NZ9/01619 and 2013/11/B/NZ9/00150 from the National Science Centre (Available online: http://www.ncn.gov.pl).

Acknowledgments: We would like to give special thanks to Aleksandra Boba for sharing the primers sequences of same previously identified flax genes analyzed in this work.

Conflicts of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Schmid, M.W.; Heichinger, C.; Coman Schmid, D.; Guthörl, D.; Gagliardini, V.; Bruggmann, R.; Aluri, S.; Aquino, C.; Schmid, B.; Turnbull, L.A.; et al. Contribution of epigenetic variation to adaptation in Arabidopsis. *Nat. Commun.* **2018**, *9*, 4446. [CrossRef] [PubMed]
2. Robertson, A.; Wolf, D. The role of epigenetics in plant adaptation. *Trends Evol. Biol.* **2012**, *4*, 19–25. [CrossRef]
3. Sahu, P.P.; Pandey, G.; Sharma, N.; Puranik, S.; Muthamilarasan, M.; Prasad, M. Epigenetic mechanisms of plant stress responses and adaptation. *Plant. Cell Rep.* **2013**, *32*, 1151–1159. [CrossRef] [PubMed]
4. Gallusci, P.; Dai, Z.; Genard, M.; Gauffretau, A.; Leblanc-Fournier, N.; Richard-Molard, C.; Vile, D.; Brunel-Muguet, S. Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. *Trends Plant. Sci.* **2017**, *22*, 610–623. [CrossRef] [PubMed]
5. Kumar, S. Epigenomics of Plant Responses to Environmental Stress. *Epigenomes* **2018**, *2*, 6. [CrossRef]
6. Pikaard, C.S.; Scheid, O.M. Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, 1–31. [CrossRef]
7. Bond, D.M.; Baulcombe, D.C. Small RNAs and heritable epigenetic variation in plants. *Trends Cell Biol.* **2014**, *24*, 100–107. [CrossRef]
8. Wojtasik, W.; Kulma, A.; Boba, A.; Szopa, J. Oligonucleotide treatment causes flax beta-glucanase up-regulation via changes in gene-body methylation. *BMC Plant. Biol.* **2014**, *14*, 1–15. [CrossRef]
9. Dzialo, M.; Szopa, J.; Czuj, T.; Zuk, M. Oligodeoxynucleotides Can Transiently Up- and Downregulate CHS Gene Expression in Flax by Changing DNA Methylation in a Sequence-Specific Manner. *Front. Plant. Sci.* **2017**, *8*, 755. [CrossRef]
10. Bhardwaj, K.; Verma, N.; Trivedi, R.; Bhardwaj, S.; Shukla, N. Significance of ratio of omega-3 and omega-6 in human health with special reference to flaxseed oil. *Int. J. Biol. Chem.* **2016**, *10*, 1–6. [CrossRef]
11. Zuk, M.; Richter, D.; Matuła, J.; Szopa, J. Linseed, the multipurpose plant. *Ind. Crop. Prod.* **2015**, *75*, 165–177. [CrossRef]
12. Kyselka, J.; Rabiej, D.; Dragoun, M.; Kreps, F.; Burčová, Z.; Němečková, I.; Smolová, J.; Bjelková, M.; Szydłowska-Czerniak, A.; Schmidt, Š. Antioxidant and antimicrobial activity of linseed lignans and phenolic acids. *Eur. Food Res. Technol.* **2017**, *243*, 1633–1644. [CrossRef]
13. Kristensen, M.; Jensen, M.G.; Aarestrup, J.; Petersen, K.E.; Søndergaard, L.; Mikkelsen, M.S.; Astrup, A. Flaxseed dietary fibers lower cholesterol and increase fecal fat excretion, but magnitude of effect depend on food type. *Nutr. Metab.* **2012**, *9*, 1–8. [CrossRef] [PubMed]
14. Han, H.; Yılmaz, H.; Gülçin, İ. Antioxidant Activity of Flaxseed (Linum usitatissimum L.) shell and Analysis of Its Polyphenol Contents by LC-MS/MS. *Rec. Nat. Prod.* **2018**, *12*, 397–402. [CrossRef]
15. Zuk, M.; Dorotkiewicz-Jach, A.; Drulis-Kawa, Z.; Arendt, M.; Kulma, A.; Szopa, J. Bactericidal activities of GM flax seedcake extract on pathogenic bacteria clinical strains. *BMC Biotechnol.* **2014**, *14*, 70. [CrossRef] [PubMed]
16. Veeramani, C.; Alsaif, M.A.; Al-Numair, K.S. Herbacetin, a flaxseed flavonoid, ameliorates high percent dietary fat induced insulin resistance and lipid accumulation through the regulation of hepatic lipid metabolizing and lipid-regulating enzymes. *Chem. Biol. Interact.* **2018**, *288*, 49–56. [CrossRef] [PubMed]

17. Czemplik, M.; Mierziak, J.; Szopa, J.; Kulma, A. Flavonoid C-glucosides Derived from Flax Straw Extracts Reduce Human Breast Cancer Cell Growth In vitro and Induce Apoptosis. *Front. Pharm.* **2016**, *7*, 282. [CrossRef]
18. Alam, M.N.; Almoyad, M.; Huq, F. Polyphenols in Colorectal Cancer: Current State of Knowledge including Clinical Trials and Molecular Mechanism of Action. *Biomed. Res. Int.* **2018**, *2018*, 1–29. [CrossRef]
19. Zuk, M.; Działo, M.; Richter, D.; Dymińska, L.; Matuła, J.; Kotecki, A.; Hanuza, J.; Szopa, J. Chalcone Synthase (CHS) Gene Suppression in Flax Leads to Changes in Wall Synthesis and Sensing Genes, Cell Wall Chemistry and Stem Morphology Parameters. *Front. Plant. Sci.* **2016**, *7*, 1–14. [CrossRef]
20. Eom, S.H.; Hyun, T.K. Genome-wide identification and transcriptional expression analysis of chalcone synthase in flax (Linum usitatissimum L.). *Gene. Rep.* **2016**, *5*, 51–56. [CrossRef]
21. Boba, A.; Kostyn, K.; Kostyn, A.; Wojtasik, W.; Dziadas, M.; Preisner, M.; Szopa, J.; Kulma, A. Methyl Salicylate Level Increase in Flax after Fusarium oxysporum Infection Is Associated with Phenylpropanoid Pathway Activation. *Front. Plant. Sci.* **2017**, *7*, 1–12. [CrossRef] [PubMed]
22. Kamthan, A.; Chaudhuri, A.; Kamthan, M.; Datta, A. Genetically modified (GM) crops: milestones and new advances in crop improvement. *Appl. Genet.* **2016**, *129*, 1639–1655. [CrossRef] [PubMed]
23. Gupta, R.M.; Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Investig.* **2014**, *124*, 4154–4161. [CrossRef] [PubMed]
24. Callaway, E. CRISPR plants now subject to tough GM laws in European Union. *Nature* **2018**, *560*, 16. [CrossRef] [PubMed]
25. Zhang, C.; Hsieh, T.-F. Heritable epigenetic variation and its potential applications for crop improvement. *Plant. Breed. Biotechnol.* **2013**, *1*, 307–319. [CrossRef]
26. Liang, Z.; Shen, L.; Cui, X.; Bao, S.; Geng, Y.; Yu, G.; Liang, F.; Xie, S.; Lu, T.; Gu, X.; et al. DNA N(6)-Adenine Methylation in Arabidopsis thaliana. *Dev. Cell* **2018**, *45*, 406–416. [CrossRef] [PubMed]
27. Dinç, E.; Tóth, S.Z.; Schansker, G.; Ayaydin, F.; Kovács, L.; Dudits, D.; Garab, G.; Bottka, S. Synthetic antisense oligodeoxynucleotides to transiently suppress different nucleus- and chloroplast-encoded proteins of higher plant chloroplasts. *Plant. Physiol.* **2011**, *157*, 1628–1641. [CrossRef]
28. Giles, R.V.; Tidd, D.M. Increased specificity for antisense oligodeoxynucleotide targeting of RNA cleavage by RNase H using chimeric methylphosphonodiester/phosphodiester structures. *Nucleic Acids Res.* **1992**, *20*, 763–770. [CrossRef]
29. Joseph, J.; Kandala, J.C.; Veerapanane, D.; Weber, K.T.; Guntaka, R.V. Antiparallel polypurine phosphorothioate oligonucleotides form stable triplexes with the rat alpha1(I) collagen gene promoter and inhibit transcription in cultured rat fibroblasts. *Nucleic Acids Res.* **1997**, *25*, 2182–2188. [CrossRef]
30. Kinoshita, T.; Seki, M. Epigenetic memory for stress response and adaptation in plants. *Plant. Cell Physiol.* **2014**, *55*, 1859–1863. [CrossRef]
31. Goyali, J.C.; Igamberdiev, A.U.; Debnath, S.C. DNA methylation in lowbush blueberry (Vaccinium angustifolium Ait.) propagated by softwood cutting and tissue culture. *Can. J. Plant. Sci.* **2018**, *98*, 1035–1044. [CrossRef]
32. Medvedeva, Y.A.; Khamis, A.M.; Kulakovskiy, I.V.; Ba-Alawi, W.; Bhuyan, M.S.I.; Kawaji, H.; Lassmann, T.; Harbers, M.; Forrest, A.R.; Bajic, V.B. Effects of cytosine methylation on transcription factor binding sites. *BMC Genom.* **2014**, *15*, 119. [CrossRef] [PubMed]
33. Besseau, S.; Hoffmann, L.; Geoffroy, P.; Lapierre, C.; Pollet, B.; Legrand, M. Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. *Plant. Cell* **2007**, *19*, 148–162. [CrossRef] [PubMed]
34. Łukaszewicz, M.; Szopa, J. Pleiotropic effect of flavonoid biosynthesis manipulation in transgenic potato plants. *Acta Physiol. Plant.* **2005**, *27*, 221–228. [CrossRef]
35. Khvorova, A.; Watts, J.K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **2017**, *35*, 238. [CrossRef] [PubMed]
36. Jones, P.A. Genome wide analysis of DNA methylation and nucleosome positioning. *Epigenetics Chromatin* **2013**, *6*, O37. [CrossRef]
37. Trotter, K.W.; Archer, T.K. Assaying chromatin structure and remodeling by restriction enzyme accessibility. *Meth. Mol. Biol.* **2012**, *833*, 89–102. [CrossRef]

38. Sun, C.; Hoglund, A.S.; Olsson, H.; Mangelsen, E.; Jansson, C. Antisense oligodeoxynucleotide inhibition as a potent strategy in plant biology: identification of SUSIBA2 as a transcriptional activator in plant sugar signalling. *Plant. J. Cell Mol. Biol.* **2005**, *44*, 128–138. [CrossRef] [PubMed]
39. Wang, Z.; Hobson, N.; Galindo, L.; Zhu, S.; Shi, D.; McDill, J.; Yang, L.; Hawkins, S.; Neutelings, G.; Datla, R.; et al. The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. *Plant. J. Cell Mol. Biol.* **2012**, *72*, 461–473. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Agrobacterium-Mediated Gene Transient Overexpression and *Tobacco Rattle Virus* (TRV)-Based Gene Silencing in Cassava

Hongqiu Zeng [1], Yanwei Xie [1], Guoyin Liu [1], Yunxie Wei [1], Wei Hu [2,*] and Haitao Shi [1,*]

[1] Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou 570228, China
[2] Key Laboratory of Biology and Genetic Resources of Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Xueyuan Road 4, Haikou 571101, China
* Correspondence: huwei2013@itbb.org.cn (W.H.); haitaoshi@hainanu.edu.cn (H.S.); Tel.: +86-898-66160721 (W.H. & H.S.)

Received: 17 July 2019; Accepted: 13 August 2019; Published: 15 August 2019

Abstract: *Agrobacterium*-mediated transient expression and virus-induced gene silencing (VIGS) are very useful in functional genomics in plants. However, whether these methods are effective in cassava (*Manihot esculenta*), one of the most important tropical crops, remains elusive. In this study, we used *green fluorescent protein* (*GFP*) and *β-glucuronidase* (*GUS*) as reporter genes in a transient expression assay. *GFP* or *GUS* could be detected in the infiltrated leaves at 2 days postinfiltration (dpi) and were evidenced by visual *GFP* and *GUS* assays, reverse-transcription PCR, and Western blot. In addition, *phytoene desaturase* (*PDS*) was used to show the silencing effect in a VIGS system. Both *Agrobacterium* GV3101 and AGL-1 with *tobacco rattle virus* (TRV)-*MePDS*-infiltrated distal leaves showed an albino phenotype at 20 dpi; in particular, the AGL-1-infiltrated plants showed an obvious albino area in the most distal leaves. Moreover, the silencing effect was validated by molecular identification. Notably, compared with the obvious cassava mosaic disease symptom infiltrated by *African-cassava-mosaic-virus*-based VIGS systems in previous studies, TRV-based VIGS-system-infiltrated cassava plants did not show obvious virus-induced disease symptoms, suggesting a significant advantage. Taken together, these methods could promote functional genomics in cassava.

Keywords: *agrobacterium*; transient expression; virus-induced gene silencing (VIGS); *tobacco rattle virus* (TRV); cassava (*Manihot esculenta*)

1. Introduction

Aiming at analyzing genomic sequences and functions, functional genomics has rapidly developed as a result of sequencing projects of different species, especially of important crops. Cassava (*Manihot esculenta*), a kind of tuber crop, is widely cultivated in the tropics and some subtropical areas. Because of its low planting cost and high efficiency, it has become one of the most important industrial crops in the world [1]. Although cassava sequencing and resequencing were accomplished years ago, cassava functional genomics has developed slowly due to the time-consuming nature and low transformation rate of obtaining stable transgenic cassava plants [2]. Therefore, it is necessary to find another way to promote rapid gene function studies of cassava.

Agrobacterium-mediated transient overexpression is widely used in gene function studies in plants [3]. The most important advantages of this method are its rapid and simple process and that it does not require complex equipment, unlike other overexpression assays [4]. The simple protocol is accomplished by creating an overexpression vector, transforming it into *Agrobacterium*, and infiltrating

the *Agrobacterium* into plant culture [5]. *Green fluorescent protein* (*GFP*) and *β-glucuronidase* (*GUS*) are two convenient reporter genes in biology which can be observed by fluorescent microscopy or direct staining [6,7]. Normally, they can continue to express, reaching their peak at 3 days postinfiltration (dpi), and exert obvious expression for at least one week [4,8]. In addition, they can be efficiently expressed in a wide range of plant species. So far, transient overexpression assays have been successfully used in many species, such as *Arabidopsis* spp. [4], tobacco [5], *Mimulus lewisii* [6], *Piper colubrinum* [7], rose [9], soybean [10], *Theobroma cacao* L. [11], cotton [12], *Brassica juncea* [13], potato [14], tomato [15], and lettuce [15]. However, its effectiveness in cassava requires further investigation.

Agrobacterium-mediated virus-induced gene silencing (VIGS) is a widely used, efficient technique in gene function studies [16]. VIGS is mostly based on RNA viruses, such as *tobacco mosaic virus* (TMV), *potato virus X* (PVX), *barley stripe mosaic virus* (BSMV), *cucumber mosaic virus* (CMV), and *tobacco rattle virus* (TRV) [17–23]. Among these viruses, TRV is widely used because of its high silencing effect, long silencing duration, wide range of infiltration host species, and the mild virus-induced disease symptoms [24,25]. Presently, TRV has been used in *Arabidopsis* [26], tobacco [27], strawberry [27,28], cotton [29], piper [30], wheat [31], maize [31], and so on. For the silencing effect, most of the abovementioned infiltrated species show the phenotype at two weeks postinfiltration and the most obvious changes around the third week [27,29,32].

In this study, *Agrobacterium*-mediated transient overexpression and VIGS were established in cassava. The phenotype and examination of these assays illustrated the feasibility of the protocols, which should help promote the rapid analysis of functional genomics in cassava.

2. Results

2.1. The Effect of the Transient Expression Assay in Cassava Leaves

To investigate the effect of *GFP* transient expression, the *35S::GFP* vector [33] was transformed into *Agrobacterium* AGL-1 and GV3101, and the transformed strains were infiltrated into cassava leaves, respectively (Figure 1). At 2 dpi, the infiltrated leaves were harvested and examined by confocal laser-scanning microscopy. *Agrobacterium* with *35S::GFP*-infiltrated leaves showed clear green fluorescence in both the cytoplasm and nucleus, while the empty *Agrobacterium*-infiltrated leaves showed no fluorescence (Figure 2A). Reverse-transcription PCR was performed to clarify the effect of transient expression. *Agrobacterium* with the *35S::GFP*-infiltrated sample showed a bright band by PCR, indicating that *GFP* was expressed in the cassava leaves (Figure 2B). Moreover, *GFP* was also detected by Western blot, which was consistent with confocal observation (Figure 2C).

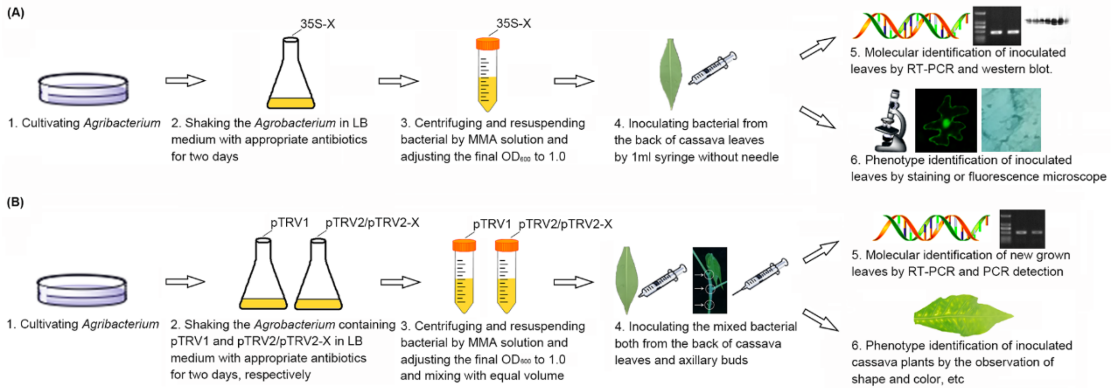


Figure 1. Schematic views of *Agrobacterium*-mediated gene transient expression and TRV based gene silencing systems in cassava. (**A**) The schematic view of *Agrobacterium*-mediated gene transient expression assay. (**B**) The schematic view of *Agrobacterium*-mediated TRV based VIGS assay. The white arrows and circles pointed the axillary buds. The figure was drew by using ChemDraw (https://chemdrawdirect.perkinelmer.cloud/js/sample/index.html) and Microsoft Office PowerPoint 2003 (Microsoft, Redmond, WA, USA).

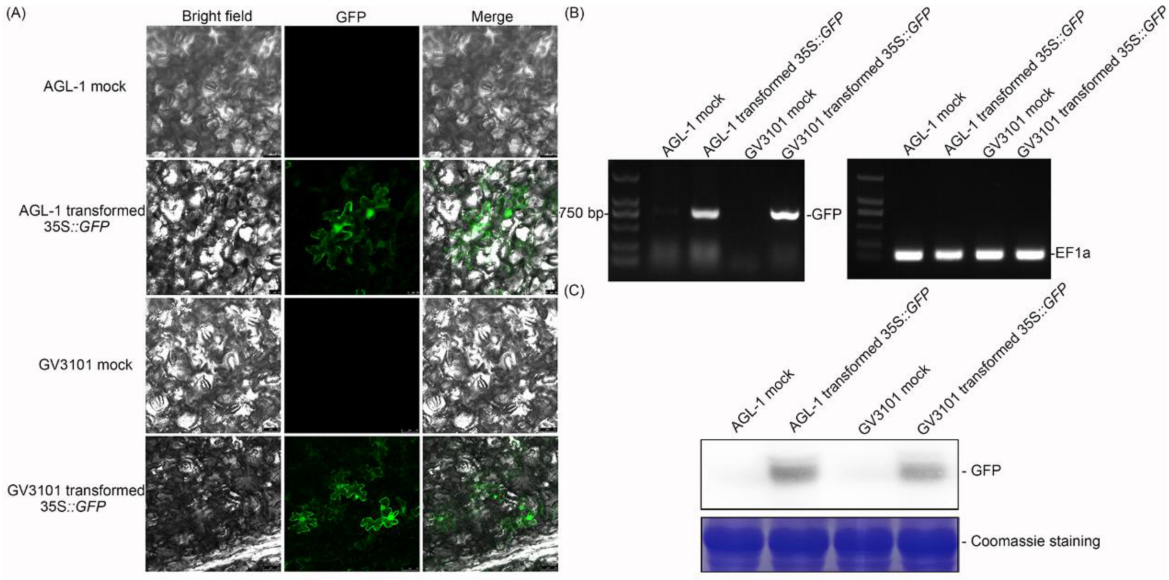


Figure 2. Analysis of transient expression of *green fluorescent protein* (*GFP*) in cassava leaves. (**A**) At 2 days postinfiltration (dpi), *Agrobacterium* with no plasmid was used as the mock, and *Agrobacterium* containing *35S::GFP* was used as the sample. Bar = 25 μm. (**B**) The relative transcript level of *GFP* is shown by reverse-transcription PCR, and *elongation factor 1* (*EF1*) was used as a reference gene. (**C**) *GFP* was detected by Western blot assay, 10 μg total protein extracted from infiltrated leaves were loaded onto SDS-PAGE gel, and Coomassie staining shows the equal protein content.

To further confirm the effect of transient expression, *GUS* was used as the second marker. After *GUS* staining, *Agrobacterium* with the *35S::GUS*-infiltrated area became light blue (Figure 3A), indicating that *GUS* was expressed at the protein level. In addition, reverse-transcription PCR showed that the expression of *35S::GUS* could be examined at the transcript level (Figure 3B), with significant *GUS* activity (Figure 3C). These results suggested that *GFP* and *GUS* were expressed with biological activity in the transient expression assay.

Based on the above results of *GFP* and *GUS* overexpression, the effects of transient overexpression showed no significant difference in *GFP* expression and *GUS* activity between AGL-1 and GV3101 (Figures 2 and 3, Table 1).

Table 1. Effects of *Agrobacterium*-mediated gene transient expression.

Report Gene	Bacterial Strain	Number	The Positive Rate of Reporter Gene
GFP	GV3101	20	60.00%
GUS	GV3101	20	75.00%
GFP	AGL-1	20	65.00%
GUS	AGL-1	20	75.00%

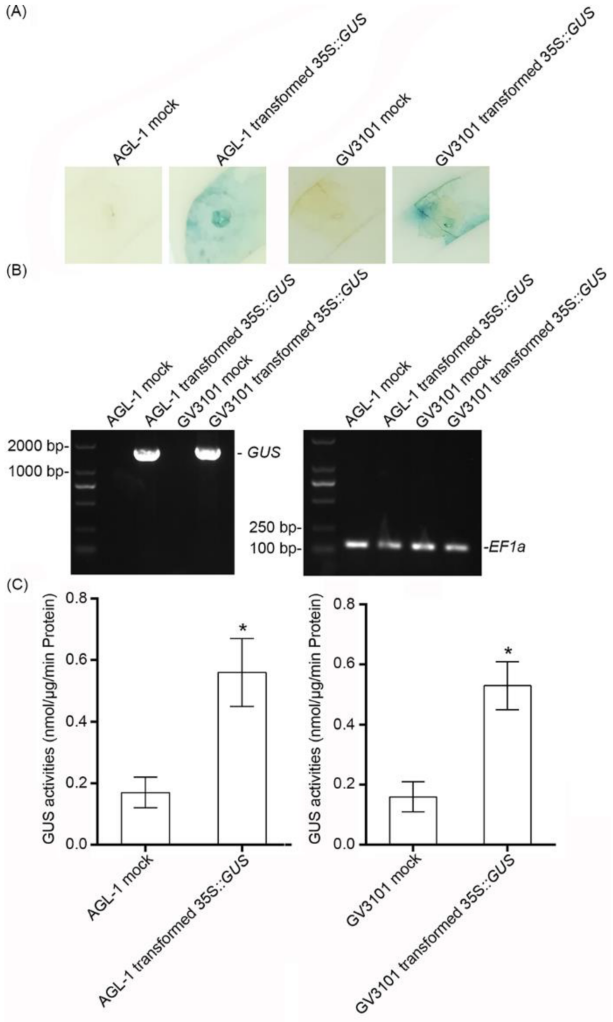


Figure 3. Analysis of transient expression of *β-glucuronidase* (*GUS*) in cassava leaves. (**A**) At 2 dpi, *Agrobacterium* with no plasmid was used as the mock, and *Agrobacterium* containing *35S::GUS* was used as the sample. The leaves were stained using *GUS* staining solution. Bar = 1 cm. (**B**) The relative transcript level of *GUS* is shown by reverse-transcription PCR, and *EF1* was used as a reference gene. (**C**) The *GUS* activities of infiltrated leaves. Statistical tests were performed using IBM SPSS (v21). Briefly, the Kolmogorov–Smirnov test and Levene's test were performed to check the normality of the data distribution and the homogeneity of variance of the data, respectively. The statistical analysis was performed using Student's *t*-test and the Tukey–Kramer test. Asterisk symbol (*) indicates significant difference at $p < 0.05$. $n = 10$ per group. All data are expressed as mean ± SD of three independent experiments.

2.2. Silencing of MePDS in Cassava

To reveal the effect of TRV-based VIGS on cassava, *phytoene desaturase* (*PDS*) was selected as a reporter gene. A 370 bp sequence of *MePDS* was amplified and inserted into the multiple cloning site of the pTRV2 vector. The pTRV1, pTRV2, and pTRV2-*MePDS* vectors were transformed into *Agrobacterium*, respectively. The cultivated *Agrobacterium* solution was infiltrated into cassava local leaves and axillary buds (Figure 1). At 20 dpi, AGL-1-infiltrated cassava showed an obvious albino area in the distal leaves, especially for the area around the main vein (Figure 4).

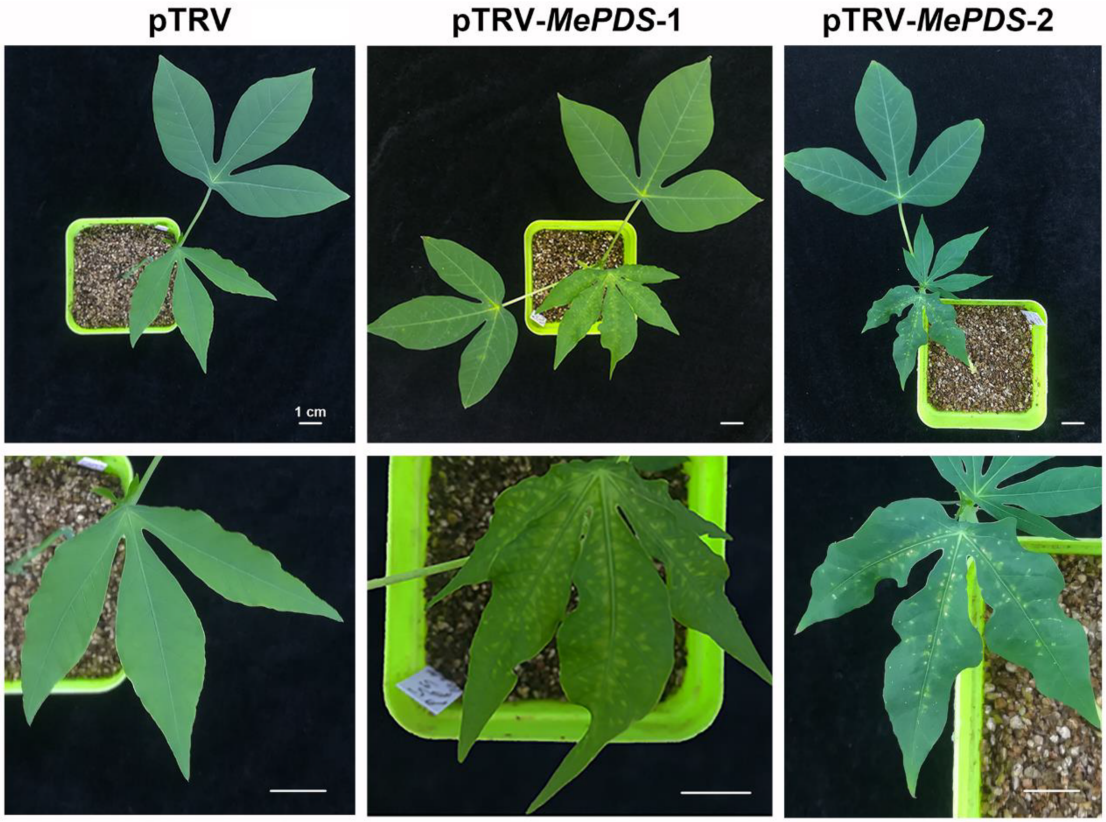


Figure 4. Phenotypes of the infiltrated albino cassava plants. The cassava plants were infiltrated with *Agrobacterium* strain AGL-1 containing the pTRV1 and pTRV2 vectors, which were used as the mock, pTRV1 and pTRV2-*MePDS*, which were used to test the albino effect. In the photos, the upper line shows the whole cassava plants and the underline shows the magnified leaves, both at 20 dpi. Bar = 1 cm.

To further investigate the level of *MePDS* silencing, DNA was extracted from distal leaves and sequences of pTRV1, pTRV2, or pTRV2-*MePDS* were detected by PCR. The result indicated that the sequences of pTRV vectors could be detected in the distal leaves of all AGL-1-infitrated cassava plants but not in the wild-type (WT) cassava (Figure 5A). In addition, the transcript levels of pTRV1 and pTRV2 sequences in the distal leaves were analyzed. Both the transcripts of pTRV1 and pTRV2 could be obviously examined in the pTRV and pTRV-*MePDS* cassava plants but showed no PCR band in WT cassava (Figure 5B). Notably, the pTRV-*MePDS* cassava leaves showed an obvious albino area around the main vein. The transcript level of *MePDS* in the distal leaves of *MePDS*-VIGS plants was 37.9% and 53.1% of that in the mock (Figure 5C), displaying significantly lower chlorophyll content than the mock (Figure 5D). Based on these results, we could conclude that the gene silencing of *MePDS* in the distal leaves of *MePDS*-VIGS plants was caused by VIGS, and the albino phenotype in the distal leaves was derived by VIGS but not the infection-induced effects, which resulted in no significant difference in local leaves.

Compared with the phenotype of *Agrobacterium* AGL-1-infiltrated distal leaves, GV3101-infiltrated distal leaves showed the albino phenotype in the part around the petiole (Supplementary Figure S1), with slightly weaker PCR detection bands, a lower *MePDS* transcript level, and lower chlorophyll content than the mock (Supplementary Figure S2). For AGL-1, 75.00% of cassava plants showed less than 60% transcript level of *MePDS* in the distal leaves, and only 37.50% of the plants exhibited the albino phenotype (Table 2). For GV3101, 62.50% of cassava plants showed a lower transcript level of *MePDS* in the distal leaves, and only 12.50% of plants showed the albino phenotype (Table 2).

These results suggested that both AGL-1 and GV3101 with TRV could induce *MePDS* silencing and regulate the color of leaves in cassava, with better effects in AGL-1-transformed TRV-based VIGS.

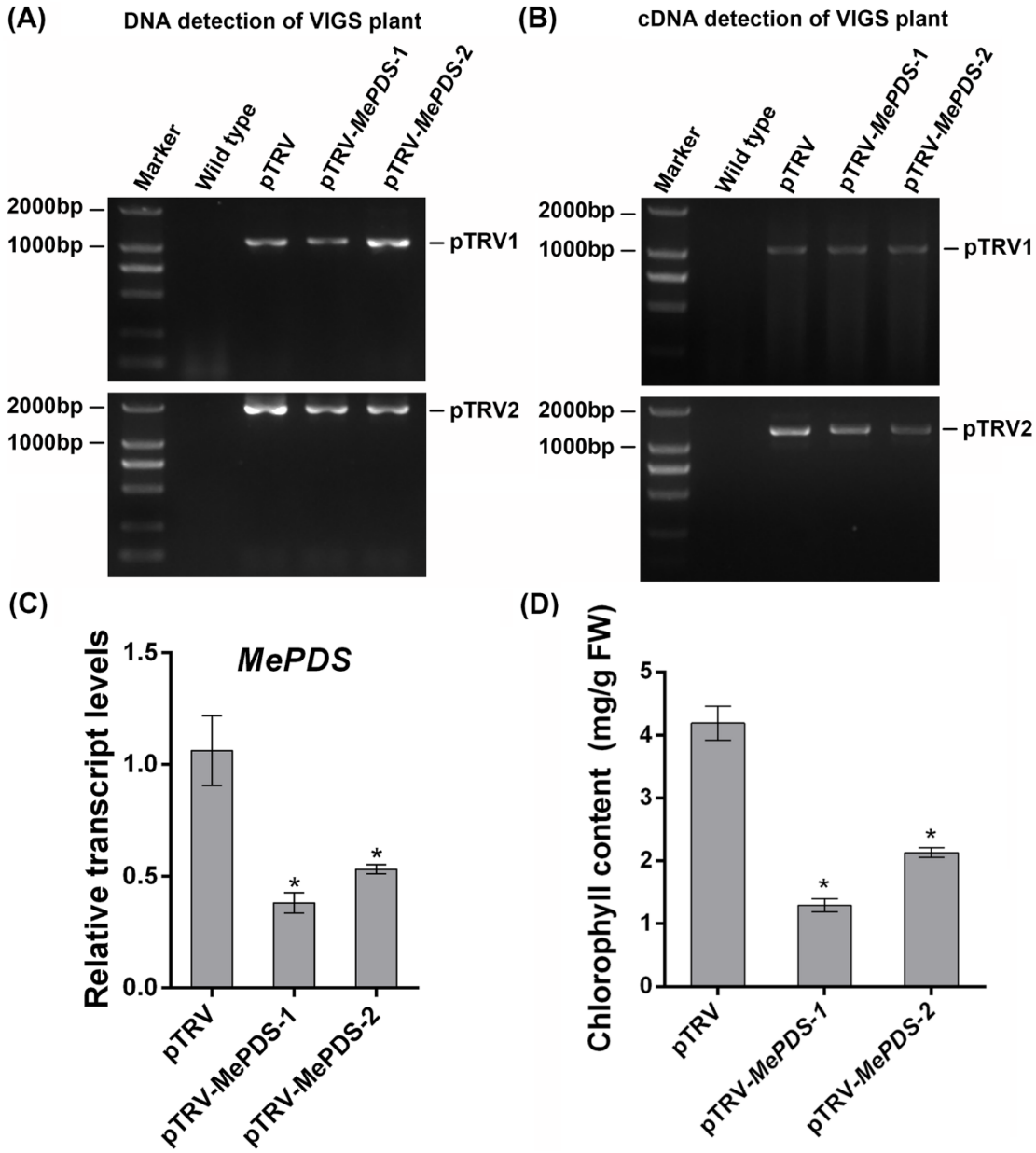


Figure 5. Detection of VIGS cassava plants. (**A**) PCR detection of pTRV1 and pTRV2 sequences from the genomic DNA of VIGS cassava plants. (**B**) Reverse-transcription PCR detection of pTRV1 and pTRV2 sequences from the cDNA of VIGS cassava plants. (**C**) The relative transcript levels of *MePDS* of the infiltrated cassava plants. *MeEF1* was used as internal control. (**D**) The chlorophyll contents in the infiltrated cassava plants. Statistical tests were performed using IBM SPSS (v21). Briefly, the Kolmogorov–Smirnov test and Levene's test were performed to check the normality of the data distribution and the homogeneity of variance of the data, respectively. The statistical analysis was performed using Student's *t*-test and the Tukey–Kramer test. Asterisk symbol (*) indicates significant difference at $p < 0.05$. All data are expressed as mean ± SD of three independent experiments.

Table 2. Effects of *Agrobacterium*-mediated TRV-based gene silencing systems in cassava.

Bacterial Strain	Number	The Positive Rate of Relative Transcript Level (<60%)	The Positive Rate of Albino Phenotype
AGL-1	8	75.00%	37.50%
GV3101	8	62.50%	12.50%

3. Discussion

Although genome sequencing of cassava has been completed for several years now, little progress has been made in determining the functional genomics of cassava [34]. In order to address this question and promote functional genomics in cassava, this study investigated *Agrobacterium*-mediated gene transient overexpression and TRV-based VIGS.

The *Agrobacterium*-mediated transient overexpression assay is a powerful tool for analyzing gene function in vivo [35]. In this study, *Agrobacterium* AGL-1 and GV3101 with *GFP* or *GUS* overexpressing plasmids were used for transient overexpression. In a previous study, GV3101 containing the *35S::GUS* vector was used in MCOL2215 and 60,444 cassava varieties to test the transient expression effect, but no *GUS* was detected. However, AGL with the *35S::GUS* vector showed a positive result [36]. In this study, both *GFP* and *GUS* were expressed by the GV3101 strain in SC124 and could be detected at the transcript (Figures 2 and 3) and protein (Figures 2 and 3) levels, indicating that they could be efficiently expressed in cassava leaves. Generally, the GV3101- and AGL-1-based transient overexpression assays had relatively high success rates (Table 1), while the negative results might have been due to the RNA cosilence process [37,38], the morphology, or the structure of the plants [39]. Therefore, whether the target genes are overexpressed in the transient overexpression assay should be analyzed first.

In addition, *Agrobacterium*-mediated VIGS is another powerful tool to investigate gene function in vivo. Different *Agrobacterium* strains (GV3101 and AGL-1) with pTRV vectors were used to silence the *MePDS* gene. AGL-1 with the transformation of pTRV1 and pTRV2-*MePDS* vector-infiltrated distal leaves showed obvious albino phenotypes at 20 dpi (Figure 4). Interestingly, the albino phenotype only showed in distal leaves and the albino area was mainly located around the leaf vein, similar to the results in *California poppy* [40] and *Solanum pseudocapsicum* [41]. From the albino leaf blade, we found that the area near the petiole became white, while the area distant from the petiole was still green (Figure 4), indicating that TRV-based *MePDS* silence might spread from the bottom to the tip of the blade in cassava. Additionally, TRV-based *Agrobacterium*-mediated VIGS was examined by genome DNA or cDNA to ensure the *Agrobacterium*-mediated transformation and expression of pTRV1 and pTRV2 or pTRV2-X [41,42]. PCR results indicated that *Agrobacterium*-mediated pTRV1 and pTRV2 or pTRV2-*MePDS* could be successfully transformed and expressed in the distal leaves of cassava plants, with a 37.9%–53.1% transcript level of *MePDS* and 58%–76.5% of chlorophyll content in the distal plant leaves (Figure 5). Because the relative transcript level of *MePDS* of the albino cassava plants was 71% of the mock and the detected bands were weak, GV3101 with pTRV1 and pTRV2-*MePDS* infiltrated cassava plants showed less obvious change, and only the part around the petiole became albino (Supplementary Figures S1 and S2). VIGS is a post-transcriptional gene silencing method [24,25]. In a VIGS system, the expressed levels of silenced marker genes were 30%–40% of control in *Solanum pseudocapsicum* L. [41], 37% of control in columbine [43], 41%–60% of control in barley [44], 28%–38% of control in tomato [45], and 30%–70% of control in rose [27]. Although these genes could not silenced completely, the albino phenotype was obvious. Consistently, the transcript level of *MePDS* in *MePDS*-VIGS plants were 37.9%–53.1% of that in the mock, with an obvious albino phenotype and a 58%–76.5% level of chlorophyll. Compared with GV3101, *Agrobacterium* AGL-1 showed better effects in the VIGS system (Table 2), which might have been due to the specific hypervirulence of AGL-1 [46]. AGL-1 is derived from pTiBo542, which has a high induction of the *vir* gene [47], which is necessary for T-DNA transfer [48].

In this study, an *Agrobacterium*-mediated transient overexpression system was established in cassava. The successful examination of the transcription, translation, and biological activity of expressed proteins suggests its application in subcellular localization experiments, bimolecular fluorescence complementation (BIFC), co-immunoprecipitation (Co-IP), enzyme activity detection, and so on. In addition, a TRV-based *Agrobacterium*-mediated VIGS system was also successfully verified in cassava (Figure 1B), as shown by the expressions of pTRV and the corresponding gene, as well as the albino phenotype. Previous studies have found that cassava could be silenced by an *African-cassava-mosaic-virus*-based VIGS system; however, this could lead to serious malformation of newly grown leaves, thereby resulting in interference in the assay of plant disease resistance [49–51]. On the contrary, TRV-based VIGS-system-infiltrated cassava plants only showed mild virus-induced disease symptoms and might be more appropriate for gene function analysis. In conclusion, this study provided two valid and efficient methods for the characterization of gene function, so as to promote functional genomics in cassava.

4. Materials and Methods

4.1. Plant Materials

The cassava plants South China 124 (SC124) were kindly provided by Dr. Wei Hu (Institute of Tropical Bioscience and Biotechnology, Haikou, China). Two-week-old tissue culture cassava plants (SC124) were transferred to small pots and grown in a chamber for two weeks until the experiment at 25 °C under a 16 h light/8 h dark cycle.

4.2. Vectors and Vector Construction

The pEGAD and pBI121 vectors were used to express *GFP* and *GUS*, respectively. The full-length sequences of *GFP* and *GUS* were driven by the 35S promoter. The VIGS assay used pTRV1 and pTRV2 vectors, which have been described previously [52]. The partial sequence of the *MePDS* gene was cloned into the multiple cloning site of the pTRV2 vector, and the primers are listed in Supplementary Table S1.

4.3. Agrobacterium Infiltration of Cassava

Agrobacterium strains GV3101 and AGL-1 were transformed by different plasmids: pEGAD, pBI121, pTRV1, pTRV2, and pTRV2-*MePDS*, respectively. GV3101 was cultivated in liquid LB medium containing 50 mg/L kanamycin, 20 mg/L rifampicin, and 50 mg/L gentamycin, while AGL-1 was cultivated in LB medium with 50 mg/L kanamycin, 20 mg/L rifampicin, and 50 mg/L carbenicillin. Both were then shook at 28 °C at 200 rpm for 2 days. The bacteria was centrifuged at 4000 rpm for 10 min, then the supernatant was discarded. The remaining bacteria was washed with double-distilled water, then centrifuged, and the supernatant was discarded again. The bacterial sediment was resuspended in MMA solution (10 mM MgCl, 10 mM MES, and 150 μM acetosyringone), the OD_{600} was adjusted to 1, and then the resuspended bacterial solution was placed in the dark for 3 h. For the transient expression assay, the standing bacterial solution was infiltrated into the second and third leaves from the top of the cassava by a 1 mL needle. For the VIGS assay, the bacterial solutions containing pTRV1 or pTRV2-*MePDS* were mixed with the same volume, and the mixture of pTRV1 or pTRV2 was used as the mock. The mixed solution was infiltrated into both the leaves and axillary buds of cassava plants to keep the silencing effect [48]. After infiltration, the cassava plants were removed to the chamber at the same light and temperature.

4.4. DNA and RNA Extraction

DNA was extracted by the Plant Genomic DNA Extraction Kit (DP305, TIANGEN, Beijing, China). Total RNA was extracted by the RNAprep Pure Plant Kit (Polysaccharides & Polyphenolics-rich) (DP441, TIANGEN, Beijing, China). The remaining DNA from the extracted RNA was digested by

RNase-free DNase I (EN0521, Thermo, Waltham, MA, USA). Then, the quality and concentration of DNA and RNA were examined by Nano Drop 2000 (Thermo, Waltham, MA, USA).

4.5. Reverse-Transcription PCR and Quantitative Real-Time PCR

The first strand cDNA was synthesized by the RevertAid First Strand cDNA Synthesis Kit (K1621, Thermo, Waltham, MA, USA). Then, the cDNA was adjusted to an equal concentration. Reverse-transcription PCR was performed by EasyTaq PCR SuperMix (AS111, TRANS, Beijing, China) with a PCR program of (1) 94 °C for 3 min; (2) 29 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min; and (3) 72 °C for 10 min and 16 °C for storage.

Quantitative real-time PCR was performed using TransStart Tip Green qPCR SuperMix (AQ141, TRANS, Beijing, China) in LightCycler®96 (Roche, Basel, Switzerland) with a PCR program of (1) 94 °C for 30 s; (2) 45 cycles of 94 °C for 5 s, 55 °C for 15 s, and 72 °C for 15 s; and (3) 95 °C for 10 s, 65 °C for 60 s, and 97 °C for 1 s as the melting curve. *Elongation factor 1* (*EF1*) was used as the reference gene, and the primers used in reverse-transcription PCR and quantitative real-time PCR are listed in Supplementary Tables S2 and S3.

4.6. Confocal Microscopy Scanning

At 2 dpi, the infiltrated area of cassava leaves was harvested, cut into small pieces, and then observed by confocal laser-scanning microscopy (TCS SP8, Leica, Heidelberg, Germany).

4.7. Protein Extraction and Western Blot

Cassava leaves were harvested and ground by liquid nitrogen and phosphate buffer solution (pH 7.4). The extracted solution was centrifuged at 12,000 rpm and 4 °C for 10 min. The supernatant was boiled with SDS-PAGE sample loading buffer (P0015, Biyotime, Shanghai, China) for 5 min. After centrifugation for 10 min, the supernatant could be used in Western blot. The Western blot assay was performed according to a previous study [53]. Briefly, the protein samples were loaded into 12% polyacrylamide gel and separated by electrophoresis. The polyacrylamide gel was transferred to a PVDF membrane (475855-1R, Millpore, Massachusetts, America) by a Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (1703940, Bio-rad, Hercules, America). Then, the PVDF membrane was blocked and incubated in 5% skim milk with anti-*GFP* antibody (AG281, Biyotime, Haimen, China).

4.8. GUS Staining and Activity Detection

GUS staining and *GUS* activity assay were carried out according to a previous study with slight modifications [48]. The infiltrated cassava leaves were harvested and immersed into *GUS* staining solution (50 mM NaH_2PO_4, 50 mM Na_2HPO_4, 10 mM EDTA-Na_2, 0.5 mM $K_4[Fe(CN)_6]$, 0.5 mM $K_3[Fe(CN)_6]$, 0.1% Triton-X100, and 2 mM X-Gluc; pH 7.0) with vacuum infiltration for 0.5 h in the dark, and then the plant leaves were incubated in the dark at 37 °C for at least 12 h. After staining, the leaves were immersed into 70% alcohol to remove chlorophyll.

The infiltrated leaves were harvested and ground by liquid nitrogen and phosphate buffer solution (pH 7.4). The extracted solution was centrifuged (12,000 rpm, 10 min, 4 °C) and the supernatant was used for further assay. The extraction was incubated in 1 mM 4-Methylumbelliferyl-b-D-glucuronide (4-MUG) at 37 °C, the reaction mixture was taken out per 5 min, and Na_2CO_3 was added to stop the reaction. The samples were detected by a microplate system (Infinite M200 Pro, TECAN, Hombrechtikon, Switzerland) with 365 nm excitation and 455 nm emission. The *GUS* activity was calculated by the standard curve made by different concentrations of 4-methylumbelliferone (4-MU).

4.9. PCR Detection of Cassava Plants

The transient expressed cassava plants were detected by PCR using genomic DNA and cDNA, with a PCR program of (1) 94 °C for 3 min; (2) 29 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min; and (3) 72 °C for 10 min and 16 °C for storage. The primers are listed in Supplementary Table S4.

4.10. Chlorophyll Content of Cassava Leaves

The chlorophyll content of the cassava leaves was quantified as previously described [54]. The harvested leaves were ground in 80% acetone. After centrifugation, the absorbance of OD_{647} and OD_{665} of the supernatant solution was detected by a microplate system (Infinite M200 Pro, TECAN, Hombrechtikon, Switzerland). The total chlorophyll content was $17.90 \times OD_{647} + 8.08 \times OD_{665}$.

4.11. Statistical Analysis

All data are expressed as mean ± SD of three independent experiments. Statistical tests were performed using IBM SPSS (v21). Briefly, the Kolmogorov–Smirnov test and Levene's test were used to check the normality of the data distribution and the homogeneity of variance of the data, respectively. Then, Student's *t*-test and the Tukey–Kramer test were used for statistical analysis. The asterisk symbol (*) indicates significant difference at $p < 0.05$.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/16/3976/s1.

Author Contributions: Conceptualization, H.Z.; Data curation, H.Z.; Formal analysis, H.Z.; Funding acquisition, H.Z., Y.X., and H.S.; Investigation, H.Z., Y.X., G.L., and Y.W.; Methodology, H.Z.; Project administration, H.S.; Supervision, W.H. and H.S.; Writing—original draft, H.Z.; Writing—review and editing, W.H. and H.S.

Funding: This research was supported by National Key R&D Program of China (No. 2018YFD1000500), the National Natural Science Foundation of China (No. 31760067), the Startup Funding and Scientific Research Foundation of Hainan University (No. kyqd1531), and the Crop Science Postgraduate Innovation Project of Hainan University Tropical Agriculture and Forestry College (College of Tropical Crops) (Nos. ZWCX2018005 and ZWCX2018018).

Acknowledgments: We thank Chris R. Somerville and Jie Zhou for sharing the vectors.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Wilson, M.C.; Mutka, A.M.; Hummel, A.W.; Berry, J.; Chauhan, R.D.; Vijayaraghavan, A.; Taylor, N.J.; Voytas, D.F.; Chitwood, D.H.; Bart, R.S. Gene expression atlas for the food security crop cassava. *New Phytol.* **2017**, *213*, 1632–1641.
2. Nyaboga, E.; Njiru, J.; Nguu, E.; Gruissem, W.; Vanderschuren, H. Unlocking the potential of tropical root crop biotechnology in east Africa by establishing a genetic transformation platform for local farmer-preferred cassava cultivars. *Front. Plant SCI.* **2013**, *4*, 526. [CrossRef]
3. Wydro, M.; Kozubek, E.; Lehmann, P. Optimization of transient *Agrobacterium*-mediated gene expression system in leaves of Nicotiana benthamiana. *Acta. Biochimica. Pol.* **2006**, *53*, 289–298.
4. Kim, M.J.; Baek, K.; Park, C.M. Optimization of conditions for transient *Agrobacterium*-mediated gene expression assays in Arabidopsis. *Plant Cell Rep.* **2009**, *28*, 1159–1167.
5. Ma, L.; Lukasik, E.; Gawehns, F.; Takken, F.L. The Use of Agroinfiltration for Transient Expression of Plant Resistance and Fungal Effector Proteins in *Nicotiana benthamiana* Leaves. *Methods Mol. Biol.* **2012**, *835*, 61.
6. Ding, B.Q.; Yuan, Y.W. Testing the utility of fluorescent proteins in *Mimulus lewisii* by an *Agrobacterium*-mediated transient assay. *Plant Cell Rep.* **2016**, *35*, 771–777.
7. Mani, T.; Manjula, S. Optimization of *Agrobacterium*-mediated transient gene expression and endogenous gene silencing in *Piper colubrinum* by vacuum infiltration. *Plant Cell Tiss. Org.* **2011**, *105*, 113–119.
8. Rakouský, S.; Kocábek, T.; Vincenciová, R.; Ondřej, M. Transient β-glucuronidase activity after infiltration of Arabidopsis thaliana by *Agrobacterium* tumefaciens. *Biol. Plant.* **1997**, *40*, 33–41. [CrossRef]

9. Lu, J.; Bai, M.; Ren, H. An efficient transient expression system for gene function analysis in rose. *Plant Methods.* **2017**, *13*, 116. [CrossRef]
10. Yang, X.F.; Yu, X.Q.; Zhou, Z.; Ma, W.J.; Tang, G.X. High-efficiency *Agrobacterium* tumefaciens mediated transformation system using cotyledonary node as explants in soybean (*Glycine max* L.). *Acta. Physiol. Plant.* **2016**, *38*, 60.
11. Fister, A.S.; Shi, Z.; Zhang, Y.F.; Helliwell, E.E.; Maximova, S.N.; Guiltinan, M.J. Protocol: Transient expression system for functional genomics in the tropical tree *Theobroma cacao* L. *Plant Methods.* **2016**, *12*, 19. [PubMed]
12. Li, H.P.; Li, K.; Guo, Y.T.; Guo, J.D.; Miao, K.T.; Botella, J.R.; Song, C.P.; Miao, Y.C. A transient transformation system for gene characterization in upland cotton (*Gossypium hirsutum*). *Plant Methods* **2018**, *14*, 50. [PubMed]
13. Kamboj, D.; Yadav, R.C.; Singh, A.; Yadav, N.R.; Singh, D. Plant regeneration and *Agrobacterium*-mediated transformation in Indian mustard (Brassica juncea). *J. Oilseed Brassica.* **2016**, *6*, 191–197.
14. Bhaskar, P.B.; Venkateshwaran, M.; Wu, L.; Ané, J.M.; Jiang, J. *Agrobacterium*-Mediated Transient Gene Expression and Silencing: A Rapid Tool for Functional Gene Assay in Potato. *PLoS ONE.* **2009**, *4*, e5812. [CrossRef] [PubMed]
15. Wroblewski, T.; Tomczak, A.; Michelmore, R. Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol. J.* **2005**, *3*, 259–273. [PubMed]
16. Lange, M.; Yellina, A.L.; Orashakova, S.; Becker, A. Virus-induced gene silencing (VIGS) in plants: An overview of target species and the virus-derived vector systems. *Methods Mol. Biol.* **2013**, *975*, 1–14. [PubMed]
17. Hiriart, J.B.; Aro, E.M.; Lehto, K. Dynamics of the VIGS-mediated chimeric silencing of the *Nicotiana benthamiana ChlH* gene and of the *tobacco mosaic virus* vector. *Mol. Plant Microbe Interact.* **2003**, *16*, 99–106. [CrossRef] [PubMed]
18. Lacomme, C.; Chapman, S. Use of *potato virus X* (PVX)-based vectors for gene expression and virus-induced gene silencing (VIGS). *Curr. Protoc. Microbiol.* **2008**, *16*, 1.
19. Zhao, D.; Zhao, J.R.; Huang, X.; Li, N.; Liu, Y.; Huang, Z.J.; Zhang, Z.Y. Functional Analysis of TNBL1 Gene in Wheat Defense Response to Barley yellow dwarf virus Using BSMV-VIGS Technique. *Acta. Agronomica. Sinica.* **2011**, *37*, 2106–2110. [CrossRef]
20. Zhou, T.; Liu, X.; Fan, Z. Use of a Virus Gene Silencing Vector for Maize Functional Genomics Research. *Methods Mol. Biol.* **2018**, *1676*, 141–150.
21. Liu, X.B.; Liu, N.; Li, F.K.; Wu, L.Z.; Zhang, J.; Wang, D.M. Establishment of TRV-mediated Transient Gene-Silencing System in Soybean. *Sci. Agri. Sin.* **2015**, 1212.
22. Fu, D.Q.; Zhu, B.Z.; Zhu, H.L.; Jiang, W.B.; Luo, Y.B. The application of TRV-mediated VIGS technique in the study of gene function in fruits and vegetables. *Plant Physiol. J.* **2017**, *43*, 299–308.
23. Unver, T.; Budak, H. Virus-Induced Gene Silencing, a Post Transcriptional Gene Silencing Method. *Int. J. Plant Genomics.* **2009**, 198680.
24. Ratcliff, F.; Martin-Hernandez, A.M.; Baulcombe, D.C. Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. *Plant J.* **2001**, *25*, 237–245. [PubMed]
25. Zhou, X.F.; Sun, J.D.; Zhao, Z.; Lv, J.; Wei, X.W.; Cai, R.; Xu, H.W. The Feasibility Analysis of PVX and TRV Vectors as the VIGS Tool for Studying the Gene Function. *Physics. Procedia.* **2002**, *23*, 46–54. [CrossRef]
26. Wang, C.C.; Cai, X.Z.; Wang, X.M.; Zheng, Z. Optimisation of tobacco rattle virus-induced gene silencing in Arabidopsis. *Funct. Plant Biol.* **2006**, *33*, 347–355.
27. Tian, J.; Pei, H.X.; Zhang, S.; Chen, J.W.; Chen, W.; Yang, R.Y.; Meng, Y.; You, J.; Gao, J.; Ma, N. TRV-GFP: A modified Tobacco rattle virus vector for efficient and visualizable analysis of gene function. *J. Exp. Bot.* **2014**, *65*, 311–322. [PubMed]
28. Tian, J.; Cheng, L.; Han, Z.Y.; Yao, Y.C. Tobacco rattle virus mediated gene silencing in strawberry plants. *Plant Cell Tiss. Org.* **2015**, *120*, 1131–1138.
29. Wang, X.Y.; LÜ, K.; Cai, C.P.; Xu, J.; Guo, W.Z. Establishment and Application of TRV-Mediated Virus-Induced Gene Silencing in Cotton. *Acta. Agronomica. Sin.* **2014**, *40*, 1356. [CrossRef]
30. Anu, K.; Jessymol, K.K.; Chidambareswaren, M.; Gayathri, G.S.; Manjula, S. Down-regulation of osmotin (PR5) gene by virus-induced gene silencing (VIGS) leads to susceptibility of resistant *Piper colubrinum* to the oomycete pathogen *Phytophthora capsici Leonian*. *Indian, J. Exp. Biol.* **2015**, *53*, 329–334.
31. Zhang, J.; Yu, D.; Zhang, Y.; Liu, K.; Xu, K.; Zhang, F.; Wang, J.; Tan, G.X.; Nie, X.H.; Ji, Q.H.; et al. Vacuum and Co-cultivation Agroinfiltration of (Germinated) Seeds Results in Tobacco Rattle Virus (TRV) Mediated

Whole-Plant Virus-Induced Gene Silencing (VIGS) in Wheat and Maize. *Front. Plant Sci.* **2017**, *8*, 393. [PubMed]
32. Hartl, M.; Merker, H.; Schmidt, D.D.; Baldwin, I.T. Optimized virus-induced gene silencing in Solanum nigrum reveals the defensive function of leucine aminopeptidase against herbivores and the shortcomings of empty vector controls. *New Phytol.* **2010**, *179*, 356–365. [CrossRef]
33. Cutler, S.R.; Ehrhardt, D.W.; Griffitts, J.S.; Somerville, C.R. Random GFP::*cDNA* fusions enable visualization of subcellular structures in cells of arabidopsis at a high frequency. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3718–3723. [CrossRef] [PubMed]
34. Apio, H.B. Production of friable embryogenic callus and regeneration of Ugandan farmer-preferred cassava genotypes. *Afr. J. Biotechnol.* **2015**, *14*, 1854–1864.
35. Tsuda, K.; Qi, Y.; Nguyen, L.V.; Bethke, G.; Tsuda, Y.; Glazebrook, J.; Katagiri, F. An efficient *Agrobacterium*-mediated transient transformation of *Arabidopsis*. *Plant J.* **2012**, *69*, 713–719. [PubMed]
36. Díaz, T.P.; Bernal, G.A.; Camilo, L.C. Transient gus gene expression in cassava (manihot esculenta crantz) using agrobacterium tumefaciens leaf infiltration. *Revista. Mvz. Cordoba* **2014**, *19*, 4338–4349.
37. Olivier, V.; Susana, R.; Pere, M.; David, B. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **2003**, *33*, 949–956.
38. Johansen, L.K.; Carrington, J.C. Silencing on the Spot. Induction and Suppression of RNA Silencing in the *Agrobacterium*-Mediated Transient Expression System. *Plant Physiol.* **2001**, *126*, 930–938. [CrossRef]
39. Andrieu, A.; Breitler, J.C.; Siré, C.; Meynard, D.; Gantet, P.; Guiderdoni, E. Anin planta, agrobacterium-mediated transient gene expression method for inducing gene silencing in rice (*oryza sativa* L.) leaves. *Rice* **2012**, *5*, 23. [CrossRef]
40. Wege, S.; Scholz, A.; Gleissberg, S.; Becker, A. Highly Efficient Virus-induced Gene Silencing (VIGS) in California Poppy (*Eschscholzia californica*): An Evaluation of VIGS as a Strategy to Obtain Functional Data from Non-model Plants. *Ann. Bot.* **2007**, *100*, 641–649.
41. Xu, H.; Xu, L.F.; Yang, P.; Cao, Y.; Tang, Y.; He, G.; Yuan, S.X.; Ming, J. Tobacco rattle virus-induced *PHYTOENE DESATURASE (PDS)* and *Mg-chelatase H subunit* (*ChlH*) gene silencing in *Solanum pseudocapsicum* L. *PeerJ.* **2018**, *6*, e4424. [PubMed]
42. Dai, W.; Wang, M.; Gong, X.; Liu, J.H. The transcription factor FcWRKY40 of *Fortunella crassifolia* functions positively in salt tolerance through modulation of ion homeostasis and proline biosynthesis by directly regulating *SOS2* and *P5CS1* homologs. *New Phytol.* **2018**, *219*, 972–989. [PubMed]
43. Gould, B.; Kramer, E.M. Virus-induced gene silencing as a tool for functional analyses in the emerging model plant Aquilegia (columbine, Ranunculaceae). *Plant Methods* **2007**, *3*, 6. [PubMed]
44. Bruun-Rasmussen, M.; Madsen, C.T.; Jessing, S.; Albrechtsen, M. Stability of Barley stripe mosaic virus-induced gene silencing in barley. Molecular plant-microbe interactions. *MPMI* **2007**, *20*, 1323. [CrossRef] [PubMed]
45. Rotenberg, D.; Thompson, T.S.; German, T.L.; Willis, D.K. Methods for effective real-time RT-PCR analysis of virus-induced gene silencing. *J. Virol. Methods* **2006**, *138*, 49–59.
46. Uçarlı, C.; Tufan, F.; Gürel, F. Expression and genomic integration of transgenes after *Agrobacterium*-mediated transformation of mature barley embryos. *Genet. Mol. Res.* **2015**, *14*, 1096–1105.
47. Lazo, G.R.; Stein, P.A.; Ludwig, R.A. A DNA transformation-competent arabidopsis genomic library in *Agrobacterium*. *Bio/Technology* **1991**, *9*, 963–967. [CrossRef]
48. Chabaud, M.; Carvalho-Niebel, F.D.; Barker, D.G. Efficient transformation of *Medicago truncatula* cv. jemalong using the hypervirulent *Agrobacterium tumefaciens* strain AGL1. *Plant Cell Rep.* **2003**, *22*, 46–51. [CrossRef]
49. Fofana, I.B.; Sangaré, A.; Collier, R.; Taylor, C.; Fauquet, C.M. A geminivirus-induced gene silencing system for gene function validation in cassava. *Plant Mol Biol.* **2004**, *56*, 613–624. [CrossRef]
50. Beyene, G.; Chauhan, R.D.; Taylor, N.J. A rapid virus-induced gene silencing (VIGS) method for assessing resistance and susceptibility to cassava mosaic disease. *Virol. J.* **2017**, *14*, 47.
51. Matias, L.E.; Joel-Elias, K. Cassava germinivirus agroclones for virus-induced gene silencing in cassava leaves and roots. *Plant Methods* **2018**. [CrossRef]
52. Liu, Y.; Schiff, M.; Dinesh-Kumar, S.P. Virus-induced gene silencing in tomato. *Plant J.* **2002**, *31*, 777–786. [CrossRef] [PubMed]

53. Cai, J.; Qin, G.; Chen, T.; Tian, S. The mode of action of remorin1 in regulating fruit ripening at transcriptional and post-transcriptional levels. *New Phytol.* **2018**, *219*, 1406–1420. [CrossRef] [PubMed]
54. Inskeep, W.P.; Bloom, P.R. Extinction coefficients of chlorophyll a and B in N, N-dimethylformamide and 80% acetone. *Plant Physiol.* **1985**, *77*, 483–485. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Depicting the Core Transcriptome Modulating Multiple Abiotic Stresses Responses in Sesame (*Sesamum indicum* L.)

Komivi Dossa [1,*], Marie A. Mmadi [1], Rong Zhou [1], Tianyuan Zhang [2], Ruqi Su [1], Yujuan Zhang [1], Linhai Wang [1], Jun You [1] and Xiurong Zhang [1,*]

[1] Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China
[2] State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China
* Correspondence: dossakomivi@gmail.com (K.D.); zhangxr@oilcrops.cn (X.Z.); Tel.: +86-27-8681-1836 (K.D. & X.Z.)

Received: 8 July 2019; Accepted: 10 August 2019; Published: 13 August 2019

Abstract: Sesame is a source of a healthy vegetable oil, attracting a growing interest worldwide. Abiotic stresses have devastating effects on sesame yield; hence, studies have been performed to understand sesame molecular responses to abiotic stresses, but the core abiotic stress-responsive genes (CARG) that the plant reuses in response to an array of environmental stresses are unknown. We performed a meta-analysis of 72 RNA-Seq datasets from drought, waterlogging, salt and osmotic stresses and identified 543 genes constantly and differentially expressed in response to all stresses, representing the sesame CARG. Weighted gene co-expression network analysis of the CARG revealed three functional modules controlled by key transcription factors. Except for salt stress, the modules were positively correlated with the abiotic stresses. Network topology of the modules showed several hub genes predicted to play prominent functions. As proof of concept, we generated over-expressing Arabidopsis lines with hub and non-hub genes. Transgenic plants performed better under drought, waterlogging, and osmotic stresses than the wild-type plants but did not tolerate the salt treatment. As expected, the hub gene was significantly more potent than the non-hub gene. Overall, we discovered several novel candidate genes, which will fuel investigations on plant responses to multiple abiotic stresses.

Keywords: stress marker genes; sesame; gene co-expression; abiotic stress tolerance; hub genes; meta-analysis

1. Introduction

Climate change causes the rising of sea levels, a decrease of available land for farming, and increased frequencies of severe droughts, intense precipitation events, elevated temperatures, as well as salt and heavy metals contamination of soils. Crop productivity and survival is tightly linked to its environment, which is being altered due to climate change [1], impairing crop yields and leading to enhanced risks of famine worldwide [2]. Being a sessile organism, plants have evolved an enormous capacity to adapt to environmental changes including heat, drought, salinity, osmotic pressure, waterlogging, etc., by modulating their physiology, growth and development. Until recently, numerous researches have focused on plant responses to individual abiotic stress [3–8]. Although these studies have potential applications in crop improvement for abiotic stress tolerance, crops growing in natural habitats are often exposed to multiple environmental stresses occurring simultaneously or at different development stages, which inflicts a more severe reduction in yields as compared to a single stress [9,10]. Hence, development of crops able to tolerate a wide range of abiotic stresses with high

productivity is imperative in order to meet various socio-economic and agro-economic challenges in the current climate change scenario [11,12]. However, the interactive improvement of multiple abiotic stress tolerances is a challenge [13], since increasing tolerance to one stress may be at the expense of tolerance to another [9,14]. Therefore, there is a need to investigate the metabolic pathways and regulatory networks of multiple abiotic stress acclimations in plants and obtain candidate genes for manipulation to improve tolerance to multiple abiotic stresses.

In recent years, it has become evident that a common battery of responses can be triggered by various stresses. For example in yeasts, it has been discovered that a set of genes are constantly activated upon exposure to various abiotic stresses and represent the core environmental stress responsive genes [15–19]. Similarly, the core environmental stress responsive genes were unraveled in plants, first in *Arabidopsis thaliana* through the AtGenExpress abiotic stress experiment [20–22] and later confirmed in various plant species, such as rice, soybean, banana, *Brachypodium distachyon* or barley [13,23–27], indicating the conservation of a core genome modulating various abiotic stresses responses between species. Thus, it is critical to compare and analyze different kinds of abiotic stress responses to find the common genes and understand how they regulate plant's adaptations to the multiple environmental stresses. This information will guide in strategies to enhance crop tolerance to multiple abiotic stresses.

Consumers become more and more health conscious, with a sharper focus on health-promoting diets. Oils and fats are an important part of a balanced and healthy diet. Among the most nutritious and healthy vegetable oils, sesame (*Sesamum indicum* L.) oil occupies a pro-eminent position because of the low level of saturated fatty acids (less than 15%) and the presence of strong antioxidants, reported to have health-promoting effects such as lowering cholesterol levels and hypertension [28,29], neuroprotective effects against hypoxia or brain damage [30] and reducing the incidence of certain cancers [31,32]. Therefore, demands for and world trade in sesame seeds have increased rapidly during the last two decades [33]. Although the global sesame planting area is extending, particularly in Africa, the productivity and yield are still very low, resulting in a huge gap between seed demand and supply [34]. In fact, sesame is cultivated in harsh environments and its growth and development are greatly affected by the adverse conditions [35]. Drought, waterlogging, heat and salt stresses represent the leading abiotic factors impairing sesame yield and productivity [36] and several of these environmental stressors often occur in combination at different growth stages. For example, in the arid and semi-arid areas of Africa, America and Asia, extreme heat and drought stresses co-occur and challenge sesame crop. In East and South Asia, waterlogging and salinity stress devastate sesame field. Over the past five years, tremendous efforts have been made to decipher the molecular basis of abiotic stress response and tolerance in sesame [6–8,37,38]. However, none of these studies has performed a meta-analysis of diverse transcriptome data to elucidate the similarities and differences among stress response pathways, and importantly, decode the core abiotic stress-responsive genes (CARG) in sesame.

Within the CARG, understanding how specific transcriptional changes are linked to stress adaptations and identifying central hubs controlling this interaction remain the main challenge. Weighted Gene Co-expression Network Analysis (WGCNA) is one of the widely used computational tools to detect co-expression modules in transcriptome data and identify hub genes playing preponderant functions based on a network construction [39]. Using WGCNA analysis of the RNA-Seq data, Shahan et al. [40] identified key regulators of flower and fruit development in strawberry. WGCNA has also been utilized to detect coexpression modules and major players for multiple biotic and abiotic stresses responses in various plant species, including Arabidopsis, rice, maize, soybean and poplar [41–45].

In this study, we conducted a meta-analysis of 72 RNA-Seq data from drought, salt, osmotic and waterlogging stresses aiming at identifying the CARGs modulating sesame responses to multiple abiotic stresses. By applying WGCNA, we revealed for the first time the co-expressed functional modules within these CARGs and highlighted the major genes to target for sesame improvement towards tolerance to multiple abiotic stresses.

2. Results

2.1. Overview of the Transcriptome Data and Gene Expression Profile under Various Stress Treatments

In this study, we analyzed the global gene expression profiles of sesame under various abiotic stress treatments based on the RNA-Sequencing (RNA-Seq) technology. In total, four datasets from salt stress [8], drought stress [7], waterlogging stress [6] and a newly generated RNA-Seq data under osmotic stress, comprising of 30, 24, 6 and 12 samples, respectively, were investigated (Table 1). In the selected studies, stress-sensitive and tolerant varieties were used. Additionally, all RNA-Seq datasets had three replicates per treatment, untreated controls for each treatment, and the tissue type was primarily the root. After cleaning and filtering the RNA-Seq datasets, we obtained clean reads ranging from 13.5 to 230 Gb with a total of 25,319 uniquely expressed genes among the 72 samples (Table 1).

Table 1. Characteristics of the RNA-Seq datasets used in this study.

Characteristics	Salt	Drought	Waterlogging	Osmotic	Total
Number of samples	30	24	6	12	72
Clean reads (Gb)	230	160	13.5	83.1	486.6
Total expressed genes	23,415	24,113	21,064	22,418	25,319

2.2. Identification of DEGs, Core Conserved DEGs in Response to Abiotic Stress

For each dataset, we compared the gene expression under control condition to the stress condition in order to identify the differentially expressed genes (DEGs). We identified more DEGs under salt stress treatment compared to the other stress treatments, indicating that salt stress triggers the most intense gene regulation in sesame (Figure 1A). A total of 12,784 DEGs were identified among different samples in the datasets. We cross-compared the identified DEGs among the four datasets aiming at identifying the core conserved DEGs in response to all the four abiotic stresses. The result showed that large numbers of DEGs were stress-specific; however, 543 genes constantly participate in sesame responses to its major abiotic stresses and represent the core abiotic stress responsive genes (CARG) (Figure 1B; Table S2; Figure S1). To confirm these CARGs, we selected 10 independent sesame genotypes and evaluated the expression fold change (FC) of 50 randomly selected genes within the CARGs under drought, salt, osmotic, heat and waterlogging stresses compared to the control. As shown in Figure S2, the expression levels of all the tested genes were highly altered under stress ($|FC|>2$), showing that the proposed CARGs are well conserved in sesame and may be functional under a more diverse arrays of abiotic stresses.

We further searched for the enriched transcription factor (TF) families within these CARGs. As shown in Figure 1C, 18 TF families were represented but the ERF, MYB, bHLH and WRKY were overrepresented, denoting that these TF families are the major regulators of sesame responses to multiple abiotic stresses.

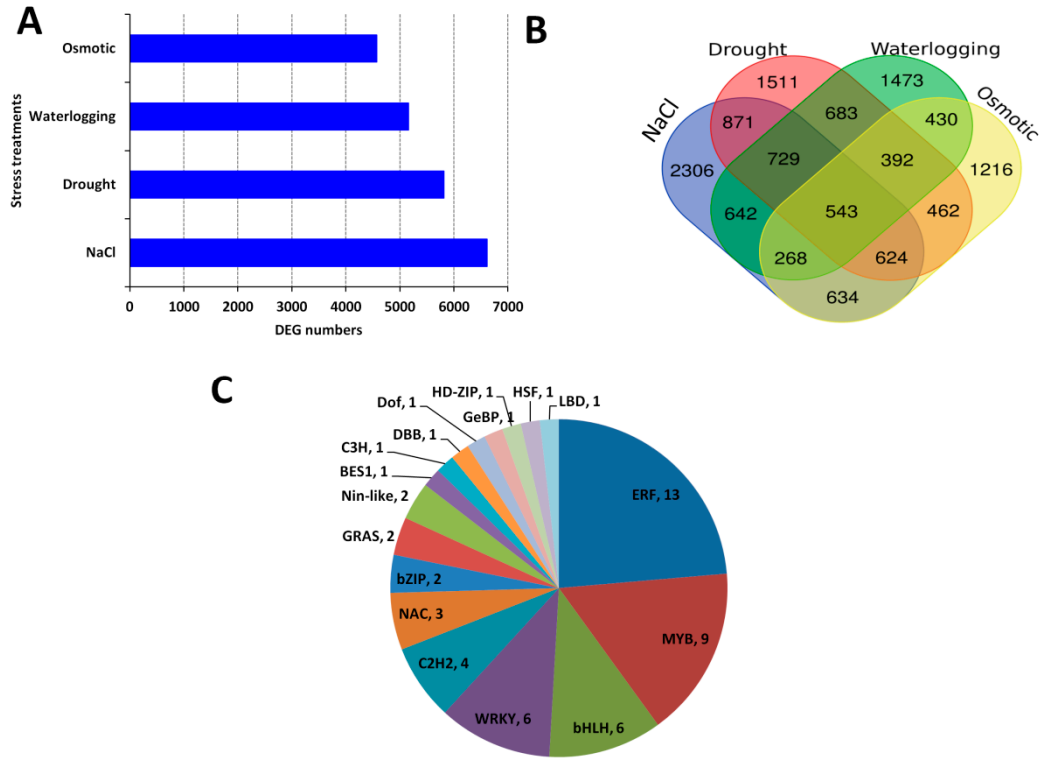


Figure 1. Identification of the core abiotic stress responsive genes (CARG) in sesame. (**A**) Differentially expressed genes detected between control and stress treatments. (**B**) Venn Diagram showing stress specific genes and CARGs. (**C**) Major transcription families enriched in the sesame CARGs.

2.3. WGCNA and Detection of Functional Modules

Weighted gene co-expression network analysis (WGCNA) was conducted on the CARGs to reveal the different modules of co-expressed genes. WGCNA divided the 543 core DEGs into three different modules named as Blue, Turquoise and Grey, containing 113, 276 and 154, genes respectively (Figure 2A,B; Table S3). Association of the detected modules and the abiotic stresses indexes showed that all the three modules respond differently to the abiotic stresses, except for salt stress. In fact, all the three modules were negatively correlated with salt stress ($r = -0.47, -0.84, -0.74$ for Blue, Turquoise and Grey, respectively), suggesting that the CARGs should be down-regulated to allow sesame survive under salt stress (Figure 2C). In addition, with all modules taken together, we could observe that the CARGs engage distinct responses according to the stress, highlighting the key roles played by the regulator genes to shape the stress-specific responses.

To explore the particular biological processes involving the three modules detected by WGCNA, we performed GO enrichment analysis. Blue module related genes represent the basal defense of sesame as evidenced by the enriched GO terms such as 'defense response' and 'response to biotic stimulus' (Figure S3A). Grey module related genes were enriched in the transporter activity role (Figure S3B). Finally, in Turquoise module, 'iron ion binding' and 'heme binding' were detected as the most enriched GO terms, which are well known to be involved in abiotic stress responses in plants [46,47] (Figure S3C). These results further support the premise that the co-expressed gene modules of the sesame CARGs play different functions in response to abiotic stresses.

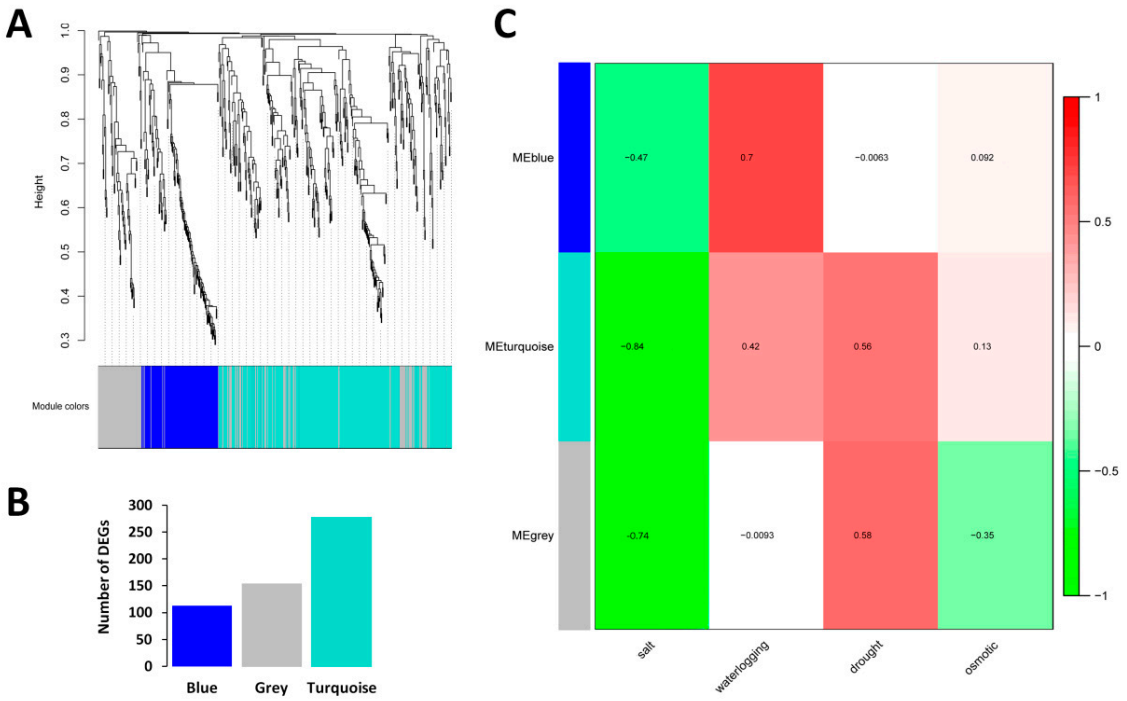


Figure 2. Detection of co-expressed modules in the sesame core abiotic stress responsive genes based on WGCNA. (**A**) Dendrogram showing the different genes clustered into co-expressed modules. (**B**) Number of assigned DEGs to the different modules. (**C**) Association between co-expressed modules and abiotic stresses in sesame. The numbers represent the Pearson correlation coefficients. Positive correlation is colored in red while negative correlation is colored in green.

2.4. Networks Displaying Relationships among Genes within Co-Expressed Modules

To understand the gene interaction within each module, we constructed the gene network using the Cytoscape software. Genes in the different modules were subsequently divided into different clusters, each having network of different number of genes (Figure 3). TFs are represented with different node colors except sky blue and the size of node circle is positively correlated with the number of genes it interacts (Figure 3). Genes with biggest node sizes represent the hub genes. In Blue module, we observed two clusters of genes connected by the gene SIN_1024285 (transmembrane protein 45B-like). We identified several hub genes, including SIN_1003530 (2-aminoethanethiol dioxygenase), SIN_1011060 (pyruvate decarboxylase 1), SIN_1018022 (pentatricopeptide repeat-containing protein), SIN_1003097 (NA), SIN_1003294 (NA), SIN_1006713 (cationic amino acid transporter 1-like), SIN_1013309 (alcohol dehydrogenase 3), SIN_1014647 (squalene monooxygenase), SIN_1017721 (NA), SIN_1024789 (ATP-dependent 6-phosphofructokinase) etc. Furthermore, some key TFs were also present SIN_1005329 (ERF) and SIN_1019627 (WRKY), which may play important regulation role in this module (Figure 3). Turquoise module contains complex clusters of genes, implying diverse biological functions within this module (Figure 3). The key hub genes detected are SIN_1017899 (NA), SIN_1002615 (G-type lectin S-receptor-like serine/threonine-protein kinase), SIN_1005524 (exocyst complex component EXO70A1-like), SIN_1017189 (NA), SIN_1007389 (NA), SIN_1016615 (NA), SIN_1009465 (pectinesterase 1-like), SIN_1021953 (WRKY) and SIN_1013789 (MAPK). In the Grey module, one main cluster surrounded by several small clusters of genes was detected (Figure S4). Several hub genes, including SIN_1002211 (aspartic proteinase nepenthesin-1), SIN_1003190 (protein ECERIFERUM 1), SIN_1019848 (GDSL esterase), SIN_1008670 (18.2 kDa class I heat shock protein), SIN_1014764 (3-ketoacyl-CoA synthase 4) and SIN_1004965 (glucan endo-1,3-β-glucosidase 13), may play preponderant roles in this module.

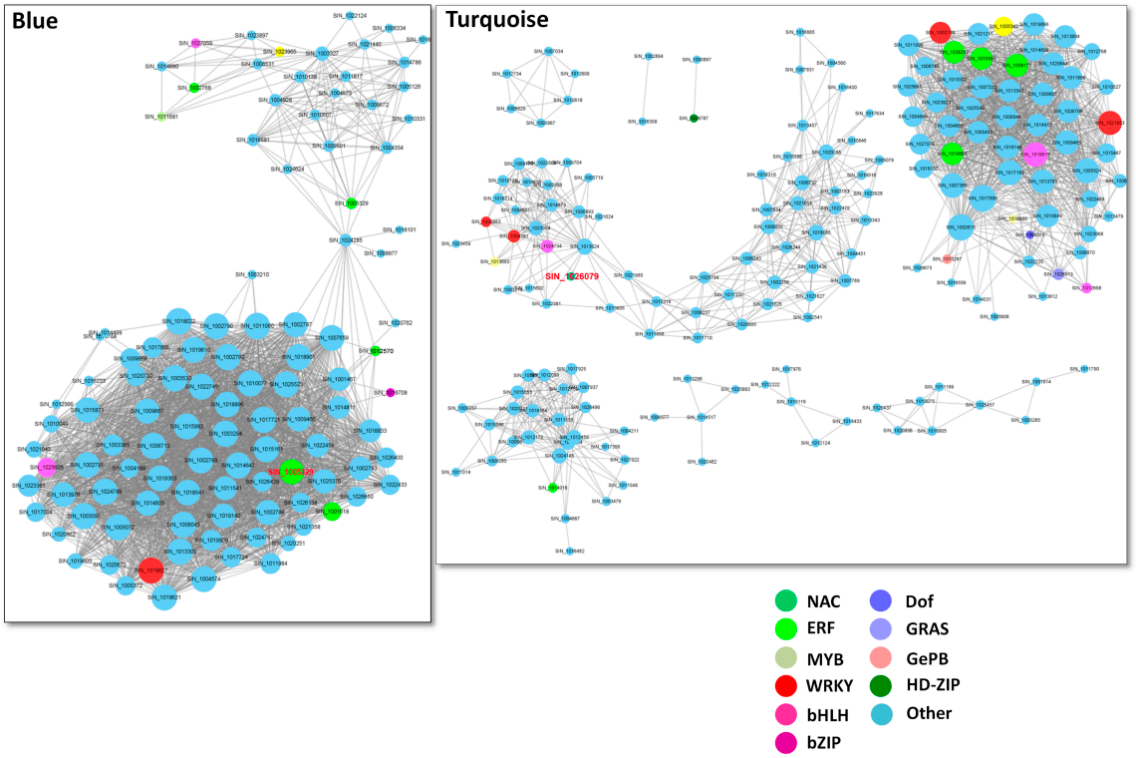


Figure 3. Co-expressed network analysis of Blue module and Turquoise module. The size of node circle is positively correlated with the number of the interacting gene partners. The gene names marked in red are those selected for validation using transgenic Arabidopsis approach.

Next, we extended our analysis to unveil the major regulators of the different co-expressed gene modules of the CARGs in sesame. First, we screened for overrepresented regulatory motifs in the 1 kb promoter regions of genes within each module. Seven TF binding motifs were enriched in the analyzed promoter regions (Table S4). Then, we constructed the gene regulatory networks predicting directional interactions between CARG regulators and targets associated with the three modules using the TF DNA binding motif information. Figure 4 presents the generated regulatory networks, in which the circular nodes represent the key regulators connected by an edge to a module. The size of the nodes is proportional to the number of the inferred regulated genes harboring the corresponding TF binding motifs in the promoter region and the nodes are colored according to the appertaining module. Our predicted networks showed the main TFs regulating gene expression within each module. An intense transcriptional activity was predicted in the Turquoise module, having the highest number of predicted regulators. The networks also highlighted the master players (SIN_1026747 (MYB), SIN_1012698 (bHLH) and SIN_1008018 (Dof)), which are predicted to regulate significant numbers of target genes in their modules and represent potential genes to exploit for the enhancement of sesame tolerance to various abiotic stresses. We infer that these regulators are the key genes that shape the sesame CARG specific responses to the major abiotic stresses.

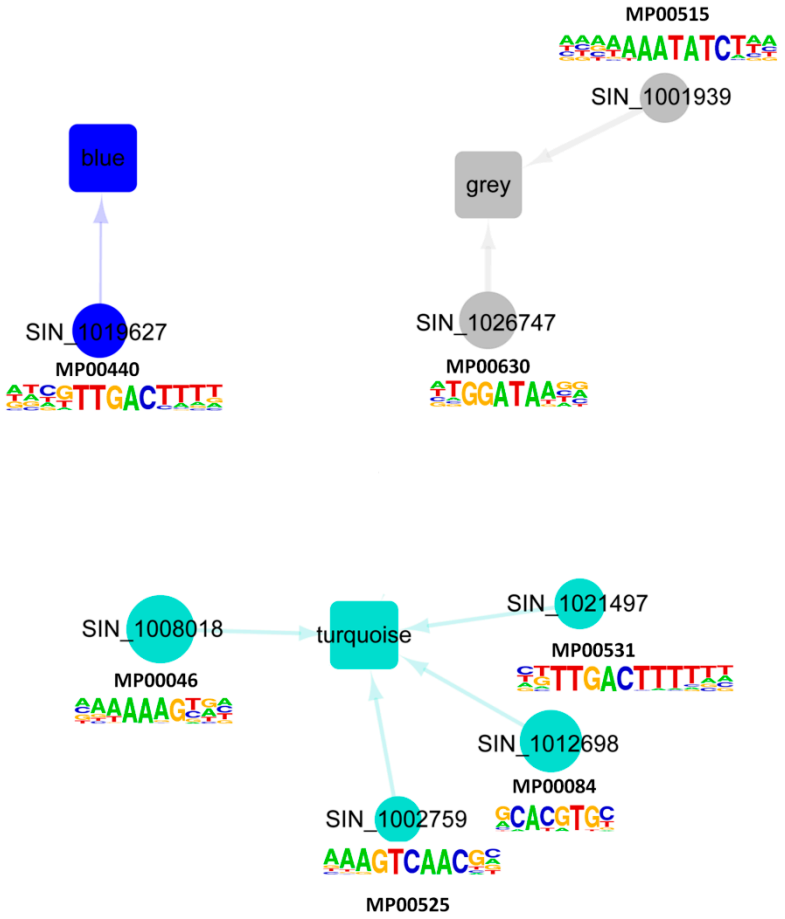


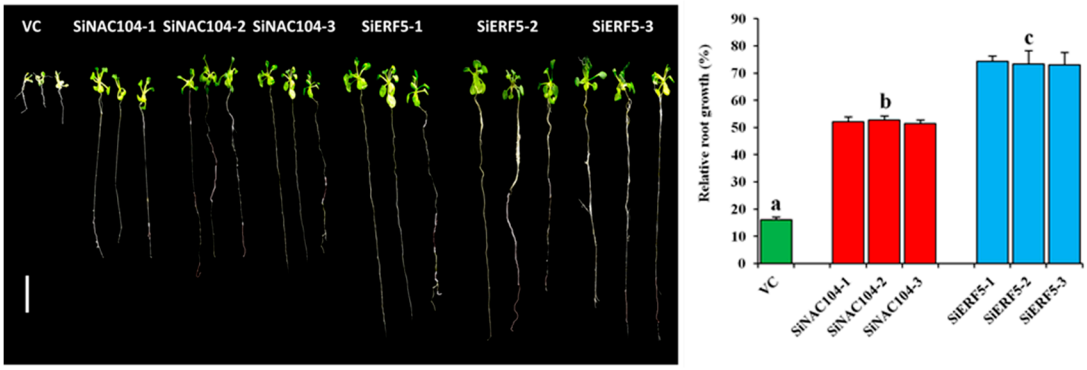
Figure 4. Predicted directional interactions of TFs and the co-expressed modules in the sesame CARGs. Network plots of inferred connections between TF and genes in the three modules. The promoter sequences of genes associated with each module were tested for overrepresentation of DNA motifs shown to be bound to TFs that are differentially transcribed following stress treatments. Each TF with a known motif is represented by a colored circle corresponding to its appertaining module. The different modules are represented by a rectangle. An edge between a TF and a module indicates significant enrichment of the corresponding binding motif in that module. The size of each TF node is proportional to the number of predicted regulated downstream genes. Logos of the seven enriched DNA binding motifs within the promoter regions of the genes belonging to each module detected by WGCNA were added.

2.5. Validation of Hub and Non-Hub TFs from the Co-Expressed Modules of the Sesame CARGs in Transgenic Arabidopsis

We selected two transcription factor encoding genes: SIN_1005329 (SiERF5) and SIN _1026079 (SiNAC104) to confirm their involvement in various abiotic stress responses using Arabidopsis system. In fact, sesame resilience to the genetic manipulation is still significant enough to justify the use of a heterologous system such as *Arabidopsis thaliana*. SiERF5 is a hub gene from Blue module while SiNAC104 is a non-hub gene from Turquoise module (Figure 3) and both were induced at different time points after stress treatments (Table S2). We hypothesized that the over-expression of SiERF5 will confer higher abiotic stress tolerance than the over-expression of SiNAC104 in Arabidopsis, given their contrasting importance in their respective modules.

We generated several T3 homozygous over-expressing Arabidopsis lines for both genes, of which three independent lines for each gene (SiERF5-1, SiERF5-2, SiERF5-3 and SiNAC104-1, SiNAC104-2, SiNAC104-3) were selected for stress applications. Overexpression of SiERF05 increased the leaf biomass which may result from a pleiotropic effect as compared to SiNAC104-overexpressing plants

and the vector control (VC) plants (Figure S5). qRT-PCR was used to confirm the integration and the expression of the transgene (Figure S5). Under osmotic stress induced by 250 mM Mannitol addition to the MS medium, the growth performance of all the transgenic lines was significantly better than the VC plants ($p < 0.001$), indicating that the transgenes confer osmotic stress tolerance (Figure 5A,B). Furthermore, we observed that the SiERF05-overexpressing lines maintained significantly better relative root growth than the SiNAC104-overexpressing lines (Figure 5A,B). Next, we evaluated the performance of the overexpressing lines and VC plants under drought (20 days), salinity (200 mM NaCl) and waterlogging (18 days). As shown in Figure 6A, SiERF5 overexpression had pleiotropic effect including delaying of flowering time and increase of the rosette biomass compared to the SiNAC104-overexpressing lines and VC plants, showing that SiERF5 participates in plant growth and development. This result hints that Blue and Turquoise modules related genes are functionally different, as demonstrated by the GO enrichment analysis. Under stress treatments, all the plants were affected, reflected by the reduced biomass (Figure 6A). However, similar to the osmotic stress treatment, the transgenic lines significantly better sustained drought, salt and waterlogging stresses than the VC plants as evidenced by their higher survival rate, relative rosette fresh weight and relative seed yield (Figure 6B–D). In addition, the results indicated that the two transgenes were more potent under drought and waterlogging stresses compared to the salt stress since we had a more pronounced biomass reduction and no seed yield under salt. Furthermore, we observed a significantly higher tolerance to the different abiotic stresses by SiERF5-overexpressing lines than the SiNAC104-overexpressing lines. Overall, these findings support the argument that the proposed sesame CARGs are functionally active under various abiotic stresses. In addition, this finding supports our formulated hypothesis, denoting that the position of a gene in the co-expressed modules (hub genes or non-hub genes) may reflect the importance of its function, which will guide the choice of high potential genes from the sesame CARGs for germplasm enhancement.



(**A**) (**B**)

Figure 5. Functional characterization of SiERF5- and SiNAC104-overexpressing lines and their counterparts vector control (VC) plants under osmotic stress induced by 250 mM Mannitol addition to the MS medium. (**A**) Phenotypes of the transgenic and VC plants under stress. The bar represents 1 cm (**B**) relative root length estimated as the ratio of the root length recorded after 10 days under stressed and control MS mediums. Values are means ±SD from two independent measurements. Bars with different letters are significantly different ($p \leq 0.05$).

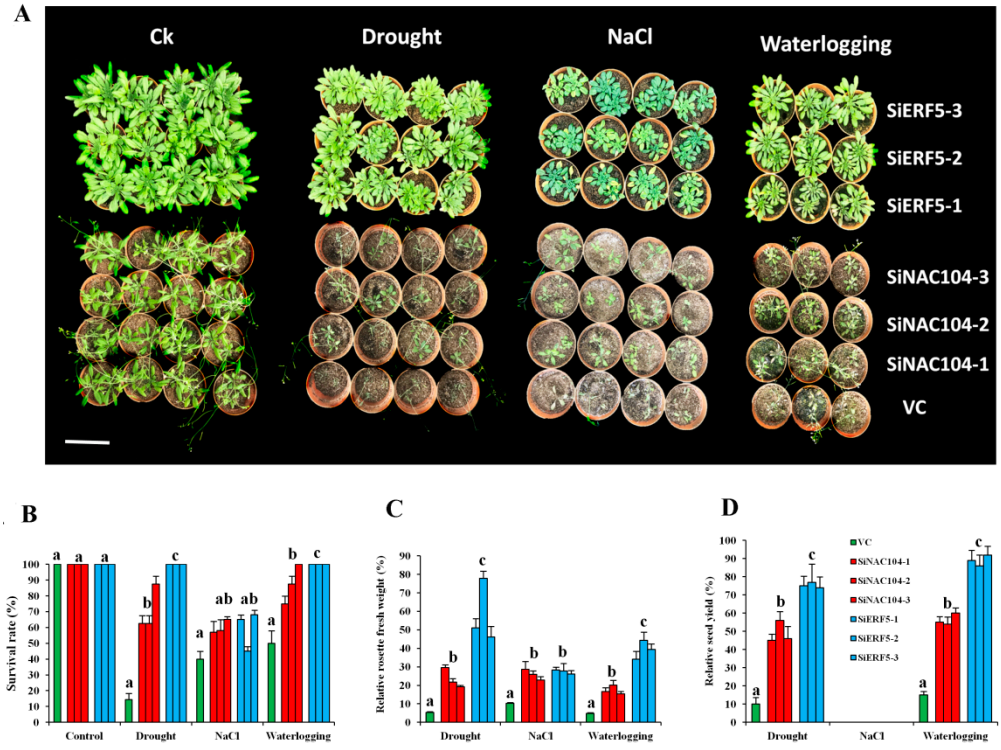


Figure 6. Functional characterization of SiERF5- and SiNAC104-overexpressing lines and their counterparts vector control (VC) plants under drought (20 days), salt (200 mM NaCl) and waterlogging (18 days). (**A**) Phenotypes of the plants. The bar represents 8 cm (**B**) Survival rate of the plants. (**C**) Relative rosette fresh weight. (**D**) Relative seed yield. The relative values are estimated as the ratio of the data recorded after stress period in stressed and control plants. Values are means ± SD from two independent measurements. For each treatment, bars with different letters are significantly different ($p \leq 0.05$).

3. Discussion

Reinforcing crop's tolerance to abiotic stress has become a hot issue in the current scenario of climate change, which is boosting extreme weathers, posing a growing threat to sustainable agriculture. Because of the multitude of environmental stresses affecting crop survival and productivity in the field, the identification of potential genes conferring tolerance to multiple abiotic stresses is highly desirable [9,48]. In this study, we employed various abiotic stresses RNA-Seq datasets (waterlogging, drought, salt and osmotic stresses) from diverse sesame genotypes with contrasting levels of tolerance. Our meta-analysis unveiled 543 genes as the core abiotic stress responsive genes (CARG) modulating sesame responses to multiple abiotic stresses. We validated a subset of these CARGs in ten independent sesame genotypes, showing that these CARGs are not genotype-dependent but are well conserved in the sesame species. The alteration of the expression levels of selected CARGs under heat stress further hints that the proposed CARGs may be functional under a more diverse set of environmental stressors. Nonetheless, our studied transcriptome data were mainly from sesame root samples; hence, additional analyses are needed to check how the proposed CARGs respond to stress in other tissues. Very recently, Cohen and Leach [27] also demonstrated that meta-analysis of diverse transcriptomic data sets in rice is a valid and robust approach to develop hypotheses for how plants respond to various stress. They discovered a list of novel candidate genes for improving rice environmental stress tolerance, which were not detected in studies focused on a single stress. Similarly, meta-analysis of transcriptome data has been performed to find out key genes in responses to abiotic stresses in other important plants [45,49–52].

The proposed sesame CARGs are functionally diversified as evidenced by the various biological pathways contributed by these genes (Figure S3). Several genes within the sesame CARGs are universally known to be abiotic stress responsive genes. For example, we detected the gene *SIN_1012768* predicted to be a member of the late embryogenesis abundant (LEA) family. Proteins encoded by the LEA are demonstrated to play defensive roles in plants during abiotic stresses, including cold, drought, salinity [53–57]. Other well-known abiotic stress marker genes are the alcohol dehydrogenase (ADH) family members, which are key enzymes responsible for catalyzing the reduction of acetaldehyde to ethanol using NADH as reductant, particularly during the periods of anaerobic stress [58,59]. ADH genes are involved in various environmental stresses such as drought cold, salinity, hypoxia and pathogen infection [60–64]. In the present study, we identified one ADH gene (*SIN_1013309*) in the sesame CARGs. Besides, we also detected several important plant abiotic stress marker genes within the sesame CARGs, including peroxidase (*SIN_1026962* and *SIN_1013457*), DREB (*SIN_1015595*), universal stress protein (*SIN_1022749* and *SIN_1012609*), glutathione *S*-transferase (*SIN_1017866* and *SIN_1002858*), 1-aminocyclopropane-1-carboxylate oxidase (*SIN_1024757*, *SIN_1007659*, *SIN_1023068*, *SIN_1009668* and *SIN_1026934*), heat shock protein (*SIN_1008669*, *SIN_1008679*, *SIN_1008672* and *SIN_1008670*) [65–81] etc. In future abiotic stress treatment experiments in sesame, we propose to select some of these well-known CARGs as abiotic stress marker genes in order to gauge the stress levels. Noteworthy, several uncharacterized genes were present in the CARGs, providing an exciting gene repertoire to further illuminate the complex mechanisms of plant responses to multiple abiotic stresses.

We employed the weighted gene co-expression network analysis (WGCNA) to break down the sesame CARGs into three functional modules (Figure 2A,B). Interestingly, the functional characterization of these three modules revealed that they are involved in distinct biological pathways in response to abiotic stresses (Figure S3). With the WGCNA package, we correlated the different abiotic stressors to the gene modules (Figure 2C). This analysis is cardinal because it allowed the identification of the synergistic and antagonistic gene modules of abiotic stress response in sesame. We found that the co-expressed modules of the sesame CARGs globally displayed positive correlations with waterlogging, drought and osmotic stresses, but they were all negatively correlated with salinity stress. This suggests that manipulation of master genes of these modules to simultaneously enhance tolerance to all the four investigated abiotic stresses may not be possible in sesame, because enhancing tolerance to waterlogging, drought and osmotic stresses will lead to an increase sensitivity to salinity stress. Our findings are not surprising, since previous studies have also shown plant antagonistic responses to some stresses [9,14].

Transcription factors (TF) are regulatory molecules that play central roles in gene transcription and promote plant adaptation to various environmental conditions. The sesame CARGs contained several TFs, with ERF, MYB, bHLH and WRKY being the more predominant families (Figure 1C). A similar meta-analysis in cotton also underscored the important role of these TF families in abiotic stress responses [52], indicating a conserved abiotic stress gene regulation mechanism in plants. ERF family has been one the most studied genes in plants. Extensive studies have shown that ERF genes are essential in responses to a wide range of abiotic stresses mediated by the plant hormone ethylene [82–86]. It has been reported that MYB TFs also play prominent roles in triggering the right response upon exposure of plants to abiotic stresses through the ABA-dependent and independent pathways (reviewed by Li et al. [87]; Roy [88]). The WRKY genes are among the top four TF families highly active in the transcriptional reprogramming during stress and act principally through the ABA mediated pathways [73,89]. Conversely, the role and regulatory mechanisms of bHLH genes in plant abiotic stresses responses are still elusive [90]. Therefore, the sesame bHLH genes detected as key regulators of abiotic stress responsive genes in this study may represent candidate genes for the elucidation of bHLH abiotic stress regulation mechanism in plants. Overall, the diversity of TFs within the sesame CARGs highlights the complex network of interacting pathways which shape the responses to multiple abiotic stresses. To further pinpoint the master players among the large number of detected TFs, we identified in the promoter of genes within each module, the enriched putative *cis*-regulatory

motifs. Previous works in yeast and human have demonstrated that genes with similar expression patterns are regulated by the same set of TFs, and therefore are likely to have similar *cis*-regulatory motifs in their upstream promoter regions [91,92]. Our study corroborated well these conclusions and unraveled for each module the master TFs that may regulate the gene expression under specific abiotic stress (Figure 4).

In gene networks, many genes only interact with a limited number of other genes, whereas a smaller subset of genes (hub genes) interacts with many other genes and therefore has a more central role [93]. Hub genes are expected to play preponderant and essential functions in organism's fitness and according to Jeong et al. [94], hub genes are three times more likely to be essential than genes with fewer interaction partners. To test this hypothesis in our predicted gene networks (Figure 3), we selected a hub gene (*SiERF5*) and a non-hub gene (*SiNAC104*), both being transcription factors. Over-expression of these two genes in *Arabidopsis thaliana* resulted in an enhanced tolerance to drought, waterlogging and osmotic stresses, but the over-expressing lines did not tolerate salinity stress (Figures 5 and 6). Furthermore, we observed that the transgenic lines over-expressing the hub gene had a stronger fitness and a higher performance under abiotic stresses compared to those transformed with the non-hub gene. It is worth mentioning that the over-expression of the hub gene had clear pleiotropic effects beyond abiotic stress responses in *Arabidopsis thaliana*, thus might play a central role in various biological pathways. This experiment therefore highlighted three key findings: (1) perturbation of a hub gene is likely to have a major fitness consequence than a non-hub gene; (2) proper manipulation of sesame CARGs may confer tolerance to multiple abiotic stresses; (3) genetic manipulation for generating sesame lines tolerant to all the four investigated abiotic stresses may be challenging due to the antagonistic response of the sesame CARGs in the face of some abiotic stresses. Although the main goal of this work was not to investigate the functional importance of hub genes versus non-hub genes, the preliminary result obtained from the Arabidopsis mutants will fuel a future study based on both sesame and Arabidopsis using over-expression and knock-out transgenic techniques and employing more hub genes and non-hub genes to comprehensively elucidate this important scientific question.

4. Materials and Methods

4.1. RNA-Sequencing Datasets of Abiotic Stressed Sesame Samples

In order to decipher the core genome involved in abiotic stress responses in plants, our group previously generated several RNA-Seq data of sesame under waterlogging [6], drought [7], salt [8] and details of the experimental procedures could be found in the respective articles. In this study, we collected the root RNA-Seq data of the waterlogging-tolerant genotype Zhongzhi13 under 3 h waterlogging stress and control condition (SRR2886790). Concerning the drought stress, we collected the root RNA-Seq data from a drought-tolerant genotype ZZM0635 (SAMN06130606) and a drought-sensitive genotype ZZM4782 (SAMN06130607) after 0, 3, 7 and 11 days drought stress. For the salt stress experiment, two contrasting genotypes (salt-tolerant WZM3063 and salt-sensitive ZZM4028) were treated with 150mM NaCl and whole seedling samples were collected at 0, 2, 6, 12 and 24 h (PRJNA524278). In addition to these released datasets, we newly generated an RNA-Seq data from root samples of osmotic stressed sesame (PRJNA552167). Two sesame genotypes (osmotic-tolerant G546 and osmotic-sensitive G259) grown in a box containing 9 L of half-strength Hoagland solution for 21 days, were treated with 2% PEG6000 for 7 days. Samples were collected in triplicate from stress and control conditions after the stress period, immediately placed in liquid nitrogen and stored at −80 °C until use.

4.2. Total RNA Isolation and Sequencing from the PEG6000-Treated Seedlings

Total RNA of the 12 PEG6000-treated samples was extracted using an EASYspin Plus kit (Aidlab, Beijing, China). The cDNA libraries generated from RNA samples were pair-end sequenced on an

Illumina Hiseq 4000 platform (San Diego, California, CA, USA.) according to the methods described by Dossa et al. [7].

4.3. RNA-Seq Data Analysis

A total of 72 RNA-Seq data, including 30, 24, 6 and 12 data from salt, drought, waterlogging and osmotic stresses, respectively, were used in this study. The raw data were first processed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to filter out adapters and low-quality sequences. The raw data was submitted to NCBI SRA (PRJNA552167). Then, the clean reads were mapped to the sesame genome v.1.0 (https://www.ncbi.nlm.nih.gov/genome/?term=sesamum) [95] using HISAT [96]. The RSEM package v1.3.0 [97] was used to calculate gene expression level for each sample expressed as fragments per kilobase of transcript per million fragments mapped (FPKM). For each treatment, the gene expression levels in the stressed samples were compared with those in the control samples in order to identify the differentially expressed genes (DEG). The DEGs were detected as described by Tarazona et al. [98] based on the parameters: Fold change ≥ 2 and Probability ≥ 0.8, with a significant false discovery rate-adjusted p value (FDR) < 0.05 based on the three biological replicates. Gene Ontology enrichment analysis for the DEGs was performed using the clusterProfiler version 3.8.

4.4. Analysis of Co-Expression Modules Based on WGCNA

Weighted Gene Co-Expression Network Analysis (WGCNA) package version: 1.61 [39] in the R software (http://www.r-project.org/) was used to construct the gene co-expression networks from the normalized log2-transformed FPKM matrix as described by Wan et al. [99] and Yang et al. [100]. This analysis was based on the core conserved DEGs between the four abiotic stress treatments. The gradient method was used to test the independence and the average connectivity degree of different modules with different power value (the power value ranging from 1 to 20). The appropriate power value was determined when the degree of independence was 0.8. Then, the adjacency was transformed into a topological overlap matrix (TOM), which could measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for network generation, and the corresponding dissimilarity (1-TOM) was calculated. To classify genes with similar expression profiles into gene modules, average linkage hierarchical clustering was conducted according to the TOM-based dissimilarity measure with a minimum size (gene group) of 50 for the genes' dendrogram.

Module-trait associations were estimated using the correlation between the module eigengene and the stress treatments. Network visualization for each module was performed using the Cytoscape software version 3.6.1 [101] with a cut off of the weight parameter (obtained from the WGCNA) set at 0.30.

4.5. Enrichment Analysis of Cis-Regulatory Motifs

To detect the enriched *cis*-regulatory motifs within the promoters of the genes belonging to each module detected by WGCNA, first, all the sesame transcription factor binding motifs were downloaded from the JASPAR website (http://jaspar.genereg.net/, [102]. Then, the sequences of 1 kb upstream from the transcription start sites of the genes belonging to each module detected by WGCNA were retrieved from the sesame genome v1.0 [95]. The promoter regions were scanned for presence/absence of the DNA binding motifs using the FIMO tool v5.0.3 [103] with a threshold ($p < 0.01$). Finally, we performed an enrichment analysis of each motif within the promoters of the genes belonging to each module using the cumulative hypergeometric distribution ($p \leq 0.05$).

4.6. Vector Construction and Arabidopsis Genetic Transformation

We selected two genes, including a hub gene (*SiERF5*, *SIN_1005239*) and a no-hub gene (*SiNAC104*, *SIN_1026079*) to demonstrate their involvement in abiotic stress responses in *Arabidopsis thaliana* following descriptions of Dossa et al. [38]. Briefly, the protein coding region were cloned by PCR

from sesame root cDNA (SiERF5-F-GCTTTCGCGAGCTCGGTACCATGAGAATGATTCTCAAGAA, SiERF5-R-CGACTCTAGAGGATCCTGTCAAGTGAGATGGTTTGA); (SiNAC104-F- GCTTTCGCG AGCTCGGTACCATGGCTGAAGGGAGGAAATG, SiNAC104-R- CGACTCTAGAGGATCCAGAT CAGCTTGCCTAACTAG) and inserted into a pCAMBIA 1301s vector (which is a modified form of the pCAMBIA1301 vector) between KPnI (5′-end) and BamHI (3′-end) sites, driven by the CaMV 35S promoter. The plasmids containing the 35S::SiERF5 and 35S::SiNAC104 constructs were transformed first into *Agrobacterium tumefaciens* strain LBA4404 and then into Arabidopsis ecotype Col-0 cv. Columbia by the floral dipping method [104]. Transgenic seeds were screened by sowing on MS medium containing 1% agar, and 1% sucrose and 50 $\mu g \cdot ml^{-1}$ hygromicin. All the putative T1 transgenic plants and vector control (VC) plants (containing an empty pCAMBIA1301 vector) were screened by PCR with genomic DNA from leaves. Furthermore, qRT-PCR was performed to confirm the expression of the transgene [105]. Three independent T3 transgenic homozygous lines were used for the stress treatments, gene expression assay and phenotypic analyses.

4.7. Evaluation of Transgenic Lines Exposed to Osmotic, Salt, Waterlogging and Drought Stresses

First, to analyze the response of the transgenic plants to osmotic stress, seeds of VC and three T3 lines for each transgene, were surface sterilized and plated on solid MS medium. The seeds were stratified for 2 days in the dark at 4 °C and then transferred to a growth chamber under a 16-h light period (long-day condition) provided by fluorescent light at 120 $\mu M \cdot m^{-2} \cdot s^{-1}$ and day/night temperatures of 22/16 °C and 60/75% relative humidity. 10 days-old seedlings were transferred into solid MS medium supplemented with 0/250 mM Mannitol. Plates were placed vertically and after 10 days, seedling root length was recorded.

Next, 10-day-old seedlings (transgenic lines and VC plants) were transferred into pots (two plants per pot) containing organic potting mix and grown in normal conditions for 15 days. Then, 1/4 of the pots were subjected to dehydration stress by withholding watering for 20 days and subsequently, plants were allowed to recover for 1 week by supplementing water [38]. Another 1/4 of the pots were watered with 200 mM NaCl solution every three days for four times and subsequently, plants were allowed to recover for 1 week by supplementing water [38]. Another 1/4 of the pots were subjected to waterlogging stress for 18 days. Pots were placed inside plastic tanks and filled with tap water up to 5 mm above the ground [106]. After waterlogging stress, plants were allowed to recover for 1 week by drainage. The remaining plants (1/4) were kept under normal growth condition throughout the experiment. Leaf samples were collected at the end of each stress treatment and in the control condition for gene expression analysis. After recovery, the plant survival rate, the above-ground rosette biomass fresh weights were recorded and pictures were captured to show visible phenotypes. We estimated the relative rosette biomass as the ratio of the records under stress and control conditions. For each treatment, eight survived plants (four pots) were kept until maturation to evaluate the seed yield. The experiment was repeated twice with four replicates in each experiment for statistical analysis.

4.8. Sesame Materials and Stress Treatments

Ten cultivars of sesame were obtained from the China National Genebank, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences and used in this experiment. The genotypes G059, G079, G207, G208, G209, G210, G212, G213, G214 and G215, all originating in China, were used in this experiment. The sesame seeds were sterilized with 3% sodium hypochlorite for 7 min and washed three times using sterile water. For the drought experiment, the seeds were sown in pots containing loam soil mixed with 10% vermiculite and plants were regularly watered. After 6 weeks, seedlings were submitted to a water stress for 7 days [7]. For the waterlogging treatment, seedlings were flooded by standing in a plastic bucket filled with tap water to 3 cm above the soil surface and maintained for 9 h according to the experimental descriptions of Dossa et al. [105]. Concerning the salt, heat and osmotic stress treatments, seedlings were hydroponically grown in a box containing half-strength Hoagland solution for 2 weeks under ambient temperature of 35 °C. Then, they were

transferred to a new nutrient solution containing 200 mM NaCl for 48 h (salt stress treatment), in a nutrient solution containing 2% PEG6000 for 5 days (osmotic stress treatment) or under 45 °C for 48 h (heat stress treatment). Root samples of stressed and control plants were collected at the same periods.

4.9. Gene Expression Analysis in Arabidopsis

The qRT-PCR was performed on RNA extracted from leaf samples (Arabidopsis) and root samples (sesame). The genes *Actin 2* (*AT3G18780*) and *Actin 7* (*SIN_1006268*) were used as the internal control for Arabidopsis and sesame, respectively. Specific primer pairs of the assayed genes were designed using the Primer5.0 software [107] (Table S1). Samples in the control condition (non-stress) were used as reference and data are presented as relative transcript level based on the $2^{-\Delta\Delta Ct}$ method [108].

4.10. Statistical Analysis

All the data were analyzed with the R software (www.r-project.org) using the one-way analysis of variance for significant difference statistical analysis. The error bars were calculated with data from two independent experiments.

5. Conclusions

Using the meta-analysis approach coupled with the weighted gene co-expression network analysis on 72 RNA-Seq datasets from drought, salt, osmotic and waterlogging treatments, we decoded the core abiotic stress responsive genes (CARG) in sesame. In total, 543 genes were detected as the sesame CARGs, some of which were experimentally validated. The CARGs were further divided into three distinct functional modules, which are involved a wide range of biological pathways. Module-traits association analysis provided insights into the synergistic and antagonistic gene modules of abiotic stress response mechanisms in sesame. The stress specific expression patterns of genes within the different modules are tightly regulated by key transcription factors from the families of ERF, WRKY, MYB and bHLH. Moreover, a set of hub genes and master regulators predicted to play prominent functions for abiotic stresses responses in sesame was identified, representing useful resources of molecular biomarkers and highly-anticipated potential candidate genes for engineering multiple stresses tolerance in sesame.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/16/3930/s1.

Author Contributions: K.D. and X.Z. conceived, designed and supervised the experiment; K.D., M.A.M., T.Z., R.S., Y.Z., R.Z., J.Y., L.W. and X.Z. performed the experiments. K.D., M.A.M. and T.Z. conducted data analyses. K.D. wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding: This research was funded by the China Agriculture Research System (CARS-14), the Agricultural Science and Technology Innovation Project of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2016-OCRI) and the Central Public-interest Scientific Institution Basal Research Fund (1610172018007).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Foley, J.A.; Ramankutty, N.; Brauman, K.A.; Cassidy, E.S.; Gerber, J.S.; Johnston, M.; Mueller, N.D.; O'Connell, C.; Ray, D.K.; West, P.C.; et al. Solutions for a cultivated planet. *Nature* **2011**, *478*, 337–342. [CrossRef] [PubMed]
2. Tester, M.; Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **2010**, *327*, 818–822. [CrossRef] [PubMed]
3. Wang, W.; Vinocur, B.; Altman, A. Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta* **2003**, *218*, 1–14. [CrossRef] [PubMed]

4. Nakashima, K.; Yamaguchi-Shinozaki, K. Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. *Physiol. Plant* **2006**, *126*, 62–71. [CrossRef]
5. Shanker, A.K.; Maheswari, M.; Yadav, S.K.; Desai, S.; Bhanu, D.; Attal, N.B.; Venkateswarlu, B. Drought stress responses in crops. *Funct. Integr. Genom.* **2014**, *14*, 11–22. [CrossRef] [PubMed]
6. Wang, L.; Li, D.; Zhang, Y.; Gao, Y.; Yu, J.; Wei, X.; Zhang, X. Tolerant and susceptible sesame genotypes reveal waterlogging stress response patterns. *PLoS ONE* **2016**, *11*, e0149912. [CrossRef] [PubMed]
7. Dossa, K.; Li, D.; Wang, L.; Zheng, X.; Liu, A.; Yu, J.; Wei, X.; Zhou, R.; Fonceka, D.; Diouf, D.; et al. Transcriptomic, biochemical and physio-anatomical investigations shed more light on responses to drought stress in two contrasting sesame genotypes. *Sci. Rep.* **2017**, *7*, 8755. [CrossRef] [PubMed]
8. Zhang, Y.; Li, D.; Zhou, R.; Wang, X.; Dossa, K.; Wang, L.; Zhang, Y.; Yu, J.; Gong, H.; Zhang, X.; et al. Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress. *BMC Plant Biol.* **2019**, *1*, 66. [CrossRef]
9. Mittler, R. Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* **2006**, *11*, 15–19. [CrossRef]
10. Prasad, P.V.V.; Pisipati, S.R.; Momcilovic, I.; Ristic, Z. Independent and combined effects of high temperature and drought stress during grain filling on plant yield and chloroplast EF-Tu expression in spring wheat. *J. Agron. Crop Sci.* **2011**, *197*, 430–441. [CrossRef]
11. Takeda, S.; Matsuoka, M. Genetic approaches to crop improvement: Responding to environmental and population changes. *Nat. Rev. Genet.* **2008**, *9*, 444–457. [CrossRef] [PubMed]
12. Newton, A.C.; Johnson, S.N.; Gregory, P.J. Implications of climate change for diseases, crop yields and food security. *Euphytica* **2011**, *179*, 3–18. [CrossRef]
13. Hu, W.; Ding, Z.; Tie, W.; Yan, Y.; Liu, Y.; Wu, C.; Liu, J.; Wang, J.; Peng, M.; Xu, B.; et al. Comparative physiological and transcriptomic analyses provide integrated insight into osmotic, cold, and salt stress tolerance mechanisms in banana. *Sci. Rep.* **2017**, *7*, 43007. [CrossRef] [PubMed]
14. Atkinson, N.J.; Urwin, P.E. The interaction of plant biotic and abiotic stresses: From genes to the field. *J. Exp. Bot.* **2012**, *63*, 3523–3543. [CrossRef] [PubMed]
15. Causton, H.C.; Ren, B.; Koh, S.S.; Harbison, C.T.; Kanin, E.; Jennings, E.G.; Lee, T.I.; True, H.L.; Lander, E.S.; Young, R.A. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **2001**, *12*, 323–337. [CrossRef] [PubMed]
16. Chen, D.; Toone, W.M.; Mata, J.; Lyne, R.; Burns, G.; Kivinen, K.; Brazma, A.; Jones, N.; Bahler, J. Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* **2003**, *14*, 214–229. [CrossRef] [PubMed]
17. Gasch, A.P. Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast* **2007**, *24*, 961–976. [CrossRef] [PubMed]
18. Lai, L.C.; Kissinger, M.T.; Burke, P.V.; Kwast, K.E. Comparison of the transcriptomic "stress response" evoked by antimycin A and oxygen deprivation in *Saccharomyces cerevisiae*. *BMC Genom.* **2008**, *9*, 627. [CrossRef]
19. Kim, D.; Kim, M.S.; Cho, K.H. The core regulation module of stress-responsive regulatory networks in yeast. *Nucleic Acids Res.* **2012**, *40*, 8793–8802. [CrossRef]
20. Kilian, J.; Whitehead, D.; Horak, J.; Wanke, D.; Weinl, S.; Batistic, O.; D'Angelo, C.; Bornberg-Bauer, E.; Kudla, J.; Harter, K. The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J. Cell Mol. Biol.* **2007**, *50*, 347–363. [CrossRef]
21. Wanke, D.; Berendzen, K.W.; Kilian, J.; Harter, K. Insights into the Arabidopsis abiotic stress response from the AtGenExpress expression profile dataset. In *Plant Stress Biology*; Hirst, H., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010. [CrossRef]
22. Hahn, A.; Kilian, J.; Mohrholz, A.; Ladwig, F.; Peschke, F.; Dautel, R.; Harter, K.; Berendzen, K.W.; Wanke, D. Plant core environmental stress response genes are systemically coordinated during abiotic stresses. *Int. J. Mol. Sci.* **2013**, *14*, 7617–7641. [CrossRef] [PubMed]
23. Breton, G.; Danyluk, J.; Charron, J.B.; Sarhan, F. Expression profiling and bioinformatic analyses of a novel stress-regulated multispanning transmembrane protein family from cereals and Arabidopsis. *Plant Physiol.* **2003**, *132*, 64–74. [CrossRef] [PubMed]

24. Mangelsen, E.; Kilian, J.; Harter, K.; Jansson, C.; Wanke, D.; Sundberg, E. Transcriptome analysis of high-temperature stress in developing barley caryopses: Early stress responses and effects on storage compound biosynthesis. *Mol. Plant* **2011**, *4*, 97–115. [CrossRef] [PubMed]
25. Le, D.T.; Nishiyama, R.; Watanabe, Y.; Tanaka, M.; Seki, M.; Ham le, H.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Tran, L.S. Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS ONE* **2012**, *7*, e49522. [CrossRef] [PubMed]
26. Priest, H.D.; Fox, S.E.; Rowley, E.R.; Murray, J.R.; Michael, T.P.; Mockler, T.C. Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress. *PLoS ONE* **2014**, *9*, e87499. [CrossRef] [PubMed]
27. Cohen, S.P.; Leach, J.E. Abiotic and biotic stresses induce a core transcriptome response in rice. *Sci. Rep.* **2019**, *9*, 6273. [CrossRef]
28. Noguchi, T.; Ikeda, K.; Sasaki, Y.; Yamamoto, J.; Yamori, Y. Effects of vitamin E and sesamin on hypertension and cerebral thrombogenesis in stroke-prone spontaneously hypertensive rats. *Hypertens. Res.* **2001**, *24*, 735–742. [CrossRef]
29. Sankar, D.; Sambandam, G.; Ramakrishna, R.M.; Pugalendi, K.V. Modulation of blood pressure, lipid profiles and redox status in hypertensive patients taking different edible oils. *Clin. Chem. Acta* **2005**, *355*, 97–104. [CrossRef]
30. Cheng, F.C.; Jinn, T.R.; Hou, R.C.W.; Tzen, J.T.C. Neuroprotective effects of sesamin and sesamolin on gerbil brain in cerebral ischemia. *Int. J. Biomed. Sci.* **2006**, *2*, 284–288.
31. Hibasami, H.; Fujikawa, T.; Takeda, H.; Nishibe, S.; Satoh, T.; Fujisawa, T.; Nakashima, K. Induction of apoptosis by Acanthopanax senticosus HARMS and its component, sesamin in human stomach cancer KATO III cells. *Oncol. Rep.* **2000**, *7*, 1213–1216. [CrossRef]
32. Miyahara, Y.; Hibasami, H.; Katsuzaki, H.; Imai, K.; Komiya, T. Sesamolin from sesame seed inhibits proliferation by inducing apoptosis in human lymphoid leukemia Molt 4B cells. *Int. J. Mol. Med.* **2001**, *7*, 369–371. [CrossRef]
33. Dossa, K.; Diouf, D.; Wang, L.; Wei, X.; Zhang, Y.; Niang, M.; Fonceka, D.; Yu, J.; Mmadi, M.A.; Yehouessi, L.W.; et al. The emerging oilseed crop *Sesamum indicum* enters the "Omics" era. Front. *Plant Sci.* **2017**, *8*, 1154.
34. Sarkar, P.K.; Khatun, A.; Singha, A. Effect of duration of water-logging on crop stand and yield of sesame. *Int. J. Innov. Appl. Stud.* **2016**, *14*, 1–6.
35. Witcombe, J.R.; Hollington, P.A.; Howarth, C.J.; Reader, S.M.; Steele, K. Breeding for abiotic stresses for sustainable agriculture. *Philos. Trans. B* **2008**, *363*, 703–716. [CrossRef]
36. Zhang, H.; Miao, H.; Ju, M. Potential for adaptation to climate change through genomic breeding in sesame. In *Genomic Designing of Climate-Smart Oilseed Crops*; Kole, C., Ed.; Springer Nature: Switzerland, Basel, 2019; pp. 371–444. [CrossRef]
37. Li, D.; Dossa, K.; Zhang, Y.; Wei, X.; Wang, L.; Zhang, Y.; Liu, A.; Zhou, R.; Zhang, X. GWAS uncovers differential genetic bases for drought and salt tolerances in sesame at the germination stage. *Genes* **2018**, *9*, 87. [CrossRef]
38. Dossa, K.; Li, D.; Yu, J.; Wang, L.; Zhang, Y.; You, J.; Zhou, R.; Mmadi, M.A.; Li, A.; Fonceka, D.; et al. The genetic basis of drought tolerance in the high oil crop *Sesamum indicum*. *Plant Biotechnol. J.* **2019**, 1–16. [CrossRef]
39. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]
40. Shahan, R.; Zawora, C.; Wight, H.; Sittmann, J.; Wang, W.; Mount, S.M.; Liu, Z. Consensus coexpression network analysis identifies key regulators of flower and fruit development in wild strawberry. *Plant Physiol.* **2018**, *178*, 202–216. [CrossRef]
41. Childs, K.L.; Davidson, R.M.; Buell, C.R. Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS ONE* **2011**, *6*, e22196. [CrossRef]
42. Weston, D.J.; Karve, A.A.; Gunter, L.E.; Jawdy, S.S.; Yang, X.; Allen, S.M.; Wullschleger, S.D. Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. *Plant Cell Environ.* **2011**, *34*, 1488–1506. [CrossRef]
43. Downs, G.S.; Bi, Y.M.; Colasanti, J.; Wu, W.; Chen, X.; Zhu, T.; Rothstein, S.J.; Lukens, L.N. A developmental transcriptional network for *Zea mays* defines coexpression modules. *Plant Physiol.* **2013**, *161*, 1830–1843. [CrossRef]

44. Ma, S.; Ding, Z.; Li, P. Maize network analysis revealed gene modules involved in development, nutrients utilization, metabolism, and stress response. *BMC Plant Biol.* **2017**, *17*, 131. [CrossRef]
45. Shen, P.; Hour, A.; Liu, L.D. Microarray meta-analysis to explore abiotic stress-specific gene expression patterns in Arabidopsis. *Bot. Stud.* **2017**, *58*, 22. [CrossRef]
46. He, H.; He, L. Heme oxygenase 1 and abiotic stresses in plants. *Acta Physiol. Plant.* **2013**, *36*, 581–588. [CrossRef]
47. Tripathi, D.K.; Singh, S.; Gaur, S.; Singh, S.; Yadav, V.; Liu, S.; Singh, V.P.; Sharma, S.; Srivastava, P.; Prasad, S.M.; et al. Acquisition and homeostasis of iron in higher plants and their probable role in abiotic stress tolerance. *Front. Environ. Sci.* **2018**, *5*, 86. [CrossRef]
48. Zhu, Y.-N.; Shi, D.-Q.; Ruan, M.-B.; Zhang, L.-L.; Meng, Z.-H.; Liu, J.; Yang, W.-C. Transcriptome analysis reveals crosstalk of responsive genes to multiple abiotic stresses in cotton (*Gossypium hirsutum* L.). *PLoS ONE* **2013**, *8*, e80218. [CrossRef]
49. Tseng, G.C.; Ghosh, D.; Feingold, E. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res.* **2012**, *40*, 3785–3799. [CrossRef]
50. Shaar-Moshe, L.; Hübner, S.; Peleg, Z. Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biol.* **2015**, *15*, 111. [CrossRef]
51. Ashrafi-Dehkordi, E.; Alemzadeh, A.; Tanaka, N.; Razi, H. Meta-analysis of transcriptomic responses to biotic and abiotic stress in tomato. *PeerJ* **2018**, *6*, e4631. [CrossRef]
52. Tahmasebi, A.E.; Ashrafi-Dehkordi, A.G.; Shahriari, S.M.; Mazloomi, E. Ebrahimie Integrative meta-analysis of transcriptomic responses to abiotic stress in cotton. *Prog. Biophys. Mol. Biol.* **2019**. [CrossRef]
53. Olveracarrillo, Y.; Campos, F.; Reyes, J.L.; Garciarrubio, A.; Covarubias, A.A. Functional analysis of the group 4 Late Embryogenesis Abundant proteins reveals their relevance in the adaptive response during water deficit in Arabidopsis. *Plant Physiol.* **2010**, *154*, 373–390. [CrossRef]
54. Wang, F.; Zhu, H.; Cheng, W.; Liu, Y.; Cheng, X.; Sun, J.; Gill, S.S.; Tuteja, N. Polyamines and abiotic stress tolerance in plants. *Plant Signal. Behav.* **2010**, *5*, 26–33.
55. Huang, Z.; Zhong, X.J.; He, J.; Jin, S.H.; Guo, H.D.; Yu, X.F.; Zhou, Y.J.; Li, X.; Ma, M.D.; Chen, Q.B.; et al. Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding lea (Late Embryogenesis Abundant) proteins in moso bamboo (*Phyllostachys edulis*). *PLoS ONE* **2016**, *11*, e0165953. [CrossRef]
56. Muvunyi, B.P.; Yan, Q.; Wu, F.; Min, X.; Yan, Z.Z.; Kanzana, G.; Wang, Y.; Zhang, J. Mining Late Embryogenesis Abundant (LEA) family genes in *Cleistogenes songorica*, a xerophyte perennial desert plant. *Int. J. Mol. Sci.* **2018**, *19*, 3430. [CrossRef]
57. Chen, Y.; Li, C.; Zhang, B.; Yi, J.; Yang, Y.; Chunyan, K.; Lei, C.; Gong, M. The role of the Late Embryogenesis Abundant (LEA) protein family in development and the abiotic stress response: A comprehensive expression analysis of potato (*Solanum tuberosum*). *Genes* **2019**, *10*, 148. [CrossRef]
58. Chang, C.; Meyerowitz, E.M. Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1408–1412. [CrossRef]
59. Chung, H.J.; Ferl, R.J. Arabidopsis alcohol dehydrogenase expression in both shoots and roots is conditioned by root growth environment. *Plant Physiol.* **1999**, *121*, 429–436. [CrossRef]
60. Matton, D.P.; Constabel, P.; Brisson, N. Alcohol dehydrogenase gene expression in potato following elicitor and stress treatment. *Plant Mol. Biol.* **1990**, *14*, 775–783. [CrossRef]
61. Christie, P.J.; Hahn, M.; Walbot, V. Low-temperature accumulation of alcohol dehydrogenase-1 mRNA and protein activity in maize and rice seedlings. *Plant Physiol.* **1991**, *95*, 699–706. [CrossRef]
62. Tesniere, C.; Torregrosa, L.; Pradal, M.; Souquet, J.-M.; Gilles, C.; Santos, K.D.; Chatelet, P.; Gunata, Z. Effects of genetic manipulation of alcohol dehydrogenase levels on the response to stress and the synthesis of secondary metabolites in grapevine leaves. *J. Exp. Bot.* **2006**, *57*, 91–99. [CrossRef]
63. Yang, C.Y. Hydrogen peroxide controls transcriptional responses of ERF73/HRE1 and ADH1 via modulation of ethylene signaling during hypoxic stress. *Planta* **2014**, *239*, 877–885. [CrossRef]
64. Shi, H.; Liu, W.; Yao, Y.; Wei, Y.; Chan, Z. Alcohol dehydrogenase 1 (ADH1) confers both abiotic and biotic stress resistance in Arabidopsis. *Plant Sci.* **2017**, *262*, 24–31. [CrossRef]
65. Sauter, M.; Rzewuski, G.; Marwedel, T.; Lorbiecke, R.A. The novel ethylene-regulated gene OsUsp1 from rice encodes a member of plant protein family related to prokaryotic universal stress proteins. *J. Exp. Bot.* **2002**, *53*, 2325–2331. [CrossRef]

66. Sun, W.; Van Montagu, M.; Verbruggen, N. Small heat shock proteins and stress tolerance in plants. *Biochim. Biophys. Acta* **2002**, *157*, 1–9. [CrossRef]
67. Dubouzet, J.G.; Sakuma, Y.; Ito, Y.; Kasuga, M.; Dubouzet, E.G.; Miura, S.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. OsDREB genes in rice, *Oryza sativa* L.; encode transcription activators that function in drought, high salt and cold responsive gene expression. *Plant J.* **2003**, *33*, 751–763. [CrossRef]
68. Kawano, T. Role of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction. *Plant Cell Rep.* **2003**, *21*, 829–837.
69. Mittova, V.; Guy, M.; Tal, M.; Volokita, M. Salinity up-regulates the antioxidative system in root mitochondria and peroxisomes of the wild salt-tolerant tomato species *Lycopersicon pennellii*. *J. Exp. Bot.* **2004**, *55*, 1105–1113. [CrossRef]
70. Valerio, L.; Meyer, M.D.; Penel, C.; Dunand, C. Expression analysis of the Arabidopsis peroxidase multigene family. *Phytochemistry* **2004**, *65*, 1331–1342. [CrossRef]
71. Passardi, F.; Cosio, C.; Penel, C.; Dunand, C. Peroxidases have more functions than a Swiss army knife. *Plant Cell Rep.* **2005**, *24*, 255–265. [CrossRef]
72. Zahur, M.; Maqbool, A.; Irfan, M.; Barozai, M.Y.K.; Rashid, B.; Riazuddin, S.; Husnain, T. Isolation and functional analysis of cotton universal stress protein promoter in response to phytohormones and abiotic stresses. *Mol. Biol.* **2009**, *43*, 578–585. [CrossRef]
73. Chen, J.H.; Jiang, H.W.; Hsieh, E.J.; Chen, H.Y.; Chien, C.T.; Hsieh, H.L.; Lin, T.P. Drought and salt stress tolerance of an Arabidopsis glutathione S-transferase U17 knockout mutant are attributed to the combined effect of glutathione and abscisic acid. *Plant Physiol.* **2012**, *158*, 340–351. [CrossRef]
74. Yoon, G.M.; Kieber, J.J. 1-Aminocyclopropane-1-carboxylic acid as a signaling molecule in plants. *AoB Plants* **2013**, *5*, plt017. [CrossRef]
75. Chen, D.; Ma, X.; Li, C.; Zhang, W.; Xia, G.; Wang, M. A wheat aminocyclopropane-1-carboxylate oxidase gene, TaACO1, negatively regulates salinity stress in *Arabidopsis thaliana*. *Plant Cell Rep.* **2014**, *33*, 1815–1827. [CrossRef]
76. Gonzali, S.; Loreti, E.; Cardarelli, F.; Novi, G.; Parlanti, S.; Pucciariello, C.; Bassolino, L.; Banti, V.; Licausi, F.; Perata, P. Universal stress protein HRU1 mediates ROS homeostasis under anoxia. *Nat. Plants* **2015**, *1*, 15151. [CrossRef]
77. Udawat, P.; Jha, R.K.; Sinha, D.; Mishra, A.; Jha, B. Overexpression of a cytosolic abiotic stress responsive universal stress protein (*SbUSP*) mitigates salt and osmotic stress in transgenic tobacco plants. *Front. Plant Sci.* **2016**, *7*, 518. [CrossRef]
78. Mishra, D.; Shekhar, S.; Singh, D.; Chakraborty, S.; Chakraborty, N. Heat shock protein and abiotic stress tolerance in plants. In *Regulation of Heat Shock Protein Responses*; Asea, A., Kaur, P., Eds.; Springer: Cham, Switzerland, 2018; Volume 13, pp. 41–69.
79. Kumar, S.; Trivedi, P.K. Gluthatione S-transferase: Role in combating abiotic stresses including arsenic detoxification in plants. *Front. Plant Sci.* **2018**, *9*, 751. [CrossRef]
80. Anirban, B. Bananas tackling drought and heat- with DREBs and more. *Physiol. Plant.* **2019**, *165*, 128–130.
81. Jangale, B.L.; Chaudhari, R.S.; Azeez, A.; Sane, P.V.; Sane, A.P.; Krishna, B. Independent and combined abiotic stresses affect the physiology and expression patterns of DREB genes differently in stress-susceptible and resistant genotypes of banana. *Physiol. Plant* **2019**, *165*, 303–318. [CrossRef]
82. Zhang, H.; Huang, Z.; Xie, B.; Chen, Q.; Tian, X.; Zhang, X.; Zhang, H.; Lu, X.; Huang, D.; Huang, R. The ethylene-, jasmonate-, abscisic acid-and NaCl-responsive tomato transcription factor JERF1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. *Planta* **2004**, *220*, 262–270. [CrossRef]
83. Mishra, M.; Das, R.; Pandey, G.K. Role of ethylene responsive factors (ERFs) in abiotic stress mediated signaling in plants. *J. Biol. Sci.* **2009**, *1*, 133–146.
84. Zhang, Z.; Li, F.; Li, D.; Zhang, H.; Huang, R. Expression of ethylene response factor JERF1 in rice improves tolerance to drought. *Planta* **2010**, *232*, 765–774. [CrossRef]
85. Müller, M.; Munné-Bosh, S. Ethylene response factors: A key regulatory hub in hormone and stress signaling. *Plant Physiol.* **2015**, *169*, 32–41. [CrossRef]
86. Klay, I.; Gouia, S.; Liu, M.; Mila, I.; Khoudi, H.; Bernadac, A.; Bouzayen, M.; Pirrello, J. Ethylene response factors (ERF) are differentially regulated by different abiotic stress types in tomato plants. *Plant Sci.* **2018**, *274*, 137–145. [CrossRef]

87. Li, C.; Ng, C.K.Y.; Fan, L.M. MYB transcription factors, active players in abiotic stress signaling. *Environ. Exp. Bot.* **2015**, *114*, 80–91. [CrossRef]
88. Roy, S. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signal Behav.* **2016**, *11*, e1117723. [CrossRef]
89. Phukan, U.J.; Jeena, G.S.; Shukla, R.K. WRKY transcription factors: Molecular regulation and stress responses in plants. *Front. Plant Sci.* **2016**, *7*, 760. [CrossRef]
90. Sun, X.; Wang, Y.; Sui, N. Transcription regulation of bHLH during plant response to stress. *Biochem. Biophys. Res. Commun.* **2018**, *503*, 397–401. [CrossRef]
91. Beer, M.A.; Tavazoie, S. Predicting gene expression from sequence. *Cell* **2004**, *117*, 185–198. [CrossRef]
92. Danko, C.G.; Pertsov, A.M. Identification of gene co-regulatory modules and associated cis-elements involved in degenerative heart disease. *BMC Med. Genom.* **2009**, *2*, 31. [CrossRef]
93. Helsen, J.; Frickel, J.; Jeller, R.; Verstrepen, K.J. Network hubs affect evolvability. *PLoS Biol.* **2019**, *17*, e3000111. [CrossRef]
94. Jeong, H.; Mason, S.P.; Barabasi, A.L.; Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **2001**, *411*, 41–42. [CrossRef]
95. Wang, L.; Yu, S.; Tong, C.; Zhao, Y.; Liu, Y.; Song, C.; Zhang, Y.; Zhang, X.; Wang, Y.; Hua, W.; et al. Genome sequencing of the high oil crop sesame. *Genome Biol.* **2014**, *15*, R39. [CrossRef]
96. Kim, D.; Landmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
97. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef]
98. Tarazona, S.; Garcia-Alcalde, F.; Dopazo, J.; Ferrer, A.; Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Res.* **2011**, *21*, 2213–2223. [CrossRef]
99. Wan, Q.; Tang, J.; Han, Y.; Wang, D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp. Eye Res.* **2018**, *166*, 13–20. [CrossRef]
100. Yang, Q.; Wang, R.; Wei, B.; Peng, C.; Wang, L.; Hu, G.; Kong, D.; Du, C. Candidate biomarkers and molecular mechanism investigation for *Glioblastoma multiforme* utilizing WGCNA. *BioMed Res. Int.* **2018**, *2018*, 4246703. [CrossRef]
101. Su, G.; Morris, J.H.; Demchak, B.; Bader, G.D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinform.* **2014**, *47*, 1–24. [CrossRef]
102. Khan, A.; Fornes, O.; Stigliani, A.; Gheorghe, M.; Castro-Mondragon, J.A.; van der Lee, R.; Bessy, A.; Chèneby, J.; Kulkarni, S.R.; Tan, G.; et al. JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **2018**, *46*, D260–D266. [CrossRef]
103. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [CrossRef]
104. Clough, S.J.; Bent, A.F. Floral dip a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 735–743. [CrossRef]
105. Dossa, K.; Mmadi, M.A.; Zhou, R.; Zhou, Q.; Yang, M.; Cisse, N.; Diouf, D.; Wang, L.; Zhang, X. The contrasting response to drought and waterlogging is underpinned by divergent DNA methylation programs associated with transcript accumulation in sesame. *Plant Sci.* **2018**, *277*, 207–217.
106. Raineri, J.; Ribichich, K.F.; Chan, R.L. The sunflower transcription factor *HaWRKY76* confers drought and flood tolerance to *Arabidopsis thaliana* plants without yield penalty. *Plant Cell Rep.* **2015**, *34*, 2065–2080. [CrossRef]
107. Lalitha, S. Primer premier 5. *Biotech Softw. Internet Rep.* **2000**, *1*, 270–272. [CrossRef]
108. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real time quantitative PCR and the 2(-Delta Delta C (T)) method. *Methods* **2001**, *25*, 402–408. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Intergeneric Relationships within the Family Salicaceae *s.l.* Based on Plastid Phylogenomics

Meng-Meng Li [1,†], De-Yan Wang [1,†], Lei Zhang [1], Ming-Hui Kang [1], Zhi-Qiang Lu [2], Ren-Bin Zhu [2], Xing-Xing Mao [1], Zhen-Xiang Xi [1] and Tao Ma [1,*]

1 Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China
2 CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China
* Correspondence: matao.yz@gmail.com
† These authors contributed equally to this study.

Received: 26 June 2019; Accepted: 31 July 2019; Published: 2 August 2019

Abstract: Many Salicaceae *s.l.* plants are recognized for their important role in the production of products such as wood, oils, and medicines, and as a model organism in life studies. However, the difference in plastid sequence, phylogenetic relationships, and lineage diversification of the family Salicaceae *s.l.* remain poorly understood. In this study, we compare 24 species representing 18 genera of the family. Simple sequence repeats (SSRs) are considered effective molecular markers for plant species identification and population genetics. Among them, a total of 1798 SSRs were identified, among which mononucleotide repeat was the most common with 1455 accounts representing 80.92% of the total. Most of the SSRs are located in the non-coding region. We also identified five other types of repeats, including 1750 tandems, 434 forward, 407 palindromic, 86 reverse, and 30 complementary repeats. The species in Salicaceae *s.l.* have a conserved plastid genome. Each plastome presented a typical quadripartite structure and varied in size due to the expansion and contraction of the inverted repeat (IR) boundary, lacking major structural variations, but we identified six divergence hotspot regions. We obtained phylogenetic relationships of 18 genera in Salicaceae *s.l.* and the 24 species formed a highly supported lineage. *Casearia* was identified as the basal clade. The divergence time between Salicaceae *s.l.* and the outgroup was estimated as ~93 Mya; *Salix*, and *Populus* diverged around 34 Mya, consistent with the previously reported time. Our research will contribute to a better understanding of the phylogenetic relationships among the members of the Salicaceae *s.l.*

Keywords: Salicaceae; phylogenetic relationship; plastid genome; comparative genomics; repeat sequences

1. Introduction

The previously defined Salicaceae *sensu stricto* (*s.s.*) only included *Salix* and *Populus* [1], but later more than 50 genera were classified into the Salicaceae *sensu lato* (*s.l.*) family, containing over 1000 species [2]. The family Salicaceae *s.l.* is a woody shrub plant, ranging in height from less than a few centimeters to tens of meters. The species in Salicaceae *s.l.* are primarily distributed in cold, tropical, and warm temperate regions and occupy extremely varied habitats [2,3]. This family's sexual systems are highly diverse. Most genera are dioecious, whereas, some are monoecious. Both XY and ZW sex determination systems have been reported in the dioecious species, indicating an amazingly diversified history of sex determination [4,5]. The species of this family have many uses. There are varied chemicals produced by the family. An abundant oil containing unsaturated fatty acids can be synthesized from *Idesia* fruits [6]. The early medicine, aspirin, was first isolated from willow and poplar bark, and the willow species have been developed as bioenergy crops [7]. *Populus* species have become

model organisms for basic research in molecular biology and genetics because of their small genome size and fast growth rates. Moreover, they play important roles in the ecosystem, plant domestication, and conservation, as well as being one of the most economically important groups of forest trees [8].

The plastid genome is widely used in plant genetic population and phylogeny analysis because of its slow rate of nucleotide substitution in gene coding genes and relatively conservative gene structure and content, which can also increase phylogenetic resolution at the lower taxonomic levels [9–14]. The plastid genome usually encodes about 80 unique proteins, 30 transfer RNAs (tRNAs), and four ribosomal RNAs (rRNAs) [15,16]. For terrestrial plants with photosynthesis, the plastid genome is 120–220 kb in size and has the typical feature of two inverted repeats (IRs), 20–28 kb in size, separated by small single copies (SSC) and large single copies (LSC) with sizes of 16–27 kb and 80–90 kb, respectively [17,18]. The size of IR in plastids varies widely for different groups, genus, family or species [13,19,20]. The IR copies recombine themselves in order to maintain or confer stability in the remaining plastome [17,18,21]. With the development of high-throughput sequencing technology in recent years, the number of complete sequenced plastid genomes has increased rapidly [22]. The whole plastid genome of 61 species in Salicaceae *s.l.* has been sequenced and stored in the GenBank.

In previous studies [23–26], their work mainly focused on the relationship of the main subclades in the genera of *Salix* and *Populus*, because the delimitation of species in these genera remains controversial. There are 29–70 species in the genus *Populus* and based on their morphological characteristics they have been grouped into the following six sections: *Abaso*, *Turanga*, *Populus*, *Leucoides*, *Aigeiros*, and *Tacamahaca* [27]. For the genus *Salix*, about 450 species have been published and two main subclades have been identified [28–30]. The most recent study used 42 species from six genera based on the complete plastomes in order to examine phylogenetic relationships of Salicaceae [31]. Although several other genera in Salicaceae were mentioned in their research, their main purpose was to determine the relationship of subclades in the genus of *Salix* and *Populus*. We know little about the phylogenetic relationships of the other 48 genera in Salicaceae *s.l.*

In this study, we sequenced 20 plastid genomes in the family, and added the following four previously published plastid genomes: *Populus euphratica*, *Salix interior*, *Idesia polycarpa maxim*, and *Poliothyrsis sinensis*. In total, there were 24 species from 18 genera of Salicaceae *s.l.*, as well as two outgroups, *Passiflora laurifolia* and *Passiflora ligularis*. We mainly aimed to: (1) determine the repeat sequence variations of plastid genomes, (2) examine structural changes in the plastomes of the Salicaceae *s.l.*, and (3) delimit intergeneric relationships within Salicaceae *s.l.*

2. Results

2.1. Characteristics of the Plastid Genomes

Twenty complete plastid genomes belonging to fourteen genera of the family Salicaceae *s.l.* were newly generated in this study (Table 1). All of the genome sequences have been submitted to the GenBank. In order to fully display the characteristics of the plastid genomes, we further collected four sequences of other species in the Salicaceae *s.l.* from the NCBI GenBank. In total, our subsequent comparative analysis included 24 species representing 18 genera of the family Salicaceae *s.l.* (Table 1). The total length of the chloroplast genome sequences ranged from 155,144 bp in *Flacourtia ramontchii* to 158,605 bp in *Bennettiodendron brevipes*. The structure of the genomes displayed a typical quadripartite structure, with a pair of inverted-repeat (IR) regions of 27,168–27,926 bp, separated by a large single copy (LSC) of 83,391–86,350 bp and a small single copy (SSC) of 16,305–17,889 bp. The LSC regions exhibited the greatest standard deviation in sequence length (s.d. = 872.58 bp), followed by the SSC regions (s.d. = 470.12 bp) and the the IR regions (s.d. = 217.50 bp). Nucleotide composition with an overall GC content of 36.8% was nearly identical in all plastid genomes.

Table 1. Characteristics of plastid genomes of 26 species.

Species	No. of Total Genes	Genome Size (bp)	LSC Length (bp)	SSC Length (bp)	IR Length (bp)	GC Content (%)	Protein-Coding Genes	No. of tRNAs Genes	No. of rRNAs Genes	Genes with Introns	GenBank Accesion Number
Populus euphratica *	131	157,839	85,858	16,649	27,666	36.5	86(8)	37(7)	8(4)	20(5)	MK267314
Salix interior *	127	156,620	85,979	16,305	27,168	37.0	80(5)	37(7)	8(4)	21(5)	NC_024681
Bennettiodendron leprosipes	131	157,872	85,643	16,439	27,895	36.7	86(8)	37(7)	8(4)	21(5)	MK281360
Bennettiodendron brevipes	131	158,605	86,350	16,501	27,877	36.7	86(8)	37(7)	8(4)	21(5)	MK046729
Idesia polycarpa Maxim *	130	157,017	84,787	16,512	27859	36.7	86(8)	36(7)	8(4)	19(5)	NC_032060
Olmediella betschleriana	131	158,171	85,783	16,588	27,900	36.7	85(8)	37(7)	8(4)	20(5)	MK044831
Carrierea calycina	131	158,085	84,663	17,570	27,926	36.8	86(8)	37(7)	8(4)	21(5)	MK263737
Itoa orientalis	131	157,875	84,495	17,742	27,819	36.7	86(8)	37(7)	8(4)	21(5)	MN078146
Poliothyrsis sinensis *	129	156,680	84,793	16,605	27641	36.8	85(8)	37(7)	8(4)	19(5)	NC_037412
Flacourtia inermis	131	155,535	83,706	16,577	27,626	36.9	86(8)	37(7)	8(4)	21(5)	MN078138
Flacourtia ramontchii	131	155,144	83,391	16,591	27,581	36.9	86(8)	37(7)	8(4)	21(5)	MN078145
Flacourtia rukam	131	155,582	83,854	16,598	27,565	36.9	86(8)	37(7)	8(4)	21(5)	MK281365
Xylosma congestum	131	155,986	84,180	16,650	27,578	36.8	86(8)	37(7)	8(4)	21(5)	MN078135
Scolopia chinensis	131	157,292	84,464	17,716	27,556	36.8	86(8)	37(7)	8(4)	21(5)	MN078144
Scolopia saeva	131	155,832	84,062	16,436	27,667	36.9	86(8)	37(7)	8(4)	21(5)	MN078143
Dovyalis caffra	131	155,182	83,593	16,683	27,453	37.0	86(8)	37(7)	8(4)	21(5)	MN078137
Homalium racemosum	131	156,961	85,818	16,547	27,298	36.6	86(8)	37(7)	8(4)	21(5)	MN078136
Homalium cochinchinense	131	155,339	83,701	16,566	27,536	36.8	86(8)	37(7)	8(4)	21(5)	MN078140
Banara guianensis	131	158,577	85,863	17,334	27,690	36.5	86(8)	37(7)	8(4)	21(5)	MH937752
Prockia crucis	131	157,406	85,622	17,024	27,380	36.7	86(8)	37(7)	8(4)	21(5)	MN078147
Azara serrata	131	158,306	85,059	17,889	27,697	36.5	86(8)	37(7)	8(4)	21(5)	MH719101
Abatia parviflora	131	156,426	84,800	16,724	27,451	36.5	86(8)	37(7)	8(4)	21(5)	MN078139
Casearia decandra Jacq	131	156,809	84,890	17,037	27,441	36.8	85(8)	37(7)	8(4)	21(5)	MN078142
Casearia velutina	131	156,008	84,446	17,220	27,171	36.8	85(8)	37(7)	8(4)	21(5)	MN078141
Passiflora laurifolia **	130	151,422	85,411	13,511	26,250	37.0	77(4)	37(7)	8(4)	20(5)	NC_038121
Passiflora ligularis **	128	150,827	85,471	13,502	25,927	37.0	77(4)	36(7)	8(4)	19(5)	NC_038122

Note: Numbers in brackets indicate genes duplicated in the IR regions. The species with * are previous published plastid genomes, and the species with ** are outgroups.

A total of 131 functional genes with the same order were annotated in each of the newly sequenced plastomes, of which 102 were unique genes, including 78 protein-coding genes, 30 tRNA genes, and four rRNA genes (Figure 1, Table 2). Most of these genes occurred as a single copy, while 19 genes were duplicated in the IR regions (Table 1). The gene *cemA* contained premature termination codons in *Casearia decandra Jacq* and *Casearia velutina*, while gene *atpF* contained premature termination codons in *Olmediella betschleriana*.

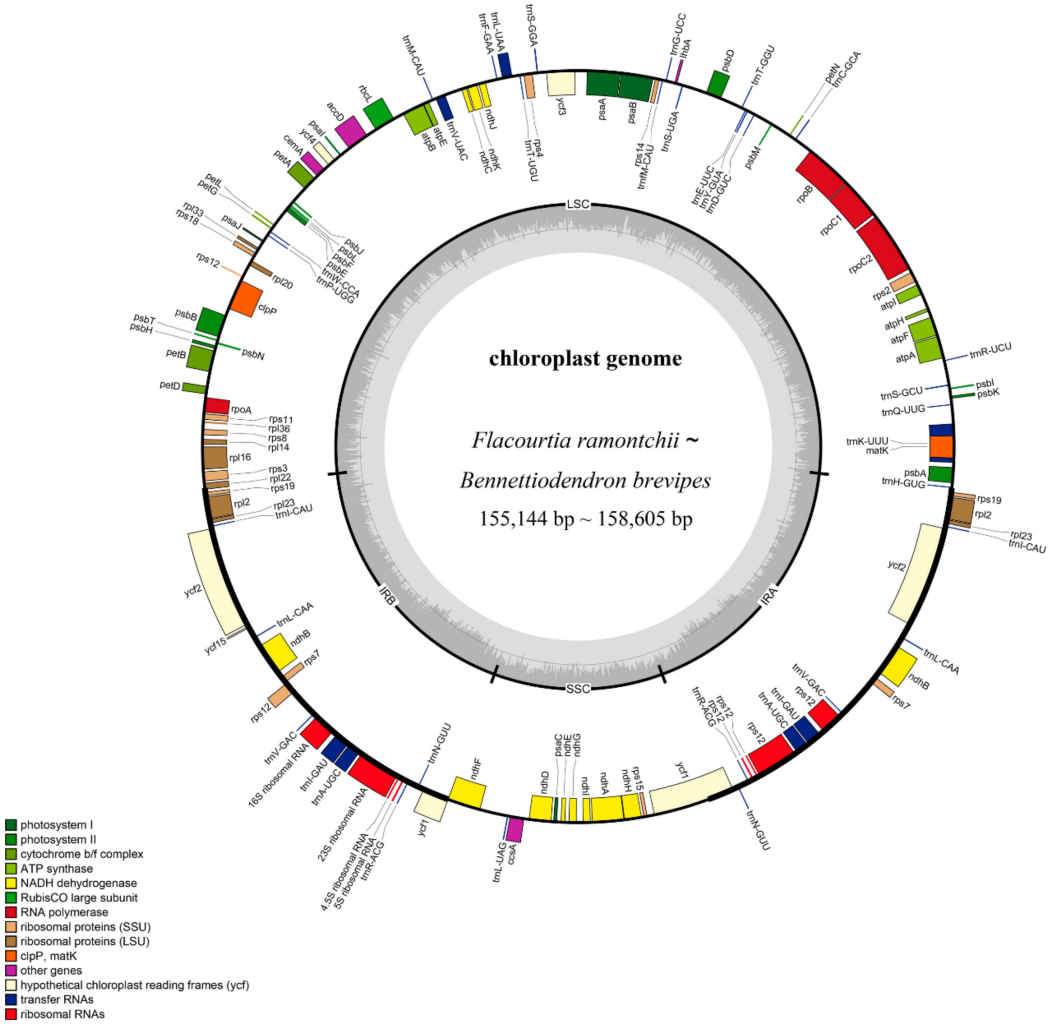


Figure 1. Gene map of 20 Salicaceae *s.l.* chloroplast genomes. Functional categories of genes are color-coded. Genes inside the circle are transcribed clockwise and genes outside the circle are transcribed counter clockwise. The dashed area in the inner circle indicates the GC content of the plastid genome.

2.2. Repeat Sequence Analysis

We identified a total of 1798 simple sequence repeat (SSR) loci across the 24 Salicaceae *s.l.* plastids (Figure 2a, Table S1). Of these, 1455 were mononucleotide repeats accounting for about 80.92% of the total SSRs, while 180 (10.01%), 98 (5.45%), 41 (2.28%), 18 (1.00%), and six (0.33%) were tetra-, di-, tri-, penta-, and hex-nucleotides repeats, respectively. The penta- and hex- nucleotides repeats were very rare across the plastid genomes. Most (74.36%) SSR loci were located in the intergenic regions, whereas, 9.07% were in intron and 16.57% were in the protein-coding regions.

Table 2. List of genes present in the plastid genomes of twenty newly sequenced Salicaceae *s.l.* species.

Category	Gene	Type Gene						
Self-replication	**Ribosomal RNA**	*rrn16*	*rrn23*	*rrn4.5*	*rrn5*			
	Transfer RNA	*trnA-UGC* * *trnC-GCA* *trnD-GUC* *trnE-UUC* *trnF-GAA*	*trnfM-CAU* *trnG-GCC* *trnG-UCC* * *trnH-GUG* *trnI-CAU*	*trnI-GAU* * *trnK-UUU* * *trnL-CAA* *trnL-UAA* * *trnL-UAG*	*trnM-CAU* *trnN-GUU* *trnY-GUA* *trnP-UGG* *trnQ-UUG*	*trnR-ACG* *trnW-CCA* *trnR-UCU* *trnS-GCU* *trnS-GGA*	*trnS-UGA* *trnT-GGU* *trnT-UGU* *trnV-GAC* *trnV-UAC* *	
	Small ribosomal units	*rps11* *rps19*	*rps12* * *rps2*	*rps14* *rps3*	*rps15* *rps4*	*rps18* *rps7*	*rps8*	
	Large ribosomal units	*rpl14* *rpl33*	*rpl16* * *rpl36*	*rpl2* *	*rpl20*	*rpl22*	*rpl23*	
	RNA polymerase subunits	*rpoA*	*rpoB*	*rpoC1* *	*rpoC2*			
Photosynthesis genes	**NADH dehydrogenase**	*ndhA* * *ndhG*	*ndhB** *ndhH*	*ndhC* *ndhI*	*ndhD* *ndhJ*	*ndhE* *ndhK*	*ndhF*	
	photosystem I	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*	*ycf3* *	*ycf4*
	photosystem II	*psbA* *psbI* *psbZ*	*psbB* *psbJ*	*psbC* *psbK*	*psbD* *psbL*	*psbE* *psbM*	*psbF* *psbN*	*psbH* *psbT*
	cytochrome b/f complex	*petA*	*petB* *	*petD*	*petG*	*petL*	*petN*	
	ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**	*atpH*	*atpI*	
	Large subunit of rubisco	*rbcL*						
Other genes	**Maturase**	*matK*						
	Protease	*clpP* *						
	Acetyl-CoA-carboxylase subunit	*accD*						
	Envelope membrane protein	*cemA*						
	Component of TIC complex	*ycf1*						
	c-type cytochrome synthesis	*ccsA*						
Unknown	**hypothetical genes reading frames**	*ycf2*						

Notes: the * symbols indicate genes with intron(s).

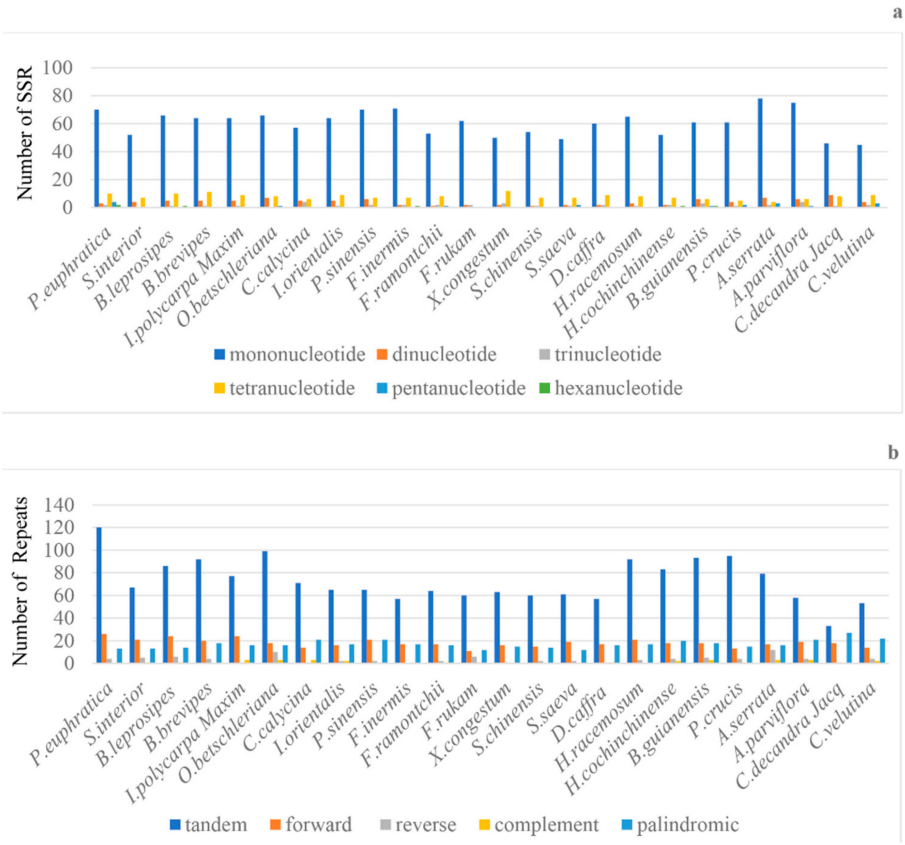


Figure 2. Comparison of repeat sequences among 24 plastomes. The *X*-axis represents the species and the *Y*-axis represents the number of repeats. (**a**) Frequency of selected motifs of simple sequence repeats (SSRs) and (**b**) number of each repeat type: tandem repeats, forward repeats, palindromic repeats, reverse repeats, and complement repeats.

We also identified 5 other types of repeats, including 1750 tandems, 434 forward, 407 palindromic, 86 reverse, and 30 complementary repeats (Figure 2b, Table S2, Table S3). Among these, tandem repeats are the most frequent type of repeats (64.65%), while complementary repeats are the least. We found that the tandem repeat sequences were mainly located in non-coding regions, whereas, only a few were located in the coding regions (*ycf2*, *rpl22*, *rpl14*, *rpoC2*, *ycf1*, *ndhD*, *ndhG*, *psbL*, *ndhF*, *rpoA*, *petD*, and *ccsA*). Most of the five types of repeats were concentrated in the intergenic regions (Table S3).

2.3. Inverted Repeats and Genome Comparison

Next, we conducted whole genome alignment using the program mVISTA (*Populus euphratica* as reference), and the results showed that both the content and order of the genes were highly conserved in the Salicaceae *s.l.* plastids (Figure S1). The IRa/SSC boundary positions for all species were located in the *ycf1* gene, while the border genes of IRa and LSC were *Rps19* and *trnH-GUG*, respectively. Only slight variations of the border structure were identified across these plastid genomes. For example, the IRb/LSC junction was located within the *rpl22* gene in all but six species (*Scolopia chinensis*, *Homalium racemosum*, *Homalium cochinchinense*, *Prockia crucis*, *Casearia velutina*, and *Casearia decandra Jacq*). The *ndhF* gene was located entirely in the SSC region of 16 species, while in the other eight species it extended into the IRb region (Figure 3).

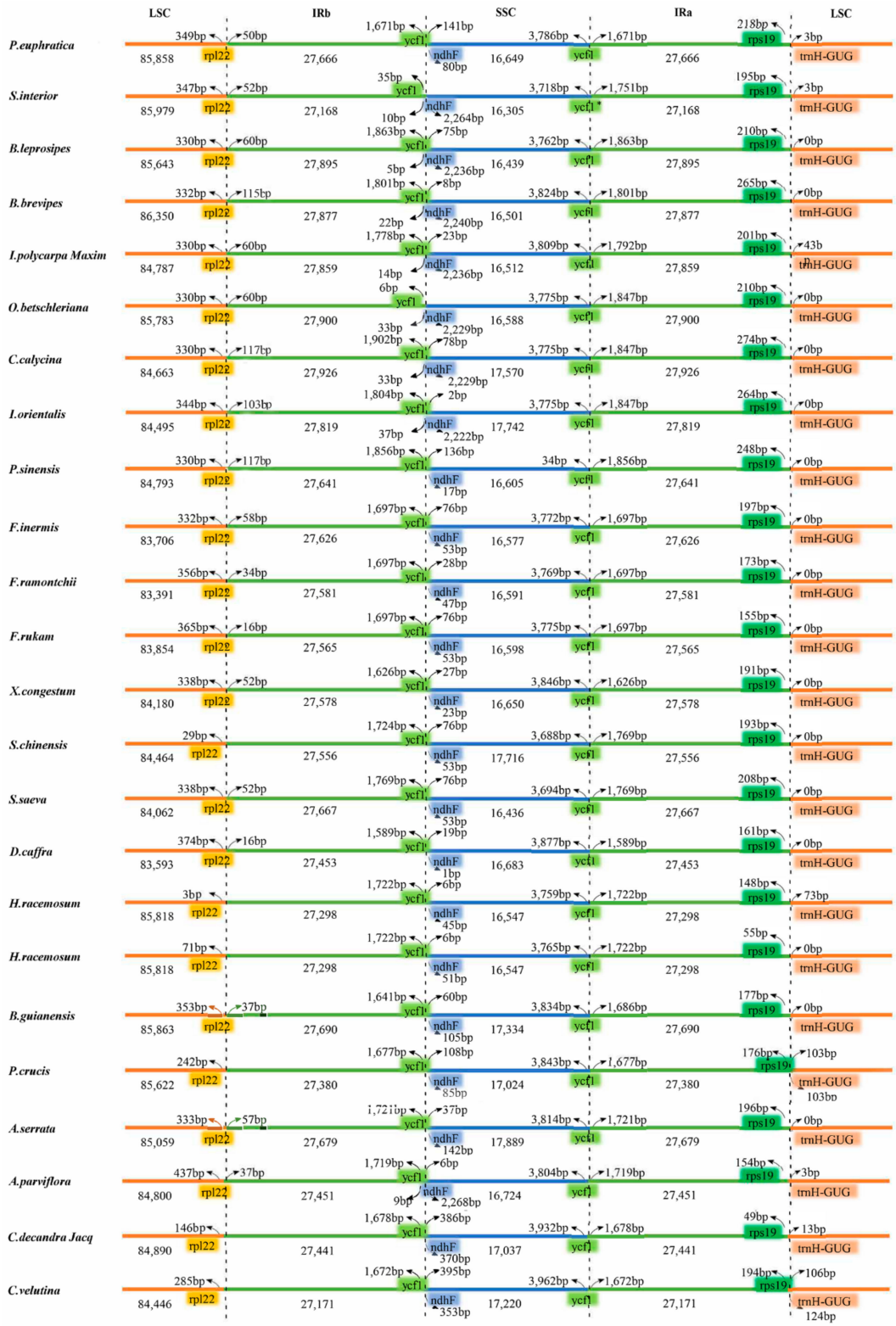


Figure 3. Comparison of the border positions of large single copy (LSC), small single copy (SSC), and inverted repeat (IR) region borders among plastid genomes of 24 Salicaceae *s.l.* species. Complete genes and portions of genes adjacent to the junctions are depicted by differently colored blocks.

2.4. Divergence Hotspots of Plastid Genomes

To evaluate the level of sequence divergence, we calculated the percentages of variation using a sliding window approach (Figure 4). Across the 24 species, the values of nucleotide variability ranged from 0 to 0.139, with an average of 0.023, suggesting a high conservation of plastid genomes within Salicaceae *s.l.* Six relatively high variable regions (divergence hotspots) were identified, which comprised one gene region (*ycf1* + *ndhF*) and the following five intergenic regions: *matK-trnQ-UUG*, *trnS-GCU-trnG-UCC*, *psbZ-trnG-GCC*, and *ndhF-trnL-UAG*, *psbE-petL*.

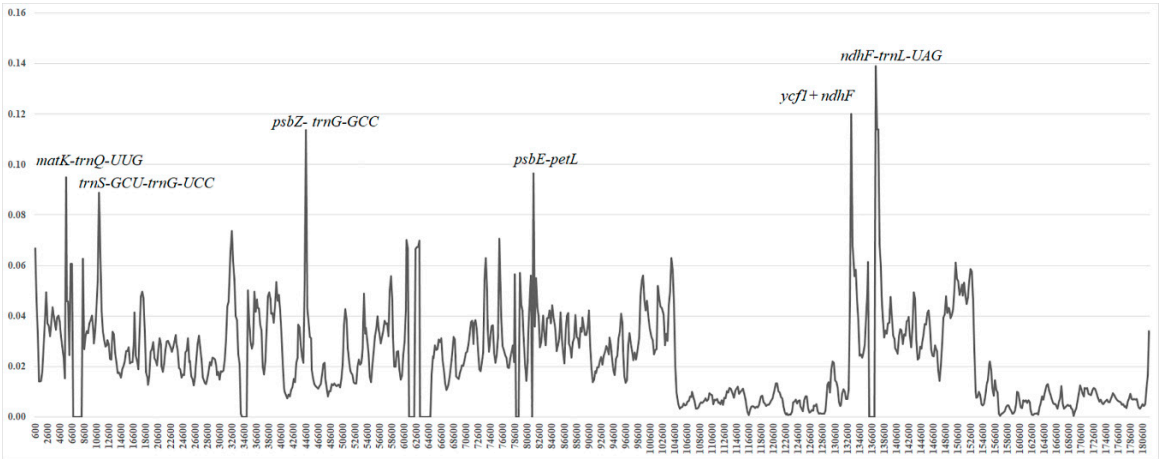


Figure 4. Comparison of the nucleotide variability (Pi) values of the plastomes. *X*-axis: position of the midpoint of a window, *Y*-axis: nucleotide diversity of each window (window length: 600 bp and step size: 200 bp).

2.5. Phylogenetic Analysis

To obtain an accurate phylogenetic relationship of Salicaceae *s.l.* species, we performed multiple sequence alignments of these 24 complete plastid genomes, including an additional two species from *Passiflora* as the outgroup. The final concatenated dataset, which included 63 protein-coding genes (Table S4) and 51,780 nucleotides, after trimming poorly aligned regions, produced a highly supported topology based on the maximum likelihood (ML) strategy (Figure 5). Two subfamilies in the phylogenetic tree, Samydoideae and Salicoideae, formed a highly supported monophyletic. *Casearia* (*Casearia decandra Jacq* and *Casearia velutina*) was the only genus from Samydoideae in this study and was identified as the basal clade. The remaining species belonged to Salicoideae and were divided into two main clades. Furthermore, we estimated divergence times from the plastome dataset using an approximate likelihood method. The Salicaceae *s.l.* was estimated to diverge from the outgroup around 93 Mya.

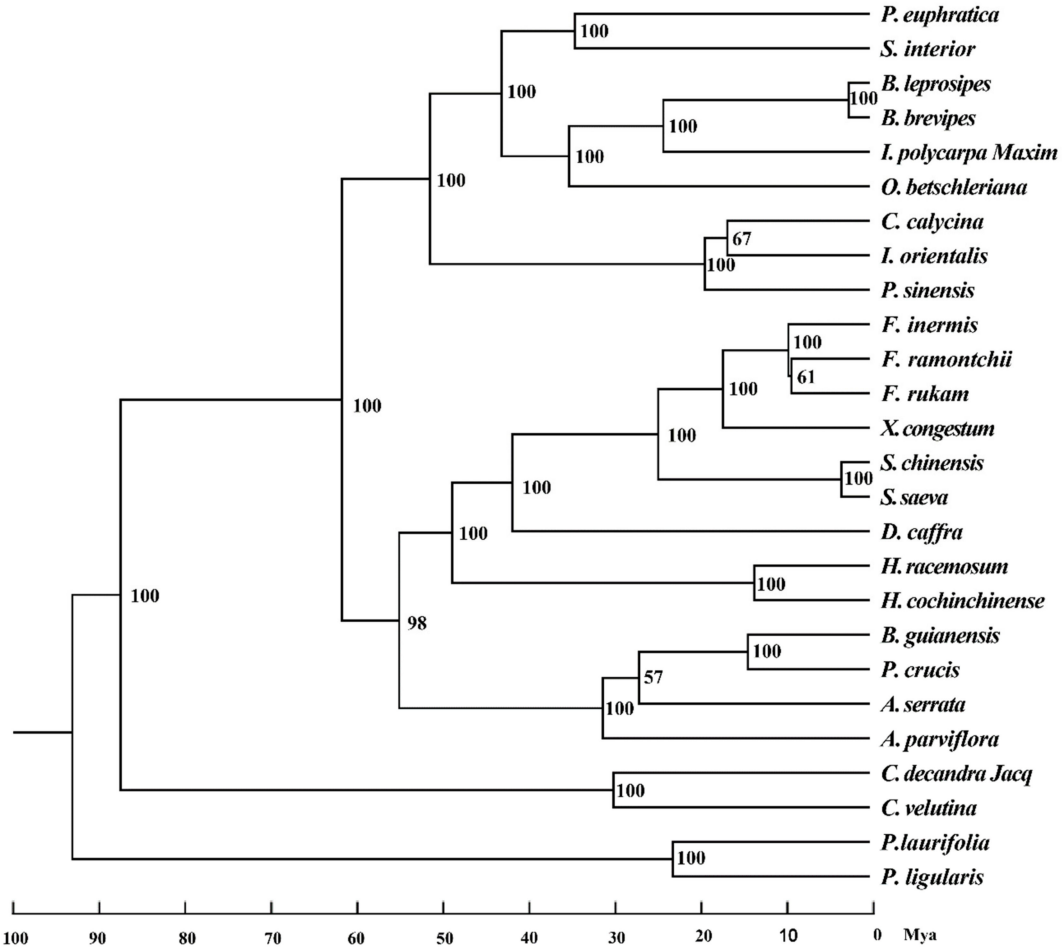


Figure 5. Phylogenetic tree of the 26 species based on 63 protein-coding genes using maximum likelihood (ML). Bootstrap values (1000 replications) between different genera are shown at the nodes. The branches are drawn in proportion to the posterior means of divergence times estimated under the GTR + I + G model with an independent relaxed clock and birth–death sampling, and the outgroup divergence constraint of 87 to 97 Mya.

3. Discussion

3.1. Features of Plastid Genomes

We used the information from complete plastid genome sequencing to research and analyze the plastomes of Salicaceae *s.l.* species. Generally, the plastids of angiosperms are considered to be highly conserved, have a typical tetragonal structure, usually contain about 110–130 unique genes, and the genome size, GC composition, as well as gene and intron content are identical in most angiosperms [32–35]. The shift of IR-SC boundaries was a common evolutionary event and played an important role in plastome size variation [36]. The IR regions of the Salicaceae *s.l.* species varied in size from 27,168 bp to 27,926 bp and the total length ranged from 155,144 bp in *Flacourtia ramontchii* to 158,605 bp in *Bennettiodendron brevipes*, as a result of the expansion and contraction of the IR borders.

Previous studies have revealed that many chloroplast genes, such as *infA*, *rpl22*, *rps19*, *rpl2* intron and *rpl23*, are transferred to the nucleus or lost [37–39]. In our study, only *infA* was not found in the plastid genome of Salicaceae *s.l.* *cemA* and contained premature termination codons in *Casearia decandra Jacq* and *Casearia velutina*, whereas, *atpF* was a pseudogene in *Olmediella betschleriana*. Gene duplication caused by IR is common in plastomes and is believed to be an important driving force in the evolution of genomes, leading to the creation of new genes and new gene functions [40]. Gene duplication has been previously reported in multiple angiosperm lineages and most of them are tRNAs [41–46].

Duplication of protein-coding genes outside of the IR is rare. In our results, seven tRNA genes and eight protein-coding genes were duplicated in the inverted-repeat region.

3.2. Repeat Sequence Variations

SSRs are a type one to seven nucleotide unit tandem repeat sequence frequently observed in plastid genome and the changes in copy number are usually polymorphic [47–50]. The SSRs are considered effective molecular markers for plant species identification and population genetics because they exhibit codominant inheritance, high repeatability, and high variation within the same species [47–50]. In this study, the mononucleotide repeats were the most common, of which 822 were T type accounting for about 57%, while 606 (41%), 16 (1%), 11 (0.7%), were A, C, and G type, respectively. Most of the SSR loci were located in the non-coding region, and only 16% of the SSR loci were found in the gene-coding region. These SSRs were located in 39 coding genes, and the genes with the highest SSR frequency are *ycf1*, *rpoc2*, and *ndhF*, which are 95, 48, and 26 times, respectively (Table S1). The plastome size variation was previously reported to contribute to tandem [36,51] and dispersed repeats [41,52–54]. We analyzed the direct repeats, inverted repeats, reverse repeats, complementary repeats, and tandem repeats in 24 Salicaceae *s.l.* species. The analysis showed that the number of tandem repeats is more than the other repeats, while complementary repeats are the least common in these species. Most of the repeats are distributed in the intergenic and intron regions, similar to those reported in other angiosperm lineages [55]. These repeat sequences and plastid SSRs provided molecular markers for studying the genetic diversity, population structure, and phylogenetics of Salicaceae *s.l.*

3.3. Comparative Analyses

Comparing chloroplast genomics not only provides insight into chloroplast evolution patterns [56], but also contributes to phylogenetic studies to understand the evolutionary relationships among taxa and apply them to species breeding and conservation [57]. We compared the 24 complete plastid genomes of 18 genera in Salicaceae *s.l.* The species in this family have a conserved plastid genome that lack major structural variations. The main cause of genomic length variation in higher plant plastids is the change in the position of the boundary between IR and SSC/LSC [58]. In this study, the location of the boundary and length of the IR regions showed that variation among the 24 species and the plastid genomes differed slightly in size, with *Bennettiodendron leprosipes*, *Bennettiodendron brevipes*, *Olmediella betschleriana*, *Carrierea calycina*, and *Ito orientalis* larger than other species, which may be related to their IR expansion. IR expansion/contraction also represents a highly variable region, which can be used to study the phylogenetic classification of plants and achieve molecular improvement of plant phenotypes. In addition, the position of the *trnH* gene, which was found in the LSC region of all 24 plastid genomes, was conserved. This result was consistent with the observation that the *trnH* gene is usually located in the IR region in monocots but is found within the LSC region in dicots.

3.4. Divergence Hotspot Analysis

The whole aligned sequences revealed relatively low divergence, however, six (*matK-trnQ-UUG*, *trnS-GCU-trnG-UCC*, *psbZ-trnG-GCC*, *ndhF-trnL-UAG*, *psbE-petL*, and *ycf1+ndhF*) displayed high variation. Further work is still necessary to determine whether these six variable loci could be used in phylogenetic analyses of related Salicaceae *s.l.* species or as potential molecular markers for population genetics and phylogenetics.

3.5. Phylogenetic Relationships

The deep relationships of mimosoids were poorly resolved by phylogenetic studies applying a few plastid markers [59,60]. Plastid phylogenomics has been proven to be efficient to resolve difficult relationships at the family level such as Orchidaceae [61] and the lower taxonomic level such as subfamilies Bambusoideae [62] and Chloridoideae of Poaceae. For the phylogenetic relationship between the species of Salicaceae, earlier studies focused on Salicaceae *s.s.* Preliminary phylogenetic

frameworks for *Salix* and *Populus* have previously been provided [23–26]. In addition, some studies of angiosperm and Malpighiales also involve the phylogenetic relationships of several genera of Salicaceae *s.l.* [63–66]. However, phylogenetic analyses focusing on Salicaceae *s.l.* have been limited to the use of one or a few genes obtained from plastid or nuclear genomes [2,67]. In this study, we obtained phylogenetic relationships of 18 genera in Salicaceae *s.l.* using 63 plastid genes from 26 species. According to the Angiosperm Phylogeny Group's most recent phylogeny, APG IV [2,68], the phylogenetic tree is divided into Salicoideae and Samydoideae. *Casearia* from Samydoideae was identified as the basal clade. The genera in one of the two main clades of Salicoideae included *Populus*, *Salix*, *Bennettiodendron*, *Idesia*, *Olmediella*, *Carrierea*, *Itoa*, and *Poliothyrsis*, all belonging to Saliceae. The other clade was divided into two subclades, one subclade included five genera, *Flacourtia*, *Xylosma*, and *Dovyalis* belonging to the Saliceae, *Scolopia* belonging to Scolopieae, and *Homolium* belonging to Homalieae. The other subclade consisted of four species of three tribes, of which *Banara* and *Prockia* belong to Prockieae, *Abatia* and *Azara* belong to Abatieae and Saliceae, respectively. The Saliceae with the most species is not monophyletic.

We further estimated divergence timescales of the major clades within the Salicaceae *s.l.* according to the calibrations of the species tree constructed on the basis of 63 protein-coding genes. The split between Salicaceae *s.l.* and the outgroup was estimated as ~93 Mya, and the basal *Casearia* was estimated to diverge from other clades around 87 Mya. The clade of Salicoideae is estimated to have originated around 61 Mya. *Salix* and *Populus* diverged around 34 Mya, consistent with the previously reported [26].

4. Materials and Methods

4.1. Sampling and Genome Sequencing

Fresh leaves and silica-gel dried materials were sampled from 20 species representing 14 genera of the family Salicaceae *s.l.* The voucher specimens for the ten fresh sampled plants collected from XiShuangBanNa Tropical Botanical Garden (Mengla, China) were deposited at the Key Laboratory of Bio-Resource and Eco-Environment of the Ministry of Education (Chengdu, Sichuan, China). The ten silica-gel-dried materials were obtained from Harvard University Herbaria and the Arnold Arboretum of Harvard University. For each species, we used the modified CTAB method [69] to extract the total genomic DNA from dry leaves. About 5 ug purified DNA was used to construct Illumina paired-end libraries with an insert size of 500 bp and sequenced using the HiSeq X Ten System by Novogene (Beijing, China). At least two gigabases (Gb) of 2 × 150 bp short reads data were generated for each sample. Reads with a Phred quality score <7 and more than 10% ambiguous nucleotides were filtered.

4.2. Plastid Genome Assembly and Annotation

The filtered reads were assembled by NOVOPlasty v2.7.1 [70], and we used the complete plastid genome *Itoa orientalis* and *Flacourtia indica* (NC_037411.1 and NC_037410.1) as the reference, and then used Geneious v11.1.5 [71] to correct the assemble. Then, we annotated the assembled plastid genome in Plann v1.1 [72]. The positions of starts, stops, introns, and exons were manually adjusted using Sequin v15.50. In addition, the physical map of the plastid genome was generated using OGDRAW v1.2 [73]. The complete plastid genome together with gene annotations was submitted to the GenBank. The accession numbers are shown in Table 1.

4.3. Repeat Sequence Analysis

The SSRs were evaluated using the online software MISA (https://webblast.ipk-gatersleben.de/misa/) [74] with thresholds: 10, 5, 4, 3, 3, 3 ten repeat units for mononucleotide SSRs, five repeat units for dinucleotide SSRs, four repeat units for trinucleotide SSRs and three repeat units for tetra-, penta-, and hexanucleotide SSRs. We used the web-based analysis tool REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [75] to detect the repeat sequences, including reverse, forward

(direct), complement, and palindromic (inverted), with a minimal repeat size of 30 bp and Hamming distance less than or equal to 3 (90% or greater sequence identity). Tandem Repeats Finder v4.09 (http://tandem.bu.edu/trf/trf.submit.options.html) [76] was used to determine the trandem repeats, alignment parameters match, mismatch. Indel were set as 2, 7, and 7.

4.4. Genome Comparative Analysis

Furter, mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) [77], a web-based program, was used to compare similarities and detect any rearrangements or inversions among different plastid genomes. We used this software (*Populus euphratica* as reference) to discover any significant interspecific and intergeneric variations among plastid genome sequences of Salicaceae *s.l.* Additionally, the IR expansion/contraction regions were compared among the 24 Salicaceae *s.l.* species (20 newly generated plastid genomes in this study and 4 previous published plastid genomes: *Populus euphratica*, *Salix interior*, *Idesia polycarpa Maxim*, and *Poliothyrsis sinensis*).

4.5. Sequence Divergence Analysis

After aligning all sequences using Mafft v7.313 [78], we located and counted the SNP sites in the plastid genomes using DnaSP v6.0 [79]. The nucleotide variability (Pi) and polymorphic sites (S) were evaluated using a sliding window analysis with a step size of 200 bp and window length of 600 bp.

4.6. Phylogenetic Analyses

There were 26 plastid genomes used for phylogenetic analyses, including 2 outgroups (*Passiflora laurifolia* and *Passiflora ligularis*) and 24 Salicaceae *s.l.* species (Table 1). We extracted 63 protein-coding genes in all 26 species using a python script. All the genes aligned by MAFFT and used RAxML v8.1.24 [80] to conduct maximum likelihood (ML) analyses with the GTR + I + G model. The best scoring ML tree was obtained with 1000 bootstrap replicates. Then, we estimated divergence times of the plastome dataset, using an approximate likelihood method as implemented in MCMCtree (in PAML version 4) [81], with an independent relaxed clock and birth–death sampling [82]. The divergence between Salicaceae *s.l.* species and the outgroup, *Passiflora*, was assigned an age constraint of 87 to 97 Mya, as has been previously used [83–86].

5. Conclusions

We compared 24 plastomes, 20 newly sequenced and four other species sequences were collected from the NCBI GenBank, all representing 18 genera of the family Salicaceae *s.l.* The SSRs are considered effective molecular markers for plant species identification and population genetics. A total of 1798 SSRs were identified. Mono-nucleotide repeats were the most common with 1455 repeats accounting for about 80.92% of the total. Most of repeats were located in the non-coding region. We also identified five other types of repeats, including 1750 tandems, 434 forward, 407 palindromic, 86 reverse, and 30 complementary repeats. The species in Salicaceae *s.l.* have a conserved plastid genome. Each plastome presented a typical quadripartite structure and varied in size due to the expansion and contraction of the IR boundary, lacking major structural variations, but we identified six divergence hotspot regions (*matK-trnQ-UUG*, *trnS-GCU-trnG-UCC*, *psbZ-trnG-GCC*, *ndhF-trnL-UAG*, *psbE-petL*, and *ycf1+ndhF*). We obtained phylogenetic relationships in 18 genera of Salicaceae *s.l.* and the 26 species formed a highly supported lineage. *Casearia* was identified as the basal clade. The divergence time between Salicaceae *s.l.* and the outgroup was estimated as ~93 Mya, and *Salix* and *Populus* diverged around 34 Mya, consistent with previously reported times. This study demonstrates the potential of plastid genome to address the genus relationship of Salicaceae *s.l.* This data will contribute to further understanding of the phylogenetic relationships among Salicaceae *s.l.*

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/15/3788/s1.

Author Contributions: T.M. designed the research and revised the manuscript; M.-M.L., D.-Y.W., L.Z., and M.-H.K. performed the experiments and analyzed the data; M.-M.L. and D.-Y.W. wrote the paper; Z.-Q.L., R.-B.Z., X.-X.M., and Z.-X.X. contributed materials and revised the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2016YFD0600101), the National Natural Science Foundation of China (41871044, 31561123001, 31500502), and the National Key Project for Basic Research (2012CB114504), and Fundamental Research Funds for the Central Universities (2018CDDY-S02-SCU, SCU2019D013).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ohashi, H. Salicaceae of Japan. *Sci. Rep.* **2001**, *40*, 269–396.
2. Chase, M.W.; Zmarzty, S.; Dolores Lledó, M.; Wurdack, K.J.; Fay, S.M.F. When in Doubt, Put It in Flacourtiaceae: A Molecular Phylogenetic Analysis Based on Plastid rbcL DNA Sequences. *Kew Bull.* **2002**, *57*, 141–181. [CrossRef]
3. Leskinen, E.; Alström-Rapaport, C. Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: Evidence from 5.8 S, ITS 1 and ITS 2 of the rDNA. *Plant Syst. Evol.* **1999**, *215*, 209–227. [CrossRef]
4. Hou, J.; Ye, N.; Zhang, D.; Chen, Y.; Fang, L.; Dai, X.; Yin, T. Different autosomes evolved into sex chromosomes in the sister genera of Salix and Populus. *Sci. Rep.* **2015**, *5*, 9076. [CrossRef] [PubMed]
5. Kersten, B.; Pakull, B.; Groppe, K.; Lueneburg, J.; Fladung, M. The sex-linked region in Populus tremuloides Turesson 141 corresponds to a pericentromeric region of about two million base pairs on P. trichocarpa chromosome 19. *Plant Biol.* **2014**, *16*, 411–418. [CrossRef]
6. Li, R.J.; Gao, X.; Li, L.M.; Liu, X.L.; Wang, Z.Y.; Lü, S.Y. De novo Assembly and Characterization of the Fruit Transcriptome of Idesia polycarpa Reveals Candidate Genes for Lipid Biosynthesis. *Front. Plant Sci.* **2016**, *7*, 801. [CrossRef] [PubMed]
7. Abrahamson, L.P.; Robison, D.J.; Volk, T.A.; White, E.H.; Neuhauser, E.F.; Benjamin, W.H.; Peterson, J.M. Sustainability and environmental issues associated with willow bioenergy development in New York (U.S.A.). *Biomass Bioenergy* **1998**, *15*, 17–22.
8. Braatne, J.H.; Hinckly, T.M.; Stettler, R.F. Influence of soil water on the physiological and morphological components of plant water balance in Populus trichocarpa, Populus deltoides and their F1 hybrids. *Tree Physiol.* **1992**, *11*, 325–340. [CrossRef]
9. Besnard, G.; Hernández, P.; Khadari, B.; Dorado, G.; Savolainen, V. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol.* **2011**, *11*, 80. [CrossRef]
10. Dexter, K.G.; Terborgh, J.W.; Cunningham, C.W. Historical effects on beta diversity and community assembly in Amazonian trees. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7787–7792. [CrossRef]
11. Howe, C.J.; Barbrook, A.C.; Lila Koumandou, V.; Nisbet, R.E.; Symington, H.A.; Wightman, T.F. Evolution of the chloroplast genome. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2003**, *358*, 99–107. [CrossRef]
12. Liu, L.X.; Li, R.; Worth, J.R.P.; Li, X.; Li, P.; Cameron, K.M.; Fu, C.X. The Complete Chloroplast Genome of Chinese Bayberry (Morella rubra, Myricaceae): Implications for Understanding the Evolution of Fagales. *Front. Plant Sci.* **2017**, *8*, 968. [CrossRef]
13. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Kai, F.M.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [CrossRef] [PubMed]
14. Khan, A.; Khan, I.A.; Heinze, B.; Azim, M.K. The Chloroplast Genome Sequence of Date Palm (Phoenix dactylifera L. cv. 'Aseel'). *Plant Mol. Biol. Rep.* **2012**, *30*, 666–678. [CrossRef]
15. Bendich, A.J. Circular chloroplast chromosomes: The grand illusion. *Plant Cell* **2004**, *16*, 1661–1666. [CrossRef]
16. Chumley, T.W.; Palmer, J.D.; Mower, J.P.; Fourcade, H.M.; Calie, P.J.; Boore, J.L.; Jansen, R.K. The complete chloroplast genome sequence of Pelargonium x hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **2006**, *23*, 2175–2190. [CrossRef] [PubMed]
17. Knox, E.B. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11097–11102. [CrossRef]
18. Palmer, J.D. Chloroplast DNA exists in two orientations. *Nature* **1983**, *301*, 92–93. [CrossRef]

19. Gurdon, C.; Maliga, P. Two Distinct Plastid Genome Configurations and Unprecedented Intraspecies Length Variation in the accD Coding Region in Medicago truncatula. *DNA Res.* **2014**, *21*, 417–427. [CrossRef]
20. Vieira, L.D.N.; Helisson, F.; Marcelo, R.; Fraga, H.P.D.F.; Cardoso, R.L.A.; Emanuel Maltempi, D.S.; Fábio, D.O.P.; Rubens Onofre, N.; Miguel Pedro, G. The complete chloroplast genome sequence of Podocarpus lambertii: Genome structure, evolutionary aspects, gene content and SSR detection. *PLoS ONE* **2014**, *9*, e90618. [CrossRef]
21. Stein, D.B.; Palmer, J.D.; Thompson, W.F. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus Osmunda. *Curr. Genet.* **1986**, *10*, 835–841. [CrossRef]
22. Tangphatsornruang, S.; Sangsrakru, D.; Chanprasert, J.; Uthaipaisanwong, P.; Yoocha, T.; Jomchai, N.; Tragoonrung, S. The Chloroplast Genome Sequence of Mungbean (Vigna radiata) Determined by High-throughput Pyrosequencing: Structural Organization and Phylogenetic Relationships. *DNA Res.* **2009**, *17*, 11–22. [CrossRef] [PubMed]
23. Huang, Y.; Jun, W.; Yongping, Y.; Chuanzhu, F.; Jiahui, C. Phylogenomic Analysis and Dynamic Evolution of Chloroplast Genomes in Salicaceae. *Front. Plant Sci.* **2017**, *8*, 1050. [CrossRef] [PubMed]
24. Zong, D.; Gan, P.; Zhou, A.; Zhang, Y.; Zou, X.; Duan, A.; Song, Y.; He, C. Plastome Sequences Help to Resolve Deep-Level Relationships of Populus in the Family Salicaceae. *Front. Plant Sci.* **2019**, *10*, 5. [CrossRef] [PubMed]
25. Zong, D.; Zhou, A.; Zhang, Y.; Zou, X.; Li, D.; Duan, A.; He, C. Characterization of the complete chloroplast genomes of five Populus species from the western Sichuan plateau, southwest China: Comparative and phylogenetic analyses. *PeerJ* **2019**, *7*, e6386. [CrossRef] [PubMed]
26. Huang, D.I.; Hefer, C.A.; Kolosova, N.; Douglas, C.J.; Cronk, Q.C. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytol.* **2014**, *204*, 693–703. [CrossRef] [PubMed]
27. Eckenwalder, E.J. Taxonomic Signal and Noise in Multivariate Interpopulational Relationships in Populus mexicana (Salicaceae). *Syst. Bot.* **1996**, *21*, 261–271. [CrossRef]
28. Chen, J.H.; Sun, H.; Wen, J.; Yang, Y.P. Molecular Phylogeny of Salix L. (Salicaceae) Inferred from Three Chloroplast Datasets and Its Systematic Implications. *Taxon* **2010**, *59*, 29–37. [CrossRef]
29. Lauron-Moreau, A.; Pitre, F.E.; Argus, G.W.; Labrecque, M.; Brouillet, L. Phylogenetic relationships of American willows (Salix L., Salicaceae). *PLoS ONE* **2015**, *10*, e0121965. [CrossRef]
30. Wu, J.; Nyman, T.; Wang, D.C.; Argus, G.W.; Yang, Y.P.; Chen, J.H. Phylogeny of Salix subgenus Salix s.l. (Salicaceae): Delimitation, biogeography, and reticulate evolution. *BMC Evol. Biol.* **2015**, *15*, 31. [CrossRef]
31. Lei, Z.; Xi, Z.X.; Wang, M.C.; Guo, X.Y.; Ma, T. Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecol. Evol.* **2018**, *8*, 7817–7823.
32. Palmer, J.D. Plastid Chromosomes; Structure and Evolution. In *The Molecular Biology of Plastids*; Bogorad, L., Vasil, I.K., Eds.; Academic Press: San Diego, CA, USA, 1911; pp. 5–53.
33. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; Depamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J. Methods for Obtaining and Analyzing Whole Chloroplast Genome Sequences. *Methods Enzym.* **2005**, *395*, 348–383.
34. Kim, K.J.; Jansen, R.K. ndhF sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10379–10383. [CrossRef]
35. Henry, R.J.; Henry, R.J. Plant diversity and evolution: Genotypic and phenotypic variation in higher plants. *Biologist* **2005**, 121.
36. Dugas, D.V.; Hernandez, D.; Koenen, E.J.M.; Schwarz, E.; Straub, S.; Hughes, C.E.; Jansen, R.K.; Nageswara-Rao, M.; Staats, M.; Trujillo, J.T. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution inclpP. *Sci. Rep.* **2015**, *5*, 16958. [CrossRef]
37. Millen, R.S.; Olmstead, R.G.; Adams, K.L.; Palmer, J.D.; Lao, N.T.; Heggie, L.; Kavanagh, T.A.; Hibberd, J.M.; Gray, J.C.; Morden, C.W. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **2001**, *13*, 645–658. [CrossRef]
38. Thomas, F.; Massenet, O.; Dorne, A.M.; Briat, J.F.; Mache, R. Expression of the rpl23, rpl2 and rps19 genes in spinach chloroplasts. *Nucleic Acids Res.* **1988**, *16*, 2461–2472. [CrossRef]
39. Downie, S.R.; Olmstead, R.G.; Zurawski, G.; Soltis, D.E.; Palmer, J.D. Six independent loss of the chloroplast DNA rpl2 intron in Dicotyledons: Molecular and phylogenetic implications. *Evolution* **1991**, *45*, 1245–1259.

40. Xiong, A.S.; Peng, R.H.; Zhuang, J.; Gao, F.; Zhu, B.; Fu, X.Y.; Xue, Y.; Jin, X.F.; Tian, Y.S.; Zhao, W. Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol. Adv.* **2009**, *27*, 340–347. [CrossRef]
41. Haberle, R.C.; Fourcade HMBoore, J.L.; Jansen, R.K. Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. *J. Mol. Evol.* **2008**, *66*, 350–361. [CrossRef]
42. Hae-Lim, L.; Jansen, R.K.; Chumley, T.W.; Ki-Joong, K. Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* **2007**, *24*, 1161–1180.
43. Hipkins, V.D.; Marshall KANeale, D.B.; Rottmann, W.H.; Strauss, S.H. A mutation hotspot in the chloroplast genome of a conifer (Douglas-fir: Pseudotsuga) is caused by variability in the number of direct repeats derived from a partially duplicated tRNA gene. *Curr. Genet.* **1995**, *27*, 572–579. [CrossRef]
44. Lenka, D.; Jan, K.; Cestmír, V.; Václav, P. TrnL- trnF intergenic spacer and trnL intron define major clades within Luzula and Juncus (Juncaceae): Importance of structural mutations. *J. Mol. Evol.* **2004**, *59*, 1–10.
45. Lin, C.P.; Wu, C.S.; Huang, Y.Y.; Chaw, S.M. The complete chloroplast genome of Ginkgo biloba reveals the mechanism of inverted repeat contraction. *Genome Biol. Evol.* **2012**, *4*, 374–381. [CrossRef]
46. Vijverberg, K.; Bachmann, K. Molecular evolution of a tandemly repeated trnF (GAA) gene in the chloroplast genomes of Microseris (Asteraceae) and the use of structural mutations in phylogenetic analyses. *Mol. Biol. Evol.* **1999**, *16*, 1329–1340. [CrossRef]
47. Jiao, Y.; Jia, H.M.; Li, X.W.; Chai, M.L.; Jia, H.J.; Chen, Z.; Wang, G.Y.; Chai, C.Y.; Weg, E.V.D.; Gao, Z.S. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (Myrica rubra). *BMC Genom.* **2012**, *13*, 201. [CrossRef]
48. Kaundun, S.S.; Matsumoto, S. Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, Camellia sinensis. *Genome* **2002**, *45*, 1041–1048. [CrossRef]
49. Powell, W.; Morgante, M.; Mcdevitt, R.; Vendramin, G.G.; Rafalski, J.A. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7759–7763. [CrossRef]
50. Pauwels, M.V.X.; Godé, C.; Frérot, H.; Castric, V.; Saumitou-Laprade, P. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, Arabidopsis halleri (Brassicaceae). *New Phytol.* **2012**, *193*, 916–928. [CrossRef]
51. Yeong Deuk, J.; Jongsun, P.; Jungeun, K.; Wonho, S.; Cheol-Goo, H.; Yong-Hwan, L.; Byoung-Cheorl, K. Complete sequencing and comparative analyses of the pepper (Capsicum annuum L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Rep.* **2011**, *30*, 217–229.
52. Cosner, M.E.; Jansen, R.K.; Palmer, J.D.; Downie, S.R. The highly rearranged chloroplast genome of Trachelium caeruleum (Campanulaceae): Multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr. Genet.* **1997**, *31*, 419–429. [CrossRef]
53. Mao-Lun, W.; Blazier, J.C.; Madhumita, G.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659.
54. Ogihara, Y.; Terachi, T.; Sasakuma, T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8573–8577. [CrossRef]
55. Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative Analysis of the Complete Chloroplast Genomes of Five Quercus Species. *Front. Plant Sci.* **2016**, *7*, 959. [CrossRef]
56. Daniell, H.; Lin, C.S.; Ming, Y.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef]
57. Wu, F.H.; Chan, M.T.; Liao, D.C.; Hsu, C.T.; Lee, Y.W.; Daniell, H.; Duvall, M.R.; Lin, C.S. Complete chloroplast genome ofOncidiumGower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. *BMC Plant Biol.* **2010**, *10*, 68. [CrossRef]
58. Ki-Joong, K.; Hae-Lim, L. Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol. Cells* **2005**, *19*, 104–113.
59. Bouchenak-Khelladi, Y.; Maurin, O.; Hurter, J.; van der Bank, M. The evolutionary history and biogeography of Mimosoideae (Leguminosae): An emphasis on African acacias. *Mol. Phylogen. Evol.* **2010**, *57*, 495–508. [CrossRef]

60. Miller, J.T.; Seigler, D. Evolutionary and taxonomic relationships of *Acacia s.l.* (Leguminosae: Mimosoideae). *Aust. Syst. Bot.* **2012**, *25*, 217–224. [CrossRef]
61. Givnish, T.J.; Spalink, D.; Ames, M.; Lyon, S.P.; Hunter, S.J.; Zuluaga, A.; Iles, W.J.D.; Clements, M.A.; Arroyo, M.T.K.; Leebens-Mack, J. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci.* **2015**, *282*, 2108–2111. [CrossRef]
62. Zhang, Y.J.; Ma, P.F.; Li, D.Z. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE* **2011**, *6*, e20596. [CrossRef]
63. Barba-Montoya, J.; Dos Reis, M.; Schneider, H.; Donoghue, P.C.J.; Yang, Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol.* **2018**, *218*, 819–834. [CrossRef]
64. Xi, Z.; Ruhfel, B.R.; Schaefer, H.; Amorim, A.M.; Sugumaran, M.; Wurdack, K.J.; Endress, P.K.; Matthews, M.L.; Stevens, P.F.; Mathews, S. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17519–17524. [CrossRef]
65. Wurdack, K.J.; Davis, C.C. Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *Am. J. Bot.* **2009**, *96*, 1551–1570. [CrossRef]
66. Cai, L.; Xi, Z.; Amorim, A.M.; Sugumaran, M.; Rest, J.S.; Liang, L.; Davis, C.C. Widespread ancient whole genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* **2017**, *221*, 565–576. [CrossRef]
67. Diffey, J.M. Phylogenetic Relationships of Salicaceae Based on Analyses of Nuclear DNA Data. *Honor. Theses* **2017**, *5*, 21.
68. Group, T.A.P.; Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Byng, J.W.; Judd, W.S.; Soltis, D.E.; Mabberley, D.J.; Sennikov, A.N.; Soltis, P.S.; et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–21.
69. Allen, G.C.; Flores-Vergara, M.A.; Krasynanski, S.; Kumar, S.; Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [CrossRef]
70. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2017**, *45*, e18.
71. Matthew, K.; Richard, M.; Amy, W.; Steven, S.H.; Matthew, C.; Shane, S.; Simon, B.; Alex, C.; Sidney, M.; Chris, D. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649.
72. Huang, D.I.; Cronk, Q.C.B. Plann: A Command-Line Application for Annotating Plastome Sequences. *Appl. Plant Sci.* **2015**, *3*, apps.1500026. [CrossRef]
73. Marc, L.; Oliver, D.; Sabine, K.; Ralph, B. OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, W575–W581.
74. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Appl. Genet.* **2003**, *106*, 411–422. [CrossRef]
75. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef]
76. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef]
77. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [CrossRef]
78. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2012**, *28*, 1647. [CrossRef]
79. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [CrossRef]
80. Alexandros, S. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313.
81. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]

82. Rannala, B.; Yang, Z. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **2007**, *56*, 453–466. [CrossRef]
83. Huang, C.H.; Sun, R.; Hu, Y.; Zeng, L.; Zhang, N.; Cai, L.; Zhang, Q.; Koch, M.A.; Al-Shehbaz, I.; Edger, P.P. Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution. *Mol. Biol. Evol.* **2016**, *33*, 394–412. [CrossRef]
84. Hermant, M.; Hennion, F.; Bartish, I.V.; Yguel, B.; Prinzing, A. Disparate relatives: Life histories vary more in genera occupying intermediate environments. *Perspect. Plant Ecol. Evol. Syst.* **2012**, *14*, 283–301. [CrossRef]
85. Baraloto, C.; Chave, J. Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *J. Ecol.* **2012**, *100*, 690–701. [CrossRef]
86. Vanneste, K.; Baele, G.; Maere, S.; Yves, V.D.P. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* **2014**, *24*, 1334–1347. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Gene-Wide Analysis of Aquaporin Gene Family in *Malus domestica* and Heterologous Expression of the Gene *MpPIP2;1* Confers Drought and Salinity Tolerance in *Arabidposis thaliana*

Haili Liu [1,†], Leilei Yang [1,†], Miaomiao Xin [1], Fengwang Ma [2] and Jingying Liu [1,*]

[1] State Key Laboratory of Crop Stress Biology for Arid Areas/Shaanxi Key Laboratory of Apple, College of Life Sciences, Northwest A&F University, Yangling 712100, China
[2] State Key Laboratory of Crop Stress Biology for Arid Areas/Shaanxi Key Laboratory of Apple, College of Horticulture, Northwest A&F University, Yangling 712100, China
* Correspondence: liujy@nwsuaf.edu.cn; Tel.: +86-029-8709-2262
† These authors contributed equally to this work.

Received: 24 June 2019; Accepted: 26 July 2019; Published: 29 July 2019

Abstract: The aquaporins (AQPs) are a family of integral membrane proteins involved in the transcellular membrane transport of water and other small molecules. A scan of the apple (*Malus domestica*) genome revealed the presence of 42 genes encoding putative AQPs. Based on a phylogenetic analysis of the deduced peptide sequences of the AQPs generated by *Arabidopsis thaliana*, poplar (*Populus trichocarpa*), and rubber (*Hevea brasiliensis*), the apple AQPs were each assigned membership of the five established AQP subfamilies, namely the PIPs (eleven members), the TIPs (thirteen members), the NIPs (eleven members), the SIPs (five members), and the XIPs (two members). The apple AQPs included asparagine-proline-alanine (NPA) motifs, an aromatic/arginine (ar/R) selectivity filter, and the Froger's positions. The heterologous expression of *MpPIP2;1* in *A. thaliana* was shown to enhance the level of tolerance exhibited against both drought and salinity.

Keywords: apple; aquaporin; functional analysis; stress tolerance

1. Introduction

The aquaporins (AQPs) represent a family of integral membrane proteins, and form channels which allow the transport of water and other small molecules across membranes [1]. These proteins are produced by species across the phylogenetic spectrum, from microbes to plants and animals [2]. A typical aquaporin feature six transmembrane (TM1–TM6) helices (H1–H6) and five connecting loops (LA–LE); both their carboxylic and amino terminals lie on the cytoplasmic side, while two half helices formed the seventh TM helix by the opposite LB and LE dipping into the membrane. Given their general conservation across many AQPs, the asparagines-proline-alanine (NPA) motifs, the aromatic/Arginine (ar/R) selectivity filter formed by four residues (F58-H182-C191-R197 in AQP1) [3], and Froger's positions (P1–P5 residues, T116-S196-A200-F212-W213 in AQP1) [4], are considered to be important for function. AQPs are tetrameric proteins and each monomer is functional independently as a channel. Furthermore, the fifth channel, which forms through the middle of tetramer array, has been suggested to conduct gases, like CO_2 [5].

The survival and growth of a plant depends on its ability to maintain a sufficient level of tissue hydration. Proteins referred to as AQPs are known to represent an important component of the regulatory machinery used by plants for this purpose. AQPs have been shown to exert control over germination, since pea seeds imbibed in the presence of the AQP inhibitor mercury do not germinate [6]. The correlation established between the elongation of the *Ricinus communis* seedling

hypocotyl and the abundance of the AQP-encoding gene *PIP2-1* has been taken to imply that AQPs also have a role in seedling growth [7]. The product of the tobacco gene *NtAQP1* has been shown to facilitate CO_2 membrane transport, and to contribute both to photosynthesis and stomatal movement [8,9]. The products of the strawberry AQP-encoding genes *FaPIP1;1* and *FaNIP1;1* both appear to be involved in the transport of water into the fruit [10,11]. A number of authors have reported that plant AQPs respond to external stress, triggering physiological adjustments which act to maintain the plant's hydration status [12–16].

Plant genomes encode a substantial number of AQPs: There are 35 such genes in the *Arabidopsis thaliana* genome [17], 72 in the soybean (*Glycine max*) genome [18], and 55 in the poplar (*Populus trichocarpa*) genome [19]. Based on their peptide sequences, higher plant AQPs have been classified into five subfamilies, namely the plasma membrane intrinsic proteins (PIPs), the tonoplast intrinsic proteins (TIPs), the Nod26-like intrinsic proteins (NIPs), the small and basic intrinsic proteins (SIPs), and the uncharacterized intrinsic proteins (XIPs). PIPs, TIPs, NIPs, and SIPs have been found in most higher plants, while to date XIPs have not been identified in *Brassicaceae* and monocots [20,21]. The most abundant of the AQPs are the PIPs and TIPs, most of which are associated with, respectively, the plasma membrane and the vacuolar membrane. Here, the family of apple (*Malus domestica*) AQPs has been characterized at the phylogenetic level, at the level of the chromosomal distribution of their encoding genes and with respect to the content of their functional domains. The effect on drought and salinity tolerance of one of these genes was explored by heterologously expressing it in *A. thaliana*.

2. Results

2.1. The Family of AQPs Encoded by the Apple Genome

A set of 67 candidate apple AQPs were identified by a key word search of the NCBI protein database. Several of these were discarded on the basis that they were either likely duplicates or represented a truncated sequence. The sequences of previously identified apple AQPs were used as queries of the whole apple genome sequence to identify additional members. The outcome was the identification of 42 putativeAQPs (Table 1). The range in length of their products was 236–309 residues and in their molecular weight was 25.1–33.2 KDa. The pI (isoelectric point) value of the presumptive AQPs varied from 4.86 to 9.97. Twenty nine of the forty two apple AQPs included six transmembrane domains.

Table 1. The information of identified apple AQPs.

Name	GeneBank accession no.	Gene ID	Size(aa)	MW(Da)	pI	TMD	Loc
MdNIP1;1	XP_008383590.1		271	28742.49	9.03	6	Plas
MdNIP1;2	XP_028954386.1		271	28743.38	8.99	6	Plas
MdNIP2;1	XP_008341577.1		290	30456.02	6.70	6	Plas
MdNIP2;2	XP_008341755.2		290	30513.07	6.70	6	Plas
MdNIP2;3		MD09G1051500 [1]	260	27306.65	6.65	6	Plas
MdNIP4;1		MD13G1025400 [1]	278	29945.16	8.97	7	Plas
MdNIP5;1	XP_008356814.1		298	30960.84	8.94	5	Vacu
MdNIP5;2	XP_008360732.1		298	31052.15	8.85	6	Vacu
MdNIP5;3	XP_008348330.2		266	27575.92	9.23	5	Plas
MdNIP6;1	XP_008343681.1		306	31641.74	8.90	6	Vacu
MdNIP7;1	XP_028952006.1		300	31457.78	8.05	6	Plas
MdPIP1;1	NP_001280914.1		289	30875.87	9.32	5	Plas
MdPIP1;2	NP_001280922.1		289	30849.79	9.30	6	Plas
MdPIP1;3	NP_001315794.1		286	30759.57	9.08	6	Plas
MdPIP1;4	NP_001280950.1		286	30660.58	9.15	6	Plas
MdPIP2;1	XP_008363507.1		281	30166.00	7.65	6	Plas
MdPIP2;2	XP_008385311.2		281	30179.06	8.25	6	Plas
MdPIP2;3	XP_008365039.2		287	30461.39	8.95	6	Plas

Table 1. *Cont.*

Name	GeneBank accession no.	Gene ID	Size(aa)	MW(Da)	pI	TMD	Loc
MdPIP2;4	XP_008367680.2		283	30050.96	8.68	6	Plas
MdPIP2;5	XP_008387595.1		285	30156.86	6.89	6	Plas
MdPIP2;6	XP_008382110.2		281	29856.60	9.08	6	Plas
MdPIP2;7	XP_008377729.1		281	29924.69	9.22	6	Plas
MdSIP1;1	XP_008357207.2		242	25827.57	9.55	6	Vacu
MdSIP1;2	XP_008348137.2		240	25347.21	9.96	5	Vacu/Plas
MdSIP1;3	XP_008342016.1		240	25427.44	9.97	5	Vacu
MdSIP2;1	XP_008354672.2		236	25496.16	9.17	5	Vacu
MdSIP2;2	XP_008338069.3		236	25364.97	9.48	6	Chlo
MdTIP1;1	XP_008387528.2		252	26029.00	5.18	6	Plas
MdTIP1;2	XP_008343557.2		252	26029.25	5.62	6	Plas
MdTIP1;3	XP_008357781.1		252	25944.08	4.96	6	Plas
MdTIP1;4	XP_008366336.1		252	25922.12	4.96	6	Plas
MdTIP2;1	XP_008342659.1		248	25234.34	5.76	7	Plas
MdTIP2;2	XP_008380900.2		248	25308.41	5.62	7	Plas
MdTIP2;3	XP_008373770.1		248	25109.15	4.86	6	Vacu
MdTIP3;1	XP_008351935.2		256	26996.43	7.10	6	Plas
MdTIP3;2		MDP0000868372 [2]	255	26710.01	7.06	5	Plas
MdTIP4;1	XP_008393878.2		248	25864.24	5.51	7	Plas
MdTIP4;2		MD08G1115300 [1]	239	25098.22	4.98	7	Plas
MdTIP5;1	XP_008354229.2		254	26270.51	7.74	6	E.R.
MdTIP5;2	XP_008376403.3		256	26196.42	6.24	6	Chlo/plas
MdXIP1;1	XP_017182198.2		304	32165.86	5.96	6	Plas
MdXIP2;1[2]	XP_028965550.1		309	33249.21	8.21	7	Plas

[1] Gene IDs with were based on Malus × domestica GDDH13 Whole Genome v1.1. [2] Gene IDs were based on Malus × domestica Whole Genome v1.0. (AQPs, aquaporins; MW, molecular weight; pI, isoelectric point; TMD, transmembrane domain; Loc, subcellular localization; Chlo, chloroplast; Plas, plasma membrane; Vacu, vacuolar membrane; E. R., endoplasmic reticulum).

2.2. Phylogenetic Analysis

The phylogenetic relationships between the set of MdAQPs with homologous proteins encoded by *A. thaliana*, poplar, and rubber (*Hevea brasiliensis*) is displayed in Figure 1. The analysis allowed the set of apple AQPs to be each assigned membership of one of the five plant AQP subfamilies, namely the MdPIPs (eleven members), the MdTIPs (thirteen members), the MdNIPs (eleven members), the MdSIPs (five members), and the MdXIPs (two members). The MdPIP members were further classified into the two subgroups, MdPIP1 and MdPIP2, the MdSIPs into the two subgroups, MdSIP1 and MdSIP2, and the MdTIPs into the five subgroups, MdTIP1-MdTIP5. The two MdXIPs belonged into the two subgroups MdXIP1 and MdXIP2, respectively, and the MdNIPs were divided into six subgroups, MdNIP1, 2, 4, 5, 6 and 7. On the basis of sharing a level of >90% similarity at the peptide level, 14 pairs of sequences were recognized, namely MdNIP1;1/MdNIP1;2, MdNIP2;1/MdNIP2;2, MdNIP5;1/MdNIP5;2, MdTIP1;3/MdTIP1;4, MdTIP4;1/MdTIP4;2, MdTIP2;1/MdTIP2;2, MdTIP3;1/MdTIP3;2, MdTIP5;1/MdTIP5;2, MdPIP1;1/MdPIP1;2, MdPIP2;1/MdPIP2;2, MdPIP2;3/MdPIP2;4, MdPIP2;6/MdPIP2;7, MdSIP1;2/MdSIP1;3, and MdSIP2;1/MdSIP2;2.

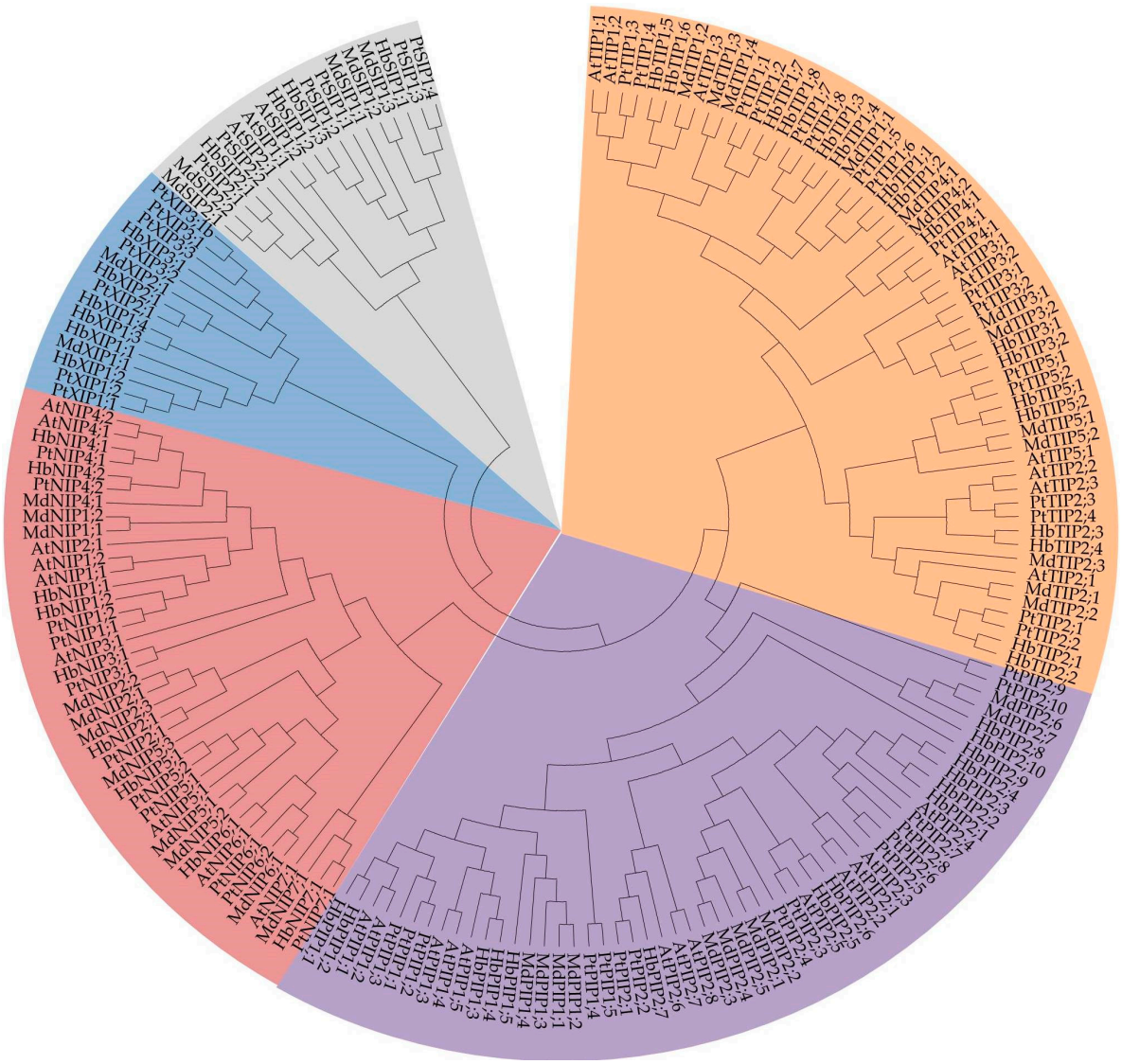


Figure 1. Phylogenetic tree of AQPs from *Arabidopsis thaliana*, *Populus trichocarpa*, *Hevea brasiliensis*, and *Malus domestica*. The protein sequences were aligned by ClustalX and the phylogenetic tree was constructed by the Neighbor-Joining method (1000 bootstrap replicates) in the MEGA6 software. The subgroups are marked by a colorful background (orange for TIPs, purple for PIPs, red for NIPs, blue for XIPs and gray for SIPs).

2.3. Chromosomal Location and Gene Structure

It was possible to map 40 of the 42 *MdAQPs* on 16 of the 17 apple chromosomes, but neither *MdNIP1;2* nor *MdTIP3;2* could be placed (Figure 2). The sequence of each of the eleven *MdPIPs* featured three introns; all but one of the thirteen *MdTIPs* featured two introns (the exception was *MdTIP1;1* in which only one intron was present); eight of the eleven *MdNIP* sequences were interrupted by four introns, with three introns present in both *MdNIP5;1* and *MdNIP5;2*, and five in *MdNIP5;3*; three of the five *MdSIPs* harbored two introns, while neither *MdSIP1;2* nor *MdSIP1;3* featured any introns; finally, *MdXIP1;1* had one intron while *MdXIP2;1* included two introns (Figure 3).

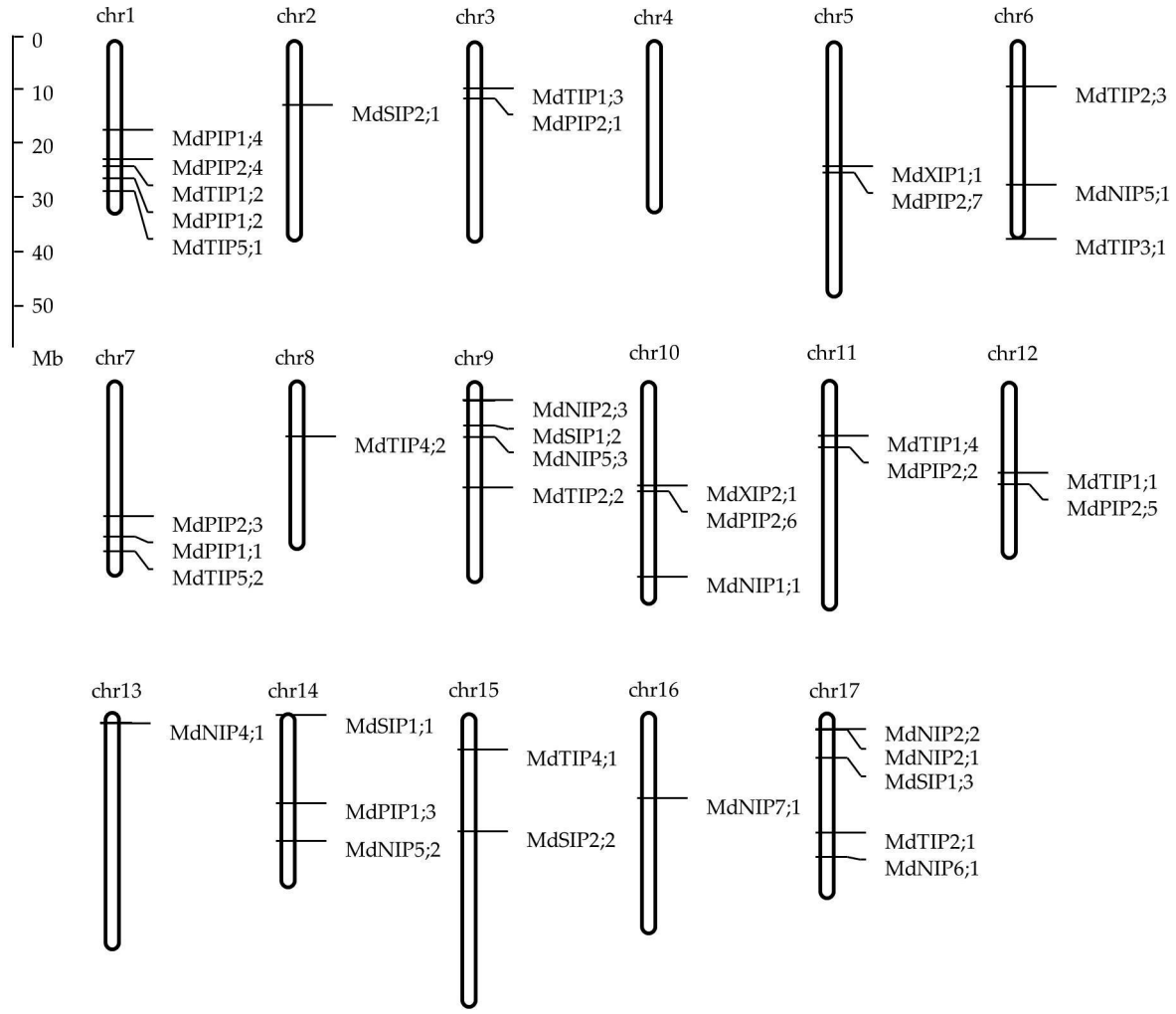


Figure 2. Distribution of *AQP* genes in apple chromosomes. Two genes (*MdNIP1;2* and *MdTIP3;2*) could not be localized on any chromosome. The scale is in megabases (Mb).

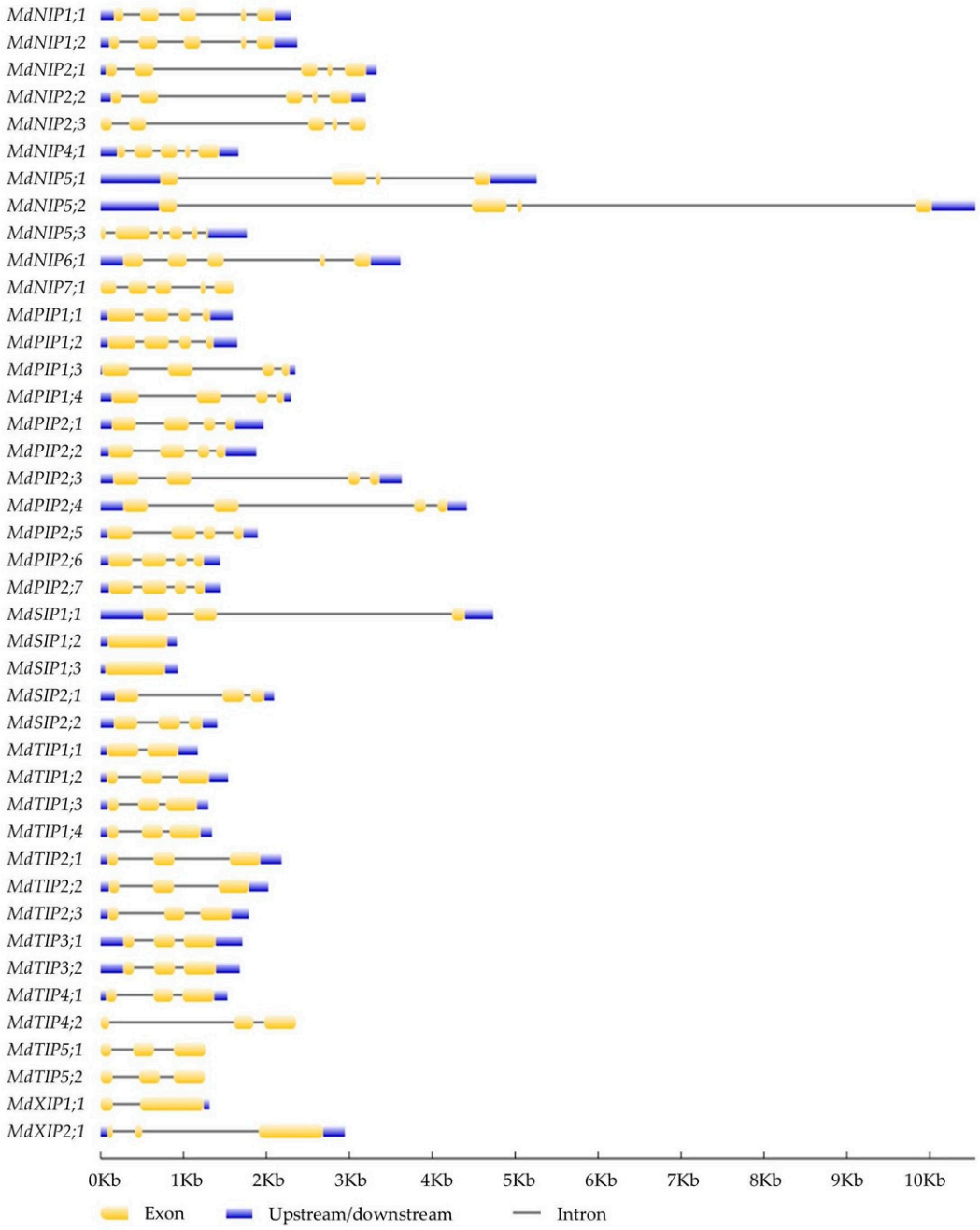


Figure 3. The exon-intron structure of apple *AQP* genes. Upstream/downstream region, exon, and intron are represented by blue box, yellow box, and grey line using GSDS software, respectively.

2.4. Conserved Residues in the Apple AQPs

The NPA motifs, ar/R filter, and Froger's positions were identified via a multiple alignment between the apple AQPs and other plant AQPs (Table 2). These conserved positions were critical for the substrate selectivity of AQPs. Both NPA domains were conserved in all MdPIP and MdTIP members, but the third residue of the first NPA in MdNIP5;1, MdNIP5;2, and MdNIP5;3 was serine rather than alanine, while in MdNIP5;1, MdNIP5;2, and MdNIP6;1, the third residue of the second NPA was valine rather than alanine; in MdNIP2;3 it was glutamate and in MdNIP5;3 it was isoleucine. The MdSIPs all carried a non-conserved third residue in the first NPA, while in addition, the first residue of the second NPA in MdSIP2;2 was serine rather than asparagine. Both the first and third residues of the first NPA of MdXIP1;1 were non-conserved and the third residues of the first NPA of MdXIP2;1 was valine. The ar/R

filter sequence was well conserved within each subfamily, but varied between the subfamilies. Each of the PIPs carried the conserved sequence phenylalanine-histidine-threonine-arginine. The greatest diversity for this motif was present among the NIPs, where each of the tetrapeptides tryptophan-valine-alanine-arginine, glycine-serine-glycine-arginine, alanine-isoleucine-glycine-arginine, threonine-isoleucine-alanine-arginine, and alanine-valine-glycine-arginine was represented. With respect to the Froger's positions, there was also high conservation within each subfamily, but variability between subfamilies. While the P1 position was the only variable residue within the PIP and XIP subfamily members, both the P1 and P2 positions varied among members of TIP and SIP subfamilies, and the P1, P2, and P5 positions were were all non-conserved for NIP subfamily members.

Table 2. NPA motifs, ar/R filter, and Froger's positons of apple AQPs.

Name	NPA motifs		Ar/R filter				Froger's Positions				
	LB	LE	H2	H5	LE1	LE2	P1	P2	P3	P4	P5
MdNIP1;1	NPA	NPA	W	V	A	R	F	S	A	Y	I
MdNIP1;2	NPA	NPA	W	V	A	R	F	S	A	Y	I
MdNIP2;1	NPA	NPA	G	S	G	R	L	T	A	Y	V
MdNIP2;2	NPA	NPA	G	S	G	R	L	T	A	Y	V
MdNIP2;3	NPA	NPE	G	S	-	R	L	T	A	Y	V
MdNIP4;1	NPA	NPA	W	V	A	R	L	S	A	Y	F
MdNIP5;1	NPS	NPV	A	I	G	R	F	T	A	Y	L
MdNIP5;2	NPS	NPV	A	I	G	R	L	T	A	Y	L
MdNIP5;3	NPS	NPI	A	I	G	R	F	T	A	Y	L
MdNIP6;1	NPA	NPV	T	I	A	R	F	T	A	Y	L
MdNIP7;1	NPA	NPA	A	V	G	R	Y	S	A	Y	I
MdPIP1;1	NPA	NPA	F	H	T	R	E	S	A	F	W
MdPIP1;2	NPA	NPA	F	H	T	R	E	S	A	F	W
MdPIP1;3	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP1;4	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;1	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;2	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;3	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;4	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;5	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;6	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdPIP2;7	NPA	NPA	F	H	T	R	Q	S	A	F	W
MdSIP1;1	NPT	NPA	V	L	P	N	M	A	A	Y	W
MdSIP1;2	NPS	NPA	S	L	P	N	M	A	A	Y	W
MdSIP1;3	NPS	NPA	S	L	P	N	M	A	A	Y	W
MdSIP2;1	NPL	NPA	S	L	G	S	F	V	A	Y	W
MdSIP2;2	NPL	SPA	S	L	G	S	F	V	A	Y	W
MdTIP1;1	NPA	NPA	H	I	A	V	T	S	A	Y	W
MdTIP1;2	NPA	NPA	H	I	A	V	T	S	A	Y	W
MdTIP1;3	NPA	NPA	H	I	A	V	T	S	A	Y	W
MdTIP1;4	NPA	NPA	H	I	A	V	T	S	A	Y	W
MdTIP2;1	NPA	NPA	H	I	G	R	T	S	A	Y	W
MdTIP2;2	NPA	NPA	H	I	G	R	T	S	A	Y	W
MdTIP2;3	NPA	NPA	H	I	G	R	T	S	A	Y	W
MdTIP3;1	NPA	NPA	H	I	A	R	T	A	A	Y	W
MdTIP3;2	NPA	NPA	H	I	A	R	T	A	A	Y	W
MdTIP4;1	NPA	NPA	H	I	A	R	S	S	A	Y	W
MdTIP4;2	NPA	NPA	H	I	A	R	S	S	A	Y	W
MdTIP5;1	NPA	NPA	N	V	G	C	T	A	A	Y	W
MdTIP5;2	NPA	NPA	N	V	G	C	I	A	A	Y	W
MdXIP1;1	SPV	NPA	V	V	V	R	M	C	A	F	W
MdXIP2;1	NPV	NPA	I	T	V	R	V	C	A	F	W

NPA, Asparagine-Proline-Alanine; Ar/R, aromatic/arginine; LE, loop E; LB, Loop B; H2, transmembrane helix 2; H5, transmembrane helix 5.

2.5. The Site of Apple PIP2 Expression

A search of the set of apple ESTs deposited in GenBank resulted in 685 hits for seven *MdPIP2s*, a subgroup of the *M. domestica AQP* gene family, the products of which are likely important regulators of water transport across the plasma membrane. The distribution of these hits among each *PIP2* gene and the organ are shown in Figure 4. The most well represented gene was *MdPIP2;4* (218 hits) and the least well represented was *MdPIP2;5* (seven hits). The site of transcription of these genes can be inferred from the frequency of their transcripts' occurrence in the 76 cDNA libraries assembled from various organs. In the bud libraries, there were 116 such ESTs out of a total of 54,099 sequences; there were 104 out of 54,120 in the leaf libraries; there were 73 out of 35,380 in the stem libraries; there was 94 out of 12,679 in the root libraries; there were 117 out of 44,772 in the flower libraries; there were 176 out of 104,341 in the fruit libraries; and finally out of libraries constructed from in vitro cultured cells, there were 5 out of 5652. When the abundance of transcripts generated from seven of the *PIP2* genes was evaluated by applying a quantitative real-time PCR (qRT-PCR) assay to RNA extracted from *M. hupehensis* root tissue, *PIP2;1* appeared to be the gene most strongly transcribed (Figure 5). Since both drought and salinity stress are sensed by roots, *PIP2;1* was chosen for a detailed functional analysis. In particular, the copy present in *Malus prunifolia* was selected, as this species provides a source of drought-tolerant rootstocks [22].

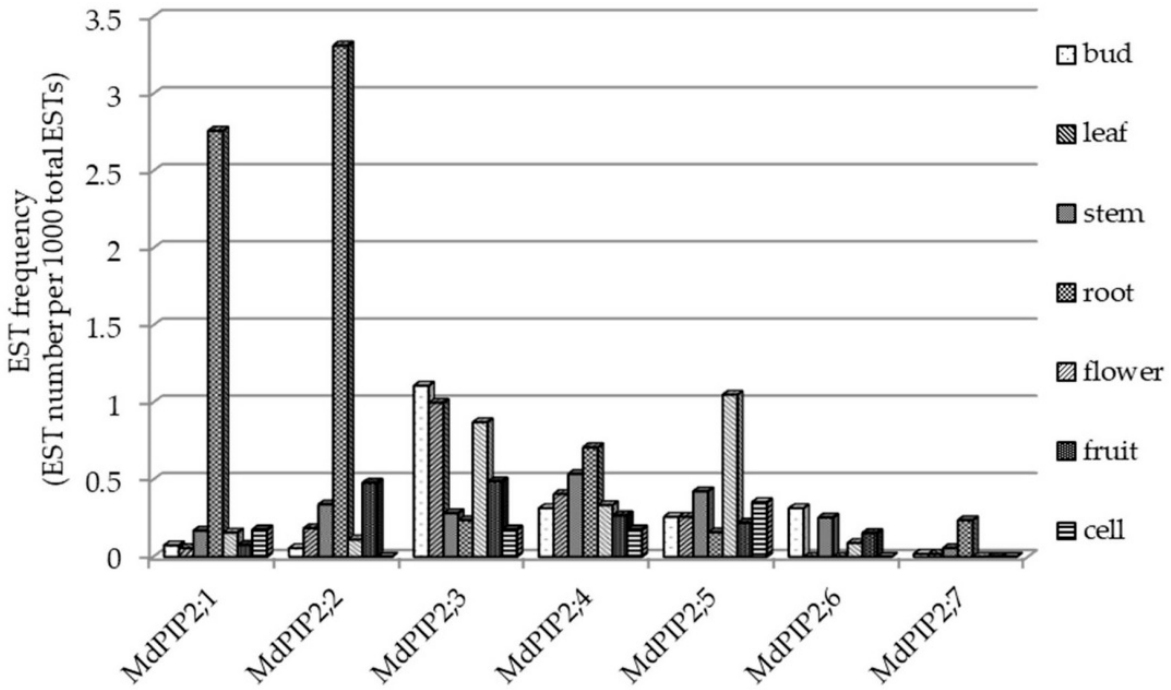


Figure 4. The frequency of the various *MdPIP2* ESTs present in cDNA libraries constructed from RNA extracted from buds, leaves, stems, roots, flowers, fruit and in vitro cultured cells.

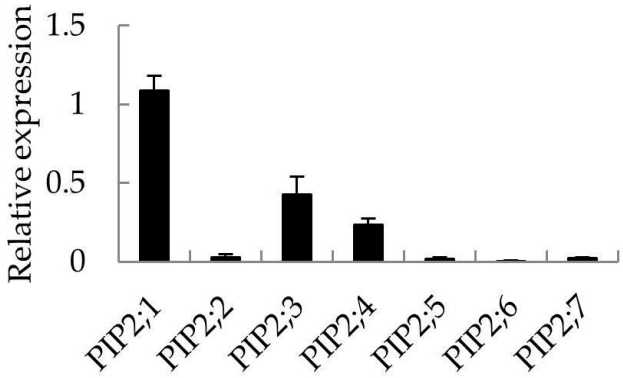


Figure 5. Transcriptional profiling of seven *PIP2* genes in the root of *M. hupehensis*. RNA was extracted from the roots of hydroponically-raised *M. hupehensis* seedlings which had formed 7–8 true leaves. Values show in the form mean ± SD (n = 3).

2.6. The Abiotic Stress Tolerance of A. thaliana Plants Heterologously Expressing MpPIP2;1

A set of ten independent *A. thaliana* transgenics heterologously expressing *MpPIP2;1* was obtained, and three of these were randomly selected to advance to the T3 generation (Figure 6). The performance of the three transgenic lines was then compared with that of wild type (WT) Col-0 plants.

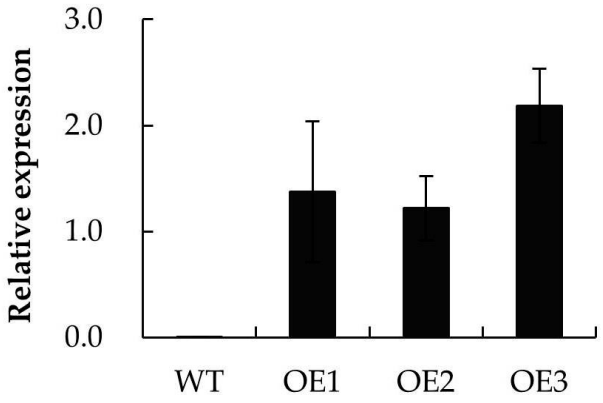


Figure 6. Relative expression of *MpPIP2;1* in three *A. thaliana* transgenic lines heterologously expressing *MpPIP2;1* using qRT-PCR with Col-0 as control. The value presented was the mean ± SD of three replicates.

The contrasting effect of drought stress on the transgenic and WT plants is illustrated in Figure 7. Under well-watered conditions, the growth of the transgenic lines was indistinguishable from that of WT plants. However, when water was withheld for 30 days, none of the WT plants remained viable as they were unable to maintain a sufficient level of leaf hydration (their relative water content fell to 10.5%); in contrast, many of transgenic plants survived, maintaining a leaf relative water content of about 20%. The post-stress survival rate of these latter plants was between 10.3% and 23.5%. A comparison of the leaf malondialdehyde (MDA) content showed that less of this stress marker accumulated in the transgenic plants than in WT plants; similarly, it was established that the relative electrolyte leakage of leaves sampled from the transgenic plants was lower than that of WT leaves. The activity of each of the enzymes superoxide dismutase (SOD), catalase (CAT), and peroxidase (POD), as well as the content of glutathione (GSH), were all greater in the transgenic plants than in the WT ones. Consistent with this result, the detached leaves of transgenic lines lost water slower than that of WT leaves. After 5 h of dehydration, the water loss rate for transgenic lines OE1, OE2, and OE3 were 17.6%, 8.8%, and 4.13% lower than that of WT, respectively.

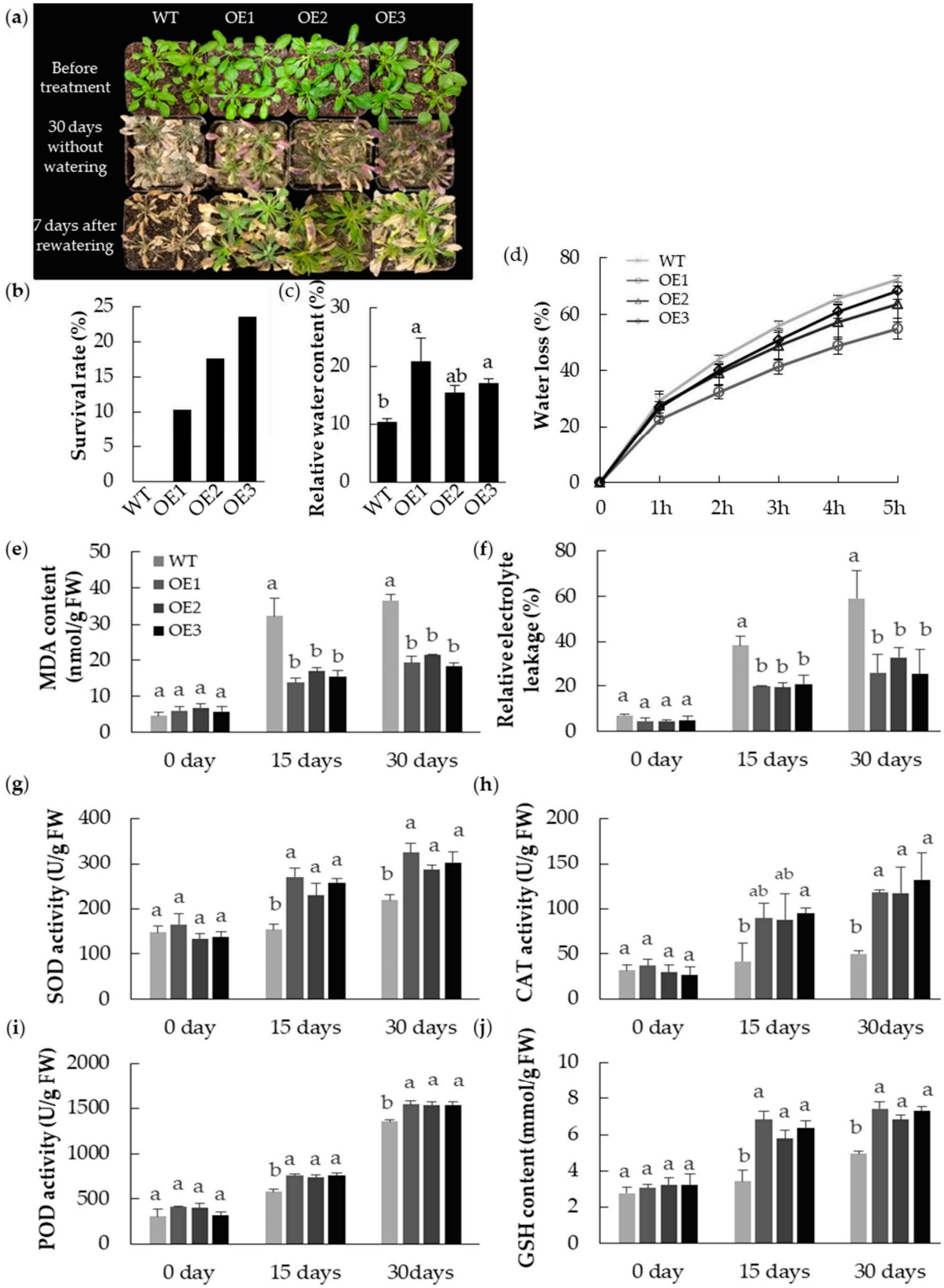


Figure 7. Heterologous expressing *MpPIP2;1* enhanced drought tolerance in *A. thaliana*. (**a**) Phenotypes of transgenic lines and wild type plants under drought stress; (**b**) survival rate after 30 days withholding of water and 7 days after rewatering; (**c**) relative water content after 30 days withholding of water; (**d**) water loss rate of detached leaves; (**e**) malondialdehyde (MDA) content; (**f**) relative electrolyte leakage; (**g**) superoxide dismutase (SOD) activity; (**h**) catalase (CAT) activity; (**i**) peroxidase (POD) activity; and (**j**) glutathione (GSH) content at 0, 15, and 30 days after water withheld. The value presented was the mean ± SD of three replicates, and the bar with different letter was significantly different between plants at $p < 0.05$.

A series of experiments were conducted to establish whether the constitutive expression of *MpPIP2;1* in *A. thaliana* had any effect on the level of tolerance to salinity stress, as imposed by exposure to 0.3 M NaCl for 14 days (Figure 8). The transgenic plants maintained a superior leaf hydration status compared to the WT plants: Their respective relative water contents were >80% and 49%. While the survival rate of WT plants was 46.4%, that of the transgenic plants was >90%. Compared to WT leaves, those of the transgenic plants accumulated less MDA, developed a lower relative electrolyte leakage, exhibited a higher activity of SOD, POD, and CAT, and their GSH content was greater.

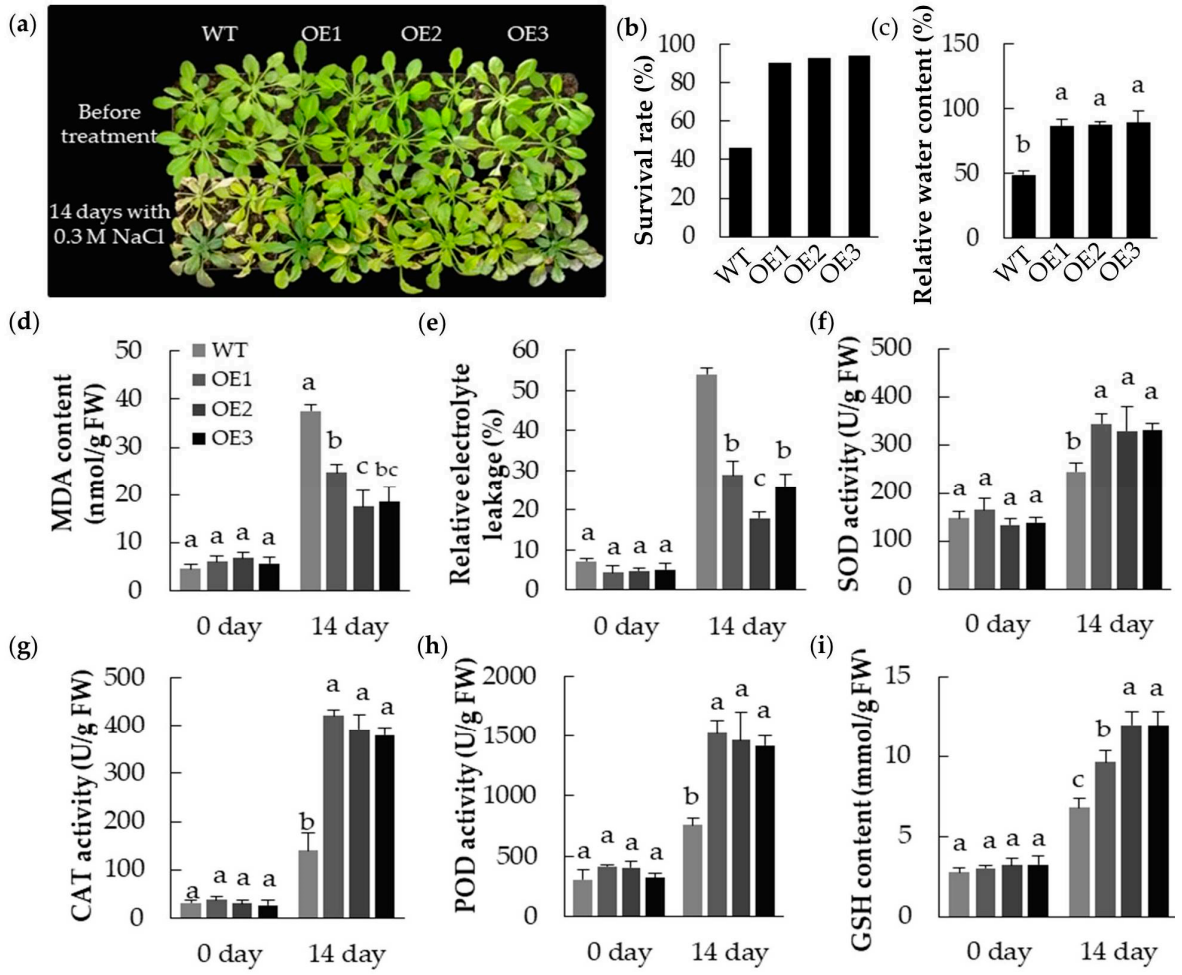


Figure 8. Heterologous expressing *MpPIP2;1* enhanced salt tolerance in *A. thaliana*. (**a**) Phenotypes of transgenic lines and wild type plants treated with 0.3 M NaCl for 14 days; (**b**) Survival rate and (**c**) relative water content after exposure to NaCl; (**d**) MDA content; (**e**) relative electrolyte leakage; (**f**) SOD activity; (**g**) CAT activity; (**h**) POD activity; and (**i**) GSH content at 0 and 14 days after NaCl treatment. The value presented was the mean ± SD of three replicates, and the bar with different letter was significantly different between plants at $p < 0.05$.

2.7. Germination and Root Elongation of MpPIP2;1 Transgenics Exposed to Either Salinity or Osmotic Stress

An experiment was conducted to establish whether the constitutive expression of *MpPIP2;1* in *A. thaliana* had any effect on germination and/or root elongation in the presence of either salinity or osmotic stress (Figure 9). When the seed was imbibed in the absence of a stress agent (mannitol or NaCl), the rate of germination of both the WT and transgenic seed was high, and there were no significant differences between the germination rates of WT and transgenic seeds. However, in the presence of either 0.25 M mannitol or 0.15 M NaCl, the rate of germination of the WT seeds fell to just 10%, while that of the transgenic seeds remained >60%. Similarly, the ability of roots to elongate was

the same for the WT and transgenic seedlings under non-stressful conditions, but the extent of its inhibition by the presence of either mannitol or NaCl differed between the transgenic and WT seedlings.

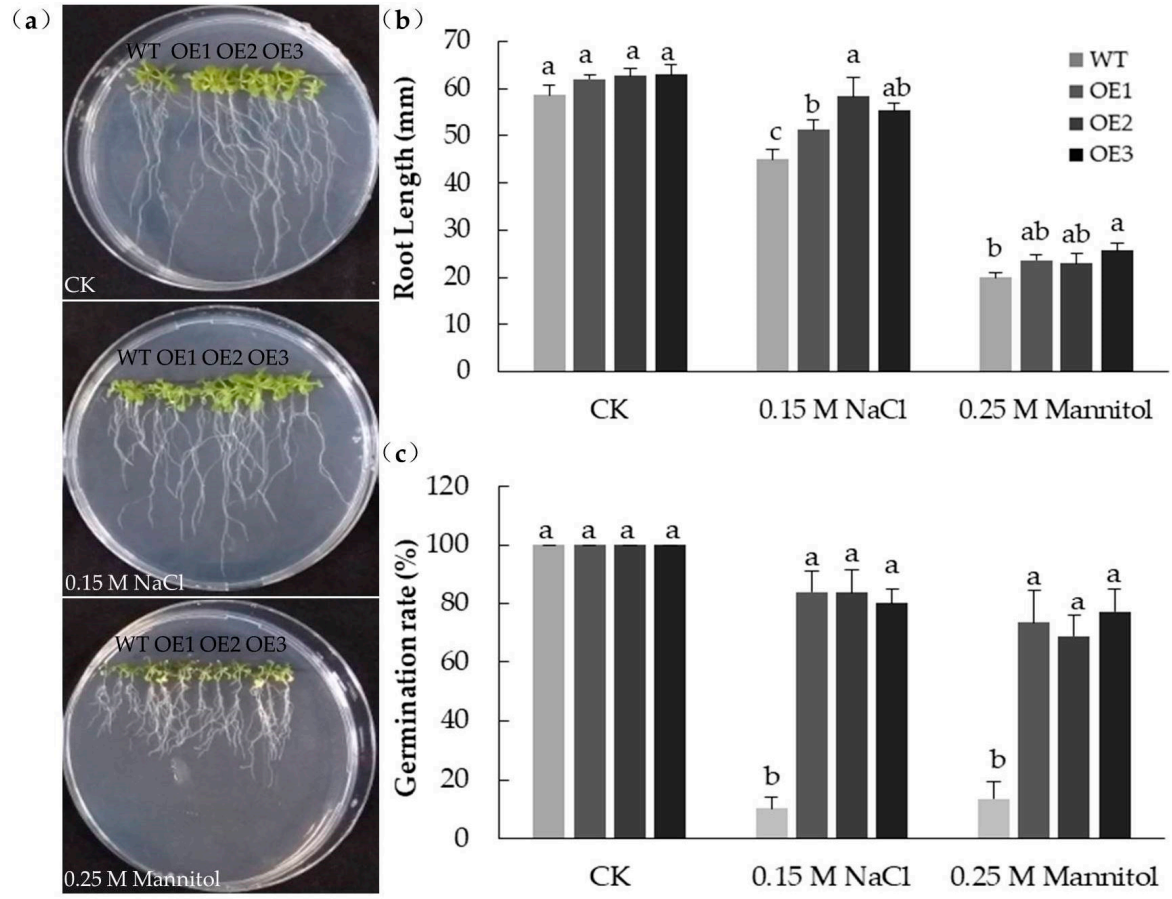


Figure 9. Heterologous expressing *MpPIP2;1* enhanced seeds germination and root elongation in *A. thaliana* either under salinity or osmotic stress. (**a**) The phenotype and (**b**) the statistical analyses of the root lengths of transgenic lines and wild type seedlings growing on MS medium (CK), or MS medium with 0.15 M NaCl or 0.25 M mannitol for 14 days; (**c**) germination rate of transgenic and wild type seeds on different mediums for 7 days. The value presented was the mean ± SD of three replicates, and the bar with different letter was significantly different between plants at $p < 0.05$.

3. Discussion

The systematic scanning of the content of AQP-encoding genes in the apple genome reported here resulted in the identification of 42 such genes. AQPs make an important contribution to the way in which plants control their uptake of water, and hence represent a key component of their response to drought and osmotic stress [23]. Thus, gaining a full understanding of how apple plants regulate their water balance and adapt to drought and osmotic stress will likely involve revealing the function of many of this set of genes. There is already some experimental evidence which supports the participation of *AQPs* in the stress response of apple. According to Hu et al. (2003), the transcription of both *MdPIP1a* and *MdPIP1b* (here renamed as, respectively, *MdPIP1;2* and *MdPIP1;1*) are up-regulated by osmotic stress [24]. The *M. zumi* homolog of *MdPIP1;1* has been shown to be inducible by salinity (as well as by low temperature) stress [25]. Meanwhile, the constitutive expression of *MzPIP2;1* (homolog of *MdPIP2;4*) in *A. thaliana* has a positive effect on drought tolerance and a small positive one on salinity tolerance [26], and the expression of *MzPIP1;3* (homolog of *MdPIP1;3*) in tomato has been shown to enhance the plants' drought tolerance [27].

Both moisture and nutrient stress are initially sensed by a plant's roots. Based on its relatively high transcript abundance (inferred both indirectly from the frequency of its representation in EST libraries and directly through a qRT-PCR analysis), *PIP2;1* is the *PIP2* gene most strongly transcribed in the roots of the apple plant. This same gene has been shown, using a suppression subtractive hybridization method, to be up-regulated in response to moisture deficit [28], while two *MdPIP2;1* ESTs have been identified in cDNA libraries developed from the roots of plants from which water had been withheld for a week (LIBEST_024527). Thus the evidence points to the conclusion that the product of *PIP2;1* is an important component of the apple plant's response to moisture stress. This evidence has been strengthened by the demonstration here that heterologously expressing the gene in *A. thaliana* had a positive effect on the plant's tolerance of both drought and salinity.

A commonly observed plant response to abiotic stress is to accumulate reactive oxygen species (ROS), which become cytotoxic when present in excess [29]. When *A. thaliana* plants heterologously expressing *MpPIP2;1* were exposed to either drought or salinity stress, both the MDA content and the relative electrolyte leakage of their leaves were below the levels shown by WT leaves; both of these traits are correlated with ROS-mediated cellular damage [30]. The maintenance of a non-damaging level of cellular ROS content is achieved both by the activity of a number of enzymes and the synthesis of antioxidant compounds [31]. Both the activity of the enzymes SOD, CAT, and POD and the content of the antioxidant compound GSH were higher in the transgenic than in the WT *A. thaliana* plants subjected to stress. The conclusion is that the product of *MpPIP2;1* likely contributes to protecting the transgenic plants experiencing drought stress by enhancing their ability to control the accumulation of ROS. Similar conclusions have been drawn with respect to AQPs in a number of plant systems [32], so it is reasonable to propose that the product of *MpPIP2;1*, a gene which is strongly transcribed in the root of *M. prunifolia*, is an important determinant of the drought stress response when expressed in its native context.

4. Materials and Methods

4.1. Identification of the Set of AQP Genes in the Apple Genome

Apple AQP sequences were recovered from the NCBI Protein database (www.ncbi.nlm.nih.gov/protein) by entering as a keyword search "(aquaporin OR MIP) AND Malus". The resulting hits were confirmed as genuine AQPs by submitting them to the NCBI Conserved Domain database (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The sequences of identified apple AQPs were used as queries to search Malus x domestica Whole Genome v1.0 and GDDH13 Whole Genome v1.1 sequence for additional members with an E value less than 0.01. A phylogenetic analysis based on the deduced peptide sequences of AQPs encoded by *A. thaliana*, poplar, and rubber [33] was used assign the apple sequences to the five established AQP subfamilies. Multiple sequence alignments were carried out using ClustalX software [34], and an unrooted phylogenetic tree was constructed using MEGA6 software [35], applying the Neighbor-Joining method and 1000 bootstrap replicates.

4.2. Chromosomal Location, Gene Structure, and Protein Properties of Apple AQPs

The GDDH13 assembly was used to reveal the chromosomal location for each of the *MdAQPs* and to determine the intron/exon structure of each gene. The latter was visualized using GSDS software (bio.tools/GSDS) [36]. The pI and molecular weight of the deduced AQPs were predicted using the ExPASY program (web.expasy.org/compute_pi/). Transmembrane regions were detected using TMHMM software (www.cbs.dtu.dk/services/TMHMM/) [37], and subcellular localizations were predicted using WoLF PSOR software (wolfpsort.hgc.jp/) [38]. Sequences representing conserved domains, NPA motifs, the ar/R filter, and the Froger positions were manually identified, based on multiple sequence alignments of apple AQPs with heterologous AQPs.

4.3. The Sites of Apple PIP2 Expression

Apple ESTs were retrieved through a BLASTN search of the GeneBank database, using as search terms each of the *MdPIP2* transcripts in turn. RNA was extracted from the roots of hydroponically-raised *M. hupehensis* seedlings (a triploid species characterized by facultative apomixis) which had formed 7–8 true leaves [39], and was processed for a series of qRT-PCR assays targeting seven *PIP2* genes. The relevant primers were designed using Beacon Designer 8 are shown in Table 3.

Table 3. Primers used in this study.

Gene Name [1]	Sequence (5′-3′)	Use	Reference
MpPIP2;1	F: ATGGCAAAAGATATTGAGGG R: TTAAGCATTGCTCCTGAAAG	Gene cloning	This study
	F: CTTGGCTCGCAAGGTTTCAC R: GTAGCCATCAGCCAACTCGT	qRT-PCR for *MpPIP2;1* in transgenic *Arabidopsis*	This study
AtACTIN	F: CCTTCTACCACCAATACATTC R: TGTTCCATTGTCGCATAC	qRT-PCR for reference gene	[40]
MhPIP2;1	F:CCTTCTACCACCAATACATTC R:TGATTATCTACAATTCCATAGCC	qRT-PCR for *MhPIP2;1*	This study
MhPIP2;2	F:GCGGTGGAACTGTAGATA R:GCTTTCTCTGGCATCAAT	qRT-PCR for *MhPIP2;2*	This study
MhPIP2;3	F:CAAGAGGAGTGCTAGAGAC R:GCCAAGTGGACAATGAAC	qRT-PCR for *MhPIP2;3*	This study
MhPIP2;4	F:CTTGGACCTGCTGTTATCT R:AATTGCTGCTCCGATGAA	qRT-PCR for *MhPIP2;4*	This study
MhPIP2;5	F:TGGATTATTCTGGAAGCAT R:GCAACATTAAGGCACATT	qRT-PCR for *MhPIP2;5*	This study
MhTIP2;6	F:GCAACCCGACCCACTAAA R:ACAACACTCTCAATACACACTACA	qRT-PCR for *MhPIP2;6*	This study
MhTIP2;7	F:CAGCAACCCAACCAACTAAA R:ATCATCATCCATCCTCTCTCAAT	qRT-PCR for *MhPIP2;7*	This study
MhACTIN	F:TTCGTTTTCGTTTTCGTTTT R:TGTTCCATTGTCGCATAC	qRT-PCR for reference gene	This study

[1] Primers were designed to target the seven *MhPIP2* based on the sequences of their homologs in *M. domestica*.

4.4. Heterologous Expression of MpPIP2;1 in A. thaliana

A full length copy of *MpPIP2;1* cDNA (JF834203.1) was PCR-amplified from an in-house p*MD19-T-MpPIP2;1* plasmid using a primer pair listed in Table 3 [41]. This was used to generate the construct p*CAMBIA2300-35S-MpPIP2;1*, which was introduced into *A. tumefaciens* strain GV3101, and from thence into *A. thaliana* (ecotype Col-0) using the floral dip method [42]. Selection for transgenic products was carried out by culturing on a medium containing 50 mg·L^{-1} kanamycin. The abundance of *MpPIP2;1* transcript produced in transgene homozygous T3 lines was evaluated using a qRT-PCR assay based on the primer pairs listed in Table 3.

4.5. Stress Tolerance Analysis

For the purpose of assaying in vitro germination, surface-sterilized WT and transgenic seeds were laid on either solidified Murashige and Skoog (1962) medium (MS) [43], MS containing 0.25 M mannitol, or MS containing 0.15 M NaCl, and held for seven days under a 16 h photoperiod at 23 °C. Root elongation was assessed by culturing pre-germinated seedlings for two weeks on vertically oriented plants containing MS, MS + 0.25 M mannitol, or MS + 0.15 M NaCl. To test for both drought and salinity tolerances, seedlings were grown in pots containing equal amounts of soil after stratified and generated in soil for 3 weeks. The water loss rate was measured by weighing the detached leaves at different time points. The plants were then either subjected to drought by the withholding of water for 30 days, or salinity stress by irrigation with 0.3 M NaCl solution [44,45]. The assays for relative leaf water content, leaf electrolyte leakage, and MDA concentration have been described elsewhere. To assay for SOD,

CAT, and POD activity and GSH content, leaves was homogenized in phosphate buffer (pH 7.5) and the resulting supernatants recovered after centrifugation were tested using commercially available kits purchased from Nanjing Jiancheng Bioengineering Institute (Nanjing, China). The SOD Assay Kit was based on hydroxylamine method, the CAT Assay Kit was based on the method of ammonium molybdate, the POD Assay Kit was based on the hydrogen peroxide oxidation reaction, and the GSH assay kit is based on reduced glutathione reacting with 5,5′-dithiobis-2-nitrobenoic acid.

4.6. Statistical Analysis

Data were statistically analyzed using routines implemented in SPSS v17.0. Means were compared using Tukey's test, applying a 0.05 probability threshold to declare significance.

5. Conclusions

The present experiment has identified that the apple genome harbors 42 genes encoding putative AQPs. It is necessary for the future to comprehensively understand the process of water and small solutes flues through the membrane and related molecular-regulating mechanisms in the apple tree. In *A. thaliana,* plants were engineered to heterologously express one of these genes (*MpPIP2;1*), a gene which is strongly transcribed in the root of *M. prunifolia*. The transgenic plants exhibited an improved level of tolerance to both drought and salinity stress, which implies that in the apple, the product of *PIP2;1* is an important component of the plant's response to moisture stress. This latter finding suggests a strategy for enhancing the stress tolerance of the apple using a molecular approach.

Author Contributions: Conceptualization, J.L.; data curation, J.L.; methodology, H.L. and L.Y.; project administration, F.M.; writing—original draft, M.X.; writing—review & editing, J.L.

Funding: This work was supported by the National Natural Science Foundation of China (31401839).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AQP	aquaporins
NPA	asparagine-proline-alanine
ar/R	aromatic/arginine
NIP	Nod26-like intrinsic proteins
PIP	plasma membrane intrinsic proteins
SIP	small and basic intrinsic proteins
TIP	tonoplast intrinsic proteins
XIP	uncharacterized intrinsic proteins
MDA	malondialdehyde
SOD	superoxide dismutase
CAT	catalase
POD	peroxidase
GSH	glutathione

References

1. Hove, R.M.; Bhave, M. Plant aquaporins with non-aqua functions: Deciphering the signature sequences. *Plant Mol. Biol.* **2011**, *75*, 413–430. [CrossRef] [PubMed]
2. Abascal, F.; Irisarri, I.; Zardoya, R. Diversity and evolution of membrane intrinsic proteins. *Biochim. Biophys. Acta.* **2014**, *1840*, 1468–1481. [CrossRef] [PubMed]
3. Sui, H.X.; Han, B.G.; Lee, J.K.; Walian, P.; Jap, B.K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **2001**, *414*, 872–878. [CrossRef] [PubMed]

4. Heymann, J.B.; Engel, A. Structural clues in the sequences of the aquaporins. *J. Mol. Biol.* **2000**, *295*, 1039–1053. [CrossRef] [PubMed]
5. Kreida, S.; Tornroth-Horsefield, S. Structural insights into aquaporin selectivity and regulation. *Curr. Opin. Struct. Biol.* **2015**, *33*, 126–134. [CrossRef] [PubMed]
6. Kshetrinayum, E.; Sahoo, D.P.; Mitra, J.; Panda, S.K. Regulation of seed germination and the role of aquaporins under abiotic stress. *Int. J. Agric. Biol.* **2017**, *2*, 607–615. [CrossRef]
7. Eisenbarth, D.A.; Weig, A.R. Dynamics of aquaporins and water relations during hypocotyl elongation in *Ricinus communis* L. seedlings. *J. Exp. Bot.* **2005**, *56*, 1831–1842. [CrossRef] [PubMed]
8. Flexas, J.; Ribas-Carbo, M.; Hanson, D.T.; Bota, J.; Otto, B.; Cifre, J.; McDowell, N.; Medrano, H.; Kaldenhoff, R. Tobacco aquaporin NtAQP1 is involved in mesophyll conductance to CO_2 *in vivo*. *Plant J.* **2006**, *48*, 427–439. [CrossRef]
9. Grondin, A.; Rodrigues, O.; Verdoucq, L.; Merlot, S.; Leonhardt, N.; Maurel, C. Aquaporins contribute to ABA-triggered stomatal closure through OST1-mediated phosphorylation. *Plant Cell.* **2015**, *27*, 1945–1954. [CrossRef]
10. Mut, P.; Bustamane, C.; Martinez, G.; Alleva, K.; Sutka, M.; Civello, M.; Amodeo, G. A fruit-specific plasma membrane aquaporin subtype *PIP1;1* is regulated druing strawberry (fragaria × ananassa) fruit ripening. *Physiol. Plant.* **2008**, *132*, 538–551. [CrossRef]
11. Molina-Hidalgo, J.F.; Medina-Puche, L.; Gelis, S.; Ramos, J. Functional characterization of *FaNIP1;1* gene, a ripening-related and receptacle-specific aquaporin in strawberry fruit. *Plant Sci* **2015**, *238*, 198–211. [CrossRef]
12. Chen, Q.; Yang, S.H.; Xiang, K.; Wang, C.T.; Xiang, N.; Yang, Y.Q.; Yang, Y.P. Molecular cloning of a plasma membrane aquaporin in *Stipa purpurea*, and exploration of its role in drought stress tolerance. *Gene* **2018**, *665*, 41–48. [CrossRef]
13. Li, W.; Qiang, X.J.; Han, R.X.; Jiang, L.L.; Zhang, S.H.; Han, J.; He, R.; Cheng, X.G. Ectopic Expression of a *Thellungiella salsuginea* aquaporin gene, *TsPIP1;1,* increased the salt tolerance of rice. *Int. J. Mol. Sci.* **2018**, *19*, 2229. [CrossRef]
14. Matsumoto, T.; Lian, H.L.; Su, W.A.; Tanaka, D.; Liu, C.W.; Iwasaki, I.; Kitagawa, Y. Role of the aquaporin PIP1 subfamily in the chilling tolerance of rice. *Plant Cell Physiol.* **2009**, *50*, 216–229. [CrossRef]
15. Shekoofa, A.; Sinclair, T.R. Aquaporin activity to improve crop drought tolerance. *Cells* **2018**, *7*, 123. [CrossRef]
16. Zhang, S.; Feng, M.; Chen, W.; Zhou, X.; Lu, J.; Wang, Y.; Li, Y.; Jiang, C.Z.; Gan, S.S.; Ma, N.; et al. In rose, transcription factor PTM balances growth and drought survival via PIP2;1 aquaporin. *Nat. Plants* **2019**, *5*, 290–299. [CrossRef]
17. Johanson, U.; Karlsson, M.; Johansson, I.; Gustavsson, S.; Sjövall, S.; Fraysse, L.; Laure Fraysse, L.; Weig, A.R.; Kjellbom, P. The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol.* **2001**, *126*, 1358–1369. [CrossRef]
18. Deshmukh, R.K.; Vivancos, J.; Guerin, V.; Sonah, H.; Labbe, C.; Belzile, F.; Belanger, R.R. Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in *Arabidopsis* and rice. *Plant Mol. Biol.* **2013**, *83*, 303–315. [CrossRef]
19. Gupta, A.B.; Sankararamakrishnan, R. Genome-wide analysis of major intrinsic proteins in the tree plant *Populus trichocarpa*: Characterization of XIP subfamily of aquaporins from evolutionary perspective. *BMC Plant Biol.* **2009**, *9*, 134. [CrossRef]
20. Danielson, J.A.; Johanson, U. Unexpected complexity of the aquaporin gene family in the moss Physcomitrella patens. *BMC Plant Biol.* **2008**, *8*, 45. [CrossRef]
21. Fox, A.R.; Maistriaux, L.C.; Chaumont, F. Toward understanding of the high number of plant aquaporin isoforms and multiple regulation mechanisms. *Plant Sci* **2017**, *264*, 179–187. [CrossRef]
22. Wang, S.C.; Liang, D.; Li, C.; Hao, Y.L.; Ma, F.W.; Shu, H.R. Influence of drought stress on the cellular ultrastructure and antioxidant system in leaves of drought-tolerant and drought-sensitive apple rootstocks. *Plant Physiol. Biochem.* **2012**, *51*, 81–97. [CrossRef]
23. Li, J.; Ban, L.; Wen, H.Y.; Wang, Z.; Dzyubenko, N.; Chapurin, V.; Gao, H.W.; Wang, X.M. An aquaporin protein is associated with drought stress tolerance. *Biochem. Biophys. Res. Commun.* **2015**, *459*, 208–213. [CrossRef]

24. Hu, C.G.; Hao, Y.J.; Honda, C.; Kita, M.; Moriguchi, T. Putative *PIP1* genes isolated from apple: Expression analyses during fruit development and under osmotic stress. *J. Exp. Bot.* **2003**, *54*, 2193–2194. [CrossRef]
25. Li, Q.T.; Zhao, Y.; Zhou, Q.Y.; Huang, L.F.; Jiang, Y.Z.; Chen, H.; Kong, J. Cloning and expression analysis of *MzPIP1;1* gene from *Malus zumi Mats*. *J.China Agric. Univ.* **2012**, *17*, 63–68.
26. Wang, L.; Li, Q.T.; Lei, Q.; Feng, C.; Gao, Y.N.; Zheng, X.D.; Zhao, Y.; Wang, Z.; Kong, J. MzPIP2;1: An aquaporin involved in Radial water movement in both water uptake and transportation, altered the drought and salt tolerance of transgenic *Arabidopsis*. *PLoS ONE* **2015**, *10*, e0142446. [CrossRef]
27. Basssett, C.L.; Baldo, A.M.; Moore, J.T.; Jenkins, R.M.; Soffe, D.S.; Wisniewski, M.E.; Norelli, J.L.; Farrell, R.E., Jr. Genes responding to water deficit in apple (*Malus X domestica* Borkh.) roots. *BMC Plant Bio.* **2014**, *14*, 182–192. [CrossRef]
28. Wang, L.; Li, Q.T.; Lei, Q.; Feng, C.; Zheng, X.D.; Zhou, F.F.; Li, L.Z.; Liu, X.; Wang, Z.; Kong, J. Ectopically expressing *MdPIP1;3*, an aquaporin gene, increased fruit size and enhanced drought tolerance of transgenic tomatoes. *BMC Plant Biol.* **2017**, *17*, 246–258. [CrossRef]
29. Foyer, C.H.; Noctor, G. Redox homeostasis and antioxidant signaling: A metabolic interface between stress perception and physiological responses. *Plant Cell* **2005**, *17*, 1866–1875. [CrossRef]
30. Gill, S.S.; Tuteja, N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* **2010**, *48*, 909–930. [CrossRef]
31. Mittler, R.; Vanderauwera, S.; Gollery, M.; Breusegem, F.V. Reactive oxygen gene network of plants. *Trends Plant Sci* **2004**, *9*, 490–498. [CrossRef]
32. Wang, L.Q.; Zhang, C.R.; Wang, Y.M.; Wang, Y.C.; Yang, C.P.; Lu, M.Z.; Wang, C. *Tamarix hispida* aquaporin *ThPIP2;5* confers salt and osmotic stress tolerance to transgenic *Tamarix* and *Arabidopsis*. *Environ. Exp. Bot.* **2018**, *152*, 158–166. [CrossRef]
33. Zou, Z.; Gong, J.; An, F.; Xie, G.S.; Wang, J.K.; Mo, Y.Y.; Yang, L.F. Genome-wide identification of rubber tree (Hevea brasiliensis Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber producing tissue. *BMC Genomics* **2015**, *16*, 1001. [CrossRef]
34. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choic. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]
35. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef]
36. Hu, B.; Jin, J.P.; Guo, A.Y.; Zhang, H.; Luo, J.C.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
37. Krogh, A.; Larsson, B.; Heijne, G.; Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef]
38. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, 585–587. [CrossRef]
39. Li, C.; Wang, P.; Wei, Z.W.; Liang, D.; Liu, C.H.; Yin, L.H.; Jia, D.F.; Fu, M.Y.; Ma, F.W. The mitigation effects of exogenous melatonin on salinity-induced stress in *Malus hupehensis*. *J. Pineal Res.* **2012**, *53*, 298–306. [CrossRef]
40. Yao, P.F.; Li, C.L.; Zhao, X.R.; Li, M.F.; Zhao, H.X.; Guo, J.Y.; Cai, Y.; Chen, H.; Wu, W. Overexpression of a tartary buchwheat gene, *FtbHLH3*, enhances drought/oxidative stress tolerance in transgenic Arabidopsis. *Front Plant Sci.* **2017**, *8*, 625–641. [CrossRef]
41. Liu, C.H.; Li, C.; Liang, D.; Ma, F.W.; Wang, S.C.; Wang, P.; Wang, R.W. Aquaporin expression in response to water-deficit stress in two Malus species: Relationship with physiological status and drought tolerance. *Plant Growth Regul.* **2013**, *70*, 187–197. [CrossRef]
42. Clough, S.J.; Bent, A.F. Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 735–743. [CrossRef]
43. Murashige, T.; Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [CrossRef]

44. Dong, Q.L.; Duan, D.Y.; Zhao, S.; Xu, B.Y.; Luo, J.W.; Wang, Q.; Liu, C.H.; Li, C.; Gong, X.Q.; Mao, K.; et al. Genome-wide analysis and cloning of the apple stress-associated protein gene family reveals MdSAP15, which confers tolerance to drought and osmotic stresses in transgenic *Arabidopsis*. *Int. J. Mol. Sci.* **2018**, *19*, 2478. [CrossRef]
45. Hou, H.M.; Jia, H.; Yan, Q.; Wang, X.P. Overexpression of a SBP-Box Gene (VpSBP16) from Chinese Wild Vitis Species in Arabidopsis Improves Salinity and Drought Stress Tolerance. *Int. J. Mol. Sci.* **2018**, *19*, 940. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

NB-LRRs Not Responding Consecutively to *Fusarium oxysporum* Proliferation Caused Replant Disease Formation of *Rehmannia glutinosa*

Aiguo Chen [1,2], Li Gu [1], Na Xu [2], Fajie Feng [1], Dexin Chen [2], Chuyun Yang [1], Bao Zhang [1], Mingjie Li [1,*] and Zhongyi Zhang [1,*]

[1] Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2] Institute of Tobacco Research, Chinese Academy of Agricultural Sciences, Qingdao 266101, China
* Correspondence: xinyuzszj@163.com (M.L.); zyzhang@fafu.edu.cn (Z.Z.); Tel.: +86-591-8374-2793 (M.L. & Z.Z.)

Received: 4 June 2019; Accepted: 25 June 2019; Published: 29 June 2019

Abstract: Consecutive monoculture practice facilitates enrichment of rhizosphere pathogenic microorganisms and eventually leads to the emergence of replant disease. However, little is known about the interaction relationship among pathogens enriched in rhizosphere soils, Nucleotide binding-leucine-rich repeats (NB-LRR) receptors that specifically recognize pathogens in effector-triggered immunity (ETI) and physiological indicators under replant disease stress in *Rehmannia glutinosa*. In this study, a controlled experiment was performed using different kinds of soils from sites never planted *R. glutinosa* (NP), replanted *R. glutinosa* (TP) and mixed by different ration of TP soils (1/3TP and 2/3TP), respectively. As a result, different levels of TP significantly promoted the proliferation of *Fusarium oxysporum* f.sp. *R. glutinosa* (FO). Simultaneously, a comparison between FO numbers and *NB-LRR* expressions indicated that *NB-LRRs* were not consecutively responsive to the FO proliferation at transcriptional levels. Further analysis found that *NB-LRRs* responded to FO invasion with a typical phenomenon of "promotion in low concentration and suppression in high concentration", and 6 *NB-LRRs* were identified as candidates for responding *R. glutinosa* replant disease. Furthermore, four critical hormones of salicylic acid (SA), jasmonic acid (JA), ethylene (ET) and abscisic acid (ABA) had higher levels in 1/3TP, 2/3TP and TP than those in NP. Additionally, increasing extents of SA contents have significantly negative trends with FO changes, which implied that SA might be inhibited by FO in replanted *R. glutinosa*. Concomitantly, the physiological indexes reacted alters of cellular process regulated by NB-LRR were affected by complex replant disease stresses and exhibited strong fluctuations, leading to the death of *R. glutinosa*. These findings provide important insights and clues into further revealing the mechanism of *R. glutinosa* replant disease.

Keywords: *Rehmannia glutinosa* L.; replant disease; rhizosphere microbes; *NB-LRR*; plant hormone

1. Introduction

Replant diseases, also known as consecutive monoculture problems or sick soil syndrome, are widespread in the production of different crops, especially in medicinal crops, fruit tree and vegetables, such as *Rehmannia glutinosa*, *Panax notoginseng*, apples, peach, strawberry, soybean etc. [1–6], which usually lead to disease aggravation, poor growth status, yield reduction and quality deterioration [7]. *Rehmannia glutinosa* Libosch, a perennial herbaceous plant, is one of the 50 traditional Chinese medicines with high value [8]. However, consecutive monoculture of *R. glutinosa* results in its abnormal growth and a significant decline in the yield and quality of tuberous roots. Moreover, it cannot be replanted on

the same land for 8–10 years [9]. The causes of problems associated with consecutive monoculture of *R. glutinosa* have become a research priority in China [10].

Allelochemicals, soil-borne diseases and soil quality deterioration are usually considered as primary factors involved in the formation of replant disease [11]. Some studies concluded that the effect of soil chemical properties is inconsistent with replant disease over time [12,13]. Furthermore, allelochemicals, such as phenolic acid and flavonoids, are quickly metabolized by soil microorganisms and shape the composition and diversity of microbial community in rhizosphere soils [14–19]. Recent studies have appointed that the healthy growth of plants is closely related to the balance of the rhizosphere microbes [20–22]. There is mounting evidence that the biotic factors mediated by rhizosphere allelochemicals is a causal agent of replant disease [10,23–25]. Previous studies have discovered that the pathogenic microorganisms in replanted *R. glutinosa* rhizosphere soil were selectively attracted by root exudates, and then colonized the root surface to proliferate, resulting in rhizosphere micro-ecology catastrophe, mainly in the transformation of rhizosphere microorganisms from "bacterial" to "fungal" types and a decline in rhizosphere microbial diversity [10,11,13,26,27]. The previous evidences have confirmed that the pathogenic *Fusarium oxysporum* f.sp. *R. glutinosa* (FO) was enriched and beneficial *Pseudomonas* spp. (PS) specifically decreased in rhizosphere soil of replanted *R. glutinosa* [28–30]. However, it remains largely unknown the mechanism of how replanted *R. glutinosa* responded to the changes in composition and diversity of microbial community in rhizosphere soil.

In natural environments, plants can regulate the rhizosphere micro-ecological status to preserve healthy growth by constantly renewing the composition and diversity of microbiome [31]. Plant immune response, containing recognition to pathogens and signal transduction, thus plays a key role in coordinating microbial communities in rhizospheres [32–34]. As we know, plant innate immunity co-evolving with pathogenic microbes have developed two strategies, including pathogen-associated molecular patterns (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) [35,36]. PTI is generally effective against non-adapted pathogens in a phenomenon called non-host resistance, and the pattern recognition receptors (PRRs) of PTI systems can recognize conserved PAMP features of different species or genera. ETI is active against adapted pathogens, and the receptors of ETI systems specifically and robustly respond to pathogen effectors through nucleotide-binding-leucine-rich repeat (NB-LRR) domain-mediated perception [37]. NB-LRRs are now believed to include the majority of plant R proteins and recognize fast-evolving effectors [38,39]. Our previous studies found that pathogenic FO was specifically enriched in the rhizosphere soil of replanted *R. glutinosa* [28], and PTI of two plant innate immunity silently responded to replant disease [40], which suggested that PTI was not sufficient to cope with fast-evolving effectors. However, some NB-LRR receptors in ETI were upregulated [41]. It remains unclear why these upregulated NB-LRRs fail to prevent the death of replant *R. glutinosa*.

Plant hormones are essential regulators for triggering plant immune resistance to cope with various pathogens [42]. The salicylic acid (SA) and jasmonic acid-ethylene (JA-ET) are believed to form the hormonal backbone of plant immune responses to pathogens, with SA involved in resistance to biotrophic pathogens and JA–ET involved in responses to necrotrophic pathogens and chewing insects [37,43]. The antagonism between SA and JA-ET often occurs through the regulatory protein non-expressor of pathogenesis-related proteins 1 (NPR1), which mediates the SA-induced expression of pathogenesis-related (PR) genes and systemic acquired resistance (SAR) [44]. Enhanced disease susceptibility 1 (EDS1) acts as a SA-pathway inducer and a JA-ET pathways repressor, and mitogen-activated protein kinase 4 (MPK4) acts as a negative regulator of systemic acquired resistance, which is controlled by the SA pathway [37]. In addition, abscisic acid (ABA) can increase susceptibility to pathogens to some extent, although it is mainly associated with responses to abiotic stresses [37,45]. Some studies have described that SA was involved in the defense against *Fusarium oxysporum* and associated with acteoside accumulation, one of important pharmacodynamic component [46,47]. Our recent study displayed that plant hormones involved in the formation mechanism of *R. glutinosa*

replant disease and are closely related to immune resistance [48]. However, a comprehensive survey on the relationship between plant hormones and replant disease in *R. glutinosa* is still unknown.

Given the interferences from complex field environment and vegetative propagation, such as various pathogens, variable light and temperature conditions, maternal resistance, size of vegetative mass, germination rate and site of shoot, pot experiment in controlled conditions with regenerated plantlets is thus an ideal research method. Here, the soils from the sites in which *R. glutinosa* had never been planted and had been consecutively planted *R. glutinosa* were proportionally mixed to control the stress levels of replant disease. The regenerated plantlets of *R. glutinosa* with feeble immunity were acclimatized and transplanted in a phytotron. The dynamic changes of pathogenic microbes in the rhizosphere soils were investigated. At same time, *NB-LRRs* expression, plant hormone contents and physiological index levels were examined in the roots of *R. glutinosa*. These research results will provide important insights for further revealing the mechanism of formation of replant disease.

2. Results

2.1. Changes in the Numbers of Pseudomonas spp. and Fusarium oxysporum in Rhizosphere Soils of Replant Disease R. glutinosa

To understand the changes in the beneficial PS and pathogenic FO in rhizospheres of replanted *R. glutinosa*, the numbers of them were showed in Figure 1. The numbers of PS were 3.87×10^8~5.99×10^8 (cell·g^{-1} soil), the numbers of FO were 0.71×10^8~2.68×10^8 (cell·g^{-1} soil), and the ratio of FO to PS were 0.16~0.55. The maximum difference ranged between 0 DAP and 9 DAP, where PS was in TP soil and FO was in NP soil (Figure 1A,B). Moreover, only the numbers of FO decreased significantly over time in control NP (Figure 1E). The inhibiting effect of NP soil on FO proliferation was thus stronger than the promoting effect of replanted soil on FO proliferation. During 0~9 DAP, the FO numbers and ratio of FO to PS in TP soil were significantly lower than that in NP soil at 0 DAP, and the reversing results with significant difference were at 9 DAP (Figure 1B,C). The PS numbers were significantly lower only in NP soil than that in TP soil at 0 DAP, and the reversed trends with no significant difference were at 9 DAP (Figure 1A). Furthermore, there were the significant positive correlations between the addition level of replant soils and the changes of the FO numbers and the FO/PS values, and between the FO numbers and the FO/PS values (Tables 1 and 2). The results evidenced that *R. glutinosa* replant disease (1/3TP, 2/3TP and TP) could significantly promote the FO proliferation in rhizospheres soils. Interestingly, at 0 DAP, the PS numbers in 1/3TP, 2/3TP and TP soil were 1.04, 1.11 and 1.21 times than that in NP soil respectively, and the FO numbers in NP soil were 1.39, 1.44 and 1.83 times than that in 1/3TP, 2/3TP and TP soil respectively. The changes were obviously different from that at 9 DAP. The results suggested that the soil had a strong ability that could restore balances among microbial communities, and this means that the appropriate ration between FO and PS in rhizosphere soil presented in *R. glutinosa* grows healthly, during fallow stage.

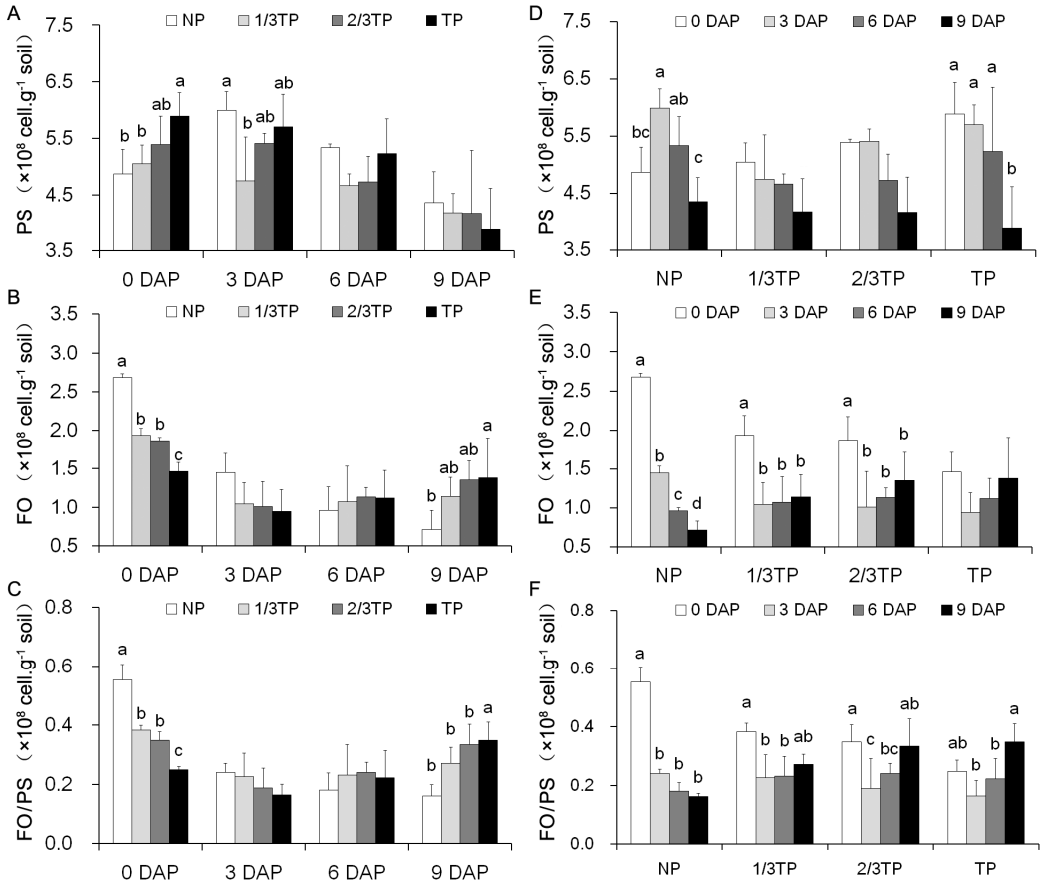


Figure 1. The influences of different replant disease stress levels on the numbers and ratio of FO to PS in *R. glutinosa* rhizosphere soils over time. NP, 1/3TP, 2/3TP and TP are the gradient treatments of replant disease stresses. Four samples in each group are compared with NP (**A**–**C**) and 0 DAP (**D**–**F**) respectively, and different lower-case letters indicate significant differences ($p < 0.05$; LSD). Data represented as the mean ± SD ($n = 3$). DAP: Days after planting. PS: *Pseudomonas* spp. FO: *Fusarium oxysporum*. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

Table 1. Pearson correlations among the different replant disease stress levels and the changes in amplitude of FO/PS and the numbers of PS and FO within 3 DAP.

	ΔFO/PS	**ΔPS**	**ΔFO**
Levels of replant disease stresses	0.7918 **	−0.5611	0.6809 *

* $p < 0.05$, ** $p < 0.01$; Δ represents the change from 0 to 3 DAP.

Table 2. Pearson correlations among the FO/PS and the numbers of PS and FO within 9 DAP.

	PS	**FO**
FO/PS	−0.2488	0.9208 **
PS		0.1331

** $p < 0.01$.

2.2. NB-LRR Lists Response to Replant Disease Stresses in R. glutinosa

To reveal the responding mode of *R. glutinosa NB-LRRs* during replant disease formation, the expression levels of *NB-LRRs* were determined by qRT-PCR. Of 35 expressed *NB-LRRs* in acclimatization stage, seven were upregulated and 11 were downregulated (Figure 2). In contrast to NP, 22 *NB-LRRs* were upregulated and six *NB-LRRs* were downregulated in TP at 3 DAP, which

accounted for 80% of the 35 *NB-LRRs*. The stage from 0 to 3 DAP was thus the key stage to *NB-LRRs* that responded to replant disease stress. In addition, only 1 *NB-LRR* was upregulated and 29 *NB-LRRs* were downregulated at 6 DAP, indicating that replant disease interference to the expression of *NB-LRRs*. In 3 DAP, the Pearson correlation analyses displayed that 35 *NB-LRRs* had no significant correlation with the PS numbers. The significant positive correlation was only between 12 *NB-LRRs* and the FO numbers in NP, 1/3TP, 2/3TP (Table 3). However, these *NB-LRRs* that have significantly positive correlation with the FO numbers, have ineffectively prevented the death of *R. glutinosa* plants at 6 DAP. The results displayed that the 12 upregulated *NB-LRRs* might not respond to the pathogenic FO in replanted *R. glutinosa*. Therefore, the six downregulated *NB-LRRs* (*RgNB5*, *RgNB14*, *RgNB26*, *RgNB29*, *RgNB34* and *RgNB35*) in TP at 3 DAP, were screened as candidates for responding to *R. glutinosa* replant disease. In addition, *RgNB14* and *RgNB26* of the 6 *NB-LRRs* were upregulated expression, and the remaining 4 *NB-LRRs* were downregulated or presented a decreasing trend at expression level in acclimatization stage. The two groups of *NB-LRRs* represented whether they responded to abiotic stress or not. Noticeably, the expression levels of *RgNB5* and *RgNB29* were continuously downregulated with addition levels of replant soils at 3 DAP.

Table 3. Pearson correlations among the expression of 35 NB-LRRs and the variation of PS and FO numbers from 0 DAP to 3 DAP.

	RgNB1	*RgNB2*	*RgNB3*	*RgNB4*	*RgNB5*	*RgNB6*	*RgNB7*
PS	−0.1156	0.0187	0.6833	−0.1978	0.5612	−0.0038	−0.3908
FO	0.9737	0.9344	0.4558	−0.3259	−0.4752	0.9219	0.9963 **
	RgNB8	*RgNB9*	*RgNB10*	*RgNB11*	*RgNB12*	*RgNB13*	*RgNB14*
PS	−0.2584	0.0342	−0.1756	0.1503	−0.4000	−0.2856	0.7931
FO	0.9950 **	0.8935	0.5160	0.8193	0.9973 **	0.9984 **	−0.7543
	RgNB15	*RgNB16*	*RgNB17*	*RgNB18*	*RgNB19*	*RgNB20*	*RgNB21*
PS	−0.3292	−0.1421	0.0665	−0.4500	−0.0412	−0.2643	−0.3687
FO	0.0526	0.9794 *	−0.4775	0.8240	0.9452	0.9970 **	0.2326
	RgNB22	*RgNB23*	*RgNB24*	*RgNB25*	*RgNB26*	*RgNB27*	*RgNB28*
PS	−0.4505	−0.4356	−0.1776	−0.3769	−0.1168	−0.0632	−0.2509
FO	0.5924	0.8712	0.9858 *	0.9845 *	−0.6370	0.9564 *	0.9953 **
	RgNB29	*RgNB30*	*RgNB31*	*RgNB32*	*RgNB33*	*RgNB34*	*RgNB35*
PS	0.4527	−0.3214	−0.3369	−0.0666	−0.4524	0.6505	0.3434
FO	−0.8157	0.9787 *	0.9965 **	−0.3699	−0.5155	−0.7435	−0.6044

* $p < 0.05$; ** $p < 0.01$.

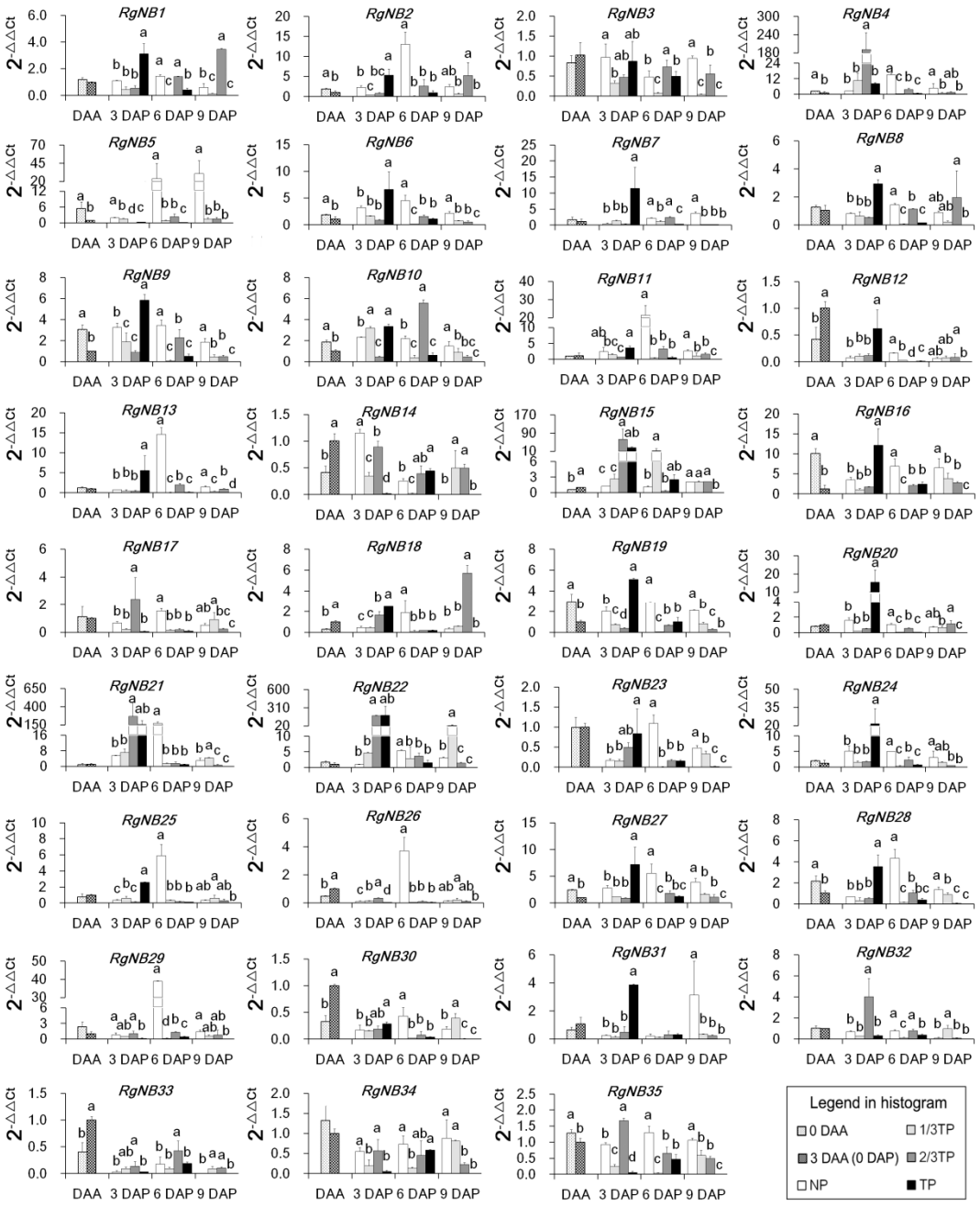


Figure 2. Nucleotide binding-leucine-rich repeats (*NB-LRRs*) expression in *R. glutinosa* roots during acclimatization and planting phase with different replant disease stress levels. The ratios of 0 DAA, 3 DAA (0 DAP), NP, 1/3TP, 2/3TP and TP gradient treatments to 3 DAA (0 DAP) are calculated and shown. Two or four samples in each group are compared with 3 DAA and NP respectively, and different lower-case letters indicate significant differences ($p < 0.05$; LSD). Data represented as the mean ± SD ($n = 3$). DAA: Days after acclimatization. DAP: Days after planting. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

2.3. Plant Hormones Response to Replant Disease Stresses in R. glutinosa

Four plant hormones including ABA, ET, JA and SA were measured, and their contents in 1/3 TP, 2/3TP and TP treatments fluctuated obviously from 0 DAA to 9 DAP (Figure 3). In the acclimatization stage (from 0 DAA to 3 DAA), the contents of plant hormones decreased significantly except for SA. The sensitive stages for rapid alteration of hormones was from 0 to 3 DAP based on the comparison

between 3 DAP and 3 DAA (0 DAP). At 3 DAP, the four hormones were significantly activated under replant disease stresses compared with the control NP by 0.56 to 1.53 times for ABA, 0.59 to 2.02 times for JA, 0.50 to 1.50 times for ET and 0.37 to 1.19 times for SA ($p < 0.05$), which showed that the four plant hormones responded strongly to replant disease stress levels. From 0 to 3 DAP, the four hormone contents in 1/3TP, 2/3TP and TP have significantly increased at 3 DAP compared with that at 0 DAP, excepting SA treated with TP soils. Of which, ABA contents increased by 65.26% to 173.64%, JA by 60.19% to 203.50%, ET by 65.26% to 173.84% and SA by 3.91% to 66.39% under replant disease stresses, but only increased by 8.50%, 0.33%, 9.73% and −24.72% in NP, respectively ($p < 0.05$). Notably, there were significant negative correlations between FO and SA ($p < 0.05$) and significant positive correlations among the four hormones ($p < 0.01$) (Table 4). The results showed that the FO in replant disease rhizosphere soil inhibited the SA biosynthesis in root. Taking together these changes showed that the complex replant disease stress, including biotic and abiotic stress, promoted the four hormones release, but FO inhibited SA biosynthesis, resulting in *R. glutinosa* death at 6 DAP.

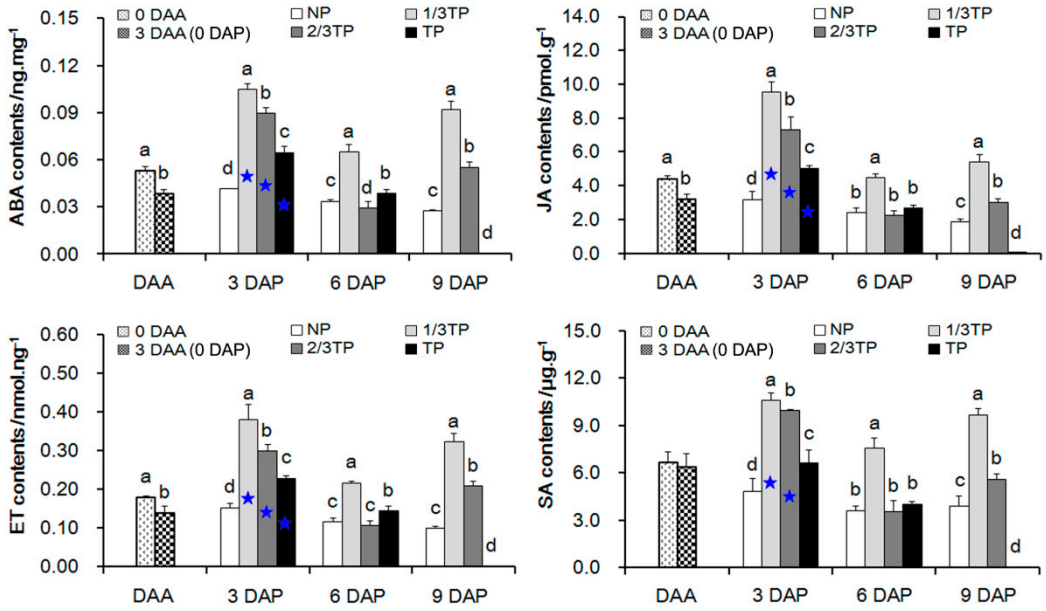


Figure 3. The four plant hormone contents in *R. glutinosa* roots during acclimatization stage and planting stage with different replant disease stress levels. 0 DAA, 3 DAA (0 DAP), NP, 1/3TP, 2/3TP and TP are gradient treatments. Two or four samples in each group are compared with 3 DAA (0 DAP) and NP respectively, and different lower-case letters indicate significant differences ($p < 0.05$; LSD). 3 DAP with NP, 1/3TP, 2/3TP and TP gradient treatments is compared with 3 DAA (0 DAP), and the asterisk indicate significant differences ($p < 0.05$; LSD). Data represented as the mean ± SD ($n = 3$). DAA: Days after acclimatization. DAP: Days after planting. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

Table 4. Pearson correlations among the variation of the PS, FO, jasmonic acid (JA), abscisic acid (ABA), ethylene (ET) and salicylic acid (SA) contents under replant disease stresses from 0 to 3 DAP.

	ΔFO	**ΔJA**	**ΔABA**	**ΔET**	**ΔSA**
ΔPS	−0.7327	0.1328	0.0254	0.1123	0.2366
ΔFO		−0.6524	−0.5981	−0.6560	−0.7475 *
ΔJA			0.9752 **	0.9839 **	0.9182 **
ΔABA				0.9650 **	0.9543 **
ΔET					0.9009 **

* $p < 0.05$, ** $p < 0.001$; Δ represents the variation in content.

2.4. Physiological Response to Replant Disease Stresses in R. glutinosa

The physiological indexes including root activity; superoxide dismutase (SOD), peroxidase (POD) and catalase (CAT) activities; and malondialdehyde (MDA) and hydrogen peroxide (H_2O_2) contents were investigated in *R. glutinosa* with different replant disease stress levels (Figure 4). In the acclimatization stage, there were no significant changes in the POD activities. The H_2O_2 contents were significantly increased, and the others that used to eliminate toxicity of H_2O_2 were decreased significantly. In contrast to 3 DAA (0 DAP), the root activities, SOD activities, POD activities and MDA contents in NP tended to increase consecutively from 3 DAP to 9 DAP, which reported the normal response of plantlets planted in healthy soil. In addition, root activity, SOD activity, CAT activity and H_2O_2 contents in NP, 1/3TP, 2/3TP and TP at 3 DAP were significantly different from those at 3 DAA (0 DAP). Of which H_2O_2 contents significantly decreased by 0.45 to 0.75 times, and root activity, SOD activity and CAT activity significantly increased by 0.36 to 2.75 times, 3.12 to 7.33 times, 2.76 to 4.26 times at 3 DAP, respectively. However, there were no gradient changes for the four physiological indexes at 3 DAP. Moreover, only the H_2O_2 contents showed a gradient change at 6 DAP. Furthermore, there were significant positive correlations between the FO numbers and the CAT activity and the MDA content only at 6 DAP (Table 5). Overall, these results displayed that these physiological indexes acting downstream of life activities were affected by complex replant disease stresses and did not show significant correlation with the FO numbers until 6 DAP.

Table 5. Pearson correlations among the variation of the two microorganisms and the six physiological indexes under replant disease stresses during three and six days after transplanting.

		ΔRoot Activity	**ΔSOD**	**ΔPOD**	**ΔCAT**	**ΔH_2O_2**	**ΔMDA**
0~3 DAP	ΔPS	0.0357	0.2189	0.3088	−0.1421	−0.3443	−0.3313
	ΔFO	0.4389	−0.1874	0.2671	0.4416	0.1959	−0.1724
0~6 DAP	ΔPS	0.1759	−0.0713	−0.0387	−0.2458	0.1394	−0.2020
	ΔFO	0.2250	0.1574	0.2263	0.7464 *	−0.5732	0.7316 *

* $p < 0.05$; Δ represents the variation in content.

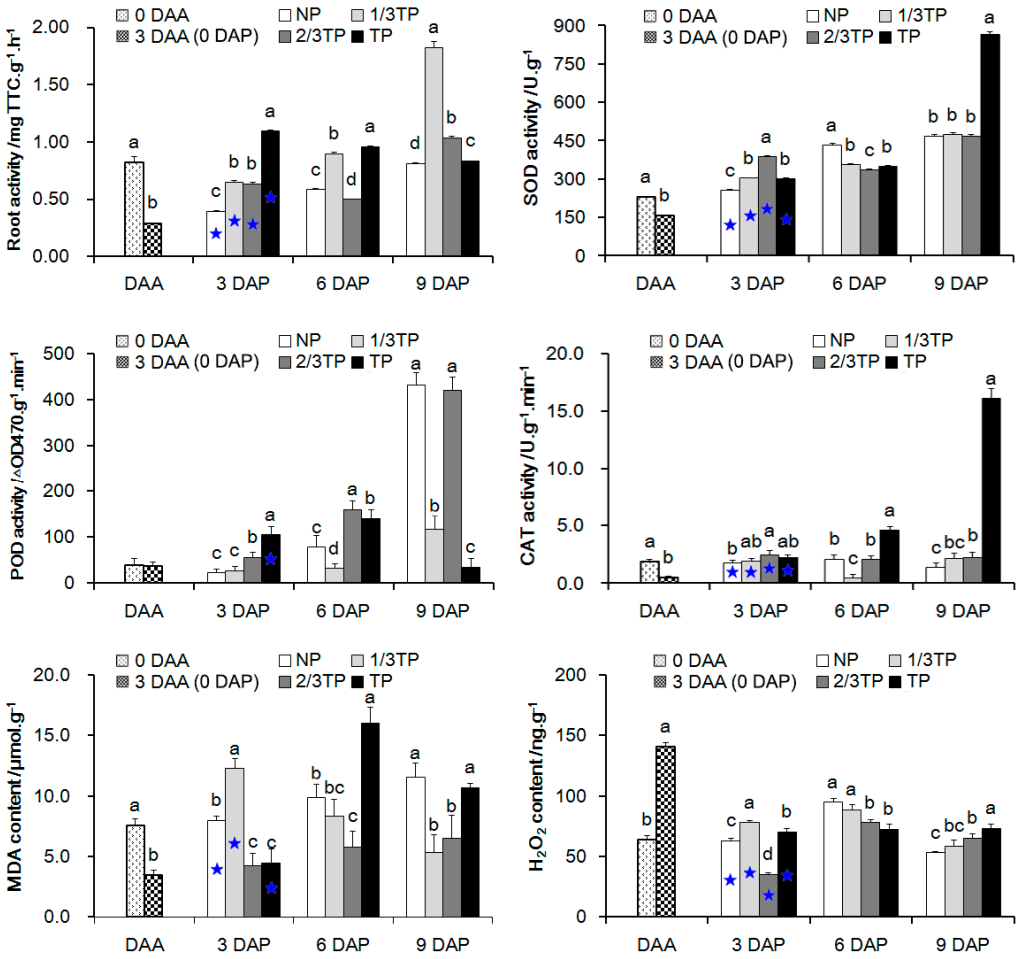


Figure 4. The contents of physiological indexes in *R. glutinosa* roots during acclimatization stage and planting stage with different replant disease stress levels. 0 DAA, 3 DAA (0 DAP), NP, 1/3TP, 2/3TP and TP are gradient treatments. Two or four samples in each group are compared with 3 DAA (0 DAP) and NP respectively, and different lower-case letters indicate significant differences ($p < 0.05$; LSD). 3 DAP with NP, 1/3TP, 2/3TP and TP gradient treatments is compared with 3 DAA (0 DAP), and the asterisk indicate significant differences ($p < 0.05$; LSD). Data represented as the mean ± SD ($n = 3$). DAA: Days after acclimatization. DAP: Days after planting. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

3. Discussion

3.1. Replant Disease Promotes Proliferation of Fusarium oxysporum in R. glutinosa Rhizospheres Soil

Previous studies have revealed that *R. glutinosa* replant disease induced rhizosphere microbes' adverse chemotaxis [49,50], resulting in the increase of pathogenic FO abundance and the decrease of beneficial PS abundance, thus the identified PS and FO were usually used as a characteristic mark in the *R. glutinosa* replant disease study [10,28]. In this study, the PS numbers decreased and the FO numbers increased in 1/3TP, 2/3TP and TP soils from 0 to 9 DAP. Moreover, the significant positive correlations were presented between the addition levels of replant soils and the changes of the FO numbers and the FO/PS values. These results revealed that replant disease promoted FO proliferation in rhizospheres soils. Interestingly, the focus on maximum difference between the PS and FO numbers displayed that the inhibiting effect of NP soil on FO proliferation was stronger than the promoting effect of replanted soil on FO proliferation. Furthermore, the number of PS in NP soil was significantly lower than that in TP soil at 0 DAP, while the number of FO in NP soil was significantly greater than that in 1/3TP, 2/3TP and TP soils at 0 DAP. The changes were obviously different from that at 9 DAP.

These results are different from those in the field [10], which might be related to the difference between the air-dried soil used in this experiment and the field soil after fallow cultivation in the literature [51]. A valuable clue was thus presented that soil may have a strong ability restored microbe balance in rhizosphere soil during the fallow stage.

3.2. NB-LRRs Failed to Respond Timely and Effectively to Pathogenic Fusarium oxysporum in Replanted R. glutinosa

Pathogens almost always occupy extracellular niches [37]. Early studies had unraveled that *Fusarium oxysporum* could inhibit the release of ATP after invading the host plant [52], which is required for NB-LRR activation [37], but low doses toxins secreted by FO could induce the biosynthesis of phytoalexins in the host plants [53]. In this study, the upregulated expressions of *NB-LRRs* were mainly appeared at 3 DAP in TP and at 6 DAP in NP, which were consistent with the changes of the FO relative values (Figure 5). According to the original data associated with the relative values, we found that the upregulated expression of *NB-LRRs* was the strongest when the number of FO was at the lowest level (about 0.95×10^8 cell·g^{-1} soil), and gradually decreased with the increasing of FO number in six days after transplanting. These results exhibited that *NB-LRR* expressions have a typical regulars with "promotion in low concentration and suppression in high concentration" when countered FO invasion, which was consistent with Scott et al. (2019) [54].

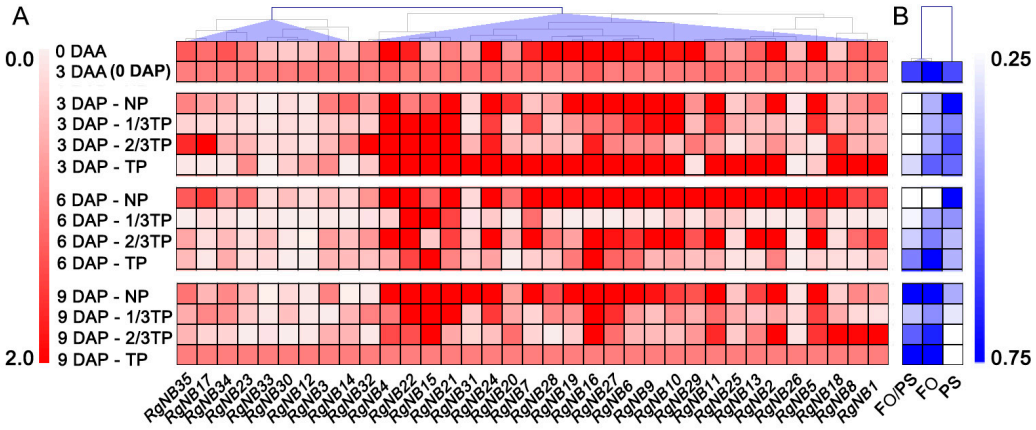


Figure 5. The heatmaps of the 35 *NB-LRR* expressions in *R. glutinosa* roots and the numbers of FO and PS and their ratios in *R. glutinosa* rhizosphere soils. (**A**) The expression profiles of 35 *NB-LRRs* in *R. glutinosa* roots over time. 3 DAA (0 DAP) is the control; (**B**) the changes in relative values of FO, PS and FO/PS over time. Calculation of the relative values: The cell numbers and the FO/PS values are first compared with that at 0 DAP (3 DAA), and then normalized (*Z*-score) for each microorganism respectively to display in the same color scale. DAA: Days after acclimatization. DAP: Days after planting. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

In antagonistic associations with microbes, plants have evolved to form two strategies of PTI and ETI for fighting microbial pathogens [37]. A well-known view is that immune receptor play a vital role in the recognition to pathogens [55,56], and constitutive downstream proteins are tightly controlled by both positive and negative regulators [39,57]. An integrated understanding for plant immune response is thereby in both the immune receptor and downstream signal transduction. Based on studies till date, there is mounting evidence that two possible interpretations were supported for the relationships of immune response and plant death. (i) Plant immunity has not been triggered [37,38]. (ii) Excessive immunity response often leads to inhibition of normal plant growth and even death [57–59]. In this study, the majority of *NB-LRRs* identified in replanted *R. glutinosa* roots were upregulated or downregulated in TP at 3 DAP in comparison to 0 DAP. At same time, there were significantly positive correlations found between 12 *NB-LRRs* and the FO numbers in NP, 1/3TP, 2/3TP and TP

during 0~3 DAP, but the FO numbers in 1/3TP, 2/3TP and TP were no different with those in NP at 3 DAP even significantly lower at 0 DAP. In addition, accompanied by the death of replanted *R. glutinosa* at 6 DAP, only 1 *NB-LRR* was upregulated (downregulated at 3 DAP) and 29 *NB-LRRs* were downregulated. Taking together these data showed that the stage of 0~3 DAP was the key stage for *NB-LRRs* to respond to replant disease stress, while the immune response was obviously inactivated to FO rather than excessive responses of NB-LRRs. Therefore, it was one of important reasons that *NB-LRRs* were not consecutively responsive to the FO proliferation at the transcriptional level in replanted *R. glutinosa* roots. These new findings provide insights into the response mechanism of *R. glutinosa* to replant disease.

One of the big gaps in our understanding of plant immunity is in the downstream signaling pathways after receptor protein activation [37]. Only two identified downstream signaling proteins of EDS1 and non-race-specific disease resistance 1 (NDR1) are required for signaling of all TIR-NB-LRRs and some CC-NB-LRRs, respectively [37]. For NB-LRR, the protein structure consists of a carboxy-terminal LRR domain for effector recognition, NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) and amino-terminal Toll/interleukin-1 receptor (TIR) or Coiled-coil (CC) domains for signal transduction to downstream proteins [41]. In this study, the six downregulated *NB-LRRs* (*RgNB5*, *RgNB14*, *RgNB26*, *RgNB29*, *RgNB34* and *RgNB35*) in TP at 3 DAP, were screened as candidates for responding to *R. glutinosa* replant disease. According to their functional conservation in these NB-LRRs, there were seven resistance in linkage group 1A (R1A), 2 resistance in linkage group 1B (R1B), 1 resistance to *Pseudomonas syringae* (RPS2), 1 target of AvrB operation 1 (TAO1) and 2 resistance to *powdery mildew* 8 (RPW8). Based on studies till date, EDS1 were required for downstream signaling of these identified NB-LRRs except for RPS2 (RgNB32) [60–63]. The results provide valuable clues for studying the signaling pathways that operate downstream of NB-LRR protein activation in replanted *R. glutinosa*.

3.3. Lower Level of SA Biosynthesis Stimulated by Fusarium oxysporum Might Be Closesly Related to the Formation of R. glutinosa Replant Disease

Generally, the stress factors associated with replant disease were complex and multiple, including pathogens, nematode and abiotic stress [7]. To resist these stress factors, plant hormones are widely involved in resistance levels as important signaling molecules [64]. ABA is mainly associated with abiotic environmental stresses [44]. SA is typically involved in the defense against biotrophs [65]. JA and ET are generally thought to act together, and to play core roles in the defense against necrotrophs [66]. The relationship between the SA and JA-ET pathway is more antagonistic other than cooperative [44]. In this study, the four hormones were significantly activated under replant disease stresses compared with the control NP at 3 DAP, and presented each other significant positive correlation ($p < 0.001$). The cooperation for SA and JA-ET was consistent with some literatures, such as Wu et al. (2018) [67] and Adie et al. (2007) [68], which was different from the antagonistic relationship between SA and JA-ET. A reasonable explanation for the cooperation of ABA with SA was that ABA increased susceptibility to pathogens in some plant-pathogen interactions [45]. However, these inferences need to be further verified. Previous studies displayed that some plant hormones including ET and ABA involved in the formation mechanism of *R. glutinosa* replant disease and closely related to immune resistance [48,69]. Some studies have demonstrated that SA was involved in the defense against *Fusarium oxysporum* and associated with acteoside accumulation, one of the important pharmacodynamic component [46,47]. More importantly, there were significant negative correlations only between ΔFO and ΔSA with the gradient changes in the stress level of replant disease (Table 4). The results revealed that the SA synthesis, which may be inhibited by *Fusarium oxysporum*, was involved in the formation of replant disease in *R. glutinosa*.

Based on the above results, a possible depiction of the immune response and its potential crosstalk with microbes and plant hormones in replanted *R. glutinosa* was drawn, and it is shown in Figure 6. These findings provide insights into the formation of replant disease.

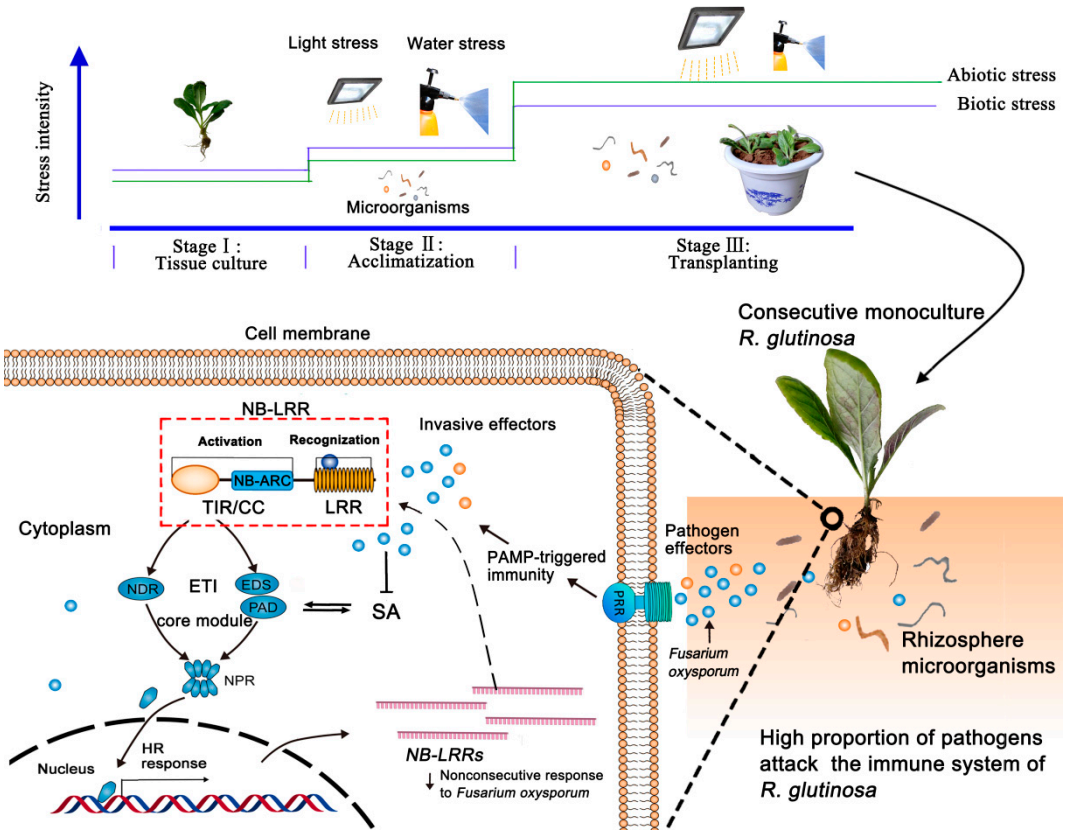


Figure 6. Schematic of effector-triggered immunity response mediated by consecutive monoculture stress in *R. glutinosa*.

4. Materials and Methods

4.1. Plant Growth and Treatments

To obtain aseptic plantlets, tuberous roots of *R. glutinosa* "Wen 85-5" were surface sterilized with 0.1% mercuric chloride solution for 17–20 min, washed five times with the sterile water, and then cultured in sterile bottles with two layers of damp gauze at the bottom. The shoots, approximately 1 cm long, were cut and cultured on hormone-free MS agar medium containing 30 g·L^{-1} sucrose and 10 g·L^{-1} agar [47]. The explants were cultured under controlled conditions (25 °C, 4000 lux, 14 h light/10 h dark photoperiod) in a growth chamber for 30 days.

To enhance the quality of plantlets for transplantation, the aseptic plantlets of *R. glutinosa* with seven to eight leaves were adapted in a phytotron (28 °C, 10,000 lux, 14 h light/10 h dark photoperiod) for 12 h, followed by unscrewing the bottle caps with a small opening to adapt for 6 h, and then removing the cap to adapt for 30 h (Figure 7A). After carefully washing away the adherent medium on the roots, the plantlets were adapted in sterile water for one day and transplanted into plastic pots.

Pot experiments were performed under controlled conditions (28 °C, 10,000 lux, 14 h light/10 h dark photoperiod) at the Institute of GAP for Chinese Medicinal Materials, Fujian Agriculture and Forestry University. *R. glutinosa* plantlets after acclimatization were transplanted on 23 August 2018 and grown in plastic pots of 18 cm diameter and 15 cm height (1.38 kg soil per pot). Three plants were planted as three replicate sub-samples in each pot. Four treatments of replant disease levels were constructed by mixing two kinds of soils in different proportions. The soils were collected from the site where *R. glutinosa* had not been planted for at least 10 years (NP) and where *R. glutinosa* had been consecutively planted for three years (TP) in Wen County, Jiaozuo City, Henan Province, in the "geo-authentic" zone of *R. glutinosa* cultivation (34°56′ N, 112°58′ E). The air-dried soil samples were taken to the laboratory for this experiment. Four treatments thus included NP, 1/3TP (mixed by 2 NP

soils and 1 TP soil), 2/3TP (mixed by 1 NP soil and 2 TP soils), and TP, NP of which was used as the control (Figure 7B).

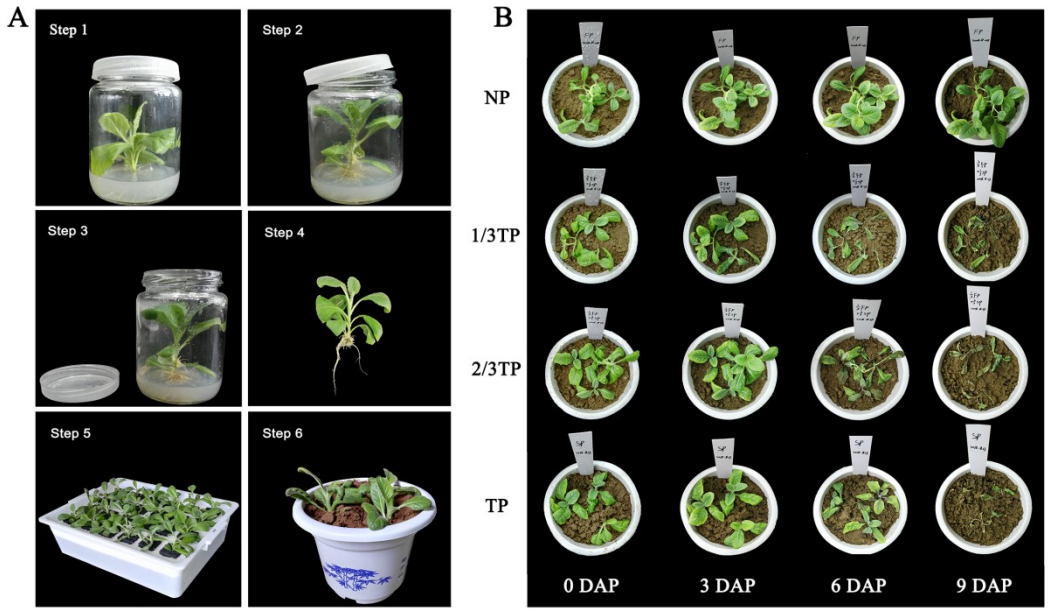


Figure 7. The operation process for acclimatization and the phenotype changes of *R. glutinosa* in the experiment. Obvious irreversible injury occurred at 6 DAP until death at 9 DAP. (**A**) The six main steps of acclimatization and transplanting. Step 1: Closed adaptation for 12 h; Step 2: Small opening for 6 h; Step 3: Completely open for 30 h; Step 4: Washing the culture medium carefully with sterile pure water; Step 5: Plantlets were adjusted to sterile pure water for 24 h; Step 6: Transplanting three plants per pot; (**B**) the phenotype changes of *R. glutinosa* under different replant disease stresses within 9 DAP. In the pot stage, 1/3TP, 2/3TP and TP are compared with NP, and 3 DAP, 6 DAP and 9 DAP are compared with 0 DAP (3 DAA). DAP: Days after planting. NP: Soil that was never planted with *R. glutinosa* for at least 10 years. TP: Soil that was consecutively planted with *R. glutinosa* in the same soils for three years.

4.2. The Collection of Fresh Root and Rhizosphere Soil Samples

On day zero, three after acclimatization (DAA), the fresh roots were collected after carefully washing with sterile water and drying with absorbent paper. The samples of 0 DAP and 3 DAA are the same. At 3, 6 and 9 DAP, the fresh roots and their rhizosphere soil were carefully collected as described in Wu et al. (2015) [10]. Briefly, the roots and the soil around the roots were carefully dug up using a sterilized fork spade and slightly shaken to remove loosely attached soil. The rhizosphere soil that was tightly attached to roots (1–3 mm zone around the root) was brushed off and collected. All of the collected samples were immediately frozen in liquid nitrogen and then stored at −80 °C for further experiments for soil DNA extraction, absolute quantification of PS and FO, qRT-PCR of NB-LRR and measurement of plant hormones and physiological index.

4.3. The Extraction of Soil DNA and Its Method Comparison

Approximately 5 g of soil of each sample was weighed for the extraction of soil DNA. Three extraction methods were compared according to the electrophoretic strips. Method I referred to the conventional cetyltrimethylammonium bromide (CTAB) method [70]. Methods II and III used different extraction methods based on the optimization for removing humic acid. PCR and gel electrophoresis were used to evaluate the different DNA extraction methods for PS and FO. For PCR, specific primers of PS (PS for: 5′-GGTCTGAGAGGATGATCAGT-3′, PS rev: 5′-TTAGCTCCACCTCGCGGC-3′) and FO (ITS1-F: 5′-CTTGGTCATTTAGAGGAAGTAA-3′, AFP308R: 5′-CGAATTAACGCGAGTCCCAA-3′) were synthesized with reference to Wu et al. (2015) [10] (SunYa Biotechnology Co., Ltd. Fuzhou,

China). Conventional PCR was performed using a Thermo cycler instrument (ThermoFisher Scientific A24812, Waltham, MA, USA) to detect transcript abundance. Each 20-μL reaction contained 0.4 μM each primer, 0.5 U 2× Es taq MasterMix enzyme, cDNA and nuclease-free water. The amplification procedure was 95 °C for 2 min, 95 °C for 5 s, and 58 °C to anneal for 30 s, for 30 cycles. Horizontal gel electrophoresis (1% gel) was used to evaluate the effect of the extraction with a DL2000 DNA Marker (Takara, Japan) in 130 V, 200 mA (Liuyi DYY-12, Beijing, China). The results showed that the extraction effect of method III was relatively best (Supplementary Figure S1A).

4.4. Absolute Quantification of Pseudomonas spp. and Fusarium oxysporum

To accurately evaluate the shift of rhizosphere microorganisms to avoid the expression changes of reference genes in different growth phases and soil samplings [71], the characteristic rhizosphere microorganisms of beneficial PS and pathogenic FO were detected based on absolute quantification PCR (AQ-PCR).

4.4.1. Construction of the Recombinant Plasmid

Amplifications of PS and FO specific primers were performed by conventional PCR (as mentioned above) and touchdown PCR (the same 20-μL reaction; 95 °C for 2 min, 95 °C for 5 s, 50–60 °C to anneal for 30 s, 72 °C for 30 s, for eight cycles; 95 °C for 2 min, 95 °C for 5 s, 58 °C to anneal for 30 s, for 30 cycles), respectively (Supplementary Figure S1B). The bright electrophoretic strips were cut and extracted using gel pure DNA kits following the manufacturer's instructions (Magen D2111-02, China). The gel extraction solutions were concentrated to over 50 ng·μL^{-1} of cDNA using a concentrator for approximately 25 min at 1400 rpm (Eppendorf Concentrator Plus AG5305, V-AQ mode). DNA fragments were inserted into a vector and ligated overnight at 16 °C using the pMD19-T vector cloning kit following the manufacturer's instructions (Takara 6013, Japan). Then, 5 μL of vector DNA solution and 50 μL of *E. coli* DH 5α were blended and incubated in ice for 30 min. After heat shock at 42 °C for 60 s, the vectors were kept on ice for 3 min. The solution with 600 μL of liquid LB culture medium was closed using Parafilm and the culture was shaken for 60 min (37 °C, 200 rpm). Then, 45 μL of 5-Bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal), 10 μL of isopropyl-L-D-thiogalactopyranoside (IPTG) and 200 μL of the culture solution were smeared evenly on solid LB culture medium containing 0.4% Ampicillin (Amp) and then incubated at 37 °C for 14 h. We selected 1–2 white single colonies and added 300 μL of liquid LB culture medium containing 0.4% Amp for co-culture (37 °C, 200 rpm) for approximately 14 h until the solutions were turbid. PCR and 1% gel electrophoresis as aforementioned were used to identify the size of DNA fragment comparing with a DL2000 DNA Marker (Takara, Japan). The solutions containing the appropriate size of DNA fragments were chosen to extract plasmids using Hipure Plasmid Micro Kit following the manufacturer's instructions (Magen, China). 10 μL solutions of plasmids were used for sequencing in Biosune, China (987 bp for PS, 433 bp for FO; Supplementary Table S1).

4.4.2. Establishment of Standard Curve

The plasmid solutions of s containing the right size of DNA fragments were amplified again using plasmids primer RV-M/M13-47 by qRT-PCR (BIO-RAD CFX96, USA) (20-μL reaction contained 0.4 μM each primer, 0.5 U SYBR Premix EX Taq II (2×), cDNA and nuclease-free water following the Takara RR820A instructions; 95 °C for 2 min, 95 °C for 5 s, 55 °C to anneal for 30 s, 72 °C for 30 s, for 30 cycles). The DNA concentration of target genes were detected by a NanoDrop2000 spectrophotometer (Thermo Scientifi, USA) and then diluted to 0, 1, 2, 3, 4 ng/μL. The standard curves were drawn based on the DNA concentration of target genes and Ct values (Supplementary Figure S2). The calculation of plasmid copy number was based on Shirima et al. (2017) [72].

$$\text{Plasmid copy number} = \frac{6.02 \times 10^{23} \times \left(\text{copies}\cdot\text{mol}^{-1}\right) \times \text{plasmid amount(g)}*}{\text{MW}} \quad (1)$$

MW = plasmid molecular weight, (=plasmid size (2692 bp) × molar mass per base (660 g·mol^{-1}·bp^{-1}); 6.02 × 10^{23} molecules/mole = Avogadro's constant; * Plasmid amount was calculated from the plasmid concentration determined by a NanoDrop2000 spectrophotometer (Thermo Scientifi, USA).

4.4.3. Determination of AQ-PCR

Ct values of recombinant plasmid containing 5 gradient concentration (as control) and soil DNA extracts were detected using specific primers of PS and FO by qRT-PCR (20-μL reaction was as mentioned above). The copy numbers of PS and FO were calculated based on the standard curve. All reactions were replicated three times.

4.5. qRT-PCR Analysis of NB-LRRs

For 35 previously identified *NB-LRRs* (Supplementary Table S2), RNA extraction of roots, reverse transcription and qRT-PCR analysis (BIO-RAD CFX96, USA) were conducted as described by Chen et al. (2018) [41]. All reactions were replicated three times. The data were normalized on the basis of the 18S rRNA threshold cycle (Ct) value. The samples with the NP treatments were used as the controls at the same sampling time, and their normalized Ct values were set to 1. The relative gene expression of the other treatments was calculated using the $2^{-\Delta\Delta CT}$ method [73].

4.6. Measurement of ABA, SA, ET and JA

The contents of abscisic acid (ABA), salicylic acid (SA), ethylene (ET) and jasmonate (JA) were determined using a one-step double-antibody sandwich enzyme-linked immunosorbent assay (ELISA). Briefly, 1.0 g of fresh root was ground in 5 mL of phosphate-buffered saline (PBS) (0.01 M, pH 7.4) with an ice-cooled mortar and centrifuged at 2500 rpm for 20 min at 4 °C to obtain a supernatant for the ELISA analysis following the protocol described in Zhao et al. (2006) [74]. The mouse monoclonal antigen and antibodies against free ABA, SA, ET and JA were provided by MLBIO Co. Ltd., Shanghai, China. The hormone content was measured at 450 nm using a microplate reader (BIO-Tek ELX800, USA). Calculations of the ELISA data were performed as described in Wang et al. (2012) [75]. The recovery percentages obtained by using internal standards during extraction and analysis were all >90%.

4.7. Measurement of Root Activity and the Physiological Index

To determine root activity, the methodology described by Zhang et al. (2013) [76] was followed with modifications. Approximately 0.5 g of fresh root was mixed with 10 mL of a half-and-half blend of 0.4% triphenyl tetrazolium chloride (TTC) and 1/15 M PBS (pH 7.4) and incubated at 37 °C for 1 h, and then the reaction was stopped by 2 mL of 1 M H_2SO_4. The roots were homogenized in ethyl acetate with a capacity of 10 mL. The absorbance of the final solution was measured at 415 nm (Pgeneral T6-1650E, China). A standard curve was used to determine the concentration of root activity in the extract.

The determination of SOD, POD and CAT activities and MDA content were as described by Li et al. (2017) [48] and Deenamo et al. (2018) [77], respectively. The pretreatment was the same, and then 0.5 g of roots was homogenized in 5 mL of precooled PBS (0.05 M, pH 7.8) with a small amount of quartz sand. Extracts were centrifuged for 15 min at 13 000 rpm. The supernatant was used for the measurement of the four indexes. The colorimetric wavelengths were 560 nm, 470 nm and 240 nm for SOD, POD and CAT, and 600 nm, 532 nm, 450 nm for MDA (Pgeneral T6-1650E, China).

The H_2O_2 content was determined by the KI method [78]. In short, 0.2 g of roots was homogenized in 0.8 mL of precooled 0.1% TCA (trichloroacetic acid) with liquid nitrogen. Extracts were centrifuged for 20 min at 19,000 rpm. Then, 0.5 mL of supernatant was added to 2 mL of KI (1 M) and 0.5 mL of PBS (100 M) for reaction at darkness for 1 h. The absorbance of the final solution was measured at 390 nm (Pgeneral T6-1650E, China). A standard curve was used to determine the concentration of H_2O_2 in the extract.

4.8. Statistical Analysis

Multiple comparisons (LSD) and Pearson correlation were analysed with SAS statistical software (V9.1 SAS Institute Inc., Cary, NC, USA). Each value with three replicates represented as the mean ± SD. $p < 0.05$ was considered as significant between any two groups. A heatmap was generated via the hierarchical clustering method using the MeV 4.9.0 tool [79].

5. Conclusions

Our results indicated that *R. glutinosa* replant disease promoted *Fusarium oxysporum* proliferation in rhizospheres soil, but *NB-LRRs* were not consecutively responsive to the FO proliferation at the transcriptional level in *R. glutinosa* roots. The analysis on the relationships between *Fusarium oxysporum* numbers and *NB-LRRs* expression showed that the *NB-LRRs* responded to the *Fusarium oxysporum* invasion with a typical phenomenon of "promotion in low concentration and suppression in high concentration", and 6 *NB-LRRs* were identified as candidates for responding to *R. glutinosa* replant disease. Salicylic acid (SA), jasmonic acid (JA), ethylene (ET) and abscisic acid (ABA) involved in the formation mechanism of *R. glutinosa* replant disease. Importantly, salicylic acid, as an important signaling molecule, which may be inhibited by *Fusarium oxysporum*, was involved in the formation of replant disease in *R. glutinosa*. Concomitantly, the physiological indexes acting downstream of life activities were affected by complex replant disease stresses and exhibited strong fluctuations, resulting in the death of *R. glutinosa*. These findings provide important insights and clues into further revealing the mechanism of replant disease.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/13/3203/s1.

Author Contributions: Conceptualization, Z.Z. and M.L.; Methodology, A.C., L.G., N.X., F.F., D.C., C.Y. and B.Z.; Validation, A.C., M.L. and Z.Z.; Formal analysis, A.C., N.X. and M.L.; Investigation, A.C., L.G. and F.F.; Resources, A.C. and M.L.; Data curation, A.C., M.L.; Writing—original draft preparation, A.C.; Writing—review and editing, M.L. and Z.Z.; Supervision, A.C., M.L. and Z.Z.; Project administration, A.C., M.L. and Z.Z.; Funding acquisition, M.L. and Z.Z.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 81573538, 81603243, 81503193 and 81403042), the National Key Research and Development Program of China (No. 2017YFC1700705), and the Natural Science Foundation of Fujian Province of China (No. 2017J01804).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DAA	Days after acclimatization
DAP	Days after planting
NP	Soil that was never planted with *R. glutinosa* for at least 10 years
TP	Soil that was consecutively planted with *R. glutinosa* in the same soils for three years
FO	*Fusarium oxysporum* f.sp. *R. glutinosa*
PS	*Pseudomonas* spp.
PR	Pathogenesis-related
PTI	Pathogen-associated molecular patterns (PAMP)-triggered immunity
ETI	Effector-triggered immunity
PRRs	Pattern recognition receptors
NB-LRR	Nucleotide-binding-leucine-rich repeat
NB-ARC	Nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4
TIR	Toll/interleukin-1 receptor
CC	Coiled-coil
NPR1	Non-expressor of pathogenesis-related proteins 1
EDS1	Enhanced disease susceptibility 1
NDR1	Non-race-specific disease resistance 1
R1A	Resistance in linkage group 1A
R1B	Resistance in linkage group 1B

RPS2	Resistance to *Pseudomonas syringae* 2
TAO1	Target of AvrB operation 1
RPW8	Resistance to *Powdery mildew* 8
MPK4	Mitogen-activated protein kinase 4
ELISA	Enzyme-linked immunosorbent assay
SA	Salicylic acid
JA	Jasmonic acid
ET	Ethylene
ABA	Abscisic acid
SOD	Superoxide dismutase
POD	Peroxidase
CAT	Catalase
MDA	Malondialdehyde
H_2O_2	Hydrogen peroxide
AQ-PCR	Absolute quantification PCR
IPTG	Isopropyl-L-D-thiogalactopyranoside
X-Gal	5-Bromo-4-chloro-3-indolyl β-D-galactoside
Amp	Ampicillin
PBS	Phosphate-buffered saline
TTC	Triphenyl tetrazolium chloride

References

1. Zhang, Z.Y.; Yin, W.J.; Li, J.; Du, J.F.; Yang, Y.H.; Chen, X.J.; Lin, W.X. Physio-ecological properties of continuous cropping *Rehmannia glutinosa*. *Chin. J. Plant Ecol.* **2010**, *34*, 547–554.
2. Yang, M.; Zhang, X.D.; Xu, Y.G.; Mei, X.Y.; Jiang, B.B.; Liao, J.J.; Yin, Z.B.; Zheng, J.F.; Zhao, Z.; Fan, L.M.; et al. Autotoxic ginsenosides in the rhizosphere contribute to the replant failure of *Panax notoginseng*. *PLoS ONE* **2015**, *10*, e0118555. [CrossRef] [PubMed]
3. Rumberger, A.; Merwin, I.A.; Thies, J.E. Microbial community development in the rhizosphere of apple trees at a replant disease site. *Soil Biol. Biochem.* **2007**, *39*, 1645–1654. [CrossRef]
4. Yang, J.I.; Ruegger, P.M.; McKenry, M.V.; Becker, J.O.; Borneman, J. Correlations between root-associated microorganisms and peach replant disease symptoms in a California soil. *PLoS ONE* **2012**, *7*, e46420. [CrossRef] [PubMed]
5. Li, W.H.; Liu, Q.Z.; Chen, P. Effect of long-term continuous cropping of strawberry on soil bacterial community structure and diversity. *J. Integr. Agric.* **2018**, *17*, 2570–2582. [CrossRef]
6. Bai, L.; Cui, J.Q.; Jie, W.G.; Cai, B.Y. Analysis of the community compositions of rhizosphere fungi in soybeans continuous cropping fields. *Microbiol. Res.* **2015**, *180*, 49–56. [CrossRef] [PubMed]
7. Chen, A.G.; Li, M.J.; Zhang, B.; Wang, F.J.; Gu, L.; Feng, F.J.; Liu, H.Y.; Zhang, Z.Y. Review on catastrophe mechanism of medicinal plant and its rhizosphere microecosystem mediated by consecutive monoculture. *Mod. Chin. Med.* **2016**, *18*, 239–245.
8. Zhang, B.; Li, X.Z.; Wang, F.Q.; Li, M.J.; Zhang, J.Y.; Gu, L.; Zhang, L.J.; Tu, W.Q.; Zhang, Z.Y. Assaying the potential autotoxins and microbial community associated with *Rehmannia glutinosa* replant problems based on its 'autotoxic circle'. *Plant Soil* **2016**, *407*, 307–322. [CrossRef]
9. Wen, X.S.; Yang, S.L.; Wei, J.H.; Zheng, J.H. Textual research on planting history of *Rehmannia glutinosa* and its cultivated varieties. *Chin. Tradit. Herb. Drugs* **2002**, *33*, 946–949.
10. Wu, L.K.; Wang, J.Y.; Huang, W.M.; Wu, H.M.; Chen, J.; Yang, Y.Q.; Zhang, Z.Y.; Lin, W.X. Plant-microbe rhizosphere interactions mediated by *Rehmannia glutinosa* root exudates under consecutive monoculture. *Sci. Rep.* **2015**, *5*, 15871. [CrossRef]
11. Zhang, Z.Y.; Lin, W.X. Continuous cropping obstacle and allelopathic autotoxicity of medicinal plants. *Chin. J. Eco-Agric.* **2009**, *17*, 189–196. [CrossRef]
12. Kuijken, R.C.P.; Snel, J.F.H.; Heddes, M.H.; Bouwmeester, H.J.; Marcelis, L.F.M. The importance of a sterile rhizosphere when phenotyping for root exudation. *Plant Soil* **2015**, *387*, 131–142. [CrossRef]

13. Wu, L.K.; Li, Z.F.; Li, J.; Khan, M.A.; Huang, W.M.; Zhang, Z.Y.; Lin, W.X. Assessment of shifts in microbial community structure and catabolic diversity in response to *Rehmannia glutinosa* monoculture. *Appl. Soil. Ecol.* **2013**, *67*, 1–9. [CrossRef]
14. Barto, E.K.; Cipollini, D. Half-lives and field soil concentrations of *Alliaria petiolata* secondary metabolites. *Chemosphere* **2009**, *76*, 71–75. [CrossRef] [PubMed]
15. Zhang, Z.Y.; Pan, L.P.; Li, H.H. Isolation, identification and characterization of soil microbes which degrade phenolic allelochemicals. *J. Appl. Microbiol.* **2010**, *108*, 1839–1849. [CrossRef] [PubMed]
16. Li, X.G.; Ding, C.F.; Hua, K.; Zhang, T.L.; Zhang, Y.N.; Zhao, L.; Yang, Y.R.; Liu, J.G.; Wang, X.X. Soil sickness of peanuts is attributable to modifications in soil microbes induced by peanut root exudates rather than to direct allelopathy. *Soil Boil. Biochem.* **2014**, *78*, 149–159. [CrossRef]
17. Venturi, V.; Keel, C. Signaling in the rhizosphere. *Trends Plant Sci.* **2016**, *21*, 187–198. [CrossRef]
18. Sasse, J.; Martinoia, E.; Northen, T. Feed your friends: Do plant exudates shape the root microbiome? *Trends Plant Sci.* **2018**, *23*, 25–41. [CrossRef]
19. Saleem, M.; Pervaiz, Z.H.; Traw, M.B. Theories, mechanisms and patterns of microbiome species coexistence in an era of climate change. In *Microbiome Community Ecology*; Springer: Cham, Switzerland, 2015; pp. 13–53.
20. Bulgarelli, D.; Schlaeppi, K.; Spaepen, S.; Ver, L.T.; Schulze-Lefert, P. Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* **2013**, *64*, 807–838. [CrossRef]
21. Lebeis, S.L. Greater than the sum of their parts: Characterizing plant microbiomes at the community-level. *Curr. Opin. Plant Biol.* **2015**, *24*, 82–86. [CrossRef]
22. Biere, A.; Goverse, A. Plant-mediated systemic interactions between pathogens, parasitic nematodes, and herbivores above- and belowground. *Annu. Rev. Phytopathol.* **2016**, *54*, 499–527. [CrossRef] [PubMed]
23. Zhao, X.S.; Zhen, W.C.; Qi, Y.Z.; Liu, X.J.; Yin, B.Z. Coordinated effects of root autotoxic substances and *Fusarium oxysporum* Schl. f. sp. *fragariae on the growth and replant disease of strawberry. Front. Agric. China* **2009**, *3*, 34. [CrossRef]
24. Mazzola, M.; Manici, L.M. Apple replant disease: Role of microbial ecology in cause and control. *Annu. Rev. Phytopathol.* **2012**, *50*, 45–65. [CrossRef] [PubMed]
25. Spath, M.; Insam, H.; Peintner, U.; Kelderer, M.; Kuhnert, R.; Franke-Whittle, I. Linking soil biotic and abiotic factors to apple replant disease: A greenhouse approach. *J. Phytopathol.* **2015**, *163*, 287–299. [CrossRef]
26. Zhang, Z.Y.; Chen, H.; Yang, Y.H.; Chen, T.; Lin, R.Y.; Chen, X.J.; Lin, W.X. Effects of continuous cropping on bacterial community diversity in rhizosphere soil of *Rehmannia glutinosa*. *Chin. J. Appl. Ecol.* **2010**, *21*, 2843–2848.
27. Zhang, Z.Y.; Lin, W.X.; Yang, Y.H.; Chen, H.; Chen, X.J. Effects of consecutively monocultured *Rehmannia glutinosa* L. on diversity of fungal community in rhizospheric soil. *Agric. Sci. China* **2011**, *10*, 1374–1384. [CrossRef]
28. Li, Z.F.; Yang, Y.Q.; Wu, L.K.; Shu, Y.; Zhao, Y.P.; Huang, W.M.; Zhang, Z.Y.; Lin, W.X. Isolation of highly pathogenic pathogens and identification of formae speciales of *Rehmannia glutinosa* L. *Chin. J. Eco-Agric.* **2013**, *21*, 1426–1433. [CrossRef]
29. Wu, L.K.; Huang, W.M.; Wang, J.Y.; Wu, H.M.; Chen, J.; Qin, X.J.; Zhang, Z.Y.; Lin, W.X. Diversity analysis of rhizosphere microflora of wild *R. glutinosa* grown in monocropping for different years. *Acta Agron. Sin.* **2015**, *41*, 308–317. [CrossRef]
30. Wu, L.K.; Wang, J.Y.; Wu, H.M.; Chen, J.; Xiao, Z.G.; Qin, X.J.; Zhang, Z.Y.; Lin, W.X. Comparative metagenomic analysis of rhizosphere microbial community composition and functional potentials under *Rehmannia glutinosa* consecutive monoculture. *Int. J. Mol. Sci.* **2018**, *19*, 2394. [CrossRef] [PubMed]
31. Berendsen, R.L.; Pieterse, C.M.J.; Bakker, P.A.H.M. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **2012**, *17*, 478–486. [CrossRef] [PubMed]
32. Santhanam, R.; Luu, V.T.; Weinhold, A.; Goldberg, J.; Oh, Y.; Baldwin, I.T. Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5013–5020. [CrossRef] [PubMed]
33. Tena, G. Immunity: NLR population control. *Nat. Plants* **2016**, *2*, 16026. [CrossRef] [PubMed]
34. Miller, R.N.; Costa Alves, G.S.; van Sluys, M.A. Plant immunity: Unravelling the complexity of plant responses to biotic stresses. *Ann. Bot.* **2017**, *119*, 681–687. [CrossRef] [PubMed]
35. Jones, J.D.G.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [CrossRef] [PubMed]

36. Stael, S.; Kmiecik, P.; Willems, P.; Kelen, K.V.D.; Coll, N.S.; Teige, M.; van Breusegem, F. Plant innate immunity-sunny side up? *Trends Plant Sci.* **2015**, *20*, 3–11. [CrossRef] [PubMed]
37. Dodds, P.N.; Rathjen, J.P. Plant immunity: Towards an integrated view of plant–pathogen interactions. *Nature* **2010**, *11*, 539–548. [CrossRef] [PubMed]
38. Kapo, P.; Devendrakumar, K.T.; Li, X. Plant NLRs: From discovery to application. *Plant Sci.* **2019**, *279*, 3–18. [CrossRef] [PubMed]
39. Saleem, M.; Ji, H.; Amirullah, A.; Traw, M.B. *Pseudomonas syringae pv. tomato DC3000* growth in multiple gene knockouts predicts interactions among hormonal, biotic and abiotic stress responses. *Eur. J. Plant Pathol.* **2017**, *149*, 779–786. [CrossRef]
40. Xie, Z.M.; Yang, C.Y.; Chen, A.G.; Li, M.J.; Gu, L.; Zhang, J.Y.; Feng, F.J.; Zhang, B.; Chen, S.Q.; Zhang, Z.Y. Identification and expression analysis of leucine-rich repeat receptor-like kinase family reveals the roles of resistance proteins during formation of replant disease in *Rehmannia glutinosa* Libosch. *Int. J. Agric. Biol.* **2019**. [CrossRef]
41. Chen, A.G.; Gu, L.; Li, M.J.; Xu, N.; Feng, F.J.; Liu, G.L.; Zhang, B.; Gong, D.P.; Zhang, J.Y.; Liu, H.Y.; et al. Identifcation of *Rehmannia glutinosa* L. NB-ARC family proteins and their typical changes under consecutive monoculture stress. *Acta Physiol. Plant.* **2018**, *40*, 95. [CrossRef]
42. Andolfo, G.; Ercolano, M.R. Plant innate immunity multicomponent model. *Front. Plant Sci.* **2015**, *6*, 987. [CrossRef] [PubMed]
43. Shigenaga, A.M.; Berens, M.L.; Tsuda, K.; Argueso, C.T. Towards engineering of hormonal crosstalk in plant immunity. *Curr. Opin. Plant Biol.* **2017**, *38*, 164–172. [CrossRef] [PubMed]
44. Derksen, H.; Rampitsch, C.; Daayf, F. Signaling cross-talk in plant disease resistance. *Plant Sci.* **2013**, *207*, 79–87. [CrossRef] [PubMed]
45. Lievens, L.; Pollier, J.; Goossens, A.; Beyaert, R.; Staal, J. Abscisic acid as pathogen effector and immune regulator. *Front. Plant Sci.* **2017**, *8*, 587. [CrossRef] [PubMed]
46. Kim, H.J.; Kim, K.J.; Lee, Y.S. Isolation and characterization of the salicylic acid induced gene in *Rehmannia glutinosa* by differential display. *Mycobiology* **2002**, *30*, 88–92. [CrossRef]
47. Wang, F.Q.; Zhi, J.Y.; Zhang, Z.Y.; Wang, L.N.; Suo, Y.F.; Xie, C.X.; Li, M.J.; Zhang, B.; Du, J.F.; Gu, L.; et al. Transcriptome analysis of salicylic acid treatment in *Rehmannia glutinosa* hairy roots using RNA-seq technique for identification of genes involved in acteoside biosynthesis. *Front. Plant Sci.* **2017**, *8*, 787. [CrossRef] [PubMed]
48. Li, M.J.; Yang, Y.H.; Feng, F.J.; Zhang, B.; Chen, S.Q.; Yang, C.Y.; Gu, L.; Wang, F.Q.; Zhang, J.Y.; Chen, A.G.; et al. Differential proteomic analysis of replanted *Rehmannia glutinosa* roots by iTRAQ reveals molecular mechanisms for formation of replant disease. *BMC Plant Biol.* **2017**, *17*, 116. [CrossRef]
49. Wu, L.K.; Chen, J.; Khan, M.U.; Wang, J.Y.; Wu, H.M.; Xiao, Z.G.; Zhang, Z.Y.; Lin, W.X. Rhizosphere fungal community dynamics associated with *Rehmannia glutinosa* replant disease in a consecutive monoculture regime. *Phytopathology* **2018**, *108*, 1493–1500. [CrossRef]
50. Wu, L.K.; Chen, J.; Xiao, Z.G.; Zhu, X.C.; Wang, J.Y.; Wu, H.M.; Wu, Y.H.; Zhang, Z.Y.; Lin, W.X. Barcoded pyrosequencing reveals a shift in the bacterial community in the rhizosphere and rhizoplane of *Rehmannia glutinosa* under consecutive monoculture. *Int. J. Mol. Sci.* **2018**, *19*, 850. [CrossRef]
51. Mendes, R.; Kruijt, M.; de Bruijn, I.; Dekkers, E.; Voort, M.V.D.; Schneider, J.H.M.; Piceno, Y.M.; DeSantis, T.Z.; Andersen, G.L.; Bakker, P.A.H.M.; et al. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **2011**, *332*, 1097–1100. [CrossRef]
52. Bouizgarne, B.; Brault, M.; Pennarun, A.M.; Rona, J.P.; Ouhdouch, Y.; Hadrami, E.I.; Bouteau, F. Electrophysiological responses to fusaric acid of root hairs from seedlings of date palm-susceptible and -resistant to *Fusarium oxysporum* f. sp. *albedinis*. *J. Phytopathol.* **2004**, *152*, 321–324. [CrossRef]
53. Ye, S.F.; Yu, J.Q.; Peng, Y.H.; Zheng, J.H.; Zou, L.Y. Incidence of Fusarium wilt in *Cucumis sativus* L. is promoted by cinnamic acid, an autotoxin in root exudates. *Plant Soil* **2004**, *263*, 143–150. [CrossRef]
54. Scott, G.; Almasrahi, A.; Mansoorkhani, F.M.; Rupar, M.; Dickinson, M.; Shama, G. Hormetic UV-C seed treatments for the control of tomato diseases. *Plant Pathol.* **2019**, *68*, 700–707. [CrossRef]
55. Li, Y.; Li, S.; Bi, D.; Cheng, Y.T.; Li, X.; Zhang, Y. SRFR1 negatively regulates plant NB-LRR resistance protein accumulation to prevent autoimmunity. *PLoS Pathog.* **2010**, *6*, e1001111. [CrossRef] [PubMed]

56. Zhu, Z.H.; Xu, F.; Zhang, Y.X.; Cheng, Y.T.; Wiermer, M.; Li, X.; Zhang, Y.L. Arabidopsis resistance protein SNC1 activates immune response through association with a transcriptional corepressor. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13960–13965. [CrossRef] [PubMed]
57. Niu, D.; Lin, X.L.; Kong, X.X.; Qu, G.P.; Cai, B.; Lee, J.Y.; Jin, J.B. SIZ1-mediated SUMOylation of TPR1 suppresses plant Immunity in *Arabidopsis*. *Mol. Plant* **2019**, *12*, 215–228. [CrossRef]
58. Richard, M.M.S.; Takken, F.L.W. Plant autoimmunity: When good things go bad. *Curr. Biol.* **2017**, *27*, 361–363. [CrossRef]
59. Wang, J.L.; Grubb, L.E.; Wang, J.Y.; Liang, X.X.; Li, L.; Gao, C.L.; Ma, M.M.; Feng, F.; Li, M.; Li, L.; et al. A regulatory module controlling homeostasis of a plant immune kinase. *Mol. Cell* **2018**, *69*, 493–504. [CrossRef]
60. Cai, X.Z.; Xu, Q.F.; Wang, C.C.; Zheng, Z. Development of a virus-induced gene-silencing system for functional analysis of the *RPS2*-dependent resistance signalling pathways in *Arabidopsis*. *Plant Mol. Biol.* **2006**, *62*, 223–232. [CrossRef]
61. Eitas, T.K.; Nimchuk, Z.L.; Dangl, J.L. *Arabidopsis* TAO1 is a TIR-NB-LRR protein that contributes to disease resistance induced by the *Pseudomonas syringae* effector AvrB. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6475–6480. [CrossRef]
62. Slootweg, E.; Roosien, J.; Spiridon, L.N.; Petrescu, A.J.; Tameling, W.; Joosten, M.; Pomp, R.; van Schaik, C.; Dees, R.; Borst, J.W.; et al. Nucleocytoplasmic distribution is required for activation of resistance by the potato NB-LRR receptor Rx1 and is balanced by its functional domains. *Plant Cell* **2010**, *22*, 4195–4215. [CrossRef] [PubMed]
63. Castel, B.; Ngou, P.M.; Cevik, V.; Redkar, A.; Kim, D.S.; Yang, Y.; Ding, P.T.; Jones, J.D.G. Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* **2019**, *222*, 966–980. [CrossRef] [PubMed]
64. Berens, M.L.; Berry, H.M.; Mine, A.; Argueso, C.T.; Tsuda, K. Evolution of hormone signaling networks in plant defense. *Annu. Rev. Phytopathol.* **2017**, *55*, 401–425. [CrossRef] [PubMed]
65. An, C.F.; Mou, Z.L. Salicylic acid and its function in plant immunity. *J. Integr. Plant Biol.* **2011**, *53*, 412–428. [CrossRef] [PubMed]
66. Dar, T.A.; Uddin, M.; Khan, M.M.A.; Hakeem, K.R.; Jaleel, H. Jasmonates counter plant stress: A review. *Environ. Exp. Bot.* **2015**, *115*, 49–57. [CrossRef]
67. Wu, L.M.; Huang, Z.Y.; Li, X.; Ma, L.M.; Gu, Q.; Wu, H.J.; Liu, J.; Borriss, R.; Wu, Z.; Gao, X.W. Stomatal closure and SA-, JA/ET-signaling pathways are essential for *Bacillus amyloliquefaciens* FZB42 to restrict leaf disease caused by *Phytophthora nicotianae* in *Nicotiana benthamiana*. *Front. Microbiol.* **2018**, *9*, 847. [CrossRef] [PubMed]
68. Adie, B.A.; Pérez-Pérez, J.; Pérez-Pérez, M.M.; Godoy, M.; Sánchez-Serrano, J.J.; Schmelz, E.A.; Solano, R. ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defense in *Arabidopsis*. *Plant Cell* **2007**, *19*, 1665–1681. [CrossRef]
69. Niu, M.M.; Li, J.; Du, J.F.; Yin, W.J.; Yang, Y.H.; Chen, X.J.; Zhang, Z.Y. Changes in source-sink relationship of photosynthate in *Rehmannia glutinosa* Libosch. and their relations with continuous cropping obstacle. *Chin. J. Ecol.* **2011**, *30*, 248–254.
70. Coyne, K.J.; Handy, S.M.; Demir, E.; Whereat, E.B.; Hutchins, D.A.; Portune, K.J.; Doblin, M.A.; Cary, C. Improved quantitative real-time PCR assays for enumeration of harmful algal species in field samples using an exogenous DNA reference standard. *Limnol. Oceanogr. Methods* **2005**, *3*, 381–391. [CrossRef]
71. Yang, L.; Lou, J.; Wang, H.Z.; Wu, L.S.; Xu, J.M. Use of an improved high-throughput absolute abundance quantification method to characterize soil bacterial community and dynamics. *Sci. Total Environ.* **2018**, *633*, 360–371. [CrossRef]
72. Shirima, R.R.; Maeda, D.G.; Kanju, E.; Ceasar, G.; Tibazarwa, F.; Legg, J.P. Absolute quantification of cassava brown streak virus mRNA by real-time qPCR. *J. Virol. Methods* **2017**, *245*, 5–13. [CrossRef] [PubMed]
73. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]
74. Zhao, J.; Li, G.; Yi, G.X.; Wang, B.M.; Deng, A.X.; Nan, T.G.; Li, Z.H.; Li, Q.X. Comparison between conventional indirect competitive enzyme-linked immunosorbent assay (icELISA) and simplified icELISA for small molecules. *Anal. Chim. Acta* **2006**, *571*, 79–85. [CrossRef] [PubMed]

75. Wang, Y.; Li, B.; Du, M.W.; Eneji, A.E.; Wang, B.M.; Duan, L.S.; Li, Z.H.; Tian, X.L. Mechanism of phytohormone involvement in feedback regulation of cotton leaf senescence induced by potassium deficiency. *J. Exp. Bot.* **2012**, *63*, 5887–5901. [CrossRef] [PubMed]
76. Zhang, X.; Huang, G.; Bian, X.; Zhao, Q. Effects of root interaction and nitrogen fertilization on the chlorophyll content, root activity, photosynthetic characteristics of intercropped soybean and microbial quantity in the rhizosphere. *Plant Soil Environ.* **2013**, *59*, 80–88. [CrossRef]
77. Deenamo, N.; Kuyyogsuy, A.; Khompatara, K.; Chanwun, T.; Ekchaweng, K.; Churngchow, N. Salicylic acid induces resistance in rubber tree against *Phytophthora palmivora*. *Int. J. Mol. Sci.* **2018**, *19*, 1883. [CrossRef] [PubMed]
78. Chakrabrty, D.; Datta, S.K. Micropropagation of gerbera: Lipid peroxidation and antioxidant enzyme activities during acclimatization process. *Acta Physiol. Plant.* **2008**, *30*, 325–331. [CrossRef]
79. Saeed, A.I.; Bhagabati, N.K.; Braisted, J.C.; Liang, W.; Sharov, V.; Howe, E.A.; Li, J.; Thiagarajan, M.; White, J.A.; Quackenbush, J. TM4 microarray software suite. *Methods Enzymol.* **2006**, *411*, 134–193. [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Identification and Expression Profiling of the Polygalacturonase (*PG*) and Pectin Methylesterase (*PME*) Genes in Grapevine (*Vitis vinifera* L.)

Nadeem Khan [1,†], Fizza Fatima [2,†], Muhammad Salman Haider [3,†], Hamna Shazadee [4], Zhongjie Liu [3], Ting Zheng [3] and Jinggui Fang [3,*]

1 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Ministry of Science and Technology/College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
2 Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada
3 Key laboratory of Genetics and Fruit Development, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
4 College of Life Science, Nanjing Agricultural University, Nanjing 210095, China
* Correspondence: fanggg@njau.edu.cn
† These authors contributed equally to this work.

Received: 19 May 2019; Accepted: 27 June 2019; Published: 28 June 2019

Abstract: In pectin regulation, polygalacturonases (PGs) and pectin methylesterases (PMEs) are critical components in the transformation, disassembly network, and remodeling of plant primary cell walls. In the current study, we identified 36 *PG* and 47 *PME* genes using the available genomic resources of grapevine. Herein, we provide a comprehensive overview of PGs and PMEs, including phylogenetic and collinearity relationships, motif and gene structure compositions, gene duplications, principal component analysis, and expression profiling during developmental stages. Phylogenetic analysis of PGs and PMEs revealed similar domain composition patterns with *Arabidopsis*. The collinearity analysis showed high conservation and gene duplications with purifying selection. The type of duplications also varied in terms of gene numbers in PGs (10 dispersed, 1 proximal, 12 tandem, and 13 segmental, respectively) and PMEs (23 dispersed, 1 proximal, 16 tandem, and 7 segmental, respectively). The tissue-specific response of *PG* and *PME* genes based on the reported transcriptomic data exhibited diverged expression patterns in various organs during different developmental stages. Among PGs, *VvPG8, VvPG10, VvPG13, VvPG17, VvPG18, VvPG19, VvPG20, VvPG22*, and *VvPG23* showed tissue- or organ-specific expression in majority of the tissues during development. Similarly, in PMEs, *VvPME3, VvPME4, VvPME5, VvPME6, VvPME19, VvPME21, VvPME23, VvPME29, VvPME31*, and *VvPME32* suggested high tissue-specific response. The gene ontology (GO), Kyoto Encyclopedia of Genes and Genomics (KEGG) enrichment, and cis-elements prediction analysis also suggested the putative functions of PGs and PMEs in plant development, such as pectin and carbohydrate metabolism, and stress activities. Moreover, qRT-PCR validation of 32 *PG* and *PME* genes revealed their role in various organs of grapevines (i.e., root, stem, tendril, inflorescence, flesh, skins, and leaves). Therefore, these findings will lead to novel insights and encourage cutting-edge research on functional characterization of PGs and PMEs in fruit crop species.

Keywords: polygalacturonase (PGs), pectin methylesterase (PMEs), collinearity analysis; gene duplications; expression profiling; grapevine

1. Introduction

Plant cell walls are mainly composed of various interacting networks of carbohydrate polymers, such as polymers of cellulose, hemicellulose, and pectins [1]. These are produced by the plant cells before being systemically delivered to the apoplast tissues [2], providing support and mechanical strength to the plants. However, the disruption of the cell wall structure exposes fruit crops to disorders, such as softening and susceptibility to diseases [3]. Until now, several studies have shown multiple associated roles linked to pectin degradation enzymes in plants. In particular, polygalacturonase (PGs; EC 3.2.1.15) hydrolytic enzymes are involved in plant organ cell separation events, reproductive developments, leaf morphology, and organ shedding [4–7]. Furthermore, expression profiling, both at a temporal and spatial level, has revealed that transcript accumulations are also responsible for cell wall softening during abscission, ripening, and dehiscence [8]. Moreover, these correlation analyses are useful for identifying cell wall sites during disassembly and studying the transcriptional patterns of key genes in cell-wall degrading enzymes [5]. In *Arabidopsis,* during different developmental stages, several members of PGs can be detected by real-time PCR (RT-PCR) in various organs, such as roots, leaves, pollen tubes, flowers, and siliques [9,10]. In addition, few members of PGs have been functionally characterized in some fruit crop species, such as grapevine (*VvPG1* and *VvPG2*), apple (*MdPG36*), banana (*MAPG1* to *MAPG4*), and pear (*PcPGl* and *PcPG3*) [11–14]. However, the silencing of *FaPG1* gene in strawberries declines the halt of the middle lamella and reduces fruit softening [15]. In pears and bananas, the softening of fruits is regulated by *Pc-PG1* and *Pc-PG2* as well as *MaPG3* and *MaPG4*, in an ethylene-dependent pattern [16,17].

Pectin methylesterases (PMEs; EC 3.1.1.11) are also hydrolytic enzymes present in plants, which play a pivotal role in the firmness and softening of the cell walls, specifically through remodeling and disassembly of pectin [18,19]. The PMEs also play a crucial role in plant development, including fruit maturity, pollen development, pollen tube growth, and other growth-related factors [20,21]. However, the activity of both PGs and PMEs during fruit softening is largely unknown and various studies have shown their contradictory actions [22,23]. PGs in plants contain three groups, namely class A, B, and C, whereas the PMEs are categorized into two main types, namely type I and type II [19,24]. PGs and PMEs are a major gene family and have extensively been studied in various crops, including *Brassica rapa, Gossypium* species, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Malus domestica*, *Cucumis sativus*, and *Citrullus lanatus* [5,11,25–28].

Grapevine (*Vitis vinifera* L.) is one of the most important woody fruit crop resource in the world [29]. They have become widely popular due to their nutritional value and health benefits. Improving the grapevine berry quality is a crucial issue, and the availability of the grapevine genome (Version 2.1) provides an excellent opportunity for its genetic study, momentously facilitating research in grapevine biology. In grapevine, the members of PGs and PMEs have not yet been comprehensively identified by genome-wide analysis and their elucidation is largely unknown. In this study, we utilized genomic resources to characterize PGs and PMEs in grapevine by various bioinformatic tools. In total, 36 *PG* and 47 *PME* genes were identified in grapevine and compared with *Arabidopsis* in order to test their domain composition. The expression patterns of genes provide valuable clues for understanding their functions. We also tested 32 *PG* and *PME* genes for various organs of grapevine, and their correlation factors were further compared. Overall, our study provides a brief understanding of the pectin-related genes in grapevine, and their substantial role in regulating organs during different developmental stages.

2. Results

2.1. Identification of PGs and PMEs in Grapevine

A total of 36 *PG* and 47 *PME* genes were identified from the grapevine genome through orthologues analysis with *Arabidopsis* (Table S1). These genes were designated as *VvPG1-VcPG36* for PGs and *VvPME1-VcPME46* for PMEs according to the previously reported nomenclature system with slight

modification. The corresponding UNIPROT gene IDs are also listed in Table 1. A brief summary of the basic information for both PGs and PMEs in grapevine, apple, peach, and citrus is shown in Tables S1 and S2, where the basic features include the protein identifier, CDS length (bp), and protein properties (i.e., protein length (aa), molecular weight (kDa), isoelectric point (PIs), and grand average of hydropathicity (GRAVY)). Additionally, gene duplication types and subcellular predictions were also explored for each protein of PGs and PMEs in grapevine. In general, the coding sequence length of these genes varied from 999–3702 bp for PGs and 768–2463 bp for PMEs, while the protein length ranged from 332–1233 aa for PGs and 255–820 aa for PMEs, respectively. Moreover, the kDa ranged from 36.38–135.75 for PGs and 28.12–90.85 for PMEs, while the PIs varied from 4.73–9.74 for PGs and 5.17–9.97 for PMEs. Additionally, the results of GRAVY reveal that both PGs and PMEs are hydrophilic and hydrophobic in nature. Although most showed hydrophobic properties with negative values, a limited number showed positive hydrophobic properties. Protein subcellular predications also confirmed that the majority of PGs and PMEs occurred in the nucleus, endoplasmic reticulum, cytoplasm, plasma membrane, and mitochondria among ohers, which are listed in Table S1. Moreover, the types of duplications also varied in terms of gene numbers in PGs (10 dispersed, 1 proximal, 12 tandem, and 13 segmental, respectively) and PMEs (23 dispersed, 1 proximal, 16 tandem, and 7 segmental, respectively). Hence, the observed variations in various properties among PGs and PMEs implies that these genes may function contrarily in a variable environment.

Table 1. Duplications of the *PG* and *PME* genes in grapevine.

Gene 1	Gene 2	*Ks*	*Ka*	*Ka/Ks*	Selection Pressure	Gene Duplications
Between *PG* genes						
VvPG1	*VvPG2*	0.029	0.022	0.77	Purifying Selection	Tandem
VvPG3	*VvPG4*	0.034	0.016	0.47	Purifying Selection	Tandem
VvPG6	*VvPG25*	0.852	0.579	0.68	Purifying Selection	Tandem
VvPG28	*VvPG31*	0.738	0.693	0.94	Purifying Selection	Tandem
VvPG33	*VvPG34*	0.663	0.197	0.30	Purifying Selection	Tandem
VvPG8	*VvPG9*	0.5	0.458	0.92	Purifying Selection	Dispersed
VvPG12	*VvPG13*	0.626	0.58	0.93	Purifying Selection	Dispersed
VvPG16	*VvPG17*	0.879	0.472	0.54	Purifying Selection	Dispersed
VvPG20	*VvPG24*	1.56	0.419	0.27	Purifying Selection	Dispersed
VvPG29	*VvPG30*	0.522	0.628	1.20	Positive Selection	Dispersed
VvPG5	*VvPG7*	0.376	0.409	1.09	Positive Selection	WGD or Segmental
VvPG10	*VvPG11*	1.266	0.231	0.18	Purifying Selection	WGD or Segmental
VvPG14	*VvPG15*	0.212	0.343	1.62	Positive Selection	WGD or Segmental
VvPG18	*VvPG21*	0.515	0.522	1.01	Positive Selection	WGD or Segmental
VvPG22	*VvPG23*	1.401	0.152	0.11	Purifying Selection	WGD or Segmental
Between *PME* genes						
VvPME2	*VvPME3*	1.312	0.421	0.32	Purifying Selection	Dispersed
VvPME5	*VvPME16*	0.527	0.625	1.19	Positive Selection	Dispersed
VvPME18	*VvPME19*	1.102	0.442	0.40	Purifying Selection	Dispersed
VvPME20	*VvPME21*	0.881	0.464	0.53	Purifying Selection	Dispersed
VvPME22	*VvPME23*	0.446	0.681	1.53	Positive Selection	Dispersed
VvPME27	*VvPME28*	1.042	0.334	0.32	Purifying Selection	Dispersed
VvPME35	*VvPME36*	0.565	0.697	1.23	Positive Selection	Dispersed
VvPME1	*VvPME4*	0.798	0.438	0.55	Purifying Selection	Tandem
VvPME8	*VvPME10*	0.18	0.019	0.11	Purifying Selection	Tandem
VvPME11	*VvPME12*	0.879	0.086	0.10	Purifying Selection	Tandem
VvPME13	*VvPME15*	1.035	0.171	0.17	Purifying Selection	Tandem
VvPME17	*VvPME25*	0.64	0.631	0.99	Purifying Selection	Tandem
VvPME6	*VvPME7*	0.271	0.41	1.51	Positive Selection	WGD or Segmental
VvPME14	*VvPME26*	0.304	0.44	1.45	Positive Selection	WGD or Segmental
VvPME33	*VvPME34*	1.284	0.209	0.16	Purifying Selection	WGD or Segmental

2.2. Phylogenetic and Collinearity Relationships, Motif Compositions and Genomic Structure of PGs and PMEs in Grapevine

The phylogenetic relationships of 36 PGs and 47 PMEs of grapevine and *Arabidopsis* were obtained using MEGA 7.0 with a maximum likelihood approach (ML). The phylogenetic tree revealed that *PG* and *PME* genes can further be divided into six and five major clades (Figure 1a,b). In the phylogenetic tree of both PGs and PMEs, we observed that clade two contained the most number of genes (15 and 14) compared to other clades in grapevine. The phylogenetic analysis suggests that both PGs and PMEs share high similarities and have close genetic relationships with *Arabidopsis*. The observed results of our phylogenetic arrangement were also consistent with previously reported studies [25,27]. In addition, we also constructed phylogenetic trees among *PG* and *PME* genes (Figures 2a and 3a). The results showed consistency among the clades of PGs compared to PMEs, which may be due to variations in tree topologies. We also analyzed the composition of motifs for both PGs and PMEs (Figures 2b and 3b). For PG and PME proteins, we obtained ten conserved motifs using the online server, Multiple Em for Motif Elicitation (MEME). The results revealed that motifs five and two frequently occurred among PG members. Similarly, for PME members, motifs six, five, four, and one were dominantly found in grapevine (Figures 2b and 3b). Hence, these results suggest that most PG and PME protein members carry unique features due to variation in their amino acid sequences. Additionally, we also obtained their LOGOS by the same online server MEME. Ten consensus sequences were acquired for both PG and PME protein members and their distribution patterns are shown in Figures S1 and S2.

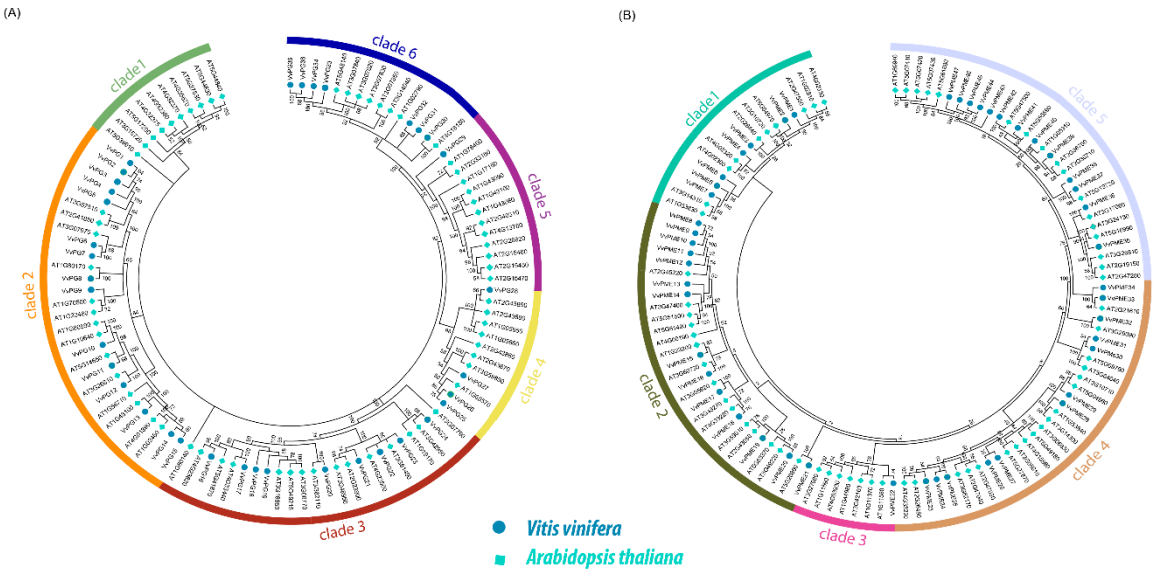


Figure 1. Phylogenetic relationship of *PG* (**A**) and *PME* (**B**) genes between grapevine and *Arabidopsis*. The phylogenetic tree was constructed by MEGA 7.0 using the Maximum Likelihood Method (1000 bootstrap).

Furthermore, based on coding sequence (CDS) and untranslated regions (UTR) of *PG* and *PME* genes in grapevine, gene structures were also resolved using TBtools software (Figures 2c and 3c). The results revealed that both PG and PME members exhibited high divergence and were largely conserved compared to each other. On the other hand, the *PG* and *PME* genes displayed more or fewer similarities among the same clades. This was also observed in a previously reported study focusing on PGs and PMEs in *Brassica rapa* [25].

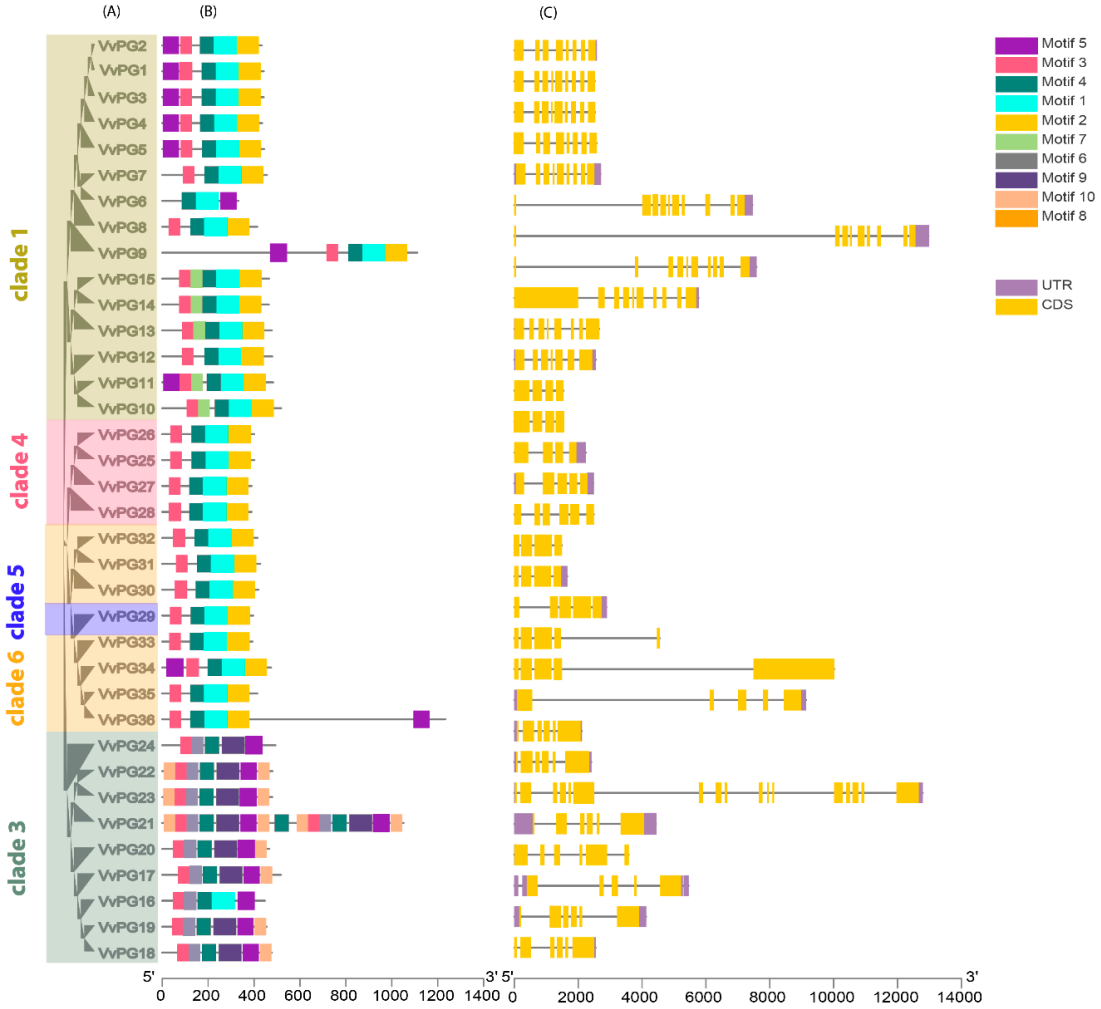


Figure 2. Phylogenetic relationship of *PGs* (**A**). The phylogenetic tree was constructed by MEGA 7.0 using the Maximum Likelihood Method (1000 bootstrap). Motif structure and upstream/downstream regions of *PGs* (**B**). The coding sequences (CDS) and untranslated regions (UTR) for PGs in grapevine (**C**). CDS and UTR are represented by yellow and green boxes. The relative position is proportionally displayed based on the kilobase scale at the bottom of the figures.

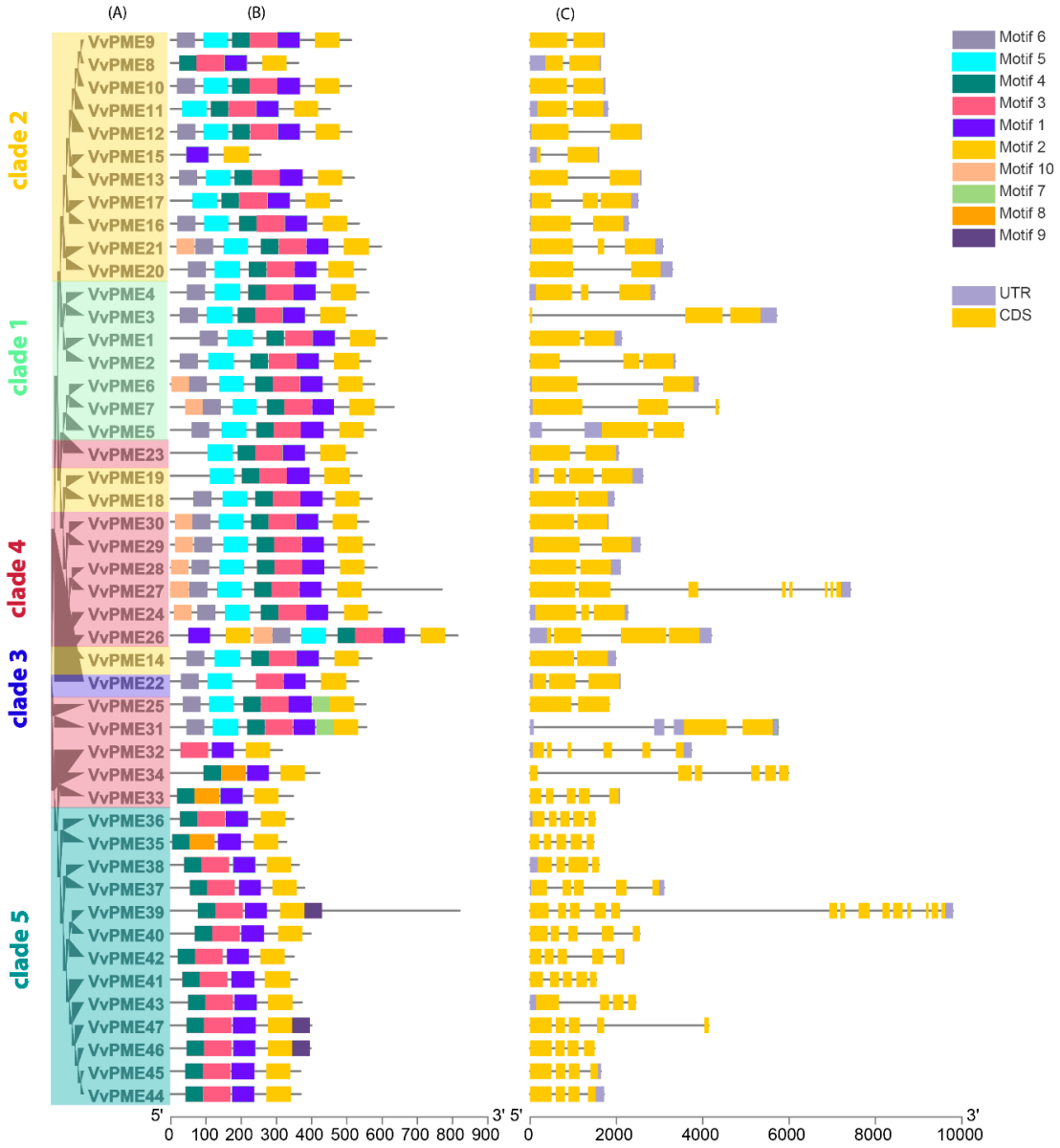


Figure 3. Phylogenetic relationship of *PMEs* (**A**). The phylogenetic tree was constructed by MEGA 7.0 using the Maximum Likelihood Method (1000 bootstrap). Motif structure and upstream/downstream regions of *PMEs* (**B**). The coding sequences (CDS) and untranslated regions (UTR) for PMEs in grapevine (**C**). CDS and UTR are represented by yellow and green boxes, respectively. The relative position is proportionally displayed based on the kilobase scale at the bottom of the figures.

2.3. Chromosomal Localization and Gene Duplication Analysis of PG and PME Genes

A total of 36 *PG* genes were distributed unevenly across different chromosomal locations of grapevine genomes (i.e., Chr01–Chr19). The majority of the chromosomes of PGs showed inconsistency in terms of genes. Both Chr05 and Chr08 exhibited the highest number (8) of genes, followed by Chr1 and Chr07 with five genes. The others varied in number (Figure 4a and Figure S3). Moreover, the chromosomal localization for PME members also displayed high variation in the number of genes. The highest number of genes (seven) was found on both Chr11 and Chr16 each, followed by Chr5 containing five genes. The others largely varied from 1–4 per chromosome (Figure 4b and Figure S3). In majority, we observed non-random distribution patterns of *PG* and *PME* genes in the grapevine genome. The 36 *PG* and 47 *PME* genes were also clustered for collinearity between

grapevine and *Arabidopsis* using Circos (Figure 4a,b). The results illustrated high conservation among PME members compared to PGs.

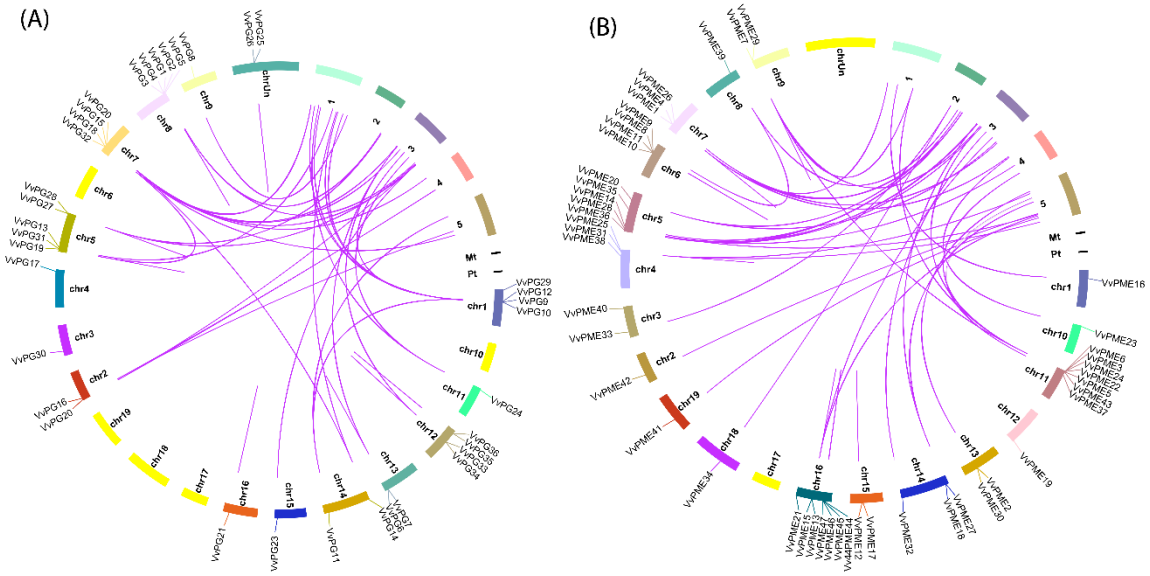


Figure 4. The collinear correlation for all genes of PGs (**A**) and PMEs (**B**) is displayed between grapevines and *Arabidopsis*. The localization of chromosomes was shown for grapevine and *Arabidopsis* in different random colors.

To study evolutionary rates and types of duplications among *PG* and *PME* genes in grapevines, we used MEGA7.0 and MCScanX. Among the 36 PGs, we identified 10 dispersed, 1 proximal, 12 tandem, and 13 segmental genes. Furthermore, we determined 23 dispersed, 1 proximal, 16 tandem, and 7 segmental genes in the 47 PMEs (Table 1). As gene duplications are vital for discovering novel biological functions, evolutions, and gene expansion [30], the segmental and dispersed duplications observed may play a major role in the expansion of PG and PME members.

To estimate the selection pressure among various types of duplications for both PGs and PMEs, we also intended their synonymous (*Ks*) and non-synonymous substitution rate (*Ka*) values. During evolutionary implications, genes are typically exposed to different types of selection processes (i.e., positive, neutral, and purifying selection). For understanding these selection pressures, we selected 15 duplicated pairs of genes among PGs and PMEs (Table 1). The *Ka/Ks* ratios for most PGs and PMEs were less than one, implicating a purifying selection and a reduction in divergence after duplications. However, we also found four pairs of PGs (i.e., *VvPG29-VvPG30*, *VvPG5-VvPG7*, *VvPG14-VvPG15*, and *VvPG18-VvPG21*) and five pairs of PMEs (*VvPME5-VvPME16*, *VvPME22-VvPME23*, *VvPME35-VvPME36*, *VvPME6-VvPME7*, and *VvPME14-VvPME26*) that indicate positive selection.

2.4. Gene Ontology Enrichment (GO) and Cis-Regulatory Elements in Grapevine

To study the regulatory functions of the 36 PGs and 47 PMEs, we performed GO annotation and GO enrichment analyses. The GO terms were largely based on three groupings, including molecular functions (MF), cellular component (CC), and biological process (BP). In brief, GO enrichments validate that PGs are enriched in various MF terms, such as "polygalacturonase activity" (GO:0004650), "hydrolase activity" (GO:0016787), and "hydrolyzing O-glycosyl compounds" (GO:0016798). The term CC was enriched in the plant-type cell wall, such as "integral component of membrane" (GO:0016021), and "extracellular region" (GO:0005576). The BP term was mainly responsive, such as "carbohydrate metabolic processes" (GO:0005975) and various other "metabolic processes" (GO:0008152), which are briefly listed in Table S3. Moreover, the PME results reveal that

MF is enriched in "enzyme inhibitor activity" (GO:0004857) "pectinesterase activity" (GO:0030599), "aspartyl esterase" (GO:0045330). The CC and BP were also found to be enriched in "membrane" (GO:0016020), "proteolysis" (GO:0006508), and "cell wall modification" (GO:0042545). Our results from the GO enrichment analysis further hinted the role of PG and PME members in grapevine.

The 36 *PG* and 47 *PME* genes were also tested for pathway enrichment analysis using the KEGG database, where the results showed enrichment in three major pathways (Table S4). These pathways include "carbohydrate metabolism" followed by "metabolism" and "pentose and glucuronate interconversions" in grapevines.

In addition, we also observed cis-acting elements by utilizing the promoter regions of both PG and PME members using the PlantCARE database. The various types of cis-regulatory elements were analyzed and are described in Figure 5 and Table S5. In brief, the majority of the genes participated in numerous signaling pathways, such as phytohormones, biotic-abiotic and other regulatory stress factors. For instance, 27.61% of the genes of PGs and PMEs were responsive to light regulations (e.g., GTI-motif, G-Box, GATA-motif, and others), followed by phytohormones (25.65%) (CGTCA, TGACG, ABRE). Other observed key regulatory elements include TC-Rich repeats, and HD-ZIP 3, which were reactive to defense stress and protein binding, respectively. These results inferred that the *PG* and *PME* genes have diverse gene functions and are indirectly involved in various biotic-abiotic/hormone signaling.

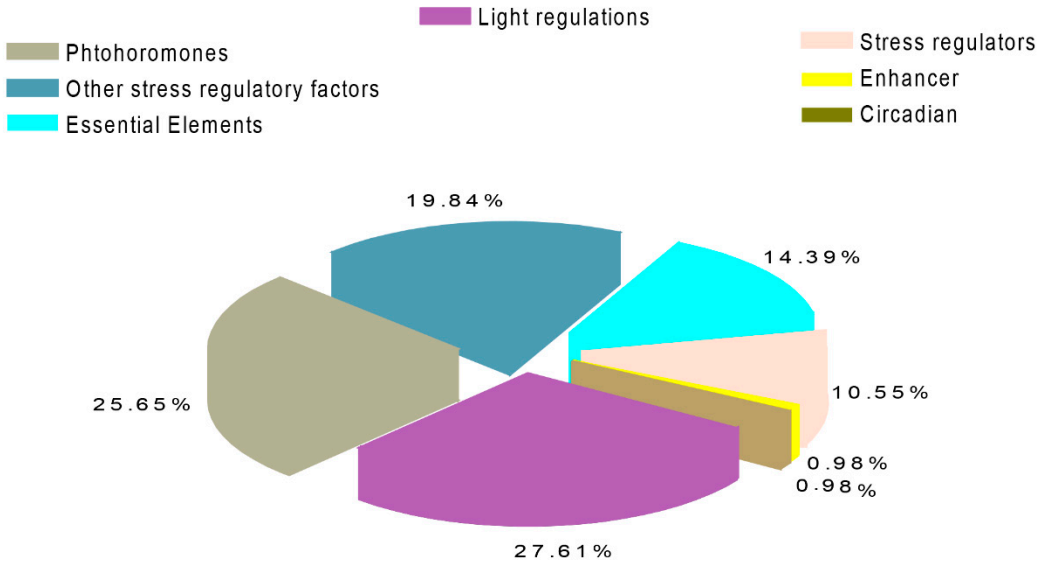


Figure 5. Various cis-elements identified in grapevine by using PlantCARE.

2.5. Tanscriptional Profiling of PGs and PMEs in Different Organs and Developmental Stages in Grapevine

To understand the spatiotemporal expression levels of *PG* and *PME* genes in grapevines, the global transcriptomic data of developmental phases of 19 different tissues and organs were retrieved from NCBI (GSE36128) [31]. Figure 6a,b represent the heat maps, indicating expression patterns of PGs and PMEs in grapevines. Among PGs, *VvPG8, VvPG10, VvPG13, VvPG17, VvPG18, VvPG19, VvPG20, VvPG22,* and *VvPG23* showed tissue- or organ-specific expression across many tissues during development. In contrast, the remaining PGs demonstrated weak tissue-specific response in any of the selected grapevine organs. Likewise, in PMEs, *VvPME3, VvPME4, VvPME5, VvPME6, VvPME19, VvPME21, VvPME23, VvPME29, VvPME31*, and *VvPME32* suggested higher tissue-specific response in all the tissues. In contrast, the others showed either moderate to weak expression or no expression in any of the grapevine tissues (Figure 6b). Overall the *PG* and *PME* genes showed enriched expression in flower (*VvPG6-7*, *VvPG31-34*, *VvPME-28*) and fruit ripening (*VvPG4-5*, *VvPME7-8*, and *VvPME44-45*), where the PGs suggested a more profound response than PMEs.

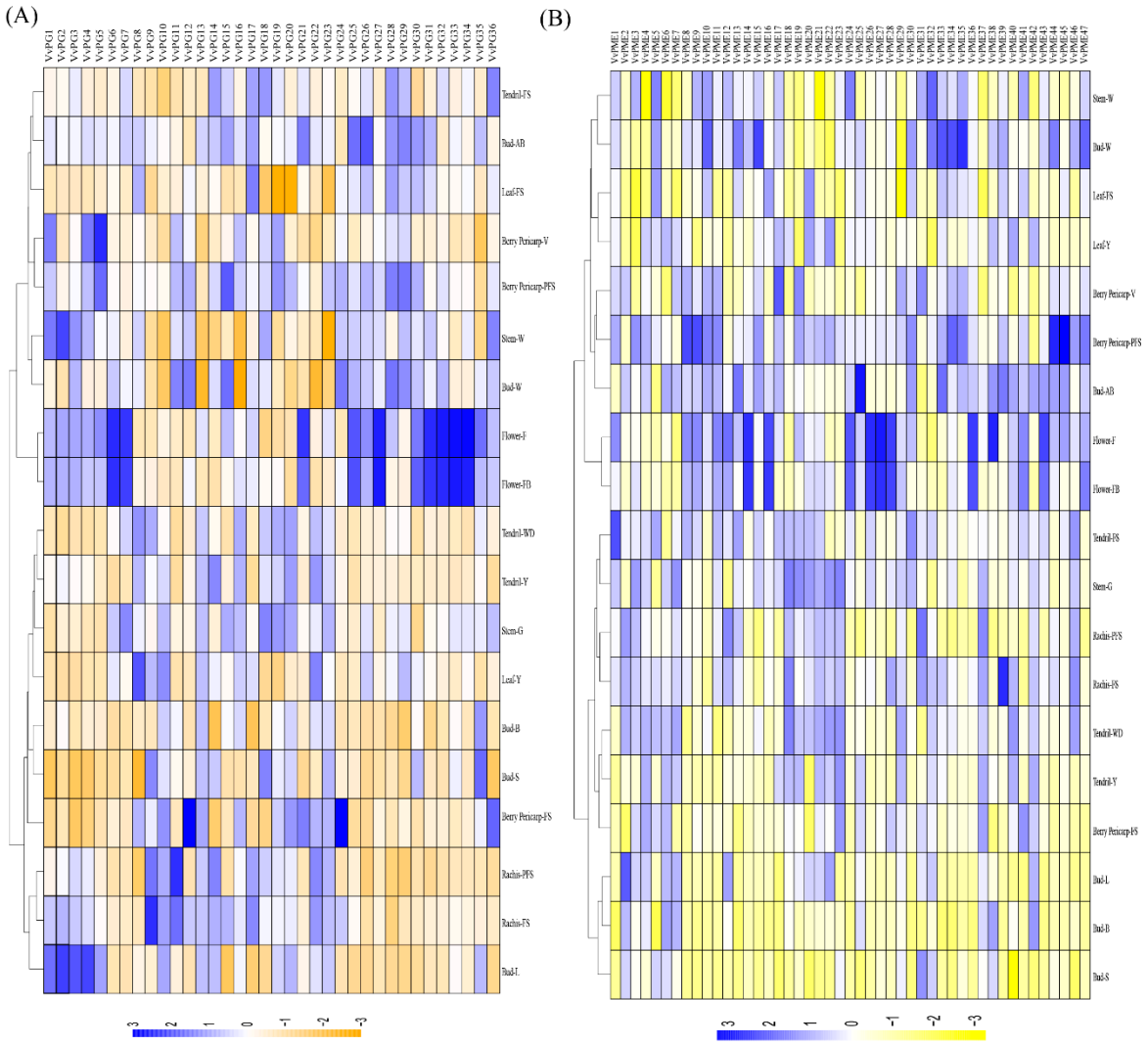


Figure 6. Expression profiles of the *PG* (**A**) and *PME* (**B**) genes in different grapevine organs, tissues, and, developmental stages. Data were normalized based on the mean expression values of each gene in all analyzed tissues. BerryPericarp-FS: berry pericarp fruit set; BerryPericarp-PFS: berry pericarp post-fruit set; BerryPericarp-V: Bud-S: bud swell; Bud-B: bud burst (green tip); Bud-AB: bud after-burst (rosette of leaf tips visible); Bud-L: latent bud; Bud-W: winter bud; Flower-FB: flowering begins (10% caps off); Flower-F: flowering (50% caps off); Leaf-Y: young leaf (pool of leaves from shoot of 5 leaves); Leaf-FS: mature leaf (pool of leaves from shoot at fruit set); Rachis-FS: rachis fruit set; Rachis-PFS: rachis post fruit set; Stem-G: green stem; Stem-W: woody stem; Tendril-Y: young tendril (pool of tendrils from shoot of 7 leaves); Tendril-WD: well developed tendril (pool of tendrils from shoot of 12 leaves); and Tendril-FS: mature tendril (pool of tendrils at fruit set).

2.6. qRT-PCR Analysis of the Candidate PG and PME Genes in Various Organs of Grapevine

To validate previous findings based on RNA-seq data, cis-element predications and gene duplication analysis, we randomly selected 32 genes for qRT-PCR, which showed either positive or purifying selection. The expression profiling of *PG* and *PME* genes was quantified in various organs, such as root, stem, tendril, inflorescence, berry flesh, berry skin, and leaf of grapevines (Figure 7a,b). The results suggest that PG and PME transcripts show distinct expression patterns, intimating that both (*PG* and *PME*) gene families have positive regulatory roles in various physiological processes in grapevine. Moreover, principal component analysis (PCA) analysis was performed to gain deeper insight into their contribution to organ development. PCA analysis of PG transcripts suggested a variation of 31.26% in PC1, 24.28% in PC2, and 15.36% in PC3, which accounted for 70.90% of the

total variation in the first three axes (Figure 7c). Among PGs, *VvPG31* (0.86) and *VvPG5* (0.85) had high positive loadings, while *VvPG21* (−0.90) and *VvPG8* (−0.77) had high negative loadings in PC1. Moreover, PCA analysis of PME transcripts suggested 83.02% in first three axes (PC1, PC2, and PC3) (Figure 7d). Among the PME transcripts, *VvPME5* (0.73) and *VvPME14* (0.61) had high positive loadings in PC1, whereas the highest negative loadings were found in *VvPME16* (−0.96), *VvPME36* (−0.82) and *VvPME1* (−0.79) (Table S6).

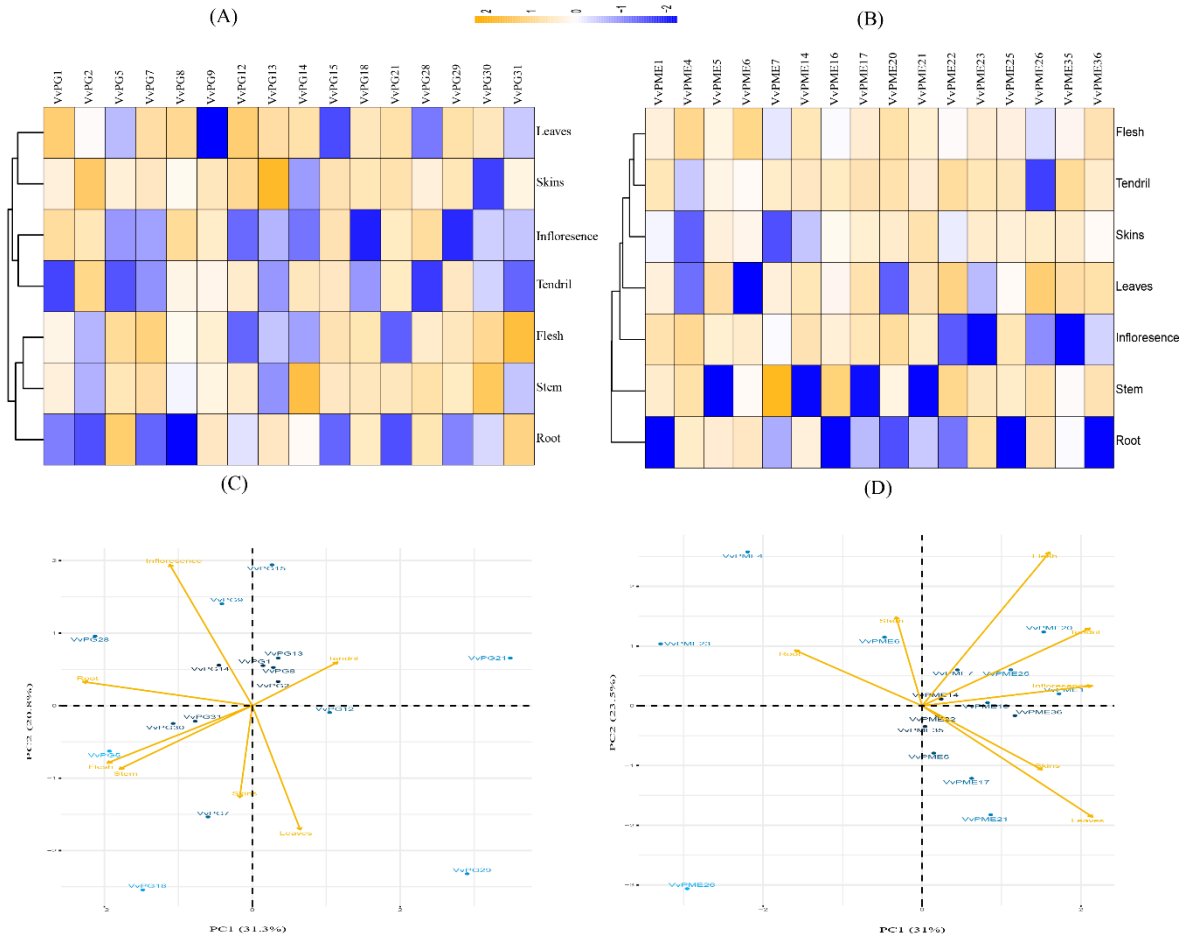


Figure 7. Relative expressions of PGs (**A**) and PMEs (**B**) in various organs, including root, stem, tendril, inflorescence, flesh, skins, and leaves and their principal component analysis for PGs (**C**) and PMEs (**D**).

3. Discussion

Realizing the significant role of PGs and PMEs in various plants observed in past studies, it is essential to systematically investigate the potential functions of these genes in grapevine. For instance, polygalacturonases (PGs) and pectin methylesterases (PMEs) have been hypothesized to play an imperative part in plant life cycles, such as cell separation and expansion, dehiscence, abscission, fruit maturity, and plant shedding [8,21,24]. In particular, PGs are a vital component of pectin disassembly, whereas PMEs play a central role in both remodeling and pectin disassembly [19,32]. Thus far, PGs and PMEs have been identified in various crops species although there is a lack of systemic analysis in grapevines. In this study, we comprehensively carried out various bioinformatics analyses by utilizing the available genomic resources of grapevines. In total, 36 *PG* and 47 *PME* genes were identified in grapevines and compared with *Arabidopsis*. For these genes, we also analyzed physicochemical properties, phylogenetic and collinearity relationships, chromosomal localization, motif and gene structure compositions, and duplication analysis. Additionally, the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomics (KEGG) enrichment, cis-regulatory elements,

and expression dynamics among various organs of grapevine have revealed extensive information related to the gene functions and their role in plant development. Subcellular predictions for maximum members of genes were largely found in diverse organelles, such as the nucleus, endoplasmic reticulum, cytoplasm, plasma membrane, mitochondria, and others. The physicochemical and protein properties (i.e., protein length (aa), molecular weight (kDa), isoelectric point (PIs), and grand average of hydropathicity (GRAVY)) drastically varied among PGs and PMEs, suggesting their variable role in micro and macro environments [33].

The selection pressure analysis (i.e., purifying, positive, and neutral selection) of gene pairs provides valuable information using the rate of divergence [34]. During evolutionary events, the values of *Ka/Ks* ratio that are less than 1.00 signify purifying selection; values equal to 1.00 specify neutral selection; and values greater than 1.00 shows positive selection [35,36]. In this study, we observed a strong selective pressure (purifying selection) among PGs compared to PMEs, which is in disagreement with the previously reported study on *Brassica rapa* [25]. We infer that PGs might duplicate earlier for their need and survival, intimating their diverse and variable functions.

In this study, we also intended *Ka/Ks* values among 15 pairs of both PGs and PMEs (i.e., tandem, dispersed, and segmental) using the MEGA7.0 software [37]. Most of the pairs had *Ka/Ks* ratios that were smaller than 1.00, inferring the purifying selection. Furthermore, only four and five pairs of PGs and PMEs have values greater than 1.00, signifying positive selection. During evolutionary processes, the higher plant underwent polyploidization events, while segmental duplications are usually responsible for larger functional divergence [38,39]. Hence, studying gene duplication is vital for understanding biological functions and expansion of gene families [30,40], Hence, the results from our study highlight the importance of segmental and dispersed duplications in the two large families of PGs and PMEs in grapevines.

Transcript expression patterns and abundance in particular organs at a given time provide clues for understanding the function of genes [28]. Transcriptional profiling and functional characterization of PGs and PMEs has been reported in several species. For example, the *PG* genes found in various species, including *Glycine max*, *Medicago truncatula*, *Zea mays,* and *O. sativa*, showed higher expression levels [41]. In strawberries, the downregulation of *FaPG1* extended their post-harvest life by reducing fruit softening [15]. Previously, PG gene family transcription patterns were also observed in *Arabidopsis* and *Populus* [24,42]. In addition, the members of the *PME* gene family have been reported as key regulators in plant–microbe interactions, cold acclimation, drought, and salt stress sensitivity [43–47]. Likewise, in grapevines, the members of the *PG* and *PME* gene family regulate the development of numerous organs at varying stages. In this study, 19 diverse tissue-specific expression patterns of *PG* and *PME* genes were examined. The results revealed that most PGs and PMEs were highly expressed in berry, tendril, and inflorescence. However, few PGs and PMEs reveal either more or similar expression patterns, signifying their unifying need and importance in plant development. In addition, qRT-PCR validation of 16 *PG* and *PME* genes revealed their vital role in various organs of grapevines (i.e., root, stem, tendril, inflorescence, flesh, skins, and leaves). Transcriptional profiling of these genes in various tissues may consequently aid the study of new adaptive functions regarding plant developmental processes in grapevine. Notably, among various *PG* and *PME* genes, visible tissue-specific expression patterns were detected, which may be correlated with their expansion, evolution and the complex nature in plant growth and development. However, their underlying molecular and evolutionary mechanisms that lead to the measurable and abrupt structural differences must be extensively investigated in the future.

Finally, GO and KEGG enrichment, and cis-element predictions in the promoter regions of PGs and PMEs revealed their key role in pectin and carbohydrate metabolism, and various stress-related activities in grapevine. Taken together, our results highlight the importance of PGs and PMEs in plants and provides a comprehensive overview of their developmental role in grapevine.

4. Materials and Methods

4.1. Mining of Grapevine PGs and PMEs

For identification of *PG* and *PME* genes in grapevine (genome version 2.1), we used BioEdit tools to obtain PGs and PMEs from all the reference sequences of *Arabidopsis*. Grapevine and *Arabidopsis* genomic sequences were retrieved from Ensembl (https://plants.ensembl.org/index.html) and TAIR (http://www.arabidopsis.org/). The sequences of other species, including apple, peach, and citrus, were downloaded from Phytozome v12.1.6 (https://phytozome.jgi.doe.gov/pz/portal.html) [48]. For domain composition analysis, we used NCBI-Conserved Domain database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and SMART databases (http://smart.embl-heidelberg.de/) [49]. In cases where PGs and PMEs domains were absent, the protein sequences were removed from the study and sequences with errors in length or having <100 aa length were also removed before analysis.

4.2. Phylogenetic Analysis of PGs and PMEs

The amino acid sequences of PGs and PMEs were aligned using MUSCLE [50] implemented in MEGA 7.0 software [51]. The phylogenetic trees were constructed using the maximum likelihood (ML) method in MEGA 7.0. In order to determine the reliability of the resulting trees, bootstrap values of 1000 replications were performed with the Jones, Taylor, and Thornton amino acid substitution model (JTT model).

4.3. Ratio of Synonymous (Ks) and Non-synonymous (Ka) for duplicated genes

The *Ka/Ks* ratios were calculated for duplicated pairs (i.e., tandem, dispersed, and segmental) using MEGA 7.0 [51]. The *Ka* and *Ks* substitution rates were calculated with the standard genetic code table by the Nei–Gojobori method (Jukes-Cantor model) in MEGA 7.0.

4.4. Gene Structure, Conserved Motifs Analysis, and Physicochemical Parameters of PG and PME Proteins

The gene structure was illustrated by TBtools software [52] by utilizing the GFF3 file of the grapevine genome. The conserved motif scanning of PG and PME proteins was carried out through local MEME Suite (Version 5.0.5) and was visualized by TBtools software. For this purpose, parameter settings were calibrated as follows: a maximum number of motifs of 10, with a minimum and maximum width of 50 and 100. The other parameters were set at default values [53]. The physicochemical properties of the PG and PME proteins (i.e., molecular weight (MW), isoelectronic points (PIs), aliphatic index and GRAVY values for each gene) were calculated using the ExPASY PROTPARAM tools (http://web.expasy.org/protparam/). The subcellular localization was predicted using the WOLF PSORT (https://wolfpsort.hgc.jp/) website.

4.5. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomics (KEGG) and Cis-Elements Predictions of PGs and PMEs

The GO enrichment was carried using an online panther server (http://pantherdb.org/) and TBtools software [52]. KEGG enrichment analysis was carried out by the online server (https://www.genome.jp/kegg/pathway.html) and their enriched pathways were further analyzed by TBtools software [52]. The promoter sequences of PGs and PMEs (i.e., selected as 1500 bp) were imported in Generic File Format (GFF) file from the grapevine genome. Subsequently, the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [54] was utilized for identifying various cis-regulatory elements for each promoter sequence of PGs and PMEs.

4.6. Chromosomal Location and Collinearity Analysis

The chromosomal locations of PGs and PMEs were mapped based on information available at the Grape Genome Database (CRIBI. Available online: http://genomes.cribi.unipd.it/grape/, V2.1).

and were illustrated using TBtools software [52]. For collinearity analysis, the relationships between grapevine and *Arabidopsis* homologs were verified and visualized by the Circos tool in TBtools software.

4.7. Principal Component Analysis (PCA)

The principal component analysis was implemented using Rstudio (R program) for qRT-PCR at a significance level of 0.05 (p-value) [37,55].

4.8. Plant Material and Methods

Six-year-old *V. vinifera* cv. Summer Black plants grown under standard field conditions were selected from Jiangsu Academy of Agricultural Sciences (JAAS), Nanjing-China. In brief, the 4th unfolded leaf was selected for tissue extraction. Various grapevine tissues/organs, such as the root, stem, tendril, inflorescence, flesh, and skin, were collected at different developmental stages. Tissue samples were immediately frozen in liquid nitrogen and stored at −80 °C for further use and RNA extraction.

4.9. RNA Isolation and Expression Profiling of PGs and PMEs in Grapevine

Total RNA was extracted from various organs using Trizol (Invitrogen, Carlsbad, CA, USA), following the manufacturer's instructions. RNA was reverse-transcribed into cDNA using the Primer Script RT reagent kit (TAKARA, Dalian, China) according to the manufacturer's instructions. Specific primers were designed using Becan Designer 7.9, and are presented in Table S7. In order to check the specificity of the primers, the BLAST tool was used against the grapevine genome for confirmation. RT-PCR was performed according to the guidelines of previous studies [56,57]. The relative fold expression was calculated with the comparative Ct-method. The expression patterns of all *PG* and *PME* genes were analyzed based on a previous study [58,59]. The housekeeping and grapevine actin gene (AB073011 and XM_010659103) was used as the reference gene for qRT-PCR.

In brief, the RT-PCR amplification reactions were performed on an ABI 7500 RT-PCR System (Applied Biosystems, CA, USA) using SYBR Green (Applied Biosystems, Carlsbad, CA, USA) with three replicates. PCR was conducted as follows: denaturation at 95 °C for 2 min, 40 cycles of denaturation at 95 °C for 10 s, annealing at 60 °C for 40 s, and extension at 72 °C for 15 s, followed by melting curve analysis (61 cycles at 65 °C for 10 s).

Transcriptomic data were utilized for various organs and developmental stages from NCBI GEO server (https://www.ncbi.nlm.nih.gov/geo/) under the series entry GSE36128. Additionally, gene expression levels were quantified by FPKM (fragments per kilobase of transcript per million fragments mapped), and heat maps were visualized by using Rstudio (R program, Boston, MA, USA).

5. Conclusions

In conclusion, we systematically carried out a genome-wide exploration of grapevines through various bioinformatic analyses, which include elucidating the physicochemical properties of PGs and PMEs, phylogenetic characterization, collinearity of PGs and PMEs, gene structure and motif composition, evolutionary rates, and gene duplications. The GO and KEGG enrichment, and cis-elements prediction analysis extended our repositories on the putative functions of PGs and PMEs in plant developments during pectin and carbohydrate metabolism, and various stress-related activities in grapevine. Additionally, expression profiling of various organs during developmental stages and their correlation by principal component analysis highlights the essential role of PGs and PMEs for plant improvements in grapevines.

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/13/3180/s1.

Author Contributions: Conceptualization, methodology, performed the experiments, N.K.; managed the research work, writing original draft, figures, review and editing, N.K., F.F., and M.S.H.; software, H.S.; validation, Z.L. and T.Z.; and supervision, funding acquisition, project administration, J.F.

Funding: This work was supported by the funding of National Key R&D Projects 2018YFD1000200, Jiangsu Agricultural Science and Technology Independent Innovation Project CX (18)2008, National Natural Science Foundation of China (31672131, 31801809).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

PGs	Polygalacturonase
PMEs	Pectin methylesterase
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
RT-PCR	Real time-PCR
PCA	Principal Component Analysis

References

1. Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; et al. Toward a systems approach to understanding plant cell walls. *Science* **2004**, *306*, 2206–2211. [CrossRef] [PubMed]
2. Keegstra, K. Plant cell walls. *Plant Physiol.* **2010**, *154*, 483–486. [CrossRef] [PubMed]
3. Seymour, G.B.; Orchard, J.; Marín-Rodríguez, M.C. Pectate lyases, cell wall degradation and fruit softening. *J. Exp. Bot.* **2002**, *53*, 2115–2119.
4. Peng, G.; Wu, J.; Lu, W.; Li, J. A polygalacturonase gene clustered into clade e involved in lychee fruitlet abscission. *Sci. Hortic.* **2013**, *150*, 244–250. [CrossRef]
5. Elliott, K.A.; González-Carranza, Z.H.; Roberts, J.A. Expression of polygalacturonases and evidence to support their role during cell separation processes in *Arabidopsis thaliana*. *J. Exp. Bot.* **2007**, *58*, 3719–3730.
6. Ogawa, M.; Kay, P.; Wilson, S.; Swain, S.M. *Arabidopsis* dehiscence zone polygalacturonase1 (adpg1), adpg2, and quartet2 are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant Cell* **2009**, *21*, 216–233. [CrossRef] [PubMed]
7. Atkinson, R.G.; Schröder, R.; Hallett, I.C.; Cohen, D.; MacRae, E.A. Overexpression of polygalacturonase in transgenic apple trees leads to a range of novel phenotypes involving changes in cell adhesion. *Plant Physiol.* **2002**, *129*, 122–133. [CrossRef] [PubMed]
8. Rose, J.K.C.; Catalá, C.; Gonzalez-Carranza, Z.H.; Roberts, J.A. Cell wall disassembly. *Annu. Plant Rev.* **2018**, *8*, 264–324.
9. Torki, M.; Mandaron, P.; Thomas, F.; Quigley, F.; Mache, R.; Falconet, D. Differential expression of a polygalacturonase gene family in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **1999**, *261*, 948–952.
10. Sander, L.; Child, R.; Ulvskov, P.; Albrechtsen, M.; Borkhardt, B. Analysis of a dehiscence zone endo-polygalacturonase in oilseed rape (*Brassica napus*) and *Arabidopsis thaliana*: Evidence for roles in cell separation in dehiscence and abscission zones, and in stylar tissues during pollen tube growth. *Plant Mol. Biol.* **2001**, *46*, 469–479. [CrossRef]
11. Chen, H.; Shao, H.; Fan, S.; Ma, J.; Zhang, D.; Han, M. Identification and phylogenetic analysis of the polygalacturonase gene family in apple. *Hortic. Plant J.* **2016**, *2*, 241–252. [CrossRef]
12. Wakasa, Y.; Kudo, H.; Ishikawa, R.; Akada, S.; Senda, M.; Niizeki, M.; Harada, T. Low expression of an endopolygalacturonase gene in apple fruit with long-term storage potential. *Postharvest Biol. Technol.* **2006**, *39*, 193–198. [CrossRef]
13. Sekine, D.; Munemura, I.; Gao, M.; Mitsuhashi, W.; Toyomasu, T.; Murayama, H. Cloning of cdnas encoding cell-wall hydrolases from pear (Pyrus *communis*) fruit and their involvement in fruit softening and development of melting texture. *Physiol. Plant.* **2006**, *126*, 163–174. [CrossRef]
14. Deytieux-Belleau, C.; Vallet, A.; Doneche, B.; Geny, L. Pectin methylesterase and polygalacturonase in the developing grape skin. *Plant Physiol. Biochem.* **2008**, *46*, 638–646. [CrossRef] [PubMed]
15. Pose, S.; Paniagua, C.; Cifuentes, M.; Blanco-Portales, R.; Quesada, M.A.; Mercado, J.A. Insights into the effects of polygalacturonase fapg1 gene silencing on pectin matrix disassembly, enhanced tissue integrity, and firmness in ripe strawberry fruits. *J. Exp. Bot.* **2013**, *64*, 3803–3815. [CrossRef] [PubMed]

16. Hiwasa, K.; Kinugasa, Y.; Amano, S.; Hashimoto, A.; Nakano, R.; Inaba, A.; Kubo, Y. Ethylene is required for both the initiation and progression of softening in pear (*Pyrus communis* L.) fruit. *J. Exp. Bot.* **2003**, *54*, 771–779. [CrossRef]
17. Asif, M.H.; Nath, P. Expression of multiple forms of polygalacturonase gene during ripening in banana fruit. *Plant Physiol. Biochem.* **2005**, *43*, 177–184. [CrossRef]
18. Atmodjo, M.A.; Hao, Z.; Mohnen, D. Evolving views of pectin biosynthesis. *Annu. Rev. Plant Biol.* **2013**, *64*, 747–779. [CrossRef]
19. Pelloux, J.; Rusterucci, C.; Mellerowicz, E.J. New insights into pectin methylesterase structure and function. *Trends Plant Sci.* **2007**, *12*, 267–277. [CrossRef]
20. Kagan-Zur, V.; Tieman, D.M.; Marlow, S.J.; Handa, A.K. Differential regulation of polygalacturonase and pectin methylesterase gene expression during and after heat stress in ripening tomato (*Lycopersicon esculentum* mill.) fruits. *Plant Mol. Biol.* **1995**, *29*, 1101–1110. [CrossRef]
21. Bosch, M.; Cheung, A.Y.; Hepler, P.K. Pectin methylesterase, a regulator of pollen tube growth. *Plant Physiol* **2005**, *138*, 1334–1346. [CrossRef]
22. Almeida, D.P.F.; Huber, D.J. In vivo pectin solubility in ripening and chill-injured tomato fruit. *Plant Sci.* **2008**, *174*, 174–182. [CrossRef]
23. Giovannoni, J.J.; DellaPenna, D.; Bennett, A.B.; Fischer, R.L. Expression of a chimeric polygalacturonase gene in transgenic rin (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. *Plant Cell* **1989**, *1*, 53–63. [PubMed]
24. Kim, J.; Shiu, S.H.; Thoma, S.; Li, W.H.; Patterson, S.E. Patterns of expansion and expression divergence in the plant polygalacturonase gene family. *Genome Biol.* **2006**, *7*, R87. [CrossRef]
25. Duan, W.; Huang, Z.; Song, X.; Liu, T.; Liu, H.; Hou, X.; Li, Y. Comprehensive analysis of the polygalacturonase and pectin methylesterase genes in *Brassica rapa* shed light on their different evolutionary patterns. *Sci. Rep.* **2016**, *6*, 25107. [CrossRef] [PubMed]
26. Li, W.; Shang, H.; Ge, Q.; Zou, C.; Cai, J.; Wang, D.; Fan, S.; Zhang, Z.; Deng, X.; Tan, Y.; et al. Genome-wide identification, phylogeny, and expression analysis of pectin methylesterases reveal their major role in cotton fiber development. *BMC Genom.* **2016**, *17*, 1000. [CrossRef]
27. Ke, X.; Wang, H.; Li, Y.; Zhu, B.; Zang, Y.; He, Y.; Cao, J.; Zhu, Z.; Yu, Y. Genome-wide identification and analysis of polygalacturonase genes in *Solanum lycopersicum*. *Int. J. Mol. Sci.* **2018**, *19*, 2290. [CrossRef]
28. Yu, Y.; Liang, Y.; Lv, M.; Wu, J.; Lu, G.; Cao, J. Genome-wide identification and characterization of polygalacturonase genes in cucumis sativus and citrullus lanatus. *Plant Physiol. Biochem.* **2014**, *74*, 263–275. [CrossRef]
29. Grimplet, J.; Adam-Blondon, A.F.; Bert, P.F.; Bitz, O.; Cantu, D.; Davies, C.; Delrot, S.; Pezzotti, M.; Rombauts, S.; Cramer, G.R. The grapevine gene nomenclature system. *BMC Genom.* **2014**, *15*, 1077. [CrossRef]
30. Hurles, M. Gene duplication: The genomic trade in spare parts. *PLoS Biol.* **2004**, *2*, E206. [CrossRef]
31. Fasoli, M.; Dal Santo, S.; Zenoni, S.; Tornielli, G.B.; Farina, L.; Zamboni, A.; Porceddu, A.; Venturini, L.; Bicego, M.; Murino, V.; et al. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* **2012**, *24*, 3489–3505. [CrossRef] [PubMed]
32. Hadfield, K.A.; Bennett, A.B. Polygalacturonases: Many genes in search of a function. *Plant Physiol* **1998**, *117*, 337–343. [CrossRef] [PubMed]
33. Khan, N.; Hu, C.M. Genome-wide identification, classification, and expression divergence of glutathione-transferase family in *Brassica rapa* under multiple hormone treatments. *BioMed Res. Int.* **2018**, *2018*, 6023457. [CrossRef] [PubMed]
34. Juretic, N.; Hoen, D.R.; Huynh, M.L.; Harrison, P.M.; Bureau, T.E. The evolutionary fate of mule-mediated duplications of host gene fragments in rice. *Genome Res.* **2005**, *15*, 1292–1297. [CrossRef] [PubMed]
35. Li, J.; Zhang, Z.; Vang, S.; Yu, J.; Wong, G.K.-S.; Wang, J. Correlation between ka/ks and ks is related to substitution model and evolutionary lineage. *J. Mol. Evol.* **2009**, *68*, 414–423. [CrossRef] [PubMed]
36. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [CrossRef] [PubMed]
37. Shazadee, H.; Khan, N. Identification and expression profiling of protein phosphatases (pp2c) gene family in *Gossypium hirsutum* L. *Int. J. Mol. Sci.* **2019**, *20*, 1395. [CrossRef] [PubMed]

38. Moghe, G.D.; Shiu, S.H. The causes and molecular consequences of polyploidy in flowering plants. *Ann. N. Y. Acad. Sci.* **2014**, *1320*, 16–34. [CrossRef] [PubMed]
39. Prince, V.E.; Pickett, F.B. Splitting pairs: The diverging fates of duplicated genes. *Nat. Rev. Genet.* **2002**, *3*, 827–837. [CrossRef] [PubMed]
40. Haider, M.S.; Khan, N.; Pervaiz, T.; Zhongjie, L.; Nasim, M.; Jogaiah, S.; Mushtaq, N.; Jiu, S.; Jinggui, F. Genome-wide identification, evolution, and molecular characterization of the pp2c gene family in woodland strawberry. *Gene* **2019**, *702*, 27–35. [CrossRef] [PubMed]
41. Liang, Y.; Yu, Y.; Cui, J.; Lyu, M.; Xu, L.; Cao, J. A comparative analysis of the evolution, expression, and cis-regulatory element of polygalacturonase genes in grasses and dicots. *Funct. Integr. Genom.* **2016**, *16*, 641–656. [CrossRef] [PubMed]
42. Yang, Z.L.; Liu, H.J.; Wang, X.R.; Zeng, Q.Y. Molecular evolution and expression divergence of the populus polygalacturonase supergene family shed light on the evolution of increasingly complex organs in plants. *New Phytol.* **2013**, *197*, 1353–1365. [CrossRef] [PubMed]
43. Chen, M.-H.; Citovsky, V. Systemic movement of a tobamovirus requires host cell pectin methylesterase. *Plant J.* **2003**, *35*, 386–392. [CrossRef] [PubMed]
44. Raiola, A.; Lionetti, V.; Elmaghraby, I.; Immerzeel, P.; Mellerowicz, E.J.; Salvi, G.; Cervone, F.; Bellincampi, D. Pectin methylesterase is induced in *Arabidopsis* upon infection and is necessary for a successful colonization by necrotrophic pathogens. *Mol. Plant-Microbe Interact.* **2011**, *24*, 432–440. [CrossRef] [PubMed]
45. Solecka, D.; Zebrowski, J.; Kacperska, A. Are pectins involved in cold acclimation and de-acclimation of winter oil-seed rape plants? *Ann. Bot.* **2008**, *101*, 521–530. [CrossRef] [PubMed]
46. An, S.H.; Sohn, K.H.; Choi, H.W.; Hwang, I.S.; Lee, S.C.; Hwang, B.K. Pepper pectin methylesterase inhibitor protein capmei1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta* **2008**, *228*, 61–78. [CrossRef]
47. Mundaya, N.J.; Owen, S.D.W.; Iain, M.; Alan, T.C.; David, H.; Balakrishnan, P. Analysis of seaweed extract-induced transcriptome leads to identification of a negative regulator of salt tolerance in *Arabidopsis*. *Hortsci. Horts* **2012**, *47*, 704–709.
48. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef]
49. Letunic, I.; Doerks, T.; Bork, P. Smart 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **2012**, *40*, D302–D305. [CrossRef]
50. Edgar, R.C. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
51. Kumar, S.; Stecher, G.; Tamura, K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
52. Chen, C.; Xia, R.; Chen, H.; He, Y. Tbtools, a toolkit for biologists integrating various hts-data handling tools with a user-friendly interface. *bioRxiv* **2018**, 289660. [CrossRef]
53. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. Meme suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef] [PubMed]
54. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. Plantcare, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef] [PubMed]
55. Haider, M.S.; Jogaiah, S.; Pervaiz, T.; Yanxue, Z.; Khan, N.; Fang, J. Physiological and transcriptional variations inducing complex adaptive mechanisms in grapevine by salt stress. *Environ. Exp. Bot.* **2019**, *162*, 455–467. [CrossRef]
56. Khan, N.; Hu, C.-M.; Amjad Khan, W.; Naseri, E.; Ke, H.; Huijie, D.; Hou, X. Evolution and expression divergence of e2 gene family under multiple abiotic and phytohormones stresses in *Brassica rapa*. *Biomed Res. Int.* **2018**, *2018*, 18. [CrossRef] [PubMed]
57. Khan, N.; Hu, C.-M.; Khan, W.A.; Wang, W.; Ke, H.; Huijie, D.; Zhishuo, Z.; Hou, X. Genome-wide identification, classification, and expression pattern of homeobox gene family in *Brassica rapa* under various stresses. *Sci. Rep.* **2018**, *8*, 16265. [CrossRef] [PubMed]

58. Shangguan, L.; Fang, X.; Chen, L.; Cui, L.; Fang, J. Genome-wide analysis of autophagy-related genes (args) in grapevine and plant tolerance to copper stress. *Planta* **2018**, *247*, 1449–1463. [CrossRef]
59. Khan, N.; Ke, H.; Hu, C.-M.; Naseri, E.; Haider, M.S.; Ayaz, A.; Amjad Khan, W.; Wang, J.; Hou, X. Genome-wide identification, evolution, and transcriptional profiling of pp2c gene family in *Brassica rapa*. *Biomed Res. Int.* **2019**, *2019*, 15. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of the MADS-Box Transcription Factor Family in *Solanum lycopersicum*

Yunshu Wang [1,†], Jianling Zhang [1,†], Zongli Hu [1], Xuhu Guo [1], Shibing Tian [2] and Guoping Chen [1,*]

1 Laboratory of molecular biology of tomato, Bioengineering College, Chongqing University, Chongqing 400044, China; wangyunshu@cqu.edu.cn (Y.W.); zhangjianling0520@126.com (J.Z.); huzongli71@163.com (Z.H.); 20131901005@cqu.edu.cn (X.G.)

2 The Institute of Vegetable Research, Chongqing Academy of Agricultural Sciences, Chongqing 401329, China; tiansbing@aliyun.com

* Correspondence: chenguoping@cqu.edu.cn; Tel./Fax: +86-2365112674

† These authors contributed equally to this work.

Received: 10 May 2019; Accepted: 15 June 2019; Published: 18 June 2019

Abstract: MADS-box family genes encode transcription factors that are involved in multiple developmental processes in plants, especially in floral organ specification, fruit development, and ripening. However, a comprehensive analysis of tomato MADS-box family genes, which is an important model plant to study flower fruit development and ripening, remains obscure. To gain insight into the MADS-box genes in tomato, 131 tomato MADS-box genes were identified. These genes could be divided into five groups (Mα, Mβ, Mγ, Mδ, and MIKC) and were found to be located on all 12 chromosomes. We further analyzed the phylogenetic relationships among Arabidopsis and tomato, as well as the protein motif structure and exon–intron organization, to better understand the tomato MADS-box gene family. Additionally, owing to the role of MADS-box genes in floral organ identification and fruit development, the constitutive expression patterns of MADS-box genes at different stages in tomato development were identified. We analyzed 15 tomato MADS-box genes involved in floral organ identification and five tomato MADS-box genes related to fruit development by qRT-PCR. Collectively, our study provides a comprehensive and systematic analysis of the tomato MADS-box genes and would be valuable for the further functional characterization of some important members of the MADS-box gene family.

Keywords: tomato; MADS-box; genome-wide analysis; floral organ; fruit development

1. Introduction

The MADS-box family genes encode transcription factors (TFs), which are widely distributed in eukaryotes and play fundamental roles in diverse biological functions [1]. The name MADS-box is derived from the initials of four transcription factors that were first discovered of this family: MINICHROMOSOME MAINTENANCE 1 (MCM1), AGAMOUS (AG), DEFICIENS (DEF), and SERUM RESPONSE FACTOR (SRF) [2]. Their N-terminal contains a highly conserved DNA-binding MADS-domain containing 56–60 amino acids [2]. Thus, a protein encoding the MADS-domain is referred to as a MADS-box protein. MADS-box genes are divided into two groups (type I and type II) throughout the eukaryotes [3]. Type I MADS-box transcription factors can be further classified into four subclasses (Mα, Mβ, Mγ, and Mδ) in view of the M domain of the encoded protein, while only a few type I genes have been characterized for their biological function [4]. Type II MADS-box transcription factors include the Myocyte Enhancer Factor 2-like (MEF2-like) group from animals and yeast and the plant-specific MIKC-type group. The name of the MIKC-type protein is derived from their four characteristic domains: MADS-box (M), intervening (I), keratin-like (K), and C-terminal (C) [5].

The MIKC type has been further subdivided into $MIKC^C$ and MIKC* types based on phylogenetic relationships and structural features [4]. The functions of the type II genes, especially the $MIKC^C$ type MADS-box genes in plants, have been reported in more detail. To date, 39 $MIKC^C$ type MADS-box genes were found in *Arabidopsis thaliana* [6], and 38 $MIKC^C$ type MADS-box genes were found in rice [7].

MADS-box genes are known to play important roles in different aspects of plants growth and development. The MADS-box genes *AGAMOUS* (*AG*) of *Arabidopsis thaliana* [8] as well as *DEFICIENS* (*DEF*) of *Antirrhinum majus* [2] were found to regulate floral organ formation two decades ago. The MADS-box genes were thought to be the main participants in floral organ specificity. With the analysis of the specification of floral organ identity determination, a genetic model (ABC model) was proposed in which the different steps of floral development were determined by three classes of genes (A, B, and C) [9]. Class A genes are necessary for the formation of the sepal. The class A genes, together with class B genes, determine the development of petals. The combination of class B and C genes is necessary for stamen identity, and class C genes function alone to form carpels [10]. However, there are many phenomena that this model cannot explain. For example, the constitutive co-expression of the B class genes *PISTILLATA* (*PI*) and *APETALA3* (*AP3*) in *Arabidopsis thaliana* [11], does not change the identity of vegetative organs. Recent studies found that the ABC model is necessary but not sufficient to provide the floral organ identity function. Therefore, the model was advanced by subdivision into five different classes (A–E). Class D genes specify ovule development, while class E genes are necessary for the specification of petals, stamens, and carpels. In addition, these five classes of genes are mainly MADS-box genes [12]. In *Arabidopsis thaliana*, *APETALA1* (*AP1*) is a typical class A gene [13], *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) belong to the class B genes [14], and *AGAMOUS* (*AG*) is a representative gene with class C function [15]. The *SEPALLATA* (*SEP*) genes are class E genes and include *SEP1*, *SEP2*, *SEP3,* and *SEP4* in *Arabidopsis thaliana* [16]. The class D gene was identified by the mutant phenotype of petunia, and the sequence similarity analysis demonstrated that the corresponding gene in Arabidopsis thaliana is *AGAMOUS-LIKE11* (*AGL11*) [17,18].

In addition, considerable evidence has revealed that MADS-box TFs share a potent effect on the regulation of fruit development and ripening [4]. There have been many studies on the function of *Solanum lycopersicum* MADS-box genes in fruit development and ripening. The tomato MADS-box gene *RIPENING INHIBITOR* (*RIN*) is an essential factor for fruit ripening that regulates many ripening-associated processes, including both ethylene-dependent and ethylene-independent ripening processes [19]. Moreover, the tomato MADS-box genes *AGAMOUSLIKE1* (*AGL1*) [20,21], *FRUITFULL1* (*FUL1*), and *FRUITFULL2* (*FUL2*) are Arabidopsis *SHATTERPROOF* (*SHP*) and *FRUITFUL* (*FULL*) homologues respectively [22,23], and their suppression results in a phenotype that is partially similar to the ripening-defective phenotype of *RIN* mutant fruits. In addition to crucial roles in the regulation of plant reproductive development, MADS-box genes have also been shown to take part in plant vegetative growth processes and some stress responses in different plants such as Arabidopsis [24], rice [7,25], wheat [26], and Chinese cabbage [27]. Thus, the MADS-box protein family is an important TFs family for plant growth and development that almost affects the whole process of plant growth and development, especially plant reproductive development.

The MADS-box family of model plant species has been widely studied, including snapdragon (*Antirrhinum majus*) [28], rice (*Oryza sativa)* [7], Chinese cabbage (*Brassica rapa*) [27], poplar (*Populus trichocarpa*) [29], bread wheat (*Triticum aestivumL*) [30], banana (*Musa acuminata*) [31], petunia (*Petunia hybrida*) [32], apple (*Malus domestica Borkh)* [33], and so on. Tomato (*Solanum lycopersicum*) is one of the most critical model plants for studying flower and fruit development, and the MADS-box genes of tomato were among the earliest to be investigated [34]. It has been reported that 24 tomato $MIKC^C$-type MADS-box genes had been identified in 2006 [35]. However, there has been no report to date concerning systematic information for the tomato MADS-box gene family, and confusion in the names of these genes is problematic for future research.

To obtain a genome-wide analysis of the tomato MAD-box gene family, we sorted 131 MADS-box genes from tomato that are highly homologous to MADS-box proteins known in other plant species, analyzed their phylogenetic relationships, gene structure, and conserved motifs, determined their exon–intron organization, and predicted their chromosomal localization. Furthermore, we obtained the predictions of the expression pattern of these tomato MADS-box genes in order to predict their expression pattern. In addition, the expression pattern of some MADS-box genes related to tomato flower organ identity, fruit development, and ripening were determined by qPCR analyses in different stages of tomato development. These results provide details of the tomato MADS-box family and may be useful for more comprehensive investigations of tomato MADS-box gene family members.

2. Results

2.1. Identification of MADS-Box Genes in Tomato

To extensively identify tomato MADS-box genes, a set of 131 tomato MADS-box genes that are highly homologous to the MADS-box proteins reported in other plants was recovered using BLAST searches against the NCBI and SGN databases. Redundant sequences were removed. Since 35 tomato MADS-box genes have been officially named, the newly identified 96 tomato MADS-box genes were designated as *SlMADS2*–*SlMADS98* (Table 1). Additionally, the molecular characteristics of the 131 MADS-box genes in tomato were analyzed. The names of the tomato MADS-box genes, the locus/gene names of SGN, the length of amino acid sequence, the molecular weight, and the isoelectric point are shown in Table 1. The statistical results showed that the amino acid sequence length of the 131 predicted tomato MADS-box proteins varied from 54 (*SlMBP19*) to 417 (*SlMADS52*), the relative molecular mass ranged from 6224.26 Da (*SlMBP19*) to 47275.1 Da (*SlMADS52*), and the isoelectric point (pI) varied from 4.41 (*SlMADS70*) to 11.03 (*SlMADS77* and *SlMADS80*). The homologous genes of the tomato MADS-box family genes in Arabidopsis and the references for the study of the functions in homologous genes are shown in Table S1. In addition, the functional reports of several genes that have been studied previously were listed.

Table 1. Overview of MADS-box genes identified in tomato. List of predicted genes and related information include gene name, gene locus, molecular details, classification of two types according to the phylogenetic analysis shown in Figure S1, homologs in Arabidopsis, as well as the reference of the gene function investigation. pI, isoelectric point; MW, molecular weight.

Gene Name	Gene Locus	Protein			Type	Reference
		Length (aa)	MW (Da)	pI		
SlMBP1/SlGLO1/PI/LePI-B	Solyc08g067230.2.1	210	24740.2	8.69	Type II	[36]
SlMBP2/SlGLO2/LePI/TPI	Solyc06g059970.2.1	214	24867.4	9.49	Type II	[37]
SlMBP3/ SlAGL11	Solyc06g064840.3.1	237	27469.6	9.25	Type II	[38]
SlMBP6/SlAGL6	Solyc01g093960.2.1	252	28608.5	8.37	Type II	[39–41]
SlMBP7/LeFUL2	Solyc03g114830.2.1	247	28614.3	9.31	Type II	[22,40,42–44]
SlMBP8	Solyc12g087830.1.1	198	22643.3	8.97	Type II	[45,46]
SlMBP10	Solyc02g065730.1.1	234	27368.1	10.12	Type II	
SlMBP11/AGL15-like	Solyc01g087990.2.1	271	30847	6.01	Type II	[47]
SlMBP13	Solyc08g080100.2.1	224	25839	9.76	Type II	
SlMBP14	*Solyc12g056460.1.1*	*206*	*23968.4*	*6.57*	Type II	
SlMBP15	Solyc12g087830.2.1	204	23667.2	8.35	Type II	
SlMBP18/SlFYFL	Solyc03g006830.2.1	222	25369.2	9.46	Type II	[48]
SlMBP19	Solyc06g035570.1.1	54	6224.26	11.29	Type I	
SlMBP20	Solyc02g089210.2.1	250	28589.2	9.87	Type II	[44]
SlMBP21	Solyc12g038510.1.1	250	28442.4	9.26	Type II	[49–51]
SlMBP22	Solyc11g005120.1.1	238	27791.6	6.96	Type II	
SlMBP23/TDR3	Solyc10g017630.2.1	166	19196.3	9.49	Type II	
SlMBP24	Solyc04g076280.3.1	235	26483.6	7.67	Type I	
SlMBP25	Solyc05g015730.1.1	80	9311.97	11.01	Type II	
TAG1	Solyc02g071730.2.1	248	28723.6	9.97	Type II	[52,53]

Table 1. *Cont.*

Gene Name	Gene Locus	Protein			Type	Reference
		Length (aa)	MW (Da)	pI		
TAGL1	Solyc07g055920.2.1	267	29940.5	9.56	Type II	[20,21,52,54]
TAGL2	Solyc05g015750.2.1	241	27579.3	9.07	Type II	[55,56]
TAGL11	Solyc11g028020.1.1	223	26051.7	9.76	Type II	[38,56]
TAGL12	Solyc11g032100.1.1	201	23104.8	6.94	Type II	[56]
TAP3/LeAP3/LeDEF	Solyc04g081000.2.1	228	26478.2	9.76	Type II	[54,57–59]
MADS-RIN	Solyc05g012020.2.1	242	27968.4	8.00	Type II	[19,60–63]
MADS-MC	Solyc05g056620.1.1	244	28660.6	8.44	Type II	[64]
JOINTLESS	Solyc11g010570.1.1	265	30426.3	7.39	Type II	[64–68]
LeAP1	Solyc05g012020.3.1	213	25077.3	5.47	Type II	[40]
TM4/TDR4/LeFUL1	Solyc06g069430.2.1	245	28290	9.57	Type II	[22,42,43,59]
TM5/TDR5/LeSEP3	Solyc05g015750.3.1	224	25999.3	9.79	Type II	[40]
TM6/TDR6	Solyc02g084630.2.1	225	26092.7	9.71	Type II	[69]
TM8/TDR8	Solyc03g019710.2.1	177	20703.9	10.44	Type II	[70]
SlMADS1	Solyc03g114840.2.1	246	28398.2	8.88	Type II	[55,71]
SlMADS2	Solyc01g060300.1.1	211	23122.6	9.36	Type I	
SlMADS3	Solyc01g060310.1.1	139	15117.5	8.99	Type I	
SlMADS4	Solyc01g060284.1.1	197	21565.5	8.63	Type I	
SlMADS5	Solyc01g066500.1.1	309	34544.4	4.75	Type I	
SlMADS6/TM29/LeSEP1	Solyc02g089200.2.1	246	28481.3	8.29	Type II	[39,41]
SlMADS7	Solyc01g103870.1.1	125	14435.6	9.54	Type I	
SlMADS8	Solyc06g071300.1.1	130	15009.1	5.14	Type I	
SlMADS9	Solyc03g007020.1.1	215	24702	9.21	Type I	
SlMADS10	Solyc04g064860.1.1	233	25871.3	7.92	Type I	
SlMADS11	Solyc04g054517.1.1	138	15663.1	7.75	Type I	
SlMADS12	Solyc04g056550.1.1	182	20838.7	5.14	Type I	
SlMADS13	Solyc04g056740.1.1	180	20521.5	7.44	Type I	
SlMADS14	Solyc06g054680.1.1	180	21152.8	7.19	Type I	
SlMADS15	Solyc06g059780.1.1	180	21057.6	6.37	Type I	
SlMADS16	Solyc10g050950.1.1	161	18602.2	6.80	Type I	
SlMADS17	Solyc10g050900.1.1	169	19796.5	7.83	Type I	
SlMADS18	Solyc10g050940.1.1	169	19623.3	7.77	Type I	
SlMADS19	Solyc01g097850.1.1	172	20073.9	7.35	Type I	
SlMADS20	Solyc03g034260.1.1	184	21441	6.29	Type I	
SlMADS21	Solyc10g018070.1.1	119	13063.9	10.14	Type I	
SlMADS22	Solyc10g018080.1.1	151	17218.5	7.04	Type I	
SlMADS23	Solyc10g018110.1.1	184	20762.4	6.55	Type I	
SlMADS24	Solyc04g025970.1.1	117	13333.4	10.16	Type I	
SlMADS25	Solyc01g010300.1.1	88	10052.7	10.79	Type I	
SlMADS26	Solyc02g032000.1.1	97	10944.7	10.32	Type I	
SlMADS27	Solyc03g119680.1.1	177	20293.2	7.26	Type I	
SlMADS28	Solyc09g061950.1.1	173	20138.1	9.33	Type I	
SlMADS29	Solyc01g106710.1.1	221	25105.7	10.23	Type I	
SlMADS30	Solyc04g025030.1.1	135	15411	10.37	Type I	
SlMADS31	Solyc04g025110.1.1	130	14313.7	10.26	Type I	
SlMADS32	Solyc01g106720.1.1	211	23531.2	10.35	Type I	
SlMADS33	Solyc01g106730.1.1	221	23888.6	10.14	Type I	
SlMADS34	Solyc04g047870.1.1	190	21695.2	10.70	Type I	
SlMADS35	Solyc10g012180.1.1	167	18824.5	5.38	Type I	
SlMADS36	Solyc10g012200.1.1	167	19113.1	8.98	Type I	
SlMADS37	Solyc11g067163.1.1	166	18919.9	7.67	Type I	
SlMADS38	Solyc01g066730.2.1	300	33023.4	5.04	Type I	
SlMADS39	Solyc01g098070.1.1	155	17200.8	10.09	Type I	
SlMADS40	Solyc01g098060.1.1	152	17191.5	9.84	Type I	
SlMADS41	Solyc03g062820.1.1	145	16350.8	9.23	Type I	
SlMADS42	Solyc01g098050.1.1	160	18178	10.45	Type I	
SlMADS43	Solyc05g013370.1.1	232	26476.7	5.46	Type I	
SlMADS44	Solyc05g046345.1.1	250	28489.8	5.47	Type I	
SlMADS45	Solyc04g025050.1.1	56	6259.52	11.07	Type I	
SlMADS46	Solyc07g017343.1.1	226	26532.7	5.2	Type I	
SlMADS47	Solyc00g179240.1.1	55	6310.11	11.16	Type I	
SlMADS48	Solyc07g052707.1.1	195	21878.8	4.86	Type I	

Table 1. *Cont.*

Gene Name	Gene Locus	Protein			Type	Reference
		Length (aa)	MW (Da)	pI		
SlMADS49	Solyc07g052700.2.1	190	21334.2	7.41	Type I	
SlMADS50	Solyc11g069770.2.1	175	20171.9	5.45	Type I	
SlMADS51	Solyc01g106700.2.1	213	24279.5	7.35	Type I	
SlMADS52	Solyc03g115910.1.1	417	47275.1	5.82	Type I	
SlMADS53	Solyc10g012380.1.1	149	17086.7	5.29	Type I	
SlMADS54	Solyc10g012390.1.1	149	17086.7	5.29	Type I	
SlMADS55	Solyc11g069770.1.1	181	20857.8	5.20	Type I	
SlMADS56	Solyc12g042967.1.1	225	25552.3	9.05	Type I	
SlMADS57	Solyc11g020620.1.1	342	38747.9	5.42	Type I	
SlMADS58	Solyc11g020660.1.1	312	35558.7	4.61	Type I	
SlMADS59	Solyc11g020320.1.1	275	31757.7	6.68	Type I	
SlMADS60	Solyc01g103550.1.1	311	35839.7	7.87	Type I	
SlMADS61	Solyc01g060310.2.1	136	15230.6	6.16	Type I	
SlMADS62	Solyc12g016170.1.1	219	25087.7	6.57	Type I	
SlMADS63	Solyc12g016180.1.1	219	25207.8	6.57	Type I	
SlMADS64	Solyc12g016150.1.1	154	17645.1	5.38	Type I	
SlMADS65	Solyc12g017300.1.1	154	17740.1	5.11	Type I	
SlMADS66	Solyc12g005210.1.1	88	9985.65	10.41	Type I	
SlMADS67	Solyc01g102260.2.1	249	27882.8	9.66	Type I	
SlMADS68	Solyc05g047712.1.1	156	18199.1	8.89	Type I	
SlMADS69	Solyc07g043080.1.1	215	24384.8	6.86	Type I	
SlMADS70	Solyc06g034317.1.1	153	16949.3	4.41	Type I	
SlMADS71	Solyc06g048380.1.1	176	20038.6	4.90	Type I	
SlMADS72	Solyc06g048380.2.1	132	15027	4.82	Type I	
SlMADS73	Solyc06g033820.1.1	196	22657.4	7.25	Type I	
SlMADS74	Solyc06g033830.1.1	113	13058.8	8.44	Type I	
SlMADS75	Solyc07g052700.3.1	197	22133.2	8.44	Type I	
SlMADS76	Solyc03g115910.2.1	434	49552.6	5.6	Type I	
SlMADS77	Solyc12g087820.1.1	106	11969.2	11.03	Type I	
SlMADS78	Solyc05g051830.2.1	389	44560	5.91	Type I	
SlMADS79	Solyc10g017640.1.1	61	7032.22	11.24	Type I	
SlMADS80	Solyc04g076680.2.1	63	70492.3	11.03	Type I	
SlMADS81	Solyc12g088080.1.1	97	11295.2	10.59	Type I	
SlMADS82	Solyc01g105800.2.1	279	32104.7	9.11	Type II	
SlMADS83	Solyc01g106170.2.1	148	17032.2	10.98	Type II	
SlMADS84	Solyc10g080030.1.1	229	25791.3	6.37	Type II	
SlMADS85	Solyc01g105810.3.1	215	24518.2	9.22	Type II	
SlMADS86	Solyc04g078300.3.1	145	16547.2	9.9	Type II	
SlMADS87	Solyc04g078300.2.1	334	38130.7	5.96	Type II	
SlMADS88	Solyc03g006830.3.1	209	24310.2	9.43	Type II	
SlMADS89	Solyc10g044967.1.1	226	26172	6.14	Type II	
SlMADS90	Solyc11g032100.2.1	201	23104.8	6.45	Type II	
SlMADS91	Solyc05g015720.2.1	204	23433.5	5.36	Type II	
SlMADS92	Solyc01g093960.3.1	242	27359.1	8.19	Type II	
SlMADS93	Solyc02g091550.2.1	249	28342.2	9.07	Type II	
SlMADS94	Solyc02g091550.1.1	235	27198.1	9.65	Type II	
SlMADS95	Solyc04g076700.3.1	226	25933.3	8.59	Type II	
SlMADS96	Solyc03g019720.3.1	193	22631.1	9.88	Type II	
SlMADS97	Solyc12g088090.2.1	239	27695.8	9.84	Type II	
SlMADS98/SlCMB1	Solyc04g005320.2.1	238	27522.1	8.60	Type II	[72]

2.2. Classification and Phylogenetic Analysis of Tomato MADS-Box Genes

To study the phylogenetic relationships among MADS-box genes in tomato and Arabidopsis [6], a phylogenetic tree was drawn by the neighbor-joining (NJ) method using MEGA 5.02 (Figure S1A). Based on previous reports on Arabidopsis, the 131 tomato MADS-box genes were classified into two types: type I (81) and type II (50). Based on the phylogenetic tree, type I and type II tomato MADS-box genes were subdivided into more detailed subgroups. Then, type I could be further divided into four

groups (Mα, Mβ, Mγ, and Mδ), while Type II (MIKC) could be further divided into MIKC* and MIKC^C. The MIKC^C-type genes included the *AP3/PI*, *SVP*, *AGL15*, *SEPALLATA* (*SEP*), *AGL6*, *AP1*, *FLOWERING LOCUS C* (*FLC*) *SOC1*, *AGAMOUS* (*AG*), *TM8*, and *DEFICIENS* (*DEF*)/*GLOBOSA* (*GLO*) subfamilies, similar to the MADS-box genes in other plants species. In addition, the phylogenetic tree of type I and type II MADS-box protein in tomato plants were constructed to identify the phylogenetic relationships of gene numbers in the two types of tomato MADS-box family more clearly (Figure S1B,C).

2.3. Conserved Motif and Gene Structure Analysis of Tomato MADS-Box Genes

The intron–exon organization was analyzed to investigate the structural diversity and evolution of the 131 tomato MADS-box genes. As shown in Figure 1, we obtained each gene's intron/exon arrangement by comparing their CDS with their genomic sequences using the program Gene Structure Display Server (GSDS). The number of introns in tomato MADS-box genes ranged from one to 11. Similar to Arabidopsis, the distribution of introns in tomato was different in type I and type II genes [6]. In our study, we found that the Mα, Mβ, and Mγ groups of the type I genes usually had no introns or one intron, which might be explained by the diversity of the reverse-transcribed origin or the differences in acquisition or loss introns by the ancestors of these three groups of genes [6]. Based on the genomic data, the Mδ group of type I and the type II genes contained multiple introns. Among the Mδ clade and the type II genes, 52 of 56 (92.9%) genes had more than five introns. Additionally, the gene structures of closely related genes in tomato MADS-box genes were more similar, and the differences were only in the lengths of introns and exons. However, some close gene pairs showed different intron/exon arrangements. For instance, *SlMBP61* has one exon, whereas its close homologs *SlMBP51* and *SlMBP10* both have two, although their phylogenetic relationship displayed a high bootstrap value (Figure S1B).

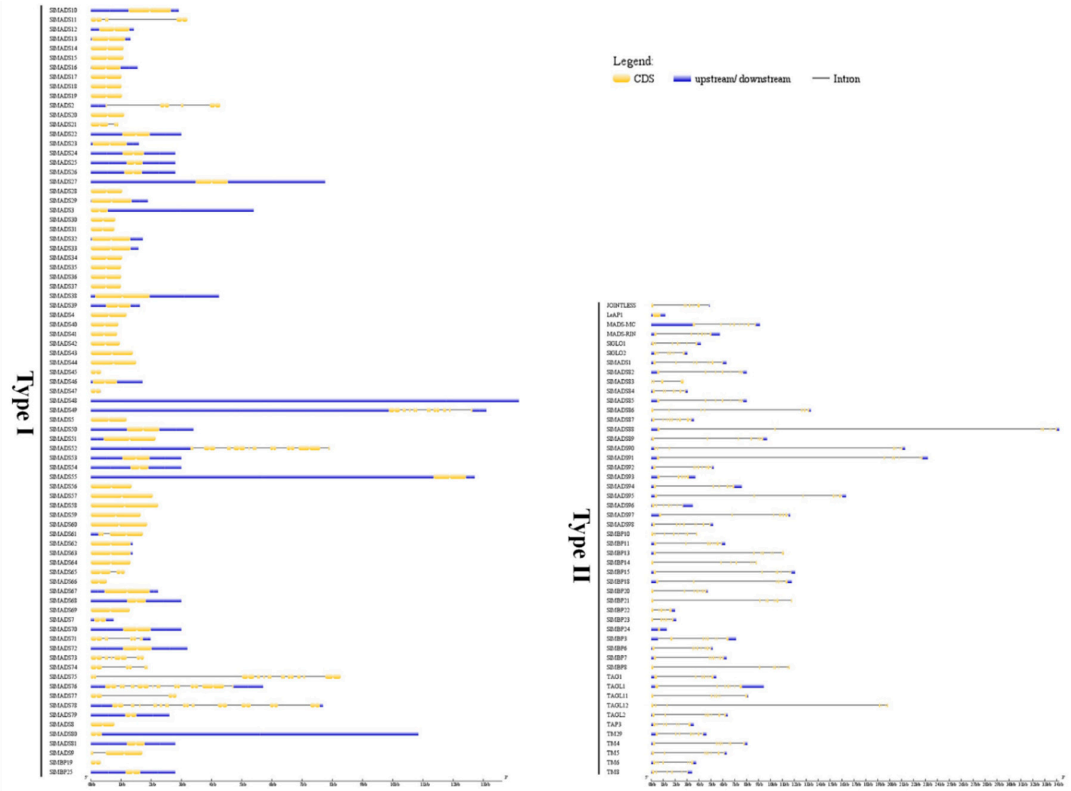


Figure 1. Gene structure analysis of MADS-box genes in tomato. The Gene Structure Display Server (GSDS) database was used to perform the exon–intron structure analyses. Lengths of exons and introns of each MADS-box gene were displayed proportionally. The blue boxes represent upstream/downstream, the yellow boxes represent exons, and the black lines represent introns.

To better analyze conserved motifs in tomato MADS-box proteins, we constructed a conserved motif figure using the Multiple EM for Motif Elicitation (MEME) program and annotated them using SMART. A total of 10 conserved motifs, named 1 to 10, were identified (Figure 2). The details of the motifs are shown in Figure S2. As expected, the same types of genes tend to possess the same motifs. Motif 1—one of the most typical MADS-box domains—comprised 42 amino acids was found in the majority of tomato MADS-box proteins. Motif 3 was also conserved across most of the tomato MADS-box proteins, including type I and type II genes. Motifs 2 and 4 represent the K domain, which plays an important role in protein–protein interactions among MADS-box proteins, and they were found only in type II MADS-box proteins. Motif 2 was identified in almost all the type II proteins except for TM8/TDR8, SlMADS86, SlMADS87, and SlMADS83. In the type II proteins, a large number of proteins had motif 4, with seven exceptions (SlGLO2, SlGLO1, SlMBP11, SlMADS84, SlMADS86, SlMADS87, and SlMADS83). Motif 9 is also a MADS-box domain that is present in a small number of tomato MADS-box proteins. However, some motifs (6, 7, 8, and 10) were shown to be weakly conserved in tomato MADS-box proteins, and they were found only in type I MADS-box proteins.

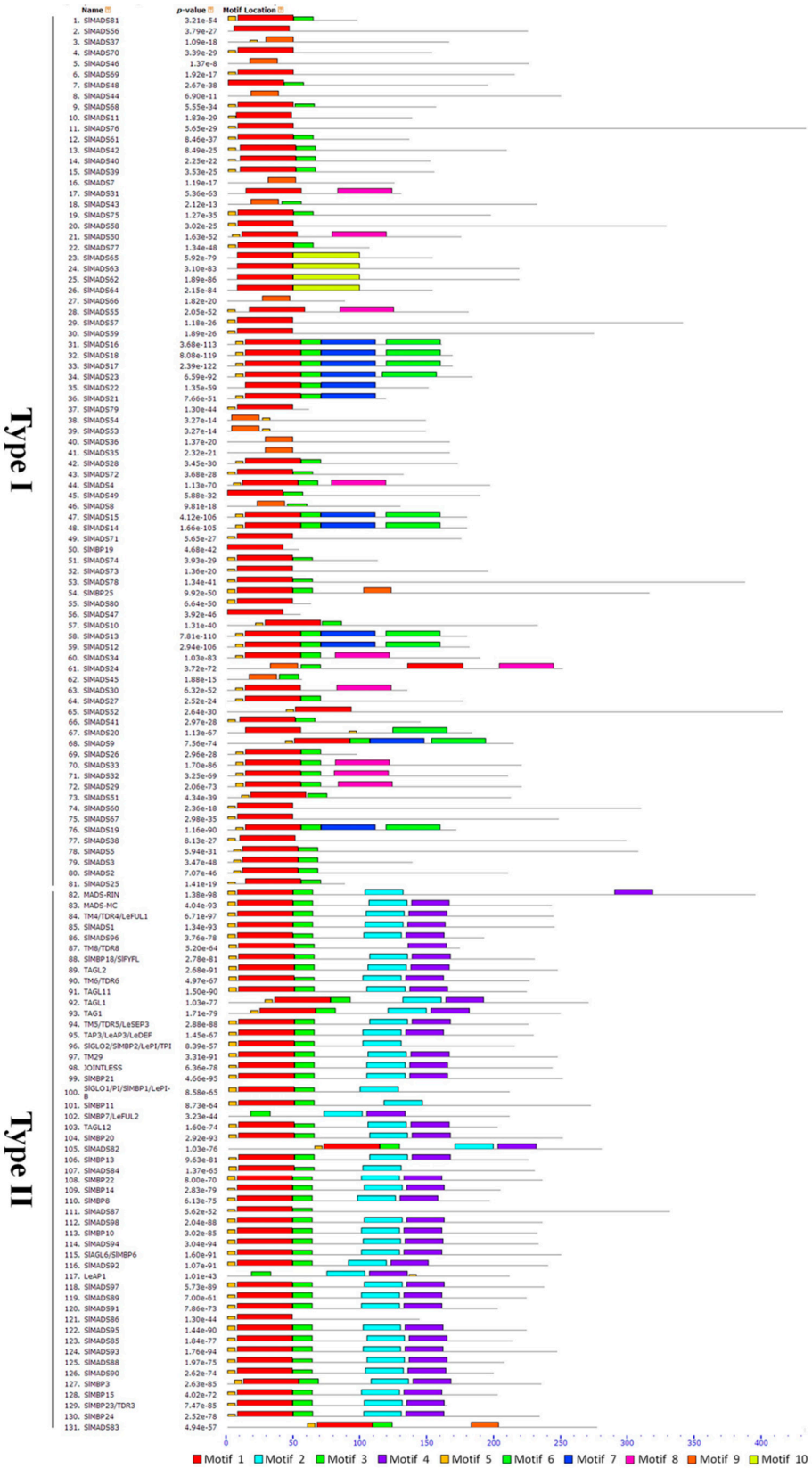


Figure 2. Conserved motif analyses of MADS-box genes in tomato. All the conserved motifs were identified by the Multiple EM for Motif Elicitation (MEME) database using the complete amino acid sequences of tomato MADS-box genes. Ten conserved different motifs were indicated by different colors. The details of motifs refer to the Supplementary Figure S2.

2.4. Chromosomal Locations of Tomato MADS-Box Genes

According to physical genome annotation files that were obtained by using genomic sequences from the SGN and NCBI databases, 131 members of the MADS-box genes were located on all 12 tomato chromosomes, implying that the MADS-box transcription factor family may have multiple functions in tomato plants.

As shown in Figure 3, the tomato MADS-box genes are distributed unevenly on each chromosome. Chromosome 1 contains the most MADS-box genes (24), while chromosome 8 and 9 contain the fewest (two). Gene duplication events have a significant function in amplifying gene family numbers and genome complexity in eukaryotes [73,74]. The tandem amplification or segmental duplication of chromosomal regions can increase gene families. In this study, the results showed that chromosomes 1, 3, 4, 10, and 12 contain gene clusters or gene hotspots; in particular, chromosome 1 contains eight MADS-box genes within a short region. Additionally, we confirmed that internal chromosome duplication (tandem duplication) occurred in these genes.

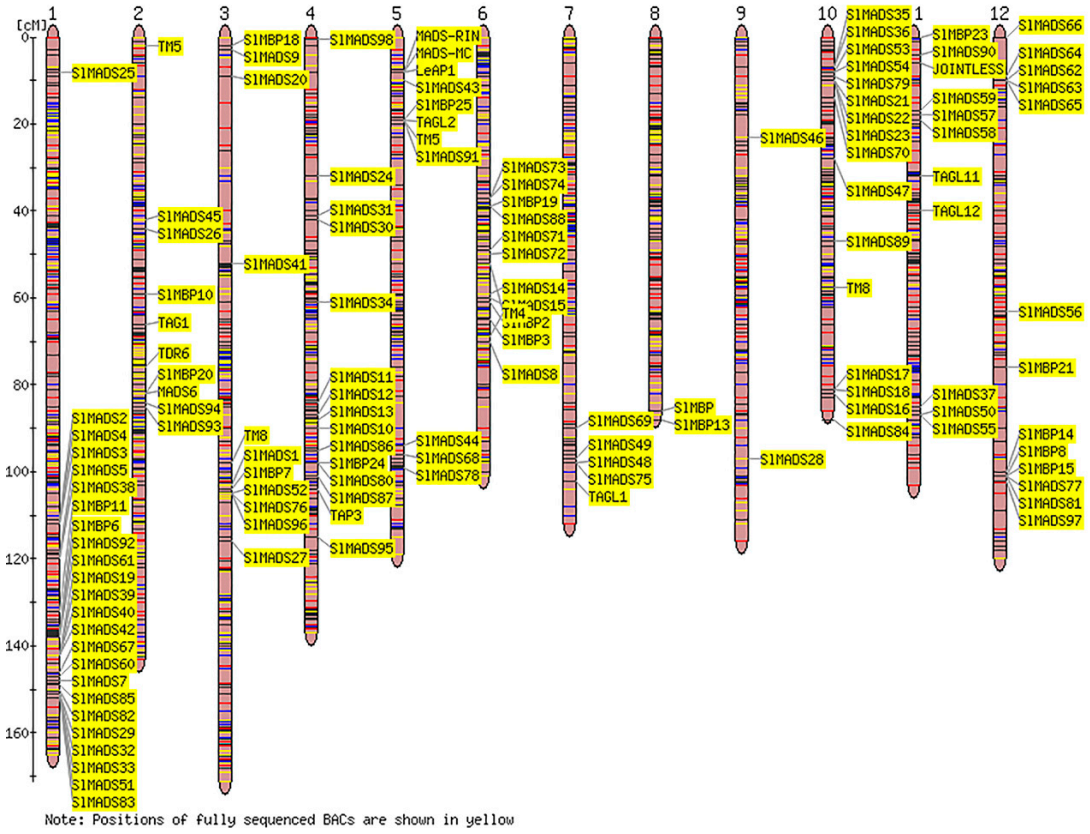


Figure 3. Chromosomal locations of tomato MADS-box genes. A total of 12 chromosomes of tomato were labeled with their names, chromosomes 1 to 12, which are indicated at the top of each bar. The position of tomato MADS-box genes on the chromosome was based on the Sol Genomics Network (SGN) and National Center for Biotechnology Information (NCBI) database and the Tomato-EXPEN 2000 was used to draw the physical map of the tomato MADS-box genes.

2.5. Predictions of Expression Profiles of Tomato MADS-Box Genes in Different Organs

To investigate the tomato MADS-box genes expression patterns in different tissues of tomato plants, we analyzed tomato transcript expression (RNA-seq) data in nine different tomato tissues at different developmental stages. This included the expression in the whole root (RT), young leaf (YL), mature leaves (ML), young flower buds (YFB), fully open flowers (F), and at the immature green (IMG), mature green (MG), break (B), and mature (MF) stages of fruit development and ripening. These datasets were searched using the locus/gene names in SGN of 124 tomato MADS-box gene sequences, except for

SlMADS4, *SlMADS11*, *SlMADS37*, *SlMADS44*, *SlMADS46*, *SlMADS56*, *SlMADS68*, *SlMADS70* and *SlMADS89*, which were not accurately found in TFGD. Then, we constructed a hierarchical clustering heat map using these datasets (Figure 4).

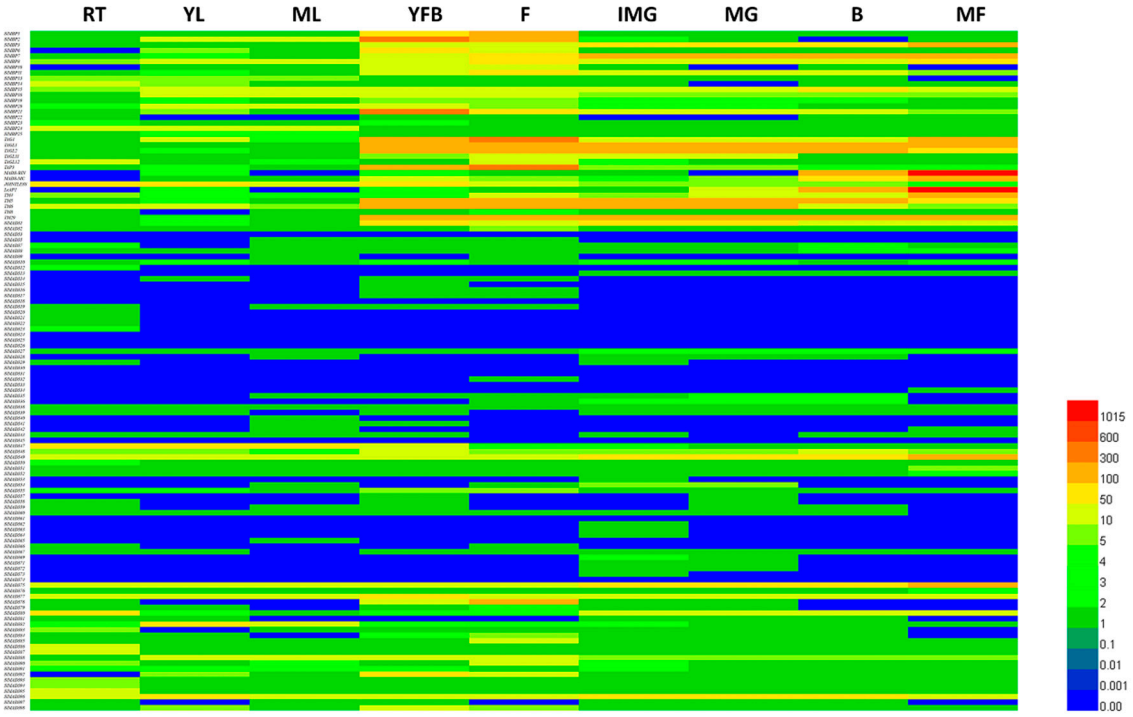


Figure 4. Heat map representation of tomato MADS-box genes in various tissues. The tissues included the whole root (Rt), young leaf (YL), mature leaves (ML), young flower buds (YFB), and fully open flowers (F), which were at the immature green (IMG), mature green (MG), break (B), and mature (MF) stages of fruit development and ripening. The bar at the bottom of the heat map represents relative gene expression values.

The expression profiles revealed that 117 genes were expressed in at least one tomato plant organ, while the other seven genes (*SlMADS24*, *SlMADS25*, *SlMADS26*, *SlMADS33*, *SlMADS45*, *SlMADS61*, and *SlMADS74*) were expressed at levels that were too low to be identified, or they had temporal and spatial specific expression patterns that showed no expression in the organs tested. Most tomato MADS-box genes displayed a broad expression range across all the organs and developmental stages, which is consistent with previous reports that the MADS-box genes may play multiple roles in plant growth and development [75,76]. However, some genes exhibited tissue-specific expression. For example, the expression of *SlMADS12*, *SlMADS20*, *SlMADS21*, *SlMADS22,* and *SlMADS23* were restricted in whole root, and the *SlMADS16, SlMADS17,* and *SlMADS132* transcripts were observed only during flower development. These results illustrate that these genes may be involved in the regulation of some biological process of the tomato root or in flower growth and development. Eight genes (*SlMBP2*, *SlMBP6*, *SlMBP10*, *SlMBP21*, *TAP3*, *SlMADS78*, *SlMADS92*, and *SlMADS98*) showed especially high expression in young flower buds (YFB) and fully open flowers (F), indicating that these genes may play important roles in floral organ development. We further discovered that most type II genes (*SlMBP3*, *SlMBP7*, *SlMBP11*, *SlMBP15*, *MADS-RIN*, *MADS-MC*, *TAGL1*, and *LeAP1*) were highly expressed during flower or fruit development; especially, the expression values of *MADS-RIN* and *MADS-MC* in fruits were more than 1000, suggesting that these genes may be associated with the reproductive growth of tomato. However, the expression of most type II genes showed no significant difference among tissues.

In short, these results indicate that the MADS-box genes had different expression levels in various tomato organs, and the predictions of the organ expression profiles of the tomato MADS-box gene

family may provide insight for future studies on the functions of MADS-box genes in tomato plant growth and development.

2.6. The Critical Tomato MADS-Box Genes Involved in Floral Organ Development

Based on the phylogenetic analysis of the MADS-box genes that participated in floral organ development in petunia, which has been reported previously [32,77] (Figure S3), 15 tomato MADS-box genes that are possibly involved in floral organ development were screened out, including tow class A genes (*MADS-MC* and *SlMBP20*), four class B genes (*TAP3*, *TM6*, *SlMBP1,* and *SlMBP2*, two class C genes (*TAG1* and *TAGL1*), two class D genes (*SlMBP3* and *SlMBP22*), and five class E genes (*TAGL2*, *TM5*, *SlMADS1*, *SlMBP21,* and *SlMBP6*), as shown in Table S2.

According to the expression profile predictions shown in Figure 4, among the 15 MADS-box genes, nine genes (*MADS-MC*, *TAP3*, *TM6*, *SlMBP1*, *SlMBP2*, *TAG1*, *SlMBP22*, *SlMBP21,* and *SlMBP6*) were extremely highly expressed in flowers. However, *TAGL1*, *SlMBP3*, *TAGL2*, *TM5*, and *SlMADS1* were mainly expressed during the stages of fruit development and ripening, and the expression of *SlMBP20* was particularly high in leaves.

To further investigate the potential role of these MADS-box genes in floral organ development, a comparison of the expression patterns of 15 tomato MADS-box genes in four whorls of floral organs (sepal, petal, stamen, and carpel) were analyzed by qPCR. As shown in Figure 5A,B, the class A genes *MADS-MC* and *SlMBP20* were highly expressed in sepals. The expression levels of class B genes (*TAP3*, *SlMBP2,* and *SlMBP1)* were notably high in the petal and stamen (Figure 5C–E), whereas the transcription of *TM6* was found to be markedly higher in carpel and stamen compared with other floral organs (Figure 5F). Two class C genes, *TAG1* and *TAGL1*, were found to be mainly expressed in stamens and carpels (Figure 5G,H). Moreover, the class D genes, *SlMBP3* and *SlMBP22,* showed organ-specific expression patterns that were exclusively expressed in carpel (Figure 5I,J). The expression patterns of class E genes indicated that the expression of *TAGL2* and TM5 were significantly higher expressed in the petal, stamen, and carpel than in the sepal (Figure 5K,L), and *SlMBP6* was highly expressed in petals and carpels (Figure 5M). In comparison, the other class E genes (*SlMADS1* and *SlMBP21*) were shown to exhibit higher expression levels in the sepal and carpel (Figure 5N,O).

In addition, the interaction networks of these 15 tomato MADS-box proteins were predicted by STRING software (Figure S4). The results showed that they established interactions with other proteins, directly or indirectly. The TAGL2, SlMBP3, TAGL1, and SlMBP22 proteins can interact directly with each other. Apart from that, the TAGL1 protein showed complex interactions with several other proteins, including the SlMBP3 and TAP3 proteins. The TM5 protein was also shown to directly interact with the TAP3 and TM6 proteins.

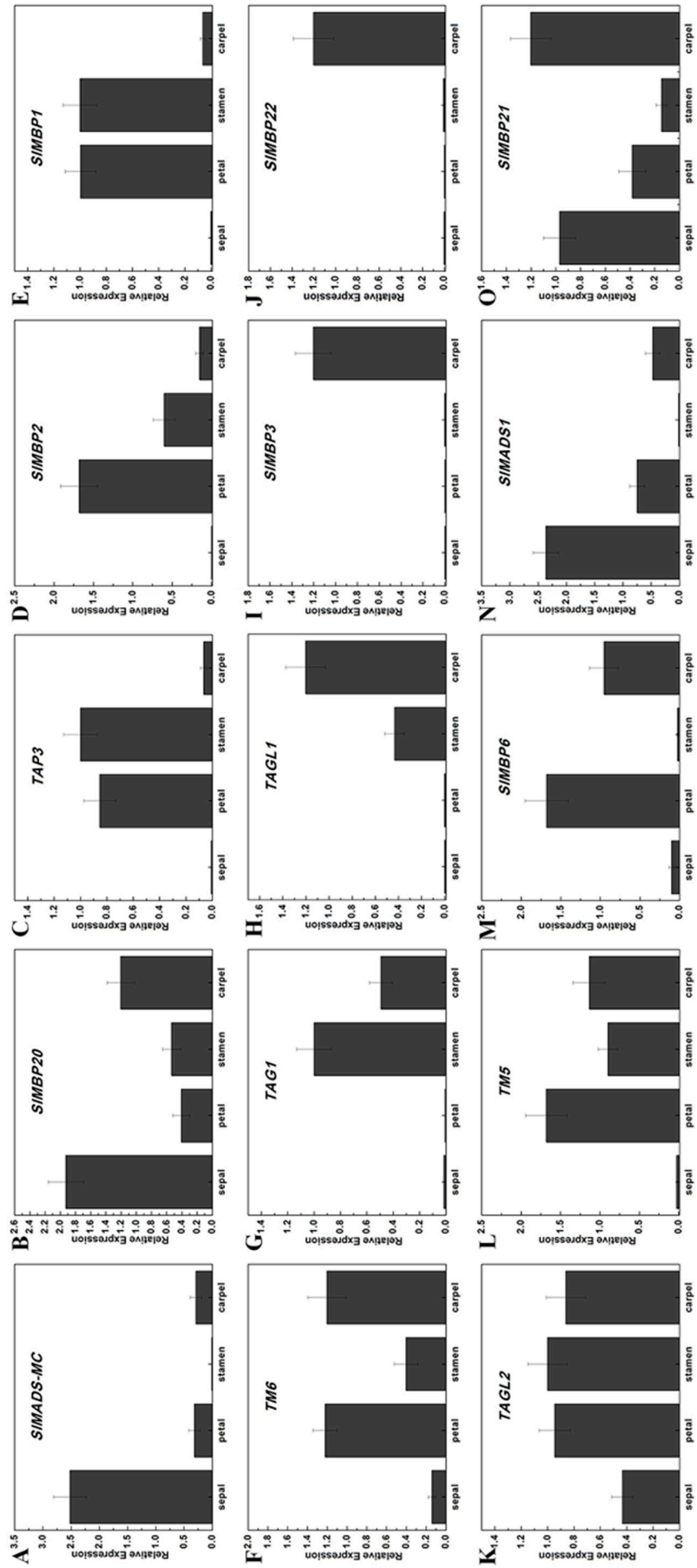


Figure 5. Expression profiles of floral organ identity genes in the four-whorl floral organs of the wild-type (WT) tomato plant for qPCR analysis. Se, sepal; Pe, petal; St, stamen; Ca, carpel. (A)–(O) Expression profiles of *SlMADS-MC*, SlMBP*21*, *TAP3*, *SlMBP2*, *SlMBP1*, *TM6*, *TAG1*, *TAGL1*, *SlMBP3*, *SlMBP22*, *TAGL2*, *TM5*, *SlMBP6*, *SlMADS1* and *SlMBP21*.Each value represents the mean ± SE of three technical replicates of a single biological sample. The *SlCAC* gene of tomato was used as the internal standard.

2.7. Differential Expression Analysis of Tomato MADS-Box Genes at Different Stages of Fruit Development and Ripening

According to the expression profiles predictions, we selected five tomato MADS-box genes (*SlMBP3*, *MADS-RIN*, *TAGL1*, *TM4*, and *SlMBP7*) that may take part in fruit development and ripening. We analyzed their expression patterns by qPCR at five different stages of fruit development and ripening, including the immature green (IMG), mature green (MG), and break (B) stages, as well as at four days after break (B+4) and seven days after break (B+7). *SlMBP3* exhibited a strikingly high expression level at the IMG stage, and showed extremely low expression at the other stages (Figure 6A). The expression levels of *MADS-RIN* and *TAGL1* exhibited an increasing tendency from the MG to the B+4 stage, and then decreased at B+7 (Figure 6B,C). *TM4* expression increased continuously during the process of fruit development and exhibited its maximum expression level at the B+7 stage (Figure 6D). Compared with the other three stages, the expression of *SlMBP7* was slightly higher at B (Figure 6E). These results indicated that the qPCR data were consistent with the predictions of expression profiles and that our predictions are suitable for investigating the expression patterns of tomato MADS-box genes.

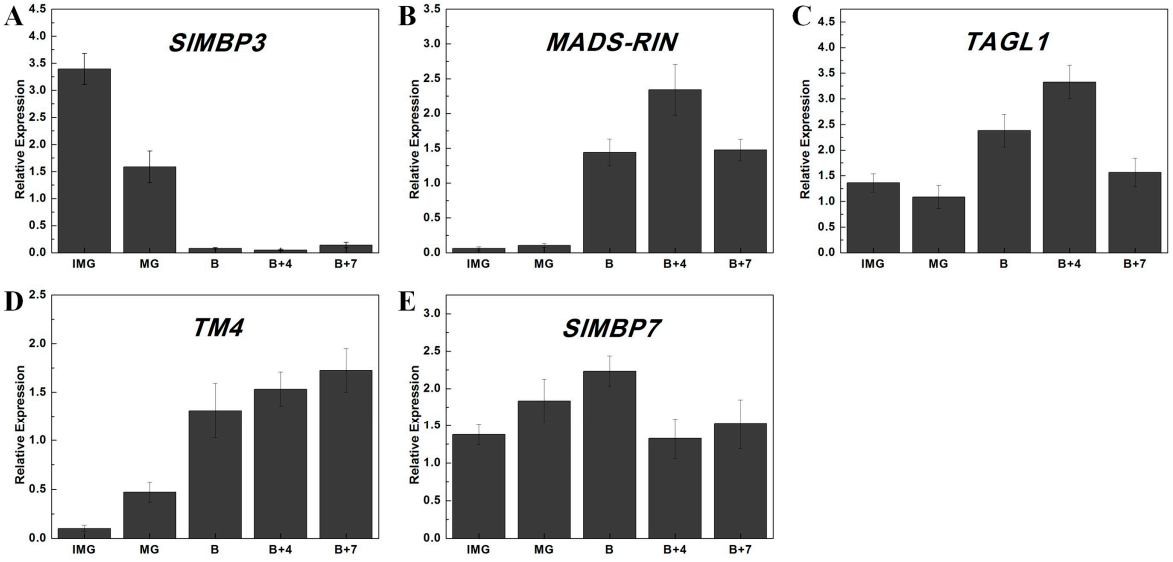


Figure 6. Relative expression of tomato MADS-box genes in difference stages of fruit development and ripening by qPCR analysis. Expression patterns of *SlMBP3* (**A**), *MADS-RIN* (**B**), *TAGL1* (**C**), *TM4* (**D**), and *SlMBP7*(**E**) in different organs, including mature green (MG), break (B), four days after break (B+4), and seven days after break (B+7). Each value represents the mean ± SE of three technical replicates of a single biological sample. The *SlCAC* gene of tomato was used as the internal standard.

3. Discussion

3.1. Characterization of MADS-Box Genes in Tomato

The MADS-box genes control diverse biological processes in plants, including vegetative growth and reproductive development. They mainly play key roles in the developmental processes of inflorescences, flowers, and fruits [78,79]. 24 MIKC^C-type MADS-box transcription factors have been identified, and their functions and evolutions in tomatoes were thoroughly studied in 2006 [35]. However, the MIKC^C-type MADS-box members are only part of the MADS-box transcription factor family, and to date, there has been no comparative report on the tomato MADS-box genes. It is well known that genome-wide analysis of gene families is a major and necessary approach to analyze the structure, evolution, and function of genes. In this study, 131 tomato MADS-box proteins were identified, and 96 new tomato MADS-box proteins with unknown functions were systemically named (Table 1). This study is the first comparative analysis of the tomato MADS-box gene family, and we

believe that the resolving confusion in naming the genes will facilitate further functional analysis of the tomato MADS-box genes.

First, we presented the phylogenetic relationships of 131 tomato MADS-box proteins with Arabidopsis MADS-box proteins to classify the tomato MADS-box proteins into five subfamilies (MIKC, Mα, Mβ, Mγ, and Mδ), as shown Figure S1A. Compared with Arabidopsis, a larger number of MADS-box proteins were found in tomato. In total, 81 tomato MADS-box genes were determined to be type I genes, including the Mα, Mβ, Mγ, and Mδ groups, which is more than that in Arabidopsis. We speculate that tomato type I MADS-box genes may have a higher duplication rate and/or a lower gene loss rate after duplication. Nevertheless, 50 tomato MADS-box genes were classified as type II genes, including MIKC^c and MIKC*, which is comparable to that in Arabidopsis. These results indicate that the genes' retention duplication have been different in various species, leading to different numbers of MAD-box genes among different species, with different evolutionary constraints [80]. Then, in order to investigate the phylogenetic relationships of MADS-box genes in tomato, a phylogenetic tree for two types of tomato MADS-box genes was constructed (Figure S1B,C). This showed that tomato MADS-box genes are conservative in subfamilies.

To obtain insight into the structural diversity of the tomato MADS-box genes, the intron–exon organization was analyzed (Figure 1). Previous studies have postulated that an intron-rich gene would lose multiple introns simultaneously by retrotransposition, thereby producing intron-less ancestral genes. In this study, we found that the Mα, Mβ and Mγ groups of the type I genes usually have no introns or one intron, which may experience the loss of multiple introns during MADS-box gene family diversification. In addition, the distribution of introns in tomato type I and type II genes were different, and the Mδ group of the type I and type II genes had more introns than the Mα, Mβ, and Mγ groups genes. Similar cases have also been detected in Arabidopsis and rice [7,13], suggesting the evolutionary conservation among plants. However, some close gene pairs showed different intron/exon arrangements, indicating that a more complicated gene structural evolution may exist in tomato MADS-box genes. The conserved motif analysis indicated that the same group contained most conserved motifs (Figure 2). The results suggested that these conserved motifs play important roles in group-specific functions. However, high structural divergence was found between the different groups. An analysis of the gene structures and conserved motifs could provide more clues about the evolutionary relationships of the MADS-box family in tomato.

Gene duplication (segmental duplication and tandem duplication) as well as transposition events were prevalent forces that result in the expansion of family members and genome complexity in eukaryotes [74]. The duplication of more than two genes located on one chromosome is confirmed as a tandem duplication event, whereas gene duplication that occurs on different chromosomes is identified as segmental duplication [73,81]. Both tandem and segmental duplication can play crucial roles in MADS-box gene expansion the tomato genome. In our study, a chromosomal location analysis of the tomato MADS-box genes showed that the MADS-box genes are distributed on 12 chromosomes (Figure 3). The tomato MADS-box genes had a high-density distribution on chromosome 1, which had 24 genes, suggesting that they might be caused by tandem duplications. The closely related tomato MADS-box genes formed tandem arrays on chromosomes 1, 3, 4, 10, and 12, which may help the tomato evolve distinct characterizations from other plants.

Since gene expression profiles can provide significant clues about gene function, the expression of tomato MADS-box genes in whole root (Rt), young leaf (YL), mature leaves (ML), young flower buds (YFB), fully open flowers (F), and five different fruit tissues were examined by transcription expression (RNA-seq) data. All the tested 124 tomato MAD-box genes that were expected to contain *SlMADS4*, *SlMADS11*, *SlMADS37*, *SlMADS44*, *SlMADS46*, *SlMADS56*, *SlMADS68*, *SlMADS70*, and *SlMADS89* showed distinct expression patterns (Figure 4). This finding may supply insight for future studies on the functions of MADS-box genes in tomato plant growth and development. For example, we found that the *SlMADS23* gene was only expressed in the root, so we speculate that the *SlMADS23* gene may play a key role in root growth and development. The roots of plants determined the

capacity of plants to acquire and distribute nutrients and water, as well as provide a means to suit the environmental conditions. Thus, the root architecture is extremely important for plant development and breeding. In the future, we will verify whether *SlMADS23* is related to root growth by constructing a *SlMADS23* overexpression vector and generating transgenic overexpression tomato plants to study the regulation of *SlMADS23* gene on root growth and development. In addition, performing *SlMADS23* gene mutagenesis with the CRISPR/Cas9 system transformation method may also prove to be a helpful strategy.

3.2. Prediction of MADS-Box Genes Involved in the Regulation of Flower Development and Floral Organ Identity

An investigation of the genetic and molecular basis of flower development and floral organ identity in Arabidopsis and petunia suggested that MADS-box genes play fundamental roles in floral organ identity and flower development [82]. It has been confirmed that five classes of MADS-box genes (A–E) were involved in specifying floral organ identity [83–85]. In Arabidopsis, the class A genes (*AP1* and *AP2*), the class B genes (*AP3* and *PI*), the class C gene (*AG*), the class D gene (*AGL11*), and the class E gene (*SEP1*, *SEP2*, *SEP3,* and *SEP4*) were MADS-box genes, which have been reported to be involved in the regulation of floral organ development [83,86]. In petunia, lots of MADS-box genes, including the class A genes PETUNIA FLOWERING GENE (*PFG*), FLORAL BINDING PROTEIN 26 (*FBP26*), and *FBP29*, the class B genes *TM6*, *PMADS1/GP*, *PMADS2,* and *FBP1*, the class C genes *PMADS3*, *FBP6,* and *FBP24*, the class D genes *FBP11* and *FBP7,* as well as *FBP2*, *FBP4*, *FBP5*, *FBP9*, and the class E genes *FBP23*, *PMADS4,* and *PMADS12* played important roles in flower development [17,87].

In this paper, we investigated the tomato MADS-box genes' phylogenetic relationships with the petunia hybrid to select 15 tomato MADS-box genes that may play specific roles in flower development (Table S2). According to the expression profile predictions shown in Figure 4, the highest expression values for most of the genes (*MADS-MC TAP3*, *TM6*, *SlMBP1*, *SlMBP2*, *TAG1*, *SlMBP22*, *SlMBP21,* and *SlMBP6*) were observed in flower development stages. Furthermore, qPCR was used to study the expression patterns of four whorls of floral organs (sepal, petal, stamen, and carpel) in these 15 tomato MADS-box genes.

AP1 is an Arabidopsis A class gene, which conferred sepal identity in the first floral [88]. In petunia, the three genes *PFG*, *FBP26,* and *FBP29* have been identified, which were orthologs of *AP1/SQUA* in Arabidopsis [17]. Our phylogenetic analysis showed that *MADS-MC* and *SlMBP20* belonged to this clade (Figure S3) and their expression were particularly high in sepal, suggesting that they might play a similar role to *AP1* (Figure 5A,B). The class B genes were involved in the identification of petal and stamen in angiosperms [89]. Regarding the class B genes, five close homologs of petunia—TOMATO MADS-BOX GENE6 (*TM6)*, GREEN PETAL (*GP*)/ PETUNIA MADS BOX GENE 1 (*PMADS1*), *PMADS2,* and *FBP1*—were found in tomato [77]. The qPCR analysis showed that *TAP3*, *SlMBP2,* and *SlMBP1* have petal and stamen specific expression, while the *TM6/TDR6* transcripts were mainly detected in the petal and carpel (Figure 5C–F). These results were similar to the homologous genes of that in petunia [77]. Two tomato MADS-box genes *TAG1* and *TAGL1*, which were involved in C functions, were from the monophyletic *AGAMOUS* (*AG*) subfamily. These two genes were mainly expressed in stamens and carpels (Figure 5G,H), which is consistent with their function in specifying stamen and carpel development [52]. *SlMBP3* and *SlMBP22*, which are highly homologous to two petunia class D MADS-box genes, *FBP11* and *FBP7*, were shown to be separately and exclusively expressed in carpel (Figure 5L,I). The result suggested that *SlMBP3* and *SlMBP22* may have similar functions to the petunia *FBP7* and *FBP11* genes, which are related to the establishment of real ovules or carpel-like structures [90]. Arabidopsis *SEP* genes, the typical class E genes, were necessary for the specification of sepal, petal, stamen, and carpel identity with interaction with the class A, B, C, and D genes [16]. In petunia, seven class E genes (*FBP2*, *FBP4*, *FBP23*, *FBP5*, *FBP9*, *pMADS12,* and *pMADS4*) have been determined that belong to the *SEPALLATA* (*SEP*) clade [32]. The tomato *TAGL2, TM5*, *SlMADS1*, *SlMBP21,* and *SlMBP6* genes were homologous to these petunia class E genes (Figure S3), and some differences in

expression patterns have been observed (Figure 5J–N), indicating that these five tomato class E genes may be involved in multiple floral organ identity. Thus, we believe that the expression patterns of the tomato MADS-box genes identified in our study will be an important tool for understanding the flower development mechanisms in tomato. Previous reports have found that the MADS-box proteins are able to form multiple homologous or heterodimeric complexes in plants, and the combinatorial MADS-box proteins are often deriving their regulatory specificity from other DNA binding or accessory factors. To understand how the tomato MADS-box genes can act in flower development and floral organ identity, it is necessary to identify the network of protein–protein interactions amongst them. Therefore, the predicted interaction networks of the 15 tomato MADS-box proteins, which are involved in floral organ development, were analyzed in our report (Figure S4). In many domesticated crops, it's an important way to select inflorescence architecture with improved flower production and yield. In tomato, SlMBP21 forms protein complexes with JOINTLESS and MACROCALYX as a transcription activator for tomato flower abscission zone development [50], because SlMBP21, J, and MC have a common function in the development of the tomato flower abscission zone. In breeding, altering any of these genes will have the function on plant growth. In this study, the predicted interaction networks may help us to understand how the tomato MADS-box genes can act in flower development and floral organ identity.

3.3. Potential Functions of Tomato MADS-Box Genes during Fruit Development

To better understand the roles of tomato MADS-box genes in fruit development and ripening, we selected five tomato MADS-box genes (*SlMBP3*, *MADS-RIN*, *TAGL1*, *TM4*, and *SlMBP7*) that were predicted to abundantly expressed at different stages of fruit development and ripening. Then, we detected their relative expression level in fruits samples from five different stages of development by qPCR (Figure 6). As shown in Figure 6A, the expression level of *SlMBP3* was found to be higher at the MG stage than at the other stages. We found that the expression levels of *MADS-RIN* and *TAGL1* exhibited an increasing tendency during the transition from the MG stage to the B+4 stage, and then decreased at the B+7 stage (Figure 6B,C). Moreover, the expression level of *TM4* was found to increase continuously in the process of fruit development (Figure 6D). The *SlMBP7* gene showed a relatively high expression level at the B stage (Figure 6E). Recent reports have identified a number of MADS-box genes that are required for the regulation of fruit development and ripening. One of the most representative is the tomato *MADS-RIN* gene, which is one of the earliest acting ripening regulators, and plays crucial roles in fruit ripening through ethylene dependent and independent ripening regulatory pathways [19,91]. In addition, *TAGL1*, *TM4*, and *SlMBP7* have been found to regulate fruit ripening in tomato [21,22]. Since our results were consistent with the functional research of these genes, the *SlMBP3* gene was predicted to be particularly high in the MG stage, indicating that it might play an important function in fruit development and ripening. These results help to advance our understanding of the function of MADS-box genes in the regulation of fruit developmental and ripening processes in tomato.

4. Materials and Methods

4.1. Plant Material and Growth Condition

In this article, tomatoes (*Solanum lycopersicum*, 'Ailsa Craig' AC^{++}) from Laboratory of molecular biology of tomato, Bioengineering College, Chongqing University, Chongqing, China, were grown in controlled greenhouse conditions of a 16-h day (25 °C)/8-h night (18 °C) cycle, 80% humidity, a 250-$\mu mol \cdot m^{-2} \cdot s^{-1}$ light intensity, and were managed routinely. The tomato flowers were tagged at anthesis and floral organs: sepals (Se), petals (Pe), stamens (St), and carpels (Ca) were collected from flowers at anthesis. The fruit color and days post-anthesis (DPA) were used to differentiate the ripening days of tomato fruits. We defined 20 DPA as the immature green (IMG), and 35 DPA as the mature green (MG), at which point the fruits are green and shiny and no obvious color change is observed.

The 38-DPA tomato fruits with color change from green to yellow were characterized as breaker (B) fruits. Besides, the samples of B+4 (4 days after breaker) fruits and B+7 (7 days after breaker) fruits were also used in our study. All the samples that we used were frozen in liquid nitrogen immediately and stored at –80 °C.

4.2. Identification of MADS-Box Genes in Tomato

The Sol Genomics Network (SGN, Available online: http://solgenomics.net/) and the National Center for Biotechnology Information (NCBI, Available online: https://www.ncbi.nlm.nih.gov/) database were used to comprehensively identify the whole MADS-box protein sequences of tomato. BLAST searches, using all the Arabidopsis and rice MADS-box protein sequences as queries, were performed to check the predicted tomato MADS-box protein sequences in the database. Subsequently, we further examined all the candidate protein sequences by the PROSITE (Available online: http://www.expasy.org/prosite/) and SMART (Available online: http://smart.embl-heidelberg.de/) databases for reliability. The tomato protein sequences, containing the typical conserved domain of the MADS-box protein family, were selected for amino acid sequence multiple alignment and phylogenetic tree analysis. Then, we obtained their DNA sequences according to their amino acid sequence from the SGN database. Additionally, the molecular weight and isoelectric points of tomato MADS-box proteins were detected by the ExPASy proteomics server.

4.3. Phylogenetic Analysis of Tomato MADS-Box Proteins

Multiple sequence alignment for the two groups of all the 131 tomato MADS-box genes (Table 1) was generated using ClustalX 1.81. The alignment results were used to conduct a phylogenetic tree by the MEGA5.02 program, and the evolutionary history was inferred using the neighbor-joining method. [92].

4.4. The Analysis of Gene Structure and Conserved Motif

The tomato MADS-box coding domain sequences (CDS) and corresponding genomic DNA sequences were collected from SGN and NCBI databases to predict gene structure. The online tool Gene Structure Display Server 2.0 (GSDS 2.0, Available online: http://gsds.cbi.pku.edu.cn/index.php), was used to construct an exon/intron map [93].

Conserved motifs of the tomato MADS-box protein sequences were identified by online software MEME Version 4.12.0 (Available online: http://meme-suite.org/tools/meme). It was performed with the following parameters: 10 different motifs, a motif width of 6–200 amino acids, and any number of repetitions. The SMART database was used to annotate the MEME motifs.

4.5. Chromosomal Locations and Identification of Interaction Network

To determine the chromosomal locations of tomato MADS-box genes, we obtained the physical genome annotation files from the SGN and NCBI database. The physical map of the tomato MADS-box genes was drawn by the Tomato-EXPEN 2000 (Available online: https://solgenomics.net/cview/map.pl).

The interaction network was conducted by STRING (functional protein association networks, Available online: https://string-db.org/) software using the search of multiple proteins sequences [94].

4.6. Digital Gene Expression Analysis of Tomato MADS-Box Genes

To obtain the expression profile of tomato MADS-box genes, the RNA-seq data based on the locus/gene names of SGN were analyzed. We downloaded the RNA-seq data from TFGD (the Tomato Functional Genomics Database), and the sequence data were obtained from various tissues in wild species LA1589 (*S. pimpinellifolium*). In addition, the data of the tomato lab and Tomato eFP Browser were also used to analyze the gene expression of tomato MADS-box genes. A heatmap was generated

by Heml 1.0 (Heatmap illustrator, Available online: http://hemi.biocuckoo.org/) using the relative expression values or ratios of each tomato MADS-box gene [95].

4.7. Total RNA Extraction and qPCR Analysis

To study the expression patterns of the MADS-box genes involved in flower organ identity and fruit development in tomato, total RNA was extracted from the sepals (Se), petals (Pe), stamens (St), and carpels (Ca) of tomato and different developmental stages of tomato fruits at different developmental stages, including IMG (immature green), MG (mature green), B (breaker), B+4 (4 days after breaker), and B+7 (7 days after breaker) using RNAiso Plus (Takara) in accordance with the instructions. After DNase digestion (Promega, Madison, WI, USA), cDNA was synthesized with oligo(dT)20 as a primer for RNA reverse-transcription using M-MLV reverse transcriptase (Promega, Madison, WI, USA). For gene expression quantification, qPCR analysis was performed with the CFX96™ Real-Time System (Bio-Rad, Hercules, CA, USA) using the GoTaq qPCR Master Mix (Promega, Madison, WI, USA). First, 1.0 µL of mixture primers, 1.0 µL of cDNA, and 3.0 µL of ddH_2O were used. NRT (no reverse transcription control) and NTC (no template control) experiments were performed to eliminate the genomic DNA and environment effects. The tomato *SlCAC* gene was used as an internal standard [96], and the $2^{-\Delta\Delta CT}$ method was used to perform the relative gene expression levels analysis [97]. In addition, all the experiments were performed in three biological triplicates with three technical replicates. The standard curves were run at the same time. All the primers used were designed by Primer 5.0 software and are shown in Table S3.

4.8. Data Analysis

In this study, the mean values of data are presented as mean ± standard deviation. The Origin 8.0 software (Available online: https://www.originlab.com/) was used to perform the data analysis, and mean differences were significant by t-test (* $p < 0.05$).

5. Conclusions

In this study, a comprehensive and systematic analysis of the tomato *MADS-box* transcription factor family was first conducted. A total of 131 genes encoding MADS-box transcription factors, including 81 type I and 50 type II genes, were extensively identified in the tomato genome. Then, we classified the genes according to their phylogenetic relationships between tomato and Arabidopsis. The phylogenetic relationships, gene structures, conserved motifs, chromosomal distribution, and expression patterns of the genes were characterized. The 131 tomato MADS-box genes showed differential expression levels in different organs. Since the MADS-box genes are the most powerful TFs that regulate floral organ identity and fruit development and ripening in plants, we showed that 15 tomato MADS-box genes were involved in floral organ development, and we studied the expression of five tomato MADS-box genes in different stages of fruit development and ripening. These results provide evidence of the relationship between MADS-box genes and floral organ and fruit development. Our study provides comprehensive information on the tomato MADS-box gene family, enables a better understanding of the structure–function relationships among the tomato MADS-box gene family members, and lays a solid foundation of comprehensive functional characteristics in the tomato MADS-box gene family. Furthermore, our bioinformatics and evolutionary analysis will be helpful for better understanding the underlying evolutionary relationship of the MADS-box family in higher plants.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/12/2961/s1.

Author Contributions: G.C. and Z.H. designed and managed the research work and improved the manuscript; Y.W. and X.G. performed the bioinformatics analysis; J.Z. performed the experiments; Y.W. wrote the manuscript; Y.W. and J.Z. prepared all figures. All the authors reviewed the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (31572129), the National Natural Science Foundation of Chongqing of China (cstc2018jcyjAX0458) and Training Program of Chongqing University Bioengineering College (0221001105301).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shore, P.; Sharrocks, A.D. The MADS-box family of transcription factors. *Eur. J. Biochem.* **1995**, *229*, 1–13. [CrossRef] [PubMed]
2. Schwarz-Sommer, Z.; Sommer, H. Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. *Science* **1990**, *250*, 931–936. [CrossRef] [PubMed]
3. Alvarezbuylla, E.R.; Pelaz, S.; Liljegren, S.J.; Gold, S.E.; Burgeff, C.; Ditta, G.S.; Ribas, d.P.L.; Martínezcastilla, L.; Yanofsky, M.F. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5328–5333. [CrossRef] [PubMed]
4. Smaczniak, C.; Immink, R.G.; Angenent, G.C.; Kaufmann, K. Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. *Development* **2012**, *139*, 3081–3098. [CrossRef] [PubMed]
5. Henschel, K.; Kofuji, R.; Hasebe, M.; Saedler, H.; Münster, T.; Theissen, G. Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. *Mol. Biol. Evol.* **2002**, *19*, 801–814. [CrossRef] [PubMed]
6. Parenicová, L.; De, F.S.; Kieffer, M.; Horner, D.S.; Favalli, C.; Busscher, J.; Cook, H.E.; Ingram, R.M.; Kater, M.M.; Davies, B. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. *Plant Cell* **2003**, *15*, 1538–1551. [CrossRef] [PubMed]
7. Arora, R.; Agarwal, P.; Ray, S.; Singh, A.K.; Singh, V.P.; Tyagi, A.K.; Kapoor, S. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom.* **2007**, *8*, 242. [CrossRef]
8. Yanofsky, M.F.; Ma, H.; Bowman, J.L.; Drews, G.N.; Feldmann, K.A.; Meyerowitz, E.M. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **1990**, *346*, 35–39. [CrossRef]
9. Coen, E.S.; Meyerowitz, E.M. The war of the whorls: Genetic interactions controlling flower development. *Nature* **1991**, *353*, 31–37. [CrossRef]
10. Causier, B.; Schwarz-Sommer, Z.; Davies, B. Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* **2010**, *21*, 73–79. [CrossRef]
11. Krizek, B.A.; Meyerowitz, E.M. The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development* **1996**, *122*, 11–22. [PubMed]
12. Theissen, G. Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **2001**, *4*, 75–85. [CrossRef]
13. Mandel, M.A.; Gustafson-Brown, C.; Savidge, B.; Yanofsky, M.F. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature* **1992**, *360*, 273–277. [CrossRef] [PubMed]
14. Jack, T.; Brockman, L.L.; Meyerowitz, E.M. The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. *Cell* **1992**, *68*, 683–697. [CrossRef]
15. Mizukami, Y.; Ma, H. Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell* **1992**, *71*, 119–131. [CrossRef]
16. Pelaz, S.; Ditta, G.S.; Baumann, E.; Wisman, E.; Yanofsky, M.F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **2000**, *405*, 200–203. [CrossRef]
17. Angenent, G.C.; Franken, J.; Busscher, M.; Van, D.A.; van Went, J.L.; Dons, H.J.; van Tunen, A.J. A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* **1995**, *7*, 1569–1582. [CrossRef]
18. Colombo, L.; Franken, J.; Koetje, E.; Went, J.V.; Dons, H.J.; Angenent, G.C.; Tunen, A.J.V. The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell* **1995**, *7*, 1859–1868. [CrossRef]
19. Vrebalov, J.; Ruezinsky, D.; Padmanabhan, V.; White, R.; Medrano, D.; Drake, R.; Schuch, W.; Giovannoni, J. A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* **2002**, *296*, 343–346. [CrossRef]

20. Itkin, M.; Seybold, H.; Breitel, D.; Rogachev, I.; Meir, S.; Aharoni, A. TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. *Plant J. Cell Mol. Biol.* **2009**, *60*, 1081–1095. [CrossRef]
21. Arroyo, A.J.M.; Mcquinn, R.; Poole, M.; Seymour, G.; Giovannoni, J.; Irish, V.F.; Vrebalov, J.; Pan, I.L.; Chung, M.; Rose, J. Fleshy fruit expansion and ripening are regulated by the Tomato SHATTERPROOF gene TAGL1. *Plant Cell* **2009**, *21*, 3041–3062.
22. Bemer, M.; Karlova, R.; Ballester, A.R.; Tikunov, Y.M.; Bovy, A.G.; Wolters-Arts, M.; Rossetto, P.B.; Angenent, G.C.; de Maagd, R.A. The Tomato FRUITFULL Homologs TDR4/FUL1 and MBP7/FUL2 Regulate Ethylene-Independent Aspects of Fruit Ripening. *Plant Cell* **2012**, *24*, 4437–4451. [CrossRef] [PubMed]
23. Shima, Y.; Kitagawa, M.; Fujisawa, M.; Nakano, T.; Kato, H.; Kimbara, J.; Kasumi, T.; Ito, Y. Tomato FRUITFULL homologues act in fruit ripening via forming MADS-box transcription factor complexes with RIN. *Plant Mol. Biol.* **2013**, *82*, 427–438. [CrossRef] [PubMed]
24. Byeong-ha Lee, D.A.H.; Zhu, J.-K. The Arabidopsis Cold-Responsive Transcriptome and Its Regulation by ICE1. *Plant Cell* **2005**, *17*, 3155–3175. [PubMed]
25. Lee, I.; Ronald, P.C. Genetic dissection of the biotic stress response using a genome-scale gene network for rice. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18548–18553. [CrossRef] [PubMed]
26. Tardif, G.; Kane, N.A.; Adam, H.; Labrie, L.; Major, G.; Gulick, P.; Sarhan, F.; Laliberté, J.F. Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Mol. Biol.* **2007**, *63*, 703–718. [CrossRef] [PubMed]
27. Duan, W.; Song, X.; Liu, T.; Huang, Z.; Ren, J.; Hou, X.; Li, Y. Genome-wide analysis of the MADS-box gene family in Brassica rapa (Chinese cabbage). *Mol. Genet. Genom.* **2015**, *290*, 239–255. [CrossRef]
28. Sommer, H.; Beltrán, J.P.; Huijser, P.; Pape, H.; Lönnig, W.E.; Saedler, H.; Schwarz-Sommer, Z. Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: The protein shows homology to transcription factors. *Embo J.* **1990**, *9*, 605–613. [CrossRef]
29. Leseberg, C.H.; Li, A.; Kang, H.; Duvall, M.; Mao, L. Genome-wide analysis of the MADS-box gene family in Populus trichocarpa. *Gene* **2006**, *378*, 84–94. [CrossRef]
30. Jian, M.; Yang, Y.; Wei, L.; Yang, C.; Ding, P.; Liu, Y.; Qiao, L.; Chang, Z.; Geng, H.; Wang, P. Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum* L.). *PLoS ONE* **2017**, *12*, e0181443.
31. Liu, J.; Zhang, J.; Zhang, J.; Miao, H.; Wang, J.; Gao, P.; Hu, W.; Jia, C.; Wang, Z.; Xu, B. Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening. *Sci. Rep.* **2017**, *7*, 3467. [CrossRef] [PubMed]
32. Busscherlange, J. Analysis of the petunia MADS-box transcription factor family. *Mol. Genet. Genom.* **2003**, *268*, 598–606.
33. Velasco, R.; Zharkikh, A.; Affourtit, J.; Dhingra, A.; Cestaro, A.; Kalyanaraman, A.; Fontana, P.; Bhatnagar, S.K.; Troggio, M.; Pruss, D.; et al. The genome of the domesticated apple (Malus x domestica Borkh.). *Nat. Genet.* **2010**, *42*, 833–839. [CrossRef] [PubMed]
34. Leseberg, C.H.; Eissler, C.L.; Wang, X.; Johns, M.A.; Duvall, M.R.; Mao, L. Interaction study of MADS-domain proteins in tomato. *J. Exp. Bot.* **2008**, *59*, 2253–2265. [CrossRef] [PubMed]
35. Hileman, L.C.; Sundstrom, J.F.; Litt, A.; Chen, M.; Shumba, T.; Irish, V.F. Molecular and phylogenetic analyses of the MADS-box gene family in tomato. *Mol. Biol. Evol.* **2006**, *23*, 2245–2258. [CrossRef] [PubMed]
36. Guo, X.; Hu, Z.; Yin, W.; Yu, X.; Zhu, Z.; Zhang, J.; Chen, G. The tomato floral homeotic protein FBP1-like gene, SlGLO1, plays key roles in petal and stamen development. *Sci. Rep.* **2016**, *6*, 20454. [CrossRef] [PubMed]
37. Martino, G.D.; Pan, I.; Emmanuel, E.; Levy, A.; Irish, V.F. Functional Analyses of Two Tomato APETALA3 Genes Demonstrate Diversification in Their Roles in Regulating Floral Development. *Plant Cell* **2006**, *18*, 1833–1845. [CrossRef]
38. Huang, B.; Routaboul, J.M.; Liu, M.; Deng, W.; Maza, E.; Mila, I.; Hu, G.; Zouine, M.; Frasse, P.; Vrebalov, J.T. Overexpression of the class D MADS-box gene Sl-AGL11 impacts fleshy tissue differentiation and structure in tomato fruits. *J. Exp. Bot.* **2017**, *68*, 4869–4884. [CrossRef]
39. Ampomahdwamena, C.; Morris, B.A.; Sutherland, P.; Veit, B.; Yao, J.L. Down-regulation of TM29, a tomato SEPALLATA homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol.* **2002**, *130*, 605–617. [CrossRef]
40. Litt, A.; Irish, V.F. Duplication and Diversification in the APETALA1/FRUITFULL Floral Homeotic Gene Lineage: Implications for the Evolution of Floral Development. *Genetics* **2003**, *165*, 821–833.

41. Yao, J.L.; Ampomah-Dwamena, C. The tomato SEPALLATA homolog TM29 has diverse functions in regulating flower and fruit development. *Plant Physiol.* **2003**, *35*, 28–33.
42. Wang, S.; Gang, L.; Zheng, H.; Luo, Z.; Wang, T.; Li, H.; Zhang, J.; Ye, Z. Members of the tomato FRUITFULL MADS-box family regulate style abscission and fruit ripening. *J. Exp. Bot.* **2014**, *65*, 3005–3014. [CrossRef] [PubMed]
43. Shima, Y.; Fujisawa, M.; Kitagawa, M.; Nakano, T.; Kimbara, J.; Nakamura, N.; Shiina, T.; Sugiyama, J.; Nakamura, T.; Kasumi, T. Tomato FRUITFULL homologs regulate fruit ripening via ethylene biosynthesis. *J. Agric. Chem. Soc. Jpn.* **2014**, *78*, 231–237.
44. Burko, Y.; Shleizer-Burko, S.; Yanai, O.; Shwartz, I.; Zelnik, I.D.; Jacob-Hirsch, J.; Kela, I.; Eshed-Williams, L.; Ori, N. A role for APETALA1/fruitfull transcription factors in tomato leaf development. *Plant Cell* **2013**, *25*, 2070–2083. [CrossRef] [PubMed]
45. Yin, W.; Hu, Z.; Hu, J.; Zhu, Z.; Yu, X.; Cui, B.; Chen, G. Tomato (*Solanum lycopersicum*) MADS-box transcription factor SlMBP8 regulates drought, salt tolerance and stress-related genes. *Plant Growth Regul.* **2017**, *83*, 55–68. [CrossRef]
46. Yin, W.; Hu, Z.; Cui, B.; Guo, X.; Hu, J.; Zhu, Z.; Chen, G. Suppression of the MADS-box gene SlMBP8 accelerates fruit ripening of tomato (*Solanum lycopersicum*). *Plant Physiol. Biochem.* **2017**, *118*, 235–244. [CrossRef] [PubMed]
47. Guo, X.; Chen, G.; Cui, B.; Gao, Q.; Guo, J.E.; Li, A.; Zhang, L.; Hu, Z. Solanum lycopersicum agamous-like MADS-box protein AGL15-like gene, SlMBP11, confers salt stress tolerance. *Mol. Breed.* **2016**, *36*, 125. [CrossRef]
48. Xie, Q.; Hu, Z.; Zhu, Z.; Dong, T.; Zhao, Z.; Cui, B.; Chen, G. Overexpression of a novel MADS-box gene SlFYFL delays senescence, fruit ripening and abscission in tomato. *Sci. Rep.* **2014**, *4*, 4367. [CrossRef]
49. Li, N.; Huang, B.; Tang, N.; Jian, W.; Zou, J.; Chen, J.; Cao, H.; Habib, S.; Dong, X.; Wei, W. The MADS-box Gene SlMBP21 Regulates Sepal Size Mediated by Ethylene and Auxin in Tomato. *Plant Cell Physiol.* **2017**, *58*, 2241–2256. [CrossRef]
50. Liu, D.; Wang, D.; Qin, Z.; Zhang, D.; Yin, L.; Wu, L.; Colasanti, J.; Li, A.; Mao, L. The SEPALLATA MADS-box protein SLMBP21 forms protein complexes with JOINTLESS and MACROCALYX as a transcription activator for development of the tomato flower abscission zone. *Plant J.* **2014**, *77*, 284–296. [CrossRef]
51. Roldan, M.V.G.; Périlleux, C.; Morin, H.; Huerga-Fernandez, S.; Latrasse, D.; Benhamed, M.; Bendahmane, A. Natural and induced loss of function mutations inSlMBP21MADS-box gene led tojointless-2phenotype in tomato. *Sci. Rep.* **2017**, *7*, 4402. [CrossRef] [PubMed]
52. Gimenez, E.; Castañeda, L.; Pineda, B.; Pan, I.L.; Moreno, V.; Angosto, T.; Lozano, R. TOMATO AGAMOUS1 and ARLEQUIN/TOMATO AGAMOUS—LIKE1 MADS-box genes have redundant and divergent functions required for tomato reproductive development. *Plant Mol. Biol.* **2016**, *91*, 513–531. [CrossRef] [PubMed]
53. Pnueli, L.; Hareven, D.; Rounsley, S.D.; Yanofsky, M.F.; Lifschitz, E. Isolation of the Tomato AGAMOUS Gene TAG1 and Analysis of Its Homeotic Role in Transgenic Plants. *Plant Cell* **1994**, *6*, 163–173. [CrossRef] [PubMed]
54. Giménez, E.; Pineda, B.; Capel, J.; Antón, M.T.; Atarés, A.; Pérezmartín, F.; Garcíasogo, B.; Angosto, T.; Moreno, V.; Lozano, R. Functional analysis of the Arlequin mutant corroborates the essential role of the ARLEQUIN/TAGL1 gene during reproductive development of tomato. *PLoS ONE* **2010**, *5*, e14427. [CrossRef] [PubMed]
55. Gaffe, J.; Lemercier, C.; Alcaraz, J.P.; Kuntz, M. Identification of three tomato flower and fruit MADS-box proteins with a putative histone deacetylase binding domain. *Gene* **2011**, *471*, 19–26. [CrossRef] [PubMed]
56. Busi, M.V.; Bustamante, C.; D'Angelo, C.; Hidalgocuevas, M.; Boggio, S.B.; Valle, E.M.; Zabaleta, E. MADS-box genes expressed during tomato seed and fruit development. *Plant Mol. Biol.* **2003**, *52*, 801–815. [PubMed]
57. Kramer, E.M.; Dorit, R.L.; Irish, V.F. Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. *Genetics* **1998**, *149*, 765–783.
58. Pracros, P.; Renaudin, J.; Eveillard, S.; Mouras, A.; Hernould, M. Tomato flower abnormalities induced by stolbur phytoplasma infection are associated with changes of expression of floral development genes. *Mol. Plant Microbe Interact.* **2006**, *19*, 62–68. [CrossRef]

59. Fujisawa, M.; Shima, Y.; Nakagawa, H.; Kitagawa, M.; Kimbara, J.; Nakano, T.; Kasumi, T.; Ito, Y. Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. *Plant Cell* **2014**, *26*, 89–101. [CrossRef]
60. Ito, Y. Regulation of Tomato Fruit Ripening by MADS-Box Transcription Factors. *Jpn. Agric. Res. Q.* **2016**, *50*, 33–38. [CrossRef]
61. Martel, C.; Vrebalov, J.; Tafelmeyer, P.; Giovannoni, J.J. The tomato (*Solanum lycopersicum*) MADS-box transcription factor RIN interacts with promoters involved in numerous ripening processes in a CNR dependent manner. *Plant Physiol.* **2011**, *157*, 1568–1579. [CrossRef] [PubMed]
62. Ito, Y.; Kitagawa, M.; Ihashi, N.; Yabe, K.; Kimbara, J.; Yasuda, J.; Ito, H.; Inakuma, T.; Hiroi, S.; Kasumi, T. DNA-binding specificity, transcriptional activation potential, and the rin mutation effect for the tomato fruit-ripening regulator RIN. *Plant J.* **2008**, *55*, 212–223. [CrossRef] [PubMed]
63. Li, L.; Zhu, B.; Fu, D.; Luo, Y. RIN transcription factor plays an important role in ethylene biosynthesis of tomato fruit ripening. *J. Sci. Food Agric.* **2011**, *91*, 2308–2314. [CrossRef] [PubMed]
64. Nakano, T.; Kimbara, J.; Fujisawa, M.; Kitagawa, M.; Ihashi, N.; Maeda, H.; Kasumi, T.; Ito, Y. MACROCALYX and JOINTLESS interact in the transcriptional regulation of tomato fruit abscission zone development. *Plant Physiol.* **2012**, *158*, 439–450. [CrossRef] [PubMed]
65. Wang, X.; Yin, J. Functional Studies of JOINTLESS and Its Interacting MADS-domain Proteins in Tomato. *Acta Hortic. Sin.* **2011**, *38*, 701–708.
66. Nakano, T.; Kato, H.; Shima, Y.; Ito, Y. Apple SVP family MADS-box proteins and the tomato pedicel abscission zone regulator JOINTLESS have similar molecular activities. *Plant Cell Physiol.* **2015**, *56*, 1097–1106. [CrossRef] [PubMed]
67. Szymkowiak, E.J.; Irish, E.E. JOINTLESS suppresses sympodial identity in inflorescence meristems of tomato. *Planta* **2006**, *223*, 646–658. [CrossRef]
68. Mao, L.; Begum, D.; Chuang, H.W.; Budiman, M.A.; Szymkowiak, E.J.; Irish, E.E.; Wing, R.A. JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development. *Nature* **2000**, *406*, 910–913. [CrossRef]
69. Pnueli, L.; Abuabeid, M.; Zamir, D.; Nacken, W.; Schwarzsommer, Z.; Lifschitz, E. The MADS box gene family in tomato: Temporal expression during floral development, conserved secondary structures and homology with homeotic genes from Antirrhinum and Arabidopsis. *Plant J.* **1991**, *1*, 255–266. [CrossRef]
70. Daminato, M.; Masiero, S.; Resentini, F.; Lovisetto, A.; Casadoro, G. Characterization of TM8, a MADS-box gene expressed in tomato flowers. *BMC Plant Biol.* **2014**, *14*, 319. [CrossRef]
71. Dong, T.; Hu, Z.; Deng, L.; Wang, Y.; Zhu, M.; Zhang, J.; Chen, G. A Tomato MADS-Box Transcription Factor, SlMADS1, Acts as a Negative Regulator of Fruit Ripening. *Plant Physiol.* **2013**, *163*, 1026–1036. [CrossRef] [PubMed]
72. Zhang, J.; Hu, Z.; Yao, Q.; Guo, X.; Nguyen, V.; Li, F.; Chen, G. A tomato MADS-box protein, SlCMB1, regulates ethylene biosynthesis and carotenoid accumulation during fruit ripening. *Sci. Rep.* **2018**, *8*, 3413. [CrossRef] [PubMed]
73. Wang, R.; Ming, M.; Li, J.; Shi, D.; Qiao, X.; Li, L.; Zhang, S.; Wu, J. Genome-wide identification of the MADS-box transcription factor family in pear (Pyrus bretschneideri) reveals evolution and functional divergence. *Peerj* **2017**, *5*, e3776. [CrossRef] [PubMed]
74. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **1994**, *256*, 119–124. [PubMed]
75. Alvarez-Buylla, E.R.; Liljegren, S.J.; Pelaz, S.; Gold, S.E.; Burgeff, C.; Ditta, G.S.; Vergara-Silva, F.; Yanofsky, M.F. MADS-box gene evolution beyond flowers: Expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J.* **2000**, *24*, 457–466. [CrossRef] [PubMed]
76. Causier, B.; Kieffer, M.; Davies, B. Plant biology. MADS-box genes reach maturity. *Science* **2002**, *296*, 275–276. [CrossRef] [PubMed]
77. Rijpkema, A.; Gerats, T.; Vandenbussche, M. Genetics of Floral Development in Petunia. *Adv. Bot. Res.* **2006**, *44*, 237–278.
78. Becker, A.; Theißen, G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* **2003**, *29*, 464–489. [CrossRef]
79. Rounsley, S.D.; Ditta, G.S.; Yanofsky, M.F. Diverse roles for MADS box genes in Arabidopsis development. *Plant Cell* **1995**, *7*, 1259–1269. [CrossRef]

80. Airoldi, C.A.; Davies, B. Gene Duplication and the Evolution of Plant MADS-box Transcription Factors. *J. Genet. Genom.* **2012**, *39*, 157–165. [CrossRef]
81. Liu, Y.; Jiang, H.Y.; Chen, W.; Qian, Y.; Ma, Q.; Cheng, B.; Zhu, S. Genome-wide analysis of the auxin response factor (ARF) gene family in maize (Zea mays). *Plant Growth Regul.* **2011**, *63*, 225–234. [CrossRef]
82. Ma, H. The unfolding drama of flower development: Recent results from genetic and molecular analyses. *Genes Dev.* **1994**, *8*, 745–756. [CrossRef] [PubMed]
83. Pelaz, S.; Tapialópez, R.; Alvarezbuylla, E.R.; Yanofsky, M.F. Conversion of leaves into petals in Arabidopsis. *Curr. Biol.* **2001**, *11*, 182–184. [CrossRef]
84. Davies, B.; Egeacortines, M.; De, E.A.S.; Saedler, H.; Sommer, H. Multiple interactions amongst floral homeotic MADS box proteins. *Embo J.* **1996**, *15*, 4330–4343. [CrossRef] [PubMed]
85. Kim, S.; Koh, J.; Yoo, M.J.; Kong, H.; Hu, Y.; Ma, H.; Soltis, P.S.; Soltis, D.E. Expression of floral MADS-box genes in basal angiosperms: Implications for the evolution of floral regulators. *Plant J. Cell Mol. Biol.* **2005**, *43*, 724–744. [CrossRef] [PubMed]
86. Theissen, G.; Becker, A.; Rosa, A.D.; Kanno, A.; Kim, J.T.; Münster, T.; Winter, K.U.; Saedler, H. A short history of MADS-box genes in plants. *Plant Mol. Biol.* **2000**, *42*, 115–149. [CrossRef] [PubMed]
87. Angenent, G.C.; Franken, J.; Busscher, M.; Colombo, L.; van Tunen, A.J. Petal and stamen formation in petunia is regulated by the homeotic gene fbp1. *Plant J.* **1993**, *4*, 101–112. [CrossRef] [PubMed]
88. Bowman, J.L.; Alvarez, J.; Weigel, D.; Meyerowitz, E.M.; Smyth, D.R. Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development* **1993**, *119*, 721–743.
89. Poupin, M.; Federici, F.; Medina, C.; Mattis, J.; Timmermann, T.J.P. Isolation of the three grape sub-lineages of B-class MADS-box TM6, PISTILLATA and APETALA3 genes which are differentially expressed during flower and fruit development. *Gene* **2007**, *404*, 10–24. [CrossRef]
90. Cheng, X.F.; Wittich, P.E.; Kieft, H.; Angenent, G.; Xuhan, X.; Lammeren, A.A.M.V. Temporal and Spatial Expression of MADS Box Genes, FBP7 and FBP11, During Initiation and Early Development of Ovules in Wild Type and Mutant Petunia hybrida. *Plant Biol.* **2000**, *2*, 693–702. [CrossRef]
91. Fujisawa, M.; Nakano, T.; Shima, Y.; Ito, Y. A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. *Plant Cell* **2013**, *25*, 371–386. [CrossRef] [PubMed]
92. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **2014**, *28*, 2731–2739. [CrossRef] [PubMed]
93. Guo, A.Y.; Zhu, Q.H.; Chen, X.; Luo, J.C. GSDS: A gene structure display server. *Hereditas* **2007**, *29*, 1023–1026. [CrossRef] [PubMed]
94. Lu, L.J.; Zhang, M. *Protein-Protein Interaction Networks*; Springer: New York, NY, USA, 2013; pp. 141–181.
95. Deng, W.; Wang, Y.; Liu, Z.; Cheng, H.; Xue, Y. HemI: A toolkit for illustrating heatmaps. *PLoS ONE* **2014**, *9*, e111988. [CrossRef] [PubMed]
96. Expósito-Rodríguez, M.; Borges, A.A.; Borges-Pérez, A.; Pérez, J.A. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* **2008**, *8*, 131. [CrossRef] [PubMed]
97. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Isolation and Characterization of CsWRKY7, a Subgroup IId WRKY Transcription Factor from *Camellia sinensis*, Linked to Development in Arabidopsis

Wei Chen [1,2], Wan-Jun Hao [1], Yan-Xia Xu [1], Chao Zheng [1], De-Jiang Ni [2], Ming-Zhe Yao [1,*] and Liang Chen [1,*]

[1] Key Laboratory of Tea Biology and Resources Utilization, Ministry of Agriculture; Tea Research Institute Chinese Academy of Agricultural Sciences, 9 South Meiling Road, Hangzhou 310008, Zhejiang, China; chenwei2551@163.com (W.C.); haowanjun@tricaas.com (W.-J.H.); xuyanxia@tricaas.com (Y.-X.X.); zhengchaotea@163.com (C.Z.)

[2] College of Horticulture and Forestry Science, Huazhong Agricultural University, 1 Shizishan Street, Hongshan District, Wuhan 430070, Hubei, China; nidj@mail.hzau.edu.cn

* Correspondence: yaomz@tricaas.com (M.-Z.Y.); liangchen@tricaas.com or chenliang@caas.cn (L.C.); Tel.: +86-571-86650444 (M.-Z.Y.); +86-571-8665-2835 (L.C.)

Received: 27 April 2019; Accepted: 6 June 2019; Published: 9 June 2019

Abstract: WRKY transcription factors (TFs) containing one or two WRKY domains are a class of plant TFs that respond to diverse abiotic stresses and are associated with developmental processes. However, little has been known about the function of *WRKY* gene in tea plant. In this study, a subgroup IId *WRKY* gene *CsWRKY7* was isolated from *Camellia sinensis*, which displayed amino acid sequence homology with Arabidopsis *AtWRKY7* and *AtWRKY15*. Subcellular localization prediction indicated that CsWRKY7 localized to nucleus. Cis-acting elements detected in the promotor region of *CsWRKY7* are mainly involved in plant response to environmental stress and growth. Consistently, expression analysis showed that *CsWRKY7* transcripts responded to NaCl, mannitol, PEG, and diverse hormones treatments. Additionally, *CsWRKY7* exhibited a higher accumulation both in old leaves and roots compared to bud. Seed germination and root growth assay indicated that overexpressed *CsWRKY7* in transgenic Arabidopsis was not sensitive to NaCl, mannitol, PEG, and low concentration of ABA treatments. *CsWRKY7* overexpressing Arabidopsis showed a late-flowering phenotype under normal conditions compared to wild type. Furthermore, gene expression analysis showed that the transcription levels of the flowering time integrator gene *FLOWERING LOCUS T* (*FT*) and the floral meristem identity genes *APETALA1* (*AP1*) and *LEAFY* (*LFY*) were lower in *WRKY7*-OE than in the WT. Taken together, these findings indicate that CsWRKY7 TF may participate in plant growth. This study provides a potential strategy to breed late-blooming tea cultivar.

Keywords: subgroup IId; CsWRKY7; flowering; Arabidopsis; *Camellia sinensis*

1. Introduction

Plants suffering from diverse abiotic stresses in the developmental process have evolved and obtained a series of mechanisms to combat with these environmental stresses. Majority of transcription factors (TFs) participate in these adaptive mechanisms [1]. WRKY TFs—a large family of transcription factors in plants—are involved in response to multiple stresses and external stimuli [2–4].

WRKY transcription factor is named after the WRKY domain, which contains one or two highly conserved WRKYGQK motifs and one zinc-finger motif [5,6]. According to the number of WRKY domains and the type of zinc finger motif, WRKY TFs can be divided into three types: Group I,

containing two WRKY domains and one C2H2 zinc-finger structure. Group II, containing one WRKY domain, and sharing the same zinc finger structure with group I. In addition, based on the amino acid sequence of the DNA-binding domain, Group II can be further divided into IIa, IIb, IIc, IId, and IIe subgroups; Group III also contains one WRKY domain, but its zinc finger structure is C2HC [4,7].

SWEET POTATO FACTOR1 (SPF1) specifically binding to W-box, the first defined WRKY TF, negatively regulates the expression of storage proteins and β-amylase in sweet potato [8]. Since then, considerable efforts have been made to unveil the role of WRKY TFs in plant response to biotic and abiotic stresses. In *Arabidopsis thaliana*, it has been demonstrated that WRKY TFs contribute to pathogen resistance, salinity, heat, and drought stress responses [7,9–11]. W-box (TTGACT/C) was present in the promoters of several stress-associated genes. Since WRKY proteins bind to W-box, they induce the expressions of these stress-associated genes [4,5,11–13]. Many studies have shown that WRKY proteins participate in secondary metabolism (phenylpropane, alkaloids, and terpenes.), and hormone signaling [10]. Moreover, WRKY proteins have been demonstrated to involve in plant growth processes, such as leaf senescence, shoot branching, trichome, and flowering. For instance, potato ScWRKY1 [14], Arabidopsis TTG2 [15], and MINISEED3 are involved in seed growth [16]; Arabidopsis AtWRKY6, -45, -53, -57, and -70 are involved in senescence regulation [17–21]; Arabidopsis AtWRKY12, -13, -75, -34, mango MIWRKY12, soybean GsWRKY20, and rice DIf1 have been reported to be involved in plant flowering regulation or pollen development [22–27]. However, few studies have reported WRKY TFs in tea plant.

Tea plant (*Camellia sinensis* L.) is an important commercial woody crop. As a nonalcoholic beverage, tea is processed from tea leaves and is widely consumed worldwide. Early blooming in tea plant and adverse environmental factors affect the quality and yields of tea. Increasing studies have demonstrated that WRKY TFs play pivotal roles in plant growth and abiotic stresses. In one previous study, two WRKY TFs—CsWRKY31 and CsWRKY48—were reported to participate in O-methylated catechin biosynthesis in tea plant (*Camellia sinensis*) [28]. Several *CsWRKY* genes were induced by diverse stresses such as temperature, ABA, and NaCl [29–31]. However, the roles of CsWRKYs in plant growth and development remain unclear. The current study aims to provide a functional characterization of *CsWRKY7*, a member of the group IId WRKY family in tea plant. CsWRKY7 was a close homolog of Arabidopsis *AtWRKY7* and *AtWRKY15*, two well-characterized Group IId WRKY proteins with important roles in plant defense, leaf senescence, and abiotic stress responses [32–34]. When it is exposed to diverse stresses, *CsWRKY7* is upregulated and localized to the nuclei in both tobacco leaves and Arabidopsis roots. *CsWRKY7*-overexpressing Arabidopsis did not respond to abiotic stresses. However, the overexpression of *CsWRKY7* delayed flowering. Gene analysis revealed the downregulation of several flowering-related genes in *CsWRKY7* overexpression lines. A better understanding of the flowering mechanisms is conducive to breeding late-blooming tea plants.

2. Results

2.1. Isolation and Characterization of CsWRKY7 from C. sinensis

CsWRKY7, one *WRKY* gene, was amplified from tea leaves cDNA by RT-PCR. This amplified *CsWRKY7* contained the complete open reading frame (ORF) of 972 bp encoding 323 amino acids. CsWRKY7 protein had a predicted molecular mass of 35.37 kDa and an isoelectric point of 9.47. Sequence analysis showed that CsWRKY7 in tea plant shared a high similarity (71.94%) to AcWRKY15 in kiwifruit (Genbank: PSS21265). Additionally, CsWRKY7 was found to have a nuclear localization signal (NLS) at 227–230 amino acid region, and have two motifs, namely, HARF structure and a shorter conserved structural motif (C-region which is known as a Ca^{2+}-dependent calmodulin-binding domain) (Figure 1A). Thus, CsWRKY7 was assigned to group IId subfamily. Phylogenetic analysis showed that CsWRKY7 was closely related to AcWRKY15, PtrWRKY7, VvWRKY7, AtWRKY7, and AtWRKY15 (Figure 1B). AtWRKY7 and AtWRKY15 TFs have been reported to be involved in plant defense response to bacterial pathogens, leaf senescence, or mitochondrial-mediated osmotic stress [32–34],

their homologous genes *CsWRKY7* was predicted to be a TF that may participate in plant development and respond to stress treatment.

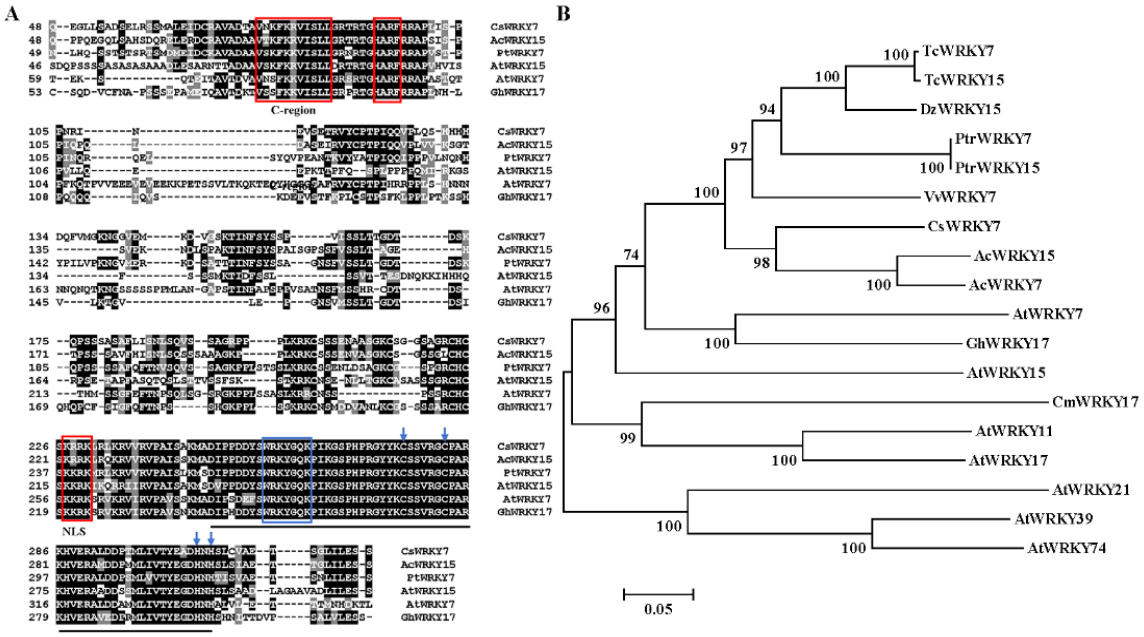


Figure 1. Protein sequence alignment and phylogenetic relationship of CsWRKY7 (**A**) Sequence alignment of the deduced CsWRKY7 protein with other Group IId WRKY homologs. Black lines highlight the conserved region of WRKY. Blue box and arrows highlight the WRKY domain and the zinc-finger motif, respectively. The conserved primary sequences—HARF motif, C-region, and putative NLS—are boxed in red. (**B**) The phylogenetic tree of CsWRKY7 and 17 other group IId WRKY subfamily members. Accession number for group IId WRKY members: AtWRKY7 (AAK28440), AtWRKY11 (AAK96194), AtWRKY15 (AF224704), AtWRKY17 (AAL13049), AtWRKY21 (AAK28441), AtWRKY39 (AAK96198), AtWRKY74 (AAL35291), AcWRKY7 (PSS36058.1), AcWRKY15 (PSS21265.1), VvWRKY7 (RVX23377.1), PtrWRKY7 (XP_006380693.1), PtrWRKY15 (XP-002310122), GhWRKY17 (HQ651068), TcWRKY7 (EOX91521), TcWRKY15 (XP-007047365), CmWRKY17 (AJF11725), and DzWRKY15 (XP-022740807).

2.2. Sequence Analysis of CsWRKY7 Promoter

According to 'Shuchazao' tea genome data, the promoter sequence of *CsWRKY7* was amplified by PCR. A 1680 bp *CsWRKY7* promoter sequence was cloned and the putative cis-elements were predicted through the PlantCARE database. Two functional elements—TATA-box and CAAT-box—were widely distributed in the promoter region. Additionally, a group of elements which respond to such environmental stresses as phytohormone stress (salicylic acid and auxin), light, plant growth (pollen), and abiotic stresses (anaerobic, sugar, wounding, NaCl, dehydration, and heat) were found (Table 1). Interestingly, many MYB-recognition sites were present in the promoter region of *CsWRKY7*, and these motifs also existed in dehydration-responsive gene *RD22A*, indicating that *CsWRKY7* might be involved in dehydration stress and be modulated by MYB members. In addition, *CsWRKY7* promoter region existed two W-boxes, which specially bind to WRKY TFs. These prediction results indicated that CsWRKY7 TF may play a vital role in stress responses and plant growth through multiple signal transduction pathways.

Table 1. Hormone-, light-, and stress-responsive elements in the 1680 bp 5′-flanking sequence of CsWRKY7 TF as predicted by the PlantCARE website.

Types	Site Name	Copy	Sequence	Function
Hormone	TCA-element	2	CCATCTTTTT	salicylic acid responsiveness
	TGA-element	1	AACGAC	auxin-responsive element
Light	Box 4	3	ATTAAT	light responsiveness
	LAMP-element	1	CTTTATCA	light responsive element
	chs-CMA1a	1	TTACTTAA	light responsive element
Stresses	ARE	1	AAACCA	anaerobic induction
	CCAAT-box	1	CAACGG	MYBHv1 binding site
	WBOXHVISO1	2	TGACT	sugar-responsive elements
	WRKY71OS	3	TGAC	A core of TGAC-containing W-box
	WBOXNTERF3	2	TGACY	W-box, wounding
	GT1GMSCAM4	6	GAAAAA	Pathogen- and NaCl-induced expression
	MYB1AT	4	WAACCA	MYB recognition site found in the promoters of *rd22*
	MYB2CONSENSUSAT2		YAACKG	MYB recognition site found in the promoters of *rd22*
	MYBCORE	4	CNGTTR	responsive to water stress
	CCAATBOX1	4	CCAAT	act cooperatively with HSEs
Development	GTGANTG10	4	GTGA	Late pollen-responsive elements
	POLLEN1LELAT52	8	AGAAA	pollen specific expression

2.3. Subcellular Localization of CsWRKY7

To examine the localization of CsWRKY7, the full-length cDNA of *CsWRKY7* was fused to enhanced green fluorescent protein (eGFP). As shown in Figure 2A, *CsWRKY7*-eGEP was localized to the nucleus when it was transiently expressed in *Nicotiana benthamiana* leaves. The green fluorescence was observed both in the cell membrane and nucleus in the GFP control vector. Similarly, green fluorescence was also observed in the nucleus of transgenic Arabidopsis seedlings (Figure 2B). These results showed that CsWRKY7 was targeted to the nucleus.

The qRT-PCR result indicated that *CsWRKY7* was expressed in all detected tissues. The expression level in old leaves was approximately 4.4 times as high as that in buds (Figure 3A). To investigate the role of CsWRKY7 TF in abiotic stresses, we analyzed the expression patterns of *CsWRKY7*, when it was exposed to temperature, sodium chloride (NaCl), sucrose (Suc), polyethylene glycol (PEG), and mannitol (Man) stresses. As shown in Figure 3B,C, in leaves, *CsWRKY7* was downregulated under high temperature stress, especially at 8 h, its expression level decreased by 65%. When exposed to low temperature, the expression level of *CsWRKY7* was not significantly changed at 4 h or 12 h. However, the expression level of *CsWRKY7* was upregulated under the treatments of NaCl, Man, and PEG. For example, its expression level increased gradually after NaCl and PEG treatment for 3 h, reaching a peak of 3.26- and 6.23- fold at 72 h, respectively. Though two sugar-responsive elements existed in the promoter region, no significant difference in the expression level of *CsWRKY7* after sucrose treatment. In order to investigate whether *CsWRKY7* expression was regulated by phytohormones, two-year-old tea plants were exposed to the major plant hormones including indolyl-3-acetic acid (IAA), naphthalene-1-acetic acid (NAA), 2,4-dichlorophenoxyacetic acid (2,4-D), abscisic acid (ABA), methyl jasmonate (MeJA), salicylic acid (SA), and gibberellins (GA). As presented in Figure 3D, the expression level of *CsWRKY7* was elevated under these hormone treatments except IAA. These results implied that *CsWRKY7* might be involved in the regulation of abiotic stress and hormones networks.

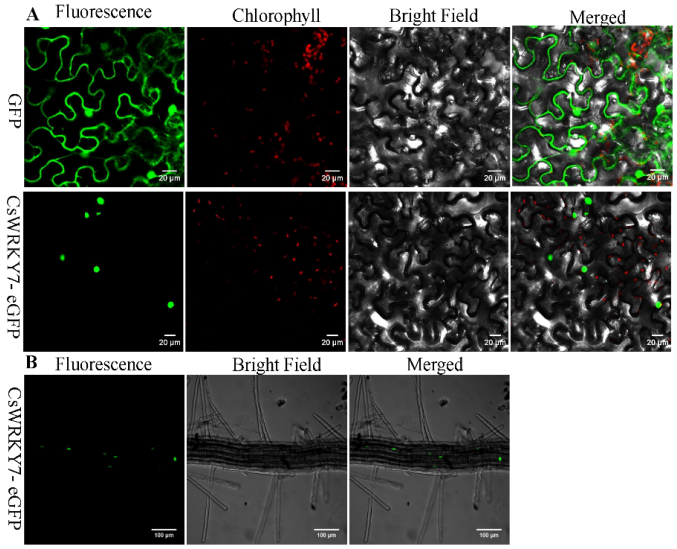


Figure 2. Subcellular localization analysis of CsWRKY7 protein: (**A**) GFP alone (upper panel) and *CsWRKY7*-eGFP (middle panel) were transiently expressed in tobacco epidermal cells. Representative images from left to right in each panel were taken under fluorescence, chlorophyll, transmitted light and an overlay of both channels. Scale bar = 20 μm. (**B**) The roots of *CsWRKY7*-eGFP (bottom panel) in transgenic Arabidopsis was used for observation of GFP fluorescence. Representative images from left to right were taken under fluorescence, bright field and an overlay of both channels. Scale bar = 100 μm.

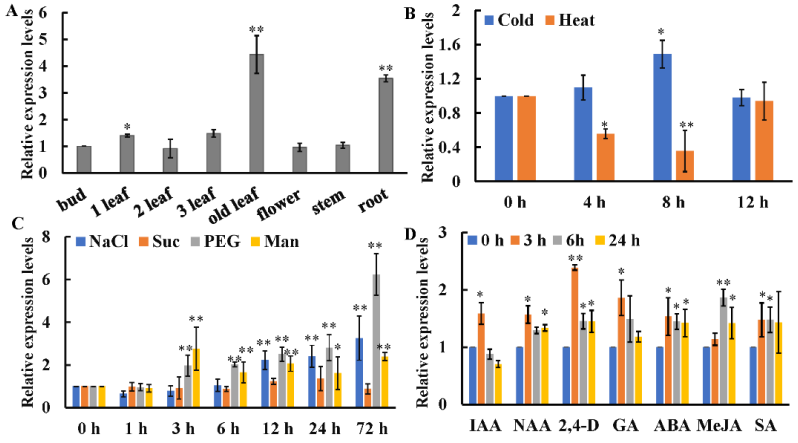


Figure 3. Relative expression levels of *CsWRKY7* in tea plants (**A**) Tissue expression profiles of *CsWRKY7* in 'Longjing 43'. Different tissues include bud, 1st leaf, 2nd leaf, 3rd leaf, old leaf, flower, stem, and root. The expression levels of *CsWRKY7* in different tissues were compared with the bud. (**B**) The transcript levels of *CsWRKY7* under cold (4 °C) and heat (38 °C). *CsWRKY7* expression levels are detected at four different time points (0, 4, 8 and 12 h) post-temperature stress treatment. (**C**) Relative expression levels of *CsWRKY7* under different abiotic stress. Two-year-old tea seedlings were treated with 150 mM NaCl, 90 mM sucrose (Suc), 10% (*w*/*v*) PEG4000 (PEG) and 90 mM mannitol (Man), and samples were harvested at the time intervals indicated. (**D**) The transcript levels of *CsWRKY7* under various phytohormone, including 100 μM indolyl-3-acetic acid (IAA), 5 μM naphthalene-1-acetic acid (NAA), 5 μM 2,4-dichlorophenoxyacetic acid (2,4-D), 100 μM abscisic acid (ABA), 50 μM methyl jasmonate (MeJA), 5 mM salicylic acid (SA), and 100 μM gibberellins (GA), which were added to the culture solution, the functional leaves were harvested at 0, 3, 6, or 24 h post-treatment. Treated samples at 0 h served as controls. Error bars represent ± S.E. for three independent experiments. The significant level is presented by the asterisks (* $p < 0.05$, ** $p < 0.01$).

2.4. Overexpression of CsWRKY7 Affects Flowering in Transgenic Arabidopsis

Seed Germination and Root Growth in Transgenic Plants under Abiotic Stresses

To assess the function of CsWRKY7 TF, an expression construct pH7FGW2.0-*CsWRKY7* was transformed into *A. thaliana*. Three homozygous transgenic lines—L8, L10, and L14—were confirmed by real-time PCR with *Actin-2* gene serving as an internal reference. As shown in Figure 4A, the *CsWRKY7* transcript levels were significantly higher in transgenic lines than in wild type.

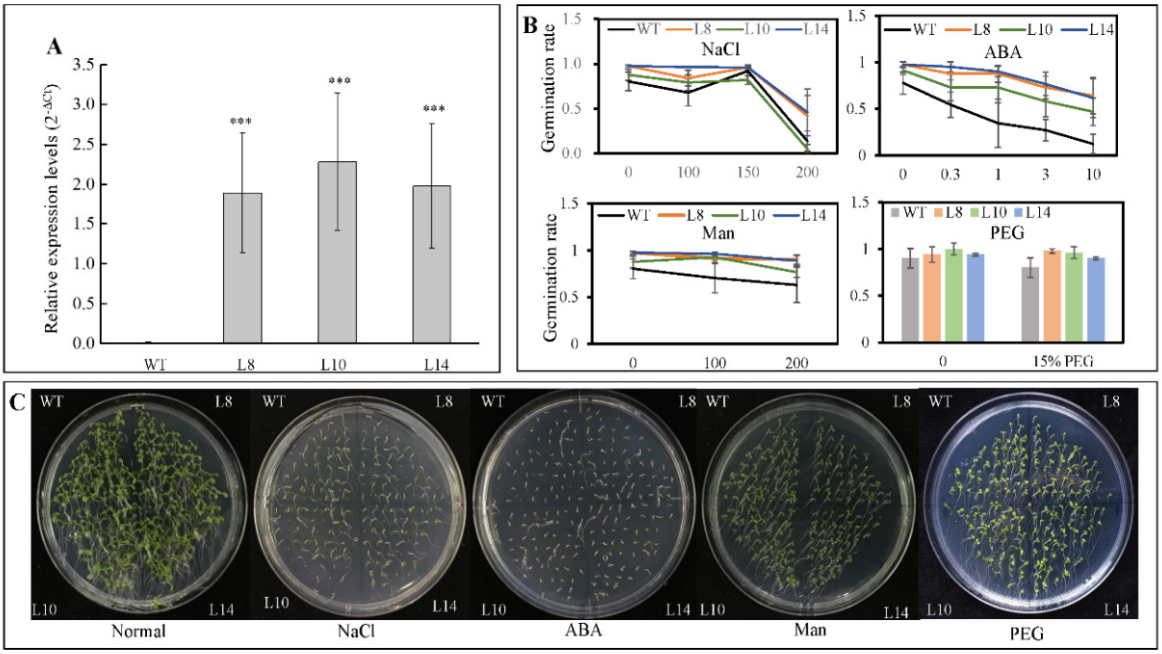


Figure 4. Germination rates of *CsWRKY7*-overexpressing Arabidopsis lines under different stress conditions (**A**) Expression of *CsWRKY7* in the leaves of WT and three transgenic lines (L8, L10, and L14), respectively. Data are shown as the mean ± S.E. (n = 3). (**B**) The germination rate of WT and transgenic lines. Their seeds grown on the 1/2 MS supplied with different concentrations of NaCl, ABA, mannitol and PEG for 4 days. Experiments were performed in five biological replicates. Fifty seeds of each WT and three transgenic lines were germinated in one replicate. Data are shown as the mean ± S.E. (**C**) Germination performance of WT and transgenic lines were taken under normal conditions, 150 mM NaCl, 0.3 μM ABA, 200 mM mannitol, and 15% PEG treatment for 7 days. Asterisk indicated that the expression level is significantly different from the value of the control ('***' $p < 0.001$).

Since *CsWRKY7* responds to abiotic stress and hormone treatments, seed germination was tested with WT and homozygous 35S::*CsWRKY7* transgenic lines to determine the specific role of CsWRKY7 TF in abiotic stress. Though the germination rate of overexpressing lines was higher than that of WT (Figure 4B,C), the difference was not statistically significant under normal growth condition or under abiotic stresses, suggesting that *CsWRKY7* overexpressing lines may not be sensitive to the above induction during seed germination. No significant difference in root growth was observed between WT and overexpressing lines in the presence of different stress media (Figure 5). However, the root elongation of 35S::*CsWRKY7* lines were higher than that of WT with 1/2 MS medium, indicating that CsWRKY7 might promote root growth at the seedling stage under normal growth condition.

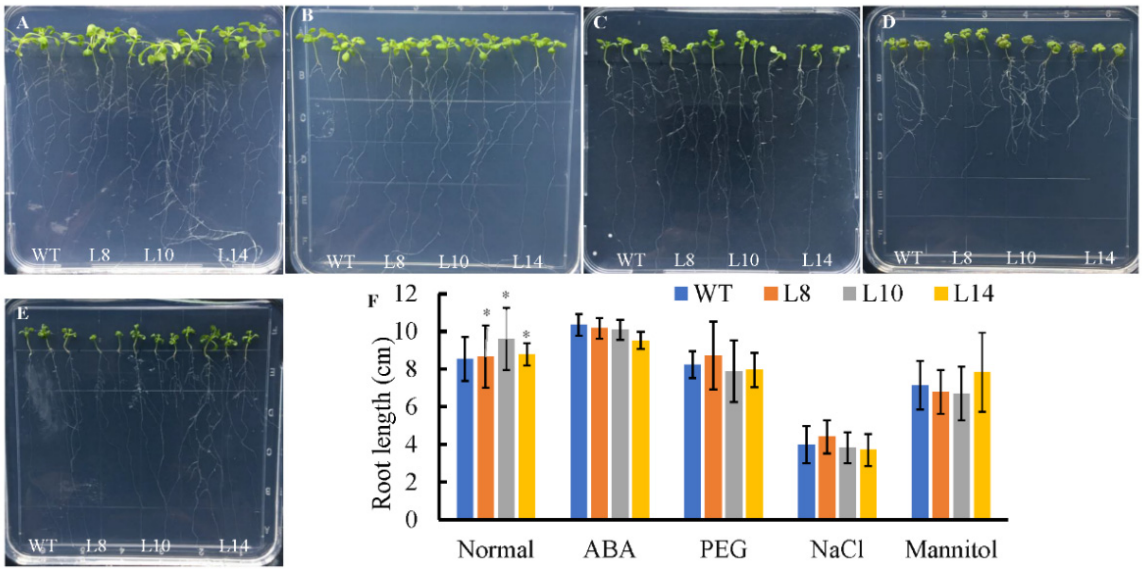


Figure 5. Root growth of *CsWRKY7*-overexpressing Arabidopsis lines under different stress conditions. Seeds were germinated for 4 days on 1/2 MS medium, and the seedlings were then transferred to 1/2 MS medium with or without different treatment for 10 days. (**A**) Normal condition, (**B**) 150 mM NaCl treatment, (**C**) 0.3 μM ABA treatment, (**D**) 200 mM mannitol treatment, and (**E**) 15%PEG6000 treatment, and (**F**) root length were measured at 10 d after the transfer, each line included three seedlings, experiments were performed in four biological replicates. Data are represented as the mean ± SE of 12 seedlings. The significant level is presented by the asterisks (* $p < 0.05$).

2.5. CsWRKY7 Overexpressing Lines Exhibit the Phenotype of Delayed Flowering

In order to investigate whether CsWRKY7 TF participated in plant growth and development, the phenotype of *CsWRKY7*-overexpressing lines and WT was observed during plant growth process. As shown in Figure 6A, B, WT plants bolt earlier than transgenic lines after 25-day growth. When wide-type plants were in silique stage, *CsWRKY7* overexpressing plants were in bolting stage or vegetative growth stage after 35-day growth. To elucidate the regulation mechanism of *CsWRKY7* in flowering, several flowering-related genes was analyzed (Figure 6C). *SUPPRESSOR OF CONSTANS 1* (*SCO1*), a floral integrator, was significantly induced in 35S::*CsWRKY7* lines compared to WT. No noticeable difference in the relative expression level of *CONSTANS* (*CO*) gene that the central regulators in the photoperiod pathway was observed between WT and transgenic lines. As a long-distance transport signal, *FLOWERING LOCUS T* (*FT*) was significantly downregulated in transgenic lines. Simultaneously, two important genes related to inflorescence meristem—*APETALA1* (*AP1*) and LEAFY (*LFY*)—were significantly suppressed in Line 8. The expression levels of *AP1* and *LFY* were downregulated by 64% and 60%, respectively. However, a flowering inhibitor, *FLOWERING LOCUS C (FLC)*, was significantly downregulated in overexpressing lines, while its homologous gene *FLOWERING LOCUS M (FLM)* was upregulated in 35S::*CsWRKY7* lines, compared to WT. Therefore, we speculate that CsWRKY7 delay flowering time might through inhibiting the transcription level of *AP1* and *LFY*.

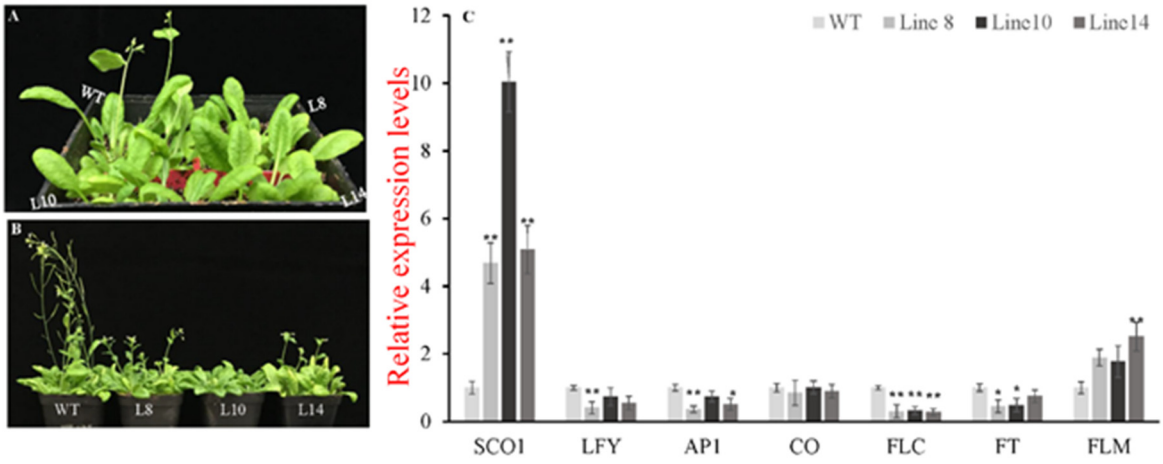


Figure 6. Identification of the *CsWRKY7* in transgenic Arabidopsis. (**A**,**B**) Overexpression of *CsWRKY7* delayed Arabidopsis flowering in different developmental stages: (**A**) Representative photographs of 25-day-old-plant of WT and transgenic lines (L8, L10, and L14) growing in normal conditions. (**B**) 35-day-old-plant. Each line included four seedlings. Experiments were performed in five biological replicates. (**C**) Expression patterns of flowering-related genes in *CsWRKY7*-overexpressing and wild type Arabidopsis. Leaf samples were harvested from 25-day-old transgenic lines and WT. The error bars indicate the means ± S.E. (n = 5), * indicates that the differences are significant ($p < 0.05$), ** indicate that the differences are highly significant ($p < 0.01$).

3. Discussion

WRKY TFs widely participate in plant stress responses and plant growth and development. This study indicated that *CsWRKY7* is involved in abiotic stresses. Constitutive overexpression of *CsWRKY7* not only promotes root growth, but also delays the flowering time in transgenic Arabidopsis plants by suppressing flowering-related genes.

CsWRKY7, homologous to AtWRKY7 and AtWRKY15, belongs to subgroup IId of WRKY family. AtWRKY7 transcription factor not only acted as a negative regulator in PAMP-mediated immune response but also participated in leaf senescence [34–36]. *CsWRKY7* was abundant in old leaves and roots in tea plant, which was somewhat similar to the expression patterns of *AtWRKY7* in Arabidopsis leaves, thus it could be speculated that CsWRKY7 might be involved in tea growth and development. Besides, *CsWRKY7* gene expression was induced by various osmotic stresses and hormones exposure (Figure 3). Sequence analysis identified several potential stress-responsive elements in the promoter region including WBOXHVISO1, CCAATBOX1, GT1GMSCAM4, and MYB1AT and pollen-specific cis-regulatory elements (GTGANTG10), which participated in regulating the response to sugar, heat, NaCl, and drought stresses, respectively (Table 1). Some previous studies have reported that these cis-regulatory elements were stress-related. For example, MYB1AT existed in the promoter region of the Arabidopsis dehydration gene *RD22*, and GT1GMSCAM4 was present in the promoter region of the pathogen and salt-inducible gene *SCaM-4*, which consisted with the cis-regulatory elements in the promoter of *CsWRKY7*, suggesting that *CsWRKY7* participated in various stress responses. (Figure 3) [37,38]. Though two pollen-specific cis-elements present in the promoter, the expression level of *CsWRKY7* in flower was not significantly different from that in bud, which might be probably due to the complicated regulation mechanisms of gene expression. Besides, two hormone response elements—SA and auxin-response elements—are also found in this region. The hormone induction experiment showed that *CsWRKY7* gene was responded to gibberellin and auxin treatment (NAA and 2,4-D), indicating that CsWRKY7 might also be involved in GA- and auxin- and abiotic stress-mediated signaling, but the corresponding mechanism remains be further explored.

Plant growth such as flowering, is a sophisticated regulated process that can be affected by diverse environmental stimuli. Previous research has reported that several WRKY members participate widely

in plant growth and development. AtWRKY44 was reported to be involved in trichome and seed coat development in Arabidopsis [15]. AtWRKY4/-6/-7/-11/-57 were involved in leaf senescence [4]. WRKY transcription factors were involved in the regulation of plants flowering. For example, AtWRKY75 directly bound to the promoter region of *FT* gene, thereby positively regulating Arabidopsis flowering [27]. Both AtWRKY12 and AtWRKY13 transcription factors regulated Arabidopsis flowering time under short day [26]. Additionally, mango MlWRKY12 [24], soybean GsWRKY20 [23], and rice OsWRKY11 [25] also had a regulatory role in flowering. However, few studies of WRKY TFs in tea plant were reported, especially their functions related to plant development. In this study, its ORF was overexpressed in Arabidopsis to determine whether or not CsWRKY7 involved in plant growth. Transgenic analysis indicated that the overexpression of CsWRKY7 altered growth and flowering time of transgenic Arabidopsis (Figures 5 and 6). Gene analysis showed that two meristem identity genes including *AP1* and *LFY* were downregulated in transgenic Arabidopsis, compared with WT (Figure 6). Interestingly, previous studies have reported that there existed one or more W-boxes in the promoter region of *AP1* and *LFY* [39]. As noted earlier, WRKY TFs could specifically bind to W-box [4,5,11–13]. Therefore, the reason why *CsWRKY7* overexpressing lines delayed flowering may lie in that *CsWRKY7* gene directly bound to the promoter regions of *AP1* and *LFY* and inhibited their transcription levels.

However, *FLC*, a suppressor in flowering, was significantly downregulated in transgenic lines, while its homologous gene *FLM* was upregulated, indicating that *FLM* and *FLC* regulated flowering through different pathways. This finding was consistent with that of Katia (2003) [40]. *CO* gene encoding a B-box protein activated the transcription of *FT*. This study indicated that no difference in the expression level of *CO* was observed between transgenic lines and wild type, whereas the downstream gene *FT* was significantly inhibited, indicating that *FT* gene might be inhibited by other complexes such as PRC2, LHP1, SMZ, and TEM [41,42]. Hence, it could be speculated that *CsWRKY7* gene may be involved in flowering regulation independent of the autonomous and vernalization pathway. Generally, these data support our hypothesis that *CsWRKY7* delays flowering. The potential role of *CsWRKY7* as a negative regulator in flowering sheds new light on the development of tea plants at transcription level, but its regulation mechanism should be further explored.

Although *CsWRKY7* was induced by some abiotic stresses in *C. sinensis*, no significant difference in seed germination rate and seedling root growth was observed between *CsWRKY7* overexpressing Arabidopsis lines and WT under abiotic stresses (salt, mannitol, PEG, and exogenous ABA) (Figures 4 and 5), which might be possibly due to the lack of correlation between the levels of mRNA and protein encoded by CsWRKY7, or due to the effect of the inserted transgene on the phenotype. However, the function of *CsWRKY7* gene in terms of osmotic stress response remains to be further characterized. The results of our study not only reveal an important role of CsWRKY7 in plant development, but also provide a foundation for breeding late-blooming tea plants.

To summarize, this study determined the response of *CsWRKY7* to various abiotic stresses and hormones treatments. Our data revealed that CsWRKY7 delayed flowering and promoted root growth in Arabidopsis. Nevertheless, it is necessary to investigate further the pathways through which CsWRKY7 regulates both the development and stress response of tea plant.

4. Materials and Methods

4.1. Plant Materials, Growth Conditions, and Stress Treatments

Two-year-old tea seedlings (*Camellia sinensis* cv. 'Longjing 43') were grown in the greenhouse of Tea Research Institute of the Chinese Academy of Agricultural Sciences (TRICAAS).

The buds, tender stem, flowers, and roots, and leaves at different developmental stages were collected for tissue-specific analysis. The methods of abiotic stresses were essentially the same with those described in our previous research [43] and the materials were collected for further analysis. *N. benthamiana* was used for protein subcellular localization, and was grown in a climate chamber (24 ± 2 °C, 70% relative humidity, and 12 h/12 h light–dark photoperiod). *A. thaliana* ecotype,

Columbia-0 (Col-0), was used as the background material and experiment control. The seeds of wild type and transgenic lines were surface sterilized with 75% ethanol and 0.01% (*v*/*v*) Tween-20 for 8–10 min, and then were washed by distilled water 3−4 times. The seeds were placed on 1/2 MS medium plate at 4 °C in darkness for 48 h, then grown in a climate incubator under a 16 h day/8 h night cycle, respectively at 22 °C/20 °C.

4.2. Cloning of CsWRKY7 Gene and Sequence Analysis

The expressed sequence tags (EST) were obtained from different transcriptome databases [44–46]. The full length of CsWRKY7 cDNA sequence was identified from different tissues of 'Longjing 43' by KOD DNA Polymerase (Toyobo, Tokyo, Japan) and RT-PCR. Gene-specific primers were as follows; 5′-ATGGCCGTCGAGCTCGTGAT-3′ (forward) and 5′-TCAAGAAGACTCTAAGATAAG-3′ (reverse). The purified RT-PCR products were inserted into pEASY-blunt simple cloning vector (TransGen Biotech, Beijing, China) and subsequently sequenced. The predicted molecular weight and theoretical isoelectric point were analyzed by ProtParam (http://web.expasy.org/protparam/, accessed on: 22 May 2017) [47]. The homology of CsWRKY7 protein with other species was analyzed by NCBI BLAST website. The amino acid sequences of WRKY II d subfamily members were analyzed by ClustalX 2.0 and DNAMAN 6.0 software. The MEGA 5.0 software was used to analyze their evolutionary relationships. CsWRKY7 genomic DNA (gDNA) sequence was searched in 'Shuchazao' tea plant genome database, and the promoter region was determined by aligning the open reading frames of the gDNA sequence and those of corresponding gene. The promoter region of CsWRKY7 (1680 bp upstream of start codon) was amplified from 'Longjing 43' gDNA by PCR using KOD DNA polymerase (Toyobo). The regulatory elements in promoter region were predicted by PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on: 9 March 2019) [48].

4.3. Expression Patterns of CsWRKY7 in Tea Plant

Total RNA was extracted from the different tissues of 'Longjing 43' using QIAGEN RNeasy Mini Kit (Qiagen, Hilden, Germany). The reverse transcription reaction was carried out by FastKing gDNA Dispelling RT SuperMix RT Reagent Kit (TIANGEN, Beijing, China). Real-time PCR was performed on an optical 384-well plate with a LightCycler 480 machine (Roche, Sussex, UK). Each reaction contained 5 μL of SYBR Green I Master Mix (Roche Diagnostics), 1.0 μL cDNA samples, and 0.4 μM of each gene specific primer in a final volume of 10 μL. The *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) gene (accession no. FS952640) was used as an internal control. The expression levels were computed by the formula $2^{-\Delta\Delta Ct}$ [49,50].

4.4. Subcellular Location Analysis

Using gateway cloning technology, *CsWRKY7* full-length cDNA sequence without terminator codon was recombined into pH7FWG2 vector, containing the enhanced green fluorescent protein (eGFP) reporter gene, to generate 35s::*CsWRKY7*-eGFP fusion construct. Then, the recombinant plasmid (35s::*CsWRKY7*-eGFP) and the control (35s::GFP) were transformed into *Agrobacterium tumefaciens EHA105* by freeze–thaw approach [51]. The confirmed bacteria were grown in yeast extract peptone (YEP) medium (pH = 7.4) at 28 °C until optical density at λ = 600 nm (OD_{600}) to 1.0. This medium added with bacteria was centrifuged at 4000 rpm for 10 min. the sediment was resuspended with suspension buffer containing 10 mM $MgCl_2$, 10 mM MES, and 100 μM acetosyringone (AS), with OD_{600} adjusted to 0.4. The suspension was infiltrated into well-developed *N. benthamiana* leaves. After infiltration, the tobacco plant was cultured in darkness for 48 h at 24 °C and then put in light for half an hour. The GFP signal was observed by a confocal microscopy. Additionally, cellular localization of the CsWRKY7-eGFP proteins in lateral roots of transgenic Arabidopsis plants was also observed by confocal microscopy.

4.5. Generation of CsWRKY7 Transgenic Plants

Gene-specific primers 5′-CACCATGGCGGTCGAGCTAG-3′ (Forward) and 5′-AGAAGACT CTAAAATGAGACCAGA-3′ (Reverse) were used to clone the open reading frame (ORF) of *CsWRKY7*. The cloned products were inserted into pENTR/D-TOPO vector (Invitrogen, Carlsbad, CA, USA) and sequenced, then the right directional sequence was inserted into pH7FWG2 vector with LR clonase II enzyme mix (Invitrogen) according to the manufacturer's instructions. The binary vector was transferred into *Agrobacterium tumefaciens strain GV3101* by freeze–thaw method [51], and then this strain was introduced into Arabidopsis by floral dip approach [52]. T4 homozygous seeds were used for the experiments.

4.6. Phenotypic Analysis of Transgenic Arabidopsis

For the germination assay, the seeds of WT and homozygous 35S::*CsWRKY7* transgenic lines were germinated on 1/2 Murashige and Skoog (MS) medium containing different concentrations of NaCl (0, 100, 150, 200 mM), mannitol (0, 100, 200 mM), ABA (0, 0.3, 1, 10 μM), and PEG (0, 15%) solution. After 48-h vernalization, seedlings were placed vertically in climate incubator, and germination rate was counted on the fourth day. The germination was defined when the roots break the seed coat by 1 mm. Photographs of the plates with 150 mM NaCl, 200 mM mannitol, 0.3 μM ABA, and 15% PEG were taken on the 7th day. Each plate consisted of WT and three transgenic lines with 50 seeds per line in 9 cm diameter plate. In root length assay, the seeds of WT and homozygous 35S::*CsWRKY7* transgenic lines were sown on 1/2 MS medium. After 48-h vernalization, the plates were placed vertically in climate incubator for 4 d. And then seedlings of similar size were transferred to different stress medium containing 150 mM NaCl, 200 mM Mannitol, 0.3 μM ABA, and 15% PEG, respectively. The root length was measured by ruler after 10-day growth. Arabidopsis seedlings were grown on 1/2 MS medium for 7–10 d, and then transplanted to sterilized soil in a growth chamber for the phenotype observation (16 h day/8 h night at 22/20 °C).

4.7. Quantitative Real-Time PCR Analysis in Transgenic Arabidopsis

To investigate the mechanism by which *CsWRKY7* delayed flowering, the expression levels of flowering-related genes in transgenic Arabidopsis lines and wild type were calculated by the $2^{-\Delta Ct}$ or $2^{-\Delta\Delta Ct}$ with the expression level of *Actin-2* gene (Gene Locus: At3g18780) as the reference control [49,50]. The primer sequence used for qRT-PCR are listed in Supplementary Table S1.

4.8. Statistical Analysis

All assays were performed with either three or more repetitive experiments. Student's *t*-test was performed to determine the significant difference, and standard error were calculated using the Built-in Functions in Excel 2016. Student's *t*-test, '*', '**', and '***' indicate significant difference at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/11/2815/s1. Table S1 Primer sequences used for qRT-PCR.

Author Contributions: Data curation, M.-Z.Y. and L.C.; Formal analysis, D.-J.N. and L.C.; Investigation, Y.-X.X.; Methodology, W.C., W.-J.H., Y.-X.X., C.Z., and D.-J.N.; Project administration, L.C.; Resources, L.C.; Software, W.C., W.-J.H., and C.Z.; Supervision, M.-Z.Y. and L.C.; Validation, L.C.; Writing—original draft, W.C.; Writing—review and editing, M.-Z.Y. and L.C.

Funding: This work was supported by the Earmarked Fund for China Agriculture Research System (CARS-019) and the Chinese Academy of Agricultural Sciences through the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2017-TRICAAS) to Liang Chen.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mitsuda, N.; Ohme-Takagi, M. Functional analysis of transcription factors in *Arabidopsis*. *Plant Cell Physiol.* **2009**, *50*, 1232–1248. [CrossRef] [PubMed]
2. Chen, L.; Song, Y.; Li, S.; Zhang, L.; Zou, C.; Yu, D. The role of WRKY transcription factors in plant abiotic stresses. *Biochim. Biophys. Acta* **2012**, *1819*, 120–128. [CrossRef] [PubMed]
3. Eulgem, T. Dissecting the WRKY web of plant defense regulators. *PLoS Pathog.* **2006**, *2*, e126. [CrossRef] [PubMed]
4. Eulgem, T.; Rushton, P.J.; Robatzek, S.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **2000**, *5*, 199–206. [CrossRef]
5. Rushton, P.J.; Macdonald, H.; Huttly, A.K.; Lazarus, C.M.; Hooley, R. Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of *α-Amy2* genes. *Plant Mol. Biol.* **1995**, *29*, 691–702. [CrossRef] [PubMed]
6. Rushton, P.J.; Torres, J.T.; Parniske, M.; Wernert, P.; Hahlbrock, K.; Somssich, I.E. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of *parsley PR1* genes. *EMBO J.* **1996**, *15*, 5690–5700. [CrossRef]
7. Bakshi, M.; Oelmüller, R. WRKY transcription factors: Jack of many trades in plants. *Plant Signal. Behav.* **2014**, *9*, e27700. [CrossRef]
8. Ishiguro, S.; Nakamura, K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and β-amylase from sweet potato. *Mol. Gen. Genet.* **1994**, *244*, 563–571. [CrossRef]
9. Miao, Y.; Smykowski, A.; Zentgraf, U. A novel upstream regulator of WRKY53 transcription during leaf senescence in *Arabidopsis thaliana*. *Plant Biol.* **2008**, *10*, 110–120. [CrossRef]
10. Phukan, U.J.; Jeena, G.S.; Shukla, R.K. WRKY transcription factors: Molecular regulation and stress responses in plants. *Front. Plant Sci.* **2016**, *7*, 760. [CrossRef]
11. Agarwal, P.; Reddy, M.P.; Chikara, J. WRKY: Its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Mol. Biol. Rep.* **2011**, *38*, 3883–3896. [CrossRef] [PubMed]
12. Ciolkowski, I.; Wanke, D.; Birkenbihl, R.P.; Somssich, I.E. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.* **2008**, *68*, 81–92. [CrossRef] [PubMed]
13. Bao, W.; Wang, X.; Chen, M.; Chai, T.; Wang, H. A WRKY transcription factor, PcWRKY33, from *Polygonum cuspidatum* reduces salt tolerance in transgenic *Arabidopsis thaliana*. *Plant Cell Rep.* **2018**, *37*, 1033–1048. [CrossRef] [PubMed]
14. Lagacé, M.; Matton, D.P. Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense*. *Planta* **2004**, *219*, 185–189. [CrossRef] [PubMed]
15. Ishida, T.; Hattori, S.; Sano, R.; Inoue, K.; Shirano, Y.; Hayashi, H.; Shibata, D.; Sato, S.; Kato, T.; Tabata, S.; et al. *Arabidopsis* TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation. *Plant Cell* **2007**, *19*, 2531–2543. [CrossRef] [PubMed]
16. Luo, M.; Dennis, E.S.; Berger, F.; Peacock, W.J.; Chaudhury, A. *MINISEED3* (*MINI3*), a WRKY family gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17531–17536. [CrossRef] [PubMed]
17. Robatzek, S.; Somssich, I.E. A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. *Plant J.* **2001**, *28*, 123–133. [CrossRef]
18. Chen, L.; Xiang, S.; Chen, Y.; Li, D.; Yu, D. *Arabidopsis* WRKY45 interacts with the DELLA protein RGL1 to positively regulate age-triggered leaf senescence. *Mol. Plant* **2017**, *10*, 1174–1189. [CrossRef]
19. Xie, Y.; Huhn, K.; Brandt, R.; Potschin, M.; Bieker, S.; Straub, D.; Doll, J.; Drechsler, T.; Zentgraf, U.; Wenkel, S. REVOLUTA and WRKY53 connect early and late leaf development in *Arabidopsis*. *Development* **2014**, *141*, 4772–4783. [CrossRef]
20. Jiang, Y.; Liang, G.; Yang, S.; Yu, D. WRKY57 functions as a node of convergence for jasmonic acid- and auxin-mediated signaling in jasmonic acid–induced leaf senescence. *Plant Cell* **2014**, *26*, 230–245. [CrossRef]
21. Besseau, S.; Li, J.; Palva, E.T. WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *J. Exp. Bot.* **2012**, *63*, 2667–2679. [CrossRef] [PubMed]

22. Jiang, W.; Yu, D.; Zou, C. Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in *Arabidopsis*. *J. Exp. Bot.* **2010**, *61*, 3901–3914. [CrossRef]
23. Zhu, D.; Cai, H.; Wu, J.; Cao, L.; Tang, L.; Hu, M.; Ji, W.; Bai, X.; Sun, X.; Liu, X.; et al. Expression of wild soybean *WRKY20* in *Arabidopsis* enhances drought tolerance and regulates ABA signalling. *J. Exp. Bot.* **2013**, *64*, 2155–2169. [CrossRef]
24. Yu, Y.; Hu, R.; Wang, H.; Cao, Y.; He, G.; Fu, C.; Zhou, G. MlWRKY12, a novel *Miscanthus* transcription factor, participates in pith secondary cell wall formation and promotes flowering. *Plant Sci.* **2013**, *212*, 1–9. [CrossRef] [PubMed]
25. Cai, Y.; Chen, X.; Xie, K.; Xing, Q.; Wu, Y.; Li, J.; Du, C.; Sun, Z.; Guo, Z. Dlf1, a WRKY transcription factor, is involved in the control of flowering time and plant height in rice. *PLoS ONE* **2014**, *9*, e102529. [CrossRef] [PubMed]
26. Li, W.; Wang, H.; Yu, D. *Arabidopsis* WRKY transcription factors WRKY12 and WRKY13 oppositely regulate flowering under short-day conditions. *Mol. Plant* **2016**, *9*, 1492–1503. [CrossRef] [PubMed]
27. Zhang, L.; Chen, L.; Yu, D. Transcription factor WRKY75 interacts with DELLA proteins to affect flowering. *Plant Physiol.* **2018**, *176*, 790–803. [CrossRef]
28. Luo, Y.; Yu, S.; Li, J.; Li, Q.; Wang, K.; Huang, J.; Liu, Z. Molecular characterization of WRKY transcription factors that act as negative regulators of O-Methylated catechin biosynthesis in tea plants (*Camellia sinensis* L.). *J. Agric. Food Chem.* **2018**, *66*, 11234–11243. [CrossRef]
29. Wu, Z.-J.; Li, X.-H.; Liu, Z.-W.; Li, H.; Wang, Y.-X.; Zhuang, J. Transcriptome-wide identification of *Camellia sinensis* WRKY transcription factors in response to temperature stress. *Mol. Genet. Genomics* **2016**, *291*, 255–269. [CrossRef]
30. Wang, Y.; Shu, Z.; Wang, W.; Jiang, X.; Li, D.; Pan, J.; Li, X. *CsWRKY2*, a novel *WRKY* gene from *Camellia sinensis*, is involved in cold and drought stress responses. *Biol. Plant.* **2016**, *60*, 443–451. [CrossRef]
31. Wang, P.; Yue, C.; Chen, D.; Zheng, Y.; Zhang, Q.; Yang, J.; Ye, N. Genome-wide identification of *WRKY* family genes and their response to abiotic stresses in tea plant (*Camellia sinensis*). *Genes Genomics* **2019**, *41*, 17–33. [CrossRef] [PubMed]
32. Kim, K.-C.; Fan, B.; Chen, Z. Pathogen-induced *Arabidopsis* WRKY7 is a transcriptional repressor and enhances plant susceptibility to *Pseudomonas syringae*. *Plant Physiol.* **2006**, *142*, 1180–1192. [CrossRef]
33. Vanderauwera, S.; Vandenbroucke, K.; Inzé, A.; van de Cotte, B.; Mühlenbock, P.; De Rycke, R.; Naouar, N.; Van Gaever, T.; Van Montagu, M.C.E.; Van Breusegem, F. AtWRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20113–20118. [CrossRef]
34. Arraño-Salinas, P.; Domínguez-Figueroa, J.; Herrera-Vásquez, A.; Zavala, D.; Medina, J.; Vicente-Carbajosa, J.; Meneses, C.; Canessa, P.; Moreno, A.A.; Blanco-Herrera, F. WRKY7, -11 and -17 transcription factors are modulators of the bZIP28 branch of the unfolded protein response during PAMP-triggered immunity in *Arabidopsis thaliana*. *Plant Sci.* **2018**, *27*, 242–250. [CrossRef] [PubMed]
35. Rogers, H.J.; Bate, N.; Combe, J.; Sullivan, J.; Sweetman, J.; Swan, C.; Lonsdale, D.M.; Twell, D. Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene *g10*. *Plant Mol. Biol.* **2001**, *45*, 577–585. [CrossRef]
36. Ko, Y.J.; Lee, S.; Song, K.; Park, S.-Y.; Ahn, I.; Bae, S.-C.; Lee, Y.H.; Hwang, D.-J. Heterologous expression of the *Brassica rapa* transcription factor BrWRKY7 enhances resistance against bacterial soft rot caused by *Pectobacterium carotovorum* in *Arabidopsis*. *Plant Biotechnol. Rep.* **2015**, *9*, 179–186. [CrossRef]
37. Abe, H.; Urao, T.; Ito, T.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **2003**, *15*, 63–78. [CrossRef]
38. Park, H.C.; Kim, M.L.; Kang, Y.H.; Jeon, J.M.; Yoo, J.H.; Kim, M.C.; Park, C.Y.; Jeong, J.C.; Moon, B.C.; Lee, J.H.; et al. Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiol.* **2004**, *135*, 2150–2161. [CrossRef]
39. Hamès, C.; Ptchelkine, D.; Grimm, C.; Thevenon, E.; Moyroud, E.; Gérard, F.; Martiel, J.L.; Benlloch, R.; Parcy, F.; Müller, C.W. Structural basis for *LEAFY* floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* **2008**, *27*, 2628–2637. [CrossRef]
40. Scortecci, K.; Michaels, S.D.; Amasino, R.M. Genetic interactions between *FLM* and other flowering-time genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* **2003**, *52*, 915–922. [CrossRef]

41. Farrona, S.; Thorpe, F.L.; Engelhorn, J.; Adrian, J.; Dong, X.; Sarid-Krebs, L.; Goodrich, J.; Turck, F. Tissue-specific expression of *FLOWERING LOCUS T* in *Arabidopsis* is maintained independently of polycomb group protein repression. *Plant Cell* **2011**, *23*, 3204–3214. [CrossRef] [PubMed]
42. Lee, J.; Lee, I. Regulation and function of *SOC1*, a flowering pathway integrator. *J. Exp. Bot.* **2010**, *61*, 2247–2254. [CrossRef] [PubMed]
43. Chen, W.; Xu, Y.-X.; Mao, J.; Hao, W.-J.; Liu, Y.-F.; Ni, D.-J.; Chen, L. Cloning and expression patterns of VQ-motif-containing proteins under abiotic stress in tea plant. *Plant Growth Regul.* **2019**, *87*, 277–286. [CrossRef]
44. Wang, D.; Li, C.-F.; Ma, C.-L.; Chen, L. Novel insights into the molecular mechanisms underlying the resistance of *Camellia sinensis* to ectropis oblique provided by strategic transcriptomic comparisons. *Sci. Hortic.* **2015**, *192*, 429–440. [CrossRef]
45. Li, C.-F.; Zhu, Y.; Yu, Y.; Zhao, Q.-Y.; Wang, S.-J.; Wang, X.-C.; Yao, M.-Z.; Luo, D.; Li, X.; Chen, L.; et al. Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC Genom.* **2015**, *16*, 560. [CrossRef] [PubMed]
46. Liu, S.-C.; Jin, J.-Q.; Ma, J.-Q.; Yao, M.-Z.; Ma, C.-L.; Li, C.-F.; Ding, Z.-T.; Chen, L. Transcriptomic analysis of tea plant responding to drought stress and recovery. *PLoS ONE* **2016**, *11*, e0147306. [CrossRef] [PubMed]
47. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.E.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **1999**, *112*, 531–552. [CrossRef]
48. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
49. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **2001**, *25*, 402–408. [CrossRef]
50. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **2008**, *3*, 1101. [CrossRef]
51. Weigel, D.; Glazebrook, J. Transformation of *Agrobacterium* using the freeze-thaw method. *CSH Protoc.* **2006**, *2006*. [CrossRef] [PubMed]
52. Clough, S.J.; Bent, A.F. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 735–743. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of TCP Family Genes in *Zea mays* L. Identified a Role for *ZmTCP42* in Drought Tolerance

Shuangcheng Ding [1,2,†], Zhenzhen Cai [1,†], Hewei Du [1,3] and Hongwei Wang [1,2,*]

1 Agricultural College, Yangtze University, Jingzhou 434025, China; shchding@yangtzeu.edu.cn (S.D.); 18339069671@163.com (Z.C.); duhewei666@163.com (H.D.)
2 Hubei Collaborative Innovation Center for Grain Industry, Yangtze University, Jingzhou 434025, China
3 College of Life Science, Yangtze University, Jingzhou 434025, China
* Correspondence: wanghw@yangtzeu.edu.cn
† These authors contributed equally to this work.

Received: 9 May 2019; Accepted: 4 June 2019; Published: 5 June 2019

Abstract: The Teosinte-branched 1/Cycloidea/Proliferating (TCP)plant-specific transcription factors (TFs)have been demonstrated to play a fundamental role in plant development and organ patterning. However, it remains unknown whether or not the *TCP* gene family plays a role in conferring a tolerance to drought stress in maize, which is a major constraint to maize production. In this study, we identified 46 *ZmTCP* genes in the maize genome and systematically analyzed their phylogenetic relationships and synteny with rice, sorghum, and *Arabidopsis TCP* genes. Expression analysis of the 46 *ZmTCP* genes in different tissues and under drought conditions, suggests their involvement in maize response to drought stress. Importantly, genetic variations in *ZmTCP32* and *ZmTCP42* are significantly associated with drought tolerance at the seedling stage. RT-qPCR results suggest that *ZmTCP32* and *ZmTCP42* RNA levels are both induced by ABA, drought, and polyethylene glycol treatments. Based on the significant association between the genetic variation of *ZmTCP42* and drought tolerance, and the inducible expression of *ZmTCP42* by drought stress, we selected *ZmTCP42*, to investigate its function in drought response. We found that overexpression of *ZmTCP42* in Arabidopsis led to a hypersensitivity to ABA in seed germination and enhanced drought tolerance, validating its function in drought tolerance. These results suggested that ZmTCP42 functions as an important TCP TF in maize, which plays a positive role in drought tolerance.

Keywords: maize; *ZmTCP*; natural variation; drought tolerance

1. Introduction

Severe drought causes a grievous decline in crop yield by negatively affecting plant growth and reproduction [1,2]. Maize (*Zea mays* L.)is considered a major crop for food, feed, and fuel, but its production is frequently hampered by water scarcity [3,4]. While the molecular mechanisms of drought stress response in plants have not been fully elucidated, many transcription factors (TFs)such as DREB1/CBF, MYB, and AREB/ABF, which regulate drought-responsive genes have been well-studied [5–7]. These reports support the idea that identification of key TFs will help us to better understand the molecular and cellular responses to drought stress.

Teosinte branched 1/Cycloidea/Proliferating (TCP) genes encode plant-specific TFs and are named after the first three functionally characterized members of this TF family—*Teosinte Branched 1* (*TB1*)in maize (*Zea mays* L.), *Cycloidea* (*CYC*)in snapdragon (*Antirrhinum majus*), and *Proliferating Cell nuclear antigen Factor* (*PCF*)in rice (*Oryza sativa*) [8]. This class of TFs share a highly conserved TCP domain, which contains a 59-amino acid, non-canonical basic-Helix-Loop-Helix (bHLH)structural motif that

allows DNA binding, protein–protein interaction, and protein nuclear localization [8–10]. More than 20 TCP TF members were identified in various plant species, such as *Arabidopsis*, rice, tomato, cotton, sorghum, and wheat [11–17]. Based on the TCP domain, these genes were divided into two classes. Class I is the PCF class; class II is subdivided into two clades—CIN (*CINCINNATA* of *Antirrhinum*) and CYC/TB1 [10,15,18]. Many members of the TCP family have been shown to be involved in the regulation of many biological processes during plant growth and development, such as leaf development, branching, floral organ morphogenesis, and senescence [8,19–22]. Possible mechanisms for the regulation were studied. These mechanisms were found to involve either a direct transcriptional control of the cell cycle genes by TCP TFs, or an indirect adjustment of the hormone activity [23]. For example, in *Arabidopsis* seeds, the DELLA proteins GAI (GA-Insensitive)and RGA (Repressor of GA)formed an unproductive complex with the class II TCP protein AtTCP14 or AtTCP15, which prevented the binding of TCPs to promoters of the core cell cycle genes [24]. In turn, GA might induce ubiquitination and degradation of DELLA proteins to relieve the constitutive inhibition of TCP transcriptional activity [24]. Interestingly, the orthologous TCP protein LANCEOLATE was found to participate in GA biosynthesis by upregulating the *SIGA-oxidase 1* gene [25]. It was also found that a subset of class I TCP proteins, such as AtTCP1, AtTCP2, AtTCP3, and AtTCP14, behaved in a similar fashion by acting as an inducer or repressor to influence the biosynthesis or signaling of several hormones during different developmental processes [26–28].

Abscisic acid (ABA)is a known stress response hormone that can mitigate physiological and environmental stresses, including drought stress, by inducing the closure of stomata, thus reducing water loss [29–32]. Although the *TCP* gene family is primarily involved in the regulation of plant growth and leaf development, other functions, such as involvement in the ABA signaling pathway, have been noted for specific *TCP* genes. For example, overexpression of *OsTCP19* in *Arabidopsis* significantly conferred both drought and heat tolerance during seedling establishment and in mature plants [33]. Furthermore, the interaction of OsTCP19 with OsABI4, which encodes a TF involved in the ABA signal transduction, suggests its function in fine-tuning drought-induced ABA signaling [33]. Additionally, creeping bentgrass plants (*Agrostis stolonifera*) that overexpresses *Osa-miR319*, in which four putative target genes, *AsPCF5*, *AsPCF6*, *AsPCF8*, and *AsTCP14* are down-regulated, significantly enhance plant tolerance to salt and drought stress associated with an increased leaf wax content and water retention [34]. It is noteworthy that AtTCP14 antagonizes ABA signaling by interacting with the DOF6 (DNA binding with one finger)TF, preventing the activation of the downstream ABA biosynthetic gene *ABA1* (*ABA deficient 1*) and other ABA-responsive genes in *Arabidopsis* seeds [35]. In contrast, *AtTCP18*, also known as *BRC1* (*Branched 1*), induces *ABF3* (*ABA responsive elements-binding factor 3*), and *ABI5* (*ABA insensitive 5*)—two key regulators of the ABA response—to maintain ABA signaling when the axillary buds enter dormancy [36,37]. These reports indicate a strong association of TCPs with ABA-mediated abiotic stress signaling. However, how the TCP TF genes function in maize, especially in response to drought stress, still remains to be elucidated. Due to the rapid linkage disequilibrium (LD)decay in the maize genome, association study is able to provide a gene-level resolution, which facilitates the genetic detection of several complex traits, such as drought tolerance [38–40]. However, limited allelic variations underlying drought tolerance have been identified [41], and rarely favorable alleles could be used for the genetic improvement of drought tolerance in maize.

Here, we comprehensively analyzed the *ZmTCP* gene family in the maize genome. Previous research has identified 29 *TCP* genes in the maize genome and subsequently, phylogeny, gene structure, chromosomal location, gene duplication, and expression levels of the 29 *ZmTCP* genes were investigated [42]. In this study, we searched against the updated maize genome B73_RefGen_v3 and identified 46 *ZmTCP* genes, and further systematically analyzed to determine their phylogenetic relationships and synteny with rice, sorghum, and *Arabidopsis TCPs*. In addition, we studied the expression profiles of these *ZmTCP*s upon drought stress. Importantly, a family-based, genome-wide association study revealed a significant association between the natural variations of *ZmTCP42* and

maize drought tolerance. Ectopic expression of *ZmTCP42* in *Arabidopsis* led to enhanced drought tolerance, validating its function in drought tolerance.

2. Results

2.1. The Maize Genome Contains 46 TCP Family Genes

We comprehensively performed a genome-wide search for putative *TCP* genes in maize. Initially, the protein sequences of the putative maize TCP TFs were retrieved from the Plant Transcription Factor Database 3.0 (available online: http://planttfdb.cbi.pku.edu.cn). Subsequently, a BLASTP analysis was performed in Phytozome V10 (available online: http://www.phytozome.net/eucalyptus.php), using all of the *Arabidopsis* and rice TCP protein sequences as queries, and the predicted protein sequences without a TCP domain (Pfam: PF03634) were excluded from the results. A Hidden Markov Model (HMM) search was also performed against the maize database using PF03634. Ultimately, 46 genes, including the 29 *ZmTCPs* previously reported by Chai et al. [42], were retrieved and further verified for the presence of a canonical TCP domain. The 17 newly identified *ZmTCPs* were named *ZmTCP30* to *ZmTCP46*. Their locus IDs, genome locations, coding sequence (CDS)lengths, and protein lengths are listed in Table 1. The 46 *ZmTCP* genes were unevenly distributed throughout the maize genome on 10 chromosomes, without any clustering. The 46 ZmTCP proteins contained a range of 98 to 778 amino acids, corresponding to 11.76 kDa to 84.78 kDa in molecular weight (Table 1). Their theoretical pI (isoelectric point)values ranged from 5.13 to 12.23, with a mean of 8.14 (Table 1), indicating that most of them were weakly alkaline (Table 1).

2.2. Maize Contains Roughly Twice as Many TCP Genes as Rice and Sorghum

In order to study the phylogenetic relationships among the predicted maize TCPs, we constructed a neighbor-joining tree for the ZmTCP proteins and their orthologs from rice, sorghum and *Arabidopsis*, based on multiple-alignment of the full-length TCP protein sequences (Figure 1). As shown on the phylogenetic tree, the 113 TCPs were classified into two main classes—class I and class II. Class II was further divided into two clades—CYC/TB1 and CIN. The phylogenetic tree based on the sequence alignments of the 46 ZmTCPs was also divided into three clades (Figure S1). The boundaries of these major clades were clearly stated by the phylogenetic locations of several canonical *TCP* genes, such as the class I genes *OsPCF1* and *OsPCF2*, the *CIN*-like class II genes *AtTCP2*, *AtTCP3*, and *AtTCP4*, and the *CYC/tb1*-like class II genes *AtTCP1*, *AtTCP12*, *AtTCP18*, and *ZmTCP2/ZmTB1*. In agreement with previous work, all *Arabidopsis*, rice, and sorghum TCPs fell in the same class or clade, as previously reported in our phylogenetic tree [11,12]. Interestingly, dicot *Arabidopsis* TCP proteins clustered separately from those of the three monocot plants in the same clad, and some proteins from three of the monocot plants displayed pairing. Furthermore, maize and sorghum TCPs were found to share a closer phylogenetic relationship than maize and rice ones, which was consistent with the notion that sorghum is a closer relative of maize than rice. Throughout the phylogenetic tree, there were 49 TCPs in the PCF clade, 34 in CIN, and 30 in CYC/TB1 (Figure 1). Among the 46 *ZmTCPs*, there were 17 *ZmTCPs* in class I and 29 *ZmTCPs* in class II; within class II, 10 in CIN and 19 in CYC/TB1 (Figure S1). Significantly, the number of maize *ZmTCPs* was roughly twice as large as the number in each of rice and sorghum. Specifically, among all 30 CYC/TB1 *TCP* genes, 19 were from maize, 4 from sorghum, and 3 each from rice and *Arabidopsis* (Figure 1). This result suggests that the *ZmTCP* gene family in the allopolyploid maize genome underwent a two-fold duplication. This expansion was biased and occurred mainly in the class II CYC/TB1 clade.

Table 1. Detailed information for 46 *ZmTCP* genes in the *Zea mays* L. genome.

Gene Name	Locus ID	Chr	CDS Length (bp)	Protein Length (aa)	Number of Exons	MW	pI	Class
ZmTCP1	GRMZM2G166687	1	723	240	1	26.11	5.15	CYC/TB1
ZmTCP2	AC233950.1_FG002	1	1131	376	1	39.86	7.8	CYC/TB1
ZmTCP3	GRMZM2G115516	1	1161	386	2	39.36	9.33	CIN
ZmTCP4	AC234521.1_FG006	2	768	255	2	27.71	10.62	CYC/TB1
ZmTCP5	GRMZM2G110242	2	831	276	1	29.9	6.26	CYC/TB1
ZmTCP6	GRMZM2G088440	2	549	182	5	21.14	8.65	CYC/TB1
ZmTCP7	GRMZM2G414114	2	1104	367	6	42.29	8.44	CYC/TB1
ZmTCP8	GRMZM2G020805	2	1323	440	2	46.22	9.46	CIN
ZmTCP9	AC205574.3_FG006	3	1698	565	2	58.83	7.1	CIN
ZmTCP10	GRMZM2G166946	3	852	283	2	30.48	8.98	CIN
ZmTCP11	GRMZM2G055024	4	1263	420	3	45.5	10.82	CYC/TB1
ZmTCP12	GRMZM2G062711	4	564	187	1	21.58	12.05	CIN
ZmTCP13	GRMZM2G060319	4	825	274	1	29.35	5.96	CYC/TB1
ZmTCP14	GRMZM2G135461	4	402	132	3	15.09	7.32	CYC/TB1
ZmTCP15	GRMZM2G078077	4	657	218	1	22.97	10.02	PCF
ZmTCP16	GRMZM2G003944	4	1164	387	1	40.76	9.33	PCF
ZmTCP17	GRMZM2G089361	5	1146	381	1	39.05	9.58	CIN
ZmTCP18	AC190734.2_FG003	5	1080	359	1	38.89	7.01	CYC/TB1
ZmTCP19	GRMZM2G445944	5	624	207	1	21.12	10.26	PCF
ZmTCP20	GRMZM2G031905	6	435	144	3	16.01	9.89	CIN
ZmTCP21	GRMZM2G142751	6	1140	379	1	39.09	8.02	PCF
ZmTCP22	GRMZM2G120151	6	897	298	2	30.9	8.07	CIN
ZmTCP23	GRMZM2G064628	7	843	280	1	30.23	6.33	CYC/TB1
ZmTCP24	GRMZM2G015037	8	1974	657	4	69.55	6.61	CIN
ZmTCP25	GRMZM2G035944	8	828	275	2	29.59	9.02	CIN
ZmTCP26	GRMZM2G113888	9	1209	402	1	40.66	8.81	PCF
ZmTCP27	GRMZM2G096610	10	531	176	1	18.33	7.43	PCF
ZmTCP28	GRMZM2G458087	10	495	164	3	18.83	8.37	CYC/TB1
ZmTCP29	GRMZM2G465091	10	624	207	1	21.01	10.01	PCF
ZmTCP30	GRMZM2G454571	1	399	132	3	15.11	7.27	CYC/TB1
ZmTCP31	AC213524.3_FG003	1	831	276	2	29.56	9.83	PCF
ZmTCP32	GRMZM2G107031	1	972	323	2	34.02	5.98	PCF
ZmTCP33	GRMZM2G416524	2	894	297	3	32.37	10.43	PCF
ZmTCP34	AC199782.5_FG003	2	1032	343	1	35.43	5.13	PCF
ZmTCP35	GRMZM5G824514	3	297	98	1	11.76	12.23	CYC/TB1
ZmTCP36	GRMZM2G089638	3	1251	416	1	42.91	5.79	PCF
ZmTCP37	GRMZM2G092214	3	975	324	2	34.35	6.7	PCF
ZmTCP38	GRMZM2G359599	4	522	173	4	19.58	8.6	CYC/TB1
ZmTCP39	GRMZM2G170232	4	396	131	4	14.38	8.87	CYC/TB1
ZmTCP40	GRMZM2G178603	5	663	220	1	22.76	10.02	PCF
ZmTCP41	GRMZM2G077755	5	1200	400	1	40.49	9.2	PCF
ZmTCP42	GRMZM2G180568	7	972	323	1	33.71	8.11	CIN
ZmTCP43	GRMZM2G148022	8	2337	778	15	84.78	7.35	CIN
ZmTCP44	GRMZM2G034638	8	948	315	2	33.08	6.57	PCF
ZmTCP45	GRMZM2G424261	10	489	162	3	17.27	5.91	CIN
ZmTCP46	GRMZM2G093895	10	1218	405	1	41.09	5.95	PCF

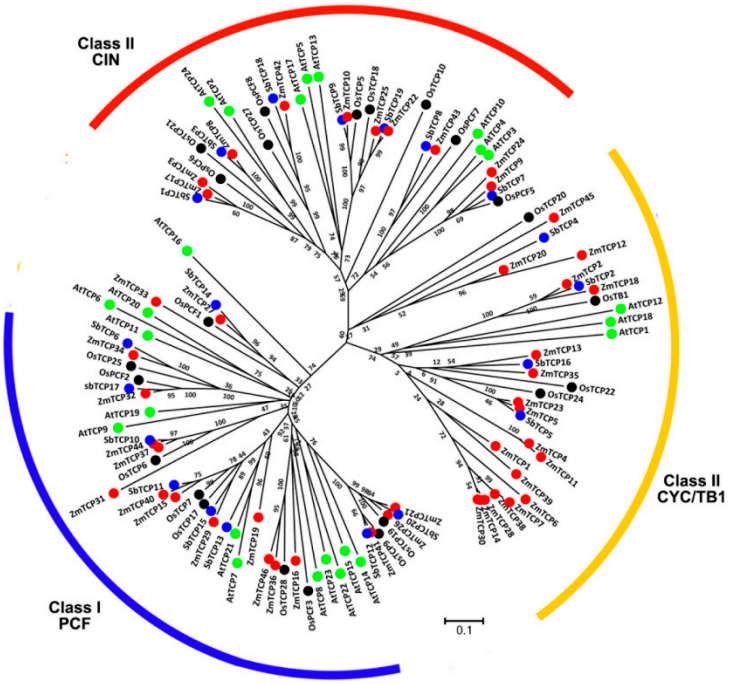


Figure 1. Phylogenetic tree of the predicted Teosinte-branched 1/Cycloidea/Proliferating (TCP)proteins from maize, rice, sorghum, and *Arabidopsis*. The phylogenetic tree was constructed based on the sequence alignment of the 113 full-length TCP protein sequences from four species. The unrooted tree was drawn by MEGA 7.0 with the neighbor-joining (NJ)method, using the following parameters—bootstrap values (1000 replicates)and the poisson model. The scale refers to the branch lengths. The gene codes and names are illustrated in red for maize, black for rice, blue for sorghum, and green for *Arabidopsis*. The names used for rice and the Arabidopsis *TCP* genes are from a previous report [8].

2.3. Many TCP Genes are Found in the Syntenic Segments of Rice, Sorghum, and Maize Genomes

As gene synteny is indicative of the homologous gene function, we explored the collinearity of rice, sorghum, and maize *TCP* genes. We collected gene collinearity data from the Plant Genome Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication, Table S1)using *ZmTCP* genes as anchors, then defined each genomic syntenic block as the chromosomal segment consisting of multiple homologous genes, across species. According to this analysis, about 20 *ZmTCP* genes were found to have syntenic members or collinear genes in rice and sorghum, as shown in Figure 2. Some chromosomal segments containing *ZmTCP* genes, including *ZmTCP2*, *3*, *8*, 9, *19*, *25*, *29*, *40*, and *ZmTCP41*, were found to have been evolutionally conserved between maize, rice, and sorghum (Figure 2). This indicated that not only the individual genes but these entire chromosomal segments were evolutionally conserved. Two segments on maize chromosomes 2 and 7, containing a duplicated gene pair *ZmTCP5* and *ZmTCP23,* share synteny with two segments on rice chromosomes 7 and 9, carrying the corresponding genes *OsTCP22* and *OsTCP24*. Two other segments on maize chromosomes 2 and 4, containing another duplicated gene pair *ZmTCP33* and *ZmTCP16,* share synteny with two segments on rice chromosomes 11 and 12, carrying corresponding genes *OsPCF3* and *OsTCP28*. Interestingly, these two paralogous gene pairs, *ZmTCP5*/*ZmTCP23* and *ZmTCP16*/*ZmTCP33*, share the same syntenic blocks on sorghum chromosomes 2 and 7, respectively, indicating that these duplicated gene pairs might have arisen from segmental duplication in maize, after maize and sorghum diverged evolutionarily (Figure 2). One rice chromosomal block, containing *OsTCP19*, shares synteny with two segments on maize chromosomes 6 and 9, with a duplicated gene pair *ZmTCP21* and *ZmTCP26.* However, only one syntenic block could be found in the sorghum genome. Another fragment on rice chromosome 1 containing *OsTCP6* and *OsTCP5* is also duplicated on maize chromosomes 3 and 8, carrying two duplicated gene pairs *ZmTCP10*/*43* and *ZmTCP37*/*44.* More interestingly, their orthologs in

sorghum are also tandemly duplicated on only a single chromosomal segment. Additionally, one maize chromosomal block containing *ZmTCP32* shares synteny with two segments on rice chromosomes 8 and 9, containing its orthologous genes *OsPCF2* and *OsTCP25*; *SbTCP* orthologs could not be found, although the syntenic segment in sorghum was identified. Based on these data, we concluded that most of the *TCP* genes existed before the species diverged, but some *ZmTCP* genes might have originated from later segmental duplication or accompanied the generation of an allotetraploid maize genome.

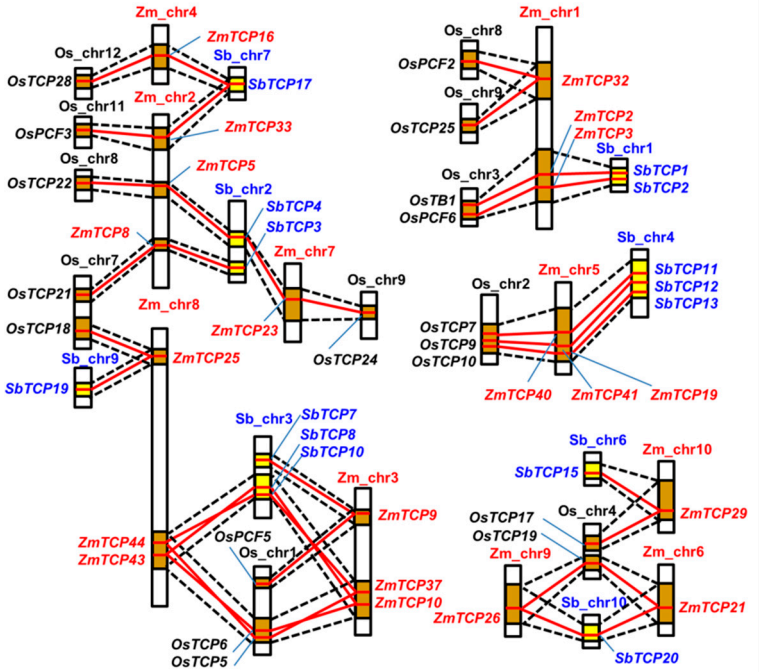


Figure 2. Schematic diagram of syntenic chromosomal segments containing *ZmTCP* genes between the rice, sorghum, and maize genomes. The maize, rice, and sorghum chromosomes are abbreviated Zm, Os, Sb, respectively. Homologous chromosomal regions between the different genomes are linked by black dotted lines and pale blue shaded regions. Each *TCP* orthologous gene pair was connected by a red line. Yellow boxes indicate homologous segments between the maize and the sorghum genomes, while the brown boxes identify the homologous regions in the maize and rice genomes.

2.4. Expression Profiles of ZmTCP Genes

According to previous studies, *TCP* genes had key roles in different aspects of plant development, as well as in response to stress [8,37]. In order to gain further insights into the roles of the *ZmTCP* genes, we assessed their tissue-specific expression profiles from the available transcriptomic data of maize B73 [43]. An expression heatmap was constructed for the 46 *ZmTCPs* in different tissues, from 15 developmental stages, under non-limiting growth conditions (Figure 3A). Results indicated that the expression patterns of different *ZmTCP* genes varied greatly. Transcripts of *ZmTCP37* and *ZmTCP44* showed relatively high levels of expression, compared to the other *ZmTCPs* examined. It is noteworthy that *ZmTCP2/TB1* and *ZmTCP18*, a homolog of *OsTB1*, were expressed relatively high in cobs and husk leaves, while *ZmTCP25* was expressed relatively high in endosperm and mature seeds. Additionally, in the B73 variety grown under non-limiting conditions, all *ZmTCP* genes in class I (PCF clade)were expressed relatively high in different tissues.

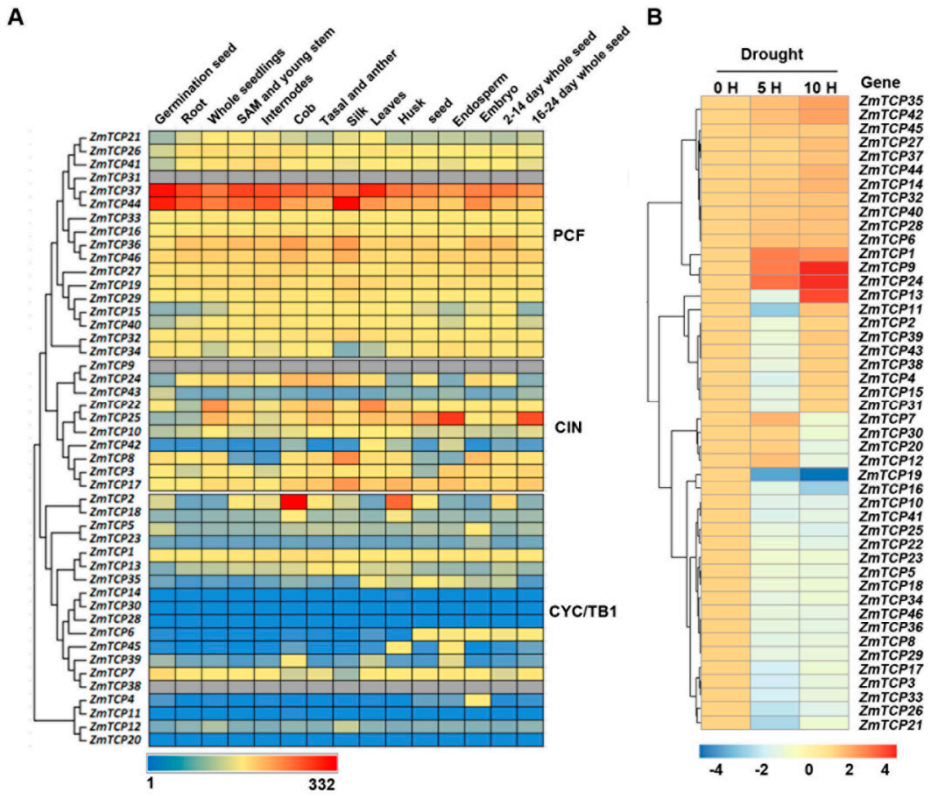


Figure 3. Expression patterns of *ZmTCP* genes in tissues and in response to drought conditions. (**A**) A heat map depicting gene expression levels of 46 *ZmTCPs* in fifteen different tissues from various developmental stages. Normalized gene expression values are shown in different colors that represent the levels of expression indicated by the scale bar. The gray color represents unavailable data. (**B**) Microarray-based expression analysis of *ZmTCP* genes. A heat map was generated based on the fold-change values in the treated samples, when compared with the unstressed control. The color scale for fold-change values is shown at the bottom. The drought-treated leaf samples were collected at two time points, 5 and 10 h, which reflected relative leaf water content (RLWC)of 70% and 60%, respectively.

Next, we investigated the expression profiles of all maize *TCP* genes, in response to drought stress, using microarray analysis. Based on their expression patterns, the maize *TCP* genes could mainly be classified into four groups (Figure 3B). The expression levels of 14 *ZmTCP* genes were continuously up-regulated (fold-change > 1)in response to both drought stress conditions (5 h and 10 h of drought treatments). Among these genes, *ZmTCP1*, *ZmTCP9*, and *ZmTCP24* showed two-fold changes (or more)in one or two drought stress conditions. In contrast, 19 *TCP* genes were continuously down-regulated (fold-change < −1)under both drought stress conditions, of which *ZmTCP19* showed about four fold change in one or two drought stress conditions. Expression of the remaining 13 genes, were either suppressed or induced under one of the drought conditions. The results clearly showed the functional divergence of *ZmTCP* genes in response to drought stress, in maize seedlings.

2.5. Association Analysis of Natural Variations in ZmTCP Genes Identified Two ZmTCP Genes Associated with Drought Tolerance in Maize

In order to further investigate whether the natural variations in any of the ZmTCP TFs are associated with the different drought tolerance levels of maize varieties, we conducted an association analysis for these genes. The drought tolerance level of each inbred line was investigated by evaluating its survival rate under severe drought stress, at the seedling stage. To assess potential associations between survival rates and *ZmTCPs*, we utilized previously reported methods and data [44,45] and previously identified single nucleotide polymorphism (SNP) markers, to characterize the presence of

genetic polymorphisms in each of these 46 *ZmTCP* genes. Among the 46 identified *ZmTCP* genes, 26 were found to be polymorphic with an average of 12 SNPs (Table 2), while the polymorphic information of the other 16 genes was currently absent (Minor Allele Frequency, MAF ≥ 0.05).

Table 2. Association analysis of the natural variation in *ZmTCP* genes with respect to drought tolerance at the seedling stage in the maize diversity panel.

Locus ID	Gene Name	Polymorphic Name	GLM	PCA	PCA + K	
			$p \leq 0.01$	$p \leq 0.01$	$p \leq 0.01$	$p \leq 0.001$
GRMZM2G166687	*ZmTCP1*	-	-	-	-	-
AC233950.1_FG002	*ZmTCP2*	3	2	0	0	0
GRMZM2G115516	*ZmTCP3*	33	0	0	0	0
AC234521.1_FG006	*ZmTCP4*	-	-	-	-	-
GRMZM2G110242	*ZmTCP5*	1	0	0	0	0
GRMZM2G088440	*ZmTCP6*	-	-	-	-	-
GRMZM2G414114	*ZmTCP7*	-	-	-	-	-
GRMZM2G020805	*ZmTCP8*	21	0	0	0	0
AC205574.3_FG006	*ZmTCP9*	19	0	0	0	0
GRMZM2G166946	*ZmTCP10*	1	0	0	0	0
GRMZM2G055024	*ZmTCP11*	-	-	-	-	-
GRMZM2G062711	*ZmTCP12*	-	-	-	-	-
GRMZM2G060319	*ZmTCP13*	1	0	0	0	0
GRMZM2G135461	*ZmTCP14*	-	-	-	-	-
GRMZM2G078077	*ZmTCP15*	5	0	0	0	0
GRMZM2G003944	*ZmTCP16*	-	-	-	-	-
GRMZM2G089361	*ZmTCP17*	28	1	0	0	0
AC190734.2_FG003	*ZmTCP18*	-	-	-	-	-
GRMZM2G445944	*ZmTCP19*	7	0	0	0	0
GRMZM2G031905	*ZmTCP20*	-	-	-	-	-
GRMZM2G142751	*ZmTCP21*	2	0	0	0	0
GRMZM2G120151	*ZmTCP22*	2	0	0	0	0
GRMZM2G064628	*ZmTCP23*	-	-	-	-	-
GRMZM2G015037	*ZmTCP24*	2	0	0	0	0
GRMZM2G035944	*ZmTCP25*	27	1	0	0	0
GRMZM2G113888	*ZmTCP26*	4	0	0	0	0
GRMZM2G096610	*ZmTCP27*	12	0	0	0	0
GRMZM2G458087	*ZmTCP28*	-	-	-	-	-
GRMZM2G465091	*ZmTCP29*	1	0	0	0	0
GRMZM2G454571	*ZmTCP30*	-	-	-	-	-
AC213524.3_FG003	*ZmTCP31*	-	-	-	-	-
GRMZM2G107031	*ZmTCP32*	19	0	1	1	0
GRMZM2G416524	*ZmTCP33*	-	-	-	-	-
AC199782.5_FG003	*ZmTCP34*	12	0	0	0	0
GRMZM5G824514	*ZmTCP35*	-	-	-	-	-
GRMZM2G089638	*ZmTCP36*	2	0	0	0	0
GRMZM2G092214	*ZmTCP37*	34	3	0	0	0
GRMZM2G359599	*ZmTCP38*	-	-	-	-	-
GRMZM2G170232	*ZmTCP39*	-	-	-	-	-
GRMZM2G178603	*ZmTCP40*	2	1	0	0	0
GRMZM2G077755	*ZmTCP41*	-	-	-	-	-
GRMZM2G180568	*ZmTCP42*	29	2	2	2	2
GRMZM2G148022	*ZmTCP43*	1	0	0	0	0
GRMZM2G034638	*ZmTCP44*	22	6	0	0	0
GRMZM2G424261	*ZmTCP45*	-	-	-	-	-
GRMZM2G093895	*ZmTCP46*	27	2	0	0	0

ZmTCP37 was found to be the most polymorphic, with 34 SNPs in this natural diversity panel. Subsequently, three statistical models were applied to identify significant genotypic and phenotypic associations. Specifically, a general linear model (GLM)with the first two principal components (PC_2)and a mixed linear model (MLM)were used to find associations (Figure 4A). The GLM method was applied to perform single-marker analysis. Then PC_2, via the first two principal components of the SNP data, was applied to correct for spurious associations caused by the population structure. The MLM method, incorporating both PC_2 and a Kinship matrix (to correct for the effect of cryptic

relatedness), was considered to be effective for controlling false positives in the association analysis (Figure 4A) [46,47]. The candidate gene association analysis detected significant associations between the genetic variations of *ZmTCP32* and *ZmTCP42* and drought tolerance, under different models, with a *p*-value ≤ 0.01 (Table 2; Figure 4B,C). However, under the standard mixed linear model (MLM), the only two significantly associated SNPs contributing to the phenotype of drought tolerance were both located at the 5′ UTR region of *ZmTCP42*, which suggested that this candidate gene was significantly associated with drought tolerance (*p*-value ≤ 0.001, $-\log_{10}p = 3.77$) (Figure 4C). We further analyzed the survival rates of maize inbred lines carrying *ZmTCP42* drought-tolerant or drought-sensitive alleles. It was found that $ZmTCP42^{AA}$ was a favorable drought tolerance allele (Figure S2). The two-fold induction of *ZmTCP42* expression by dehydration (Figure 3B)suggested that the RNA level of *ZmTCP42* was likely more correlated with drought tolerance than other potential variations among different maize varieties.

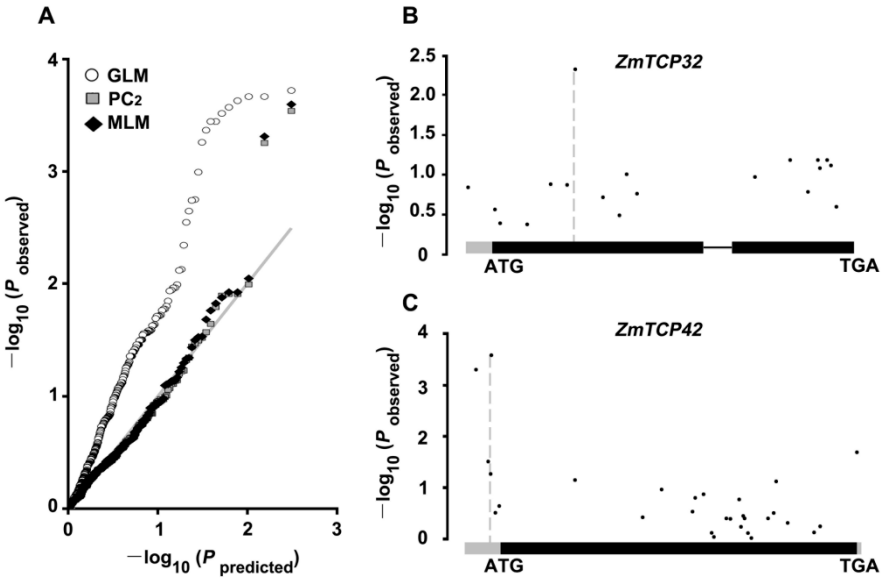


Figure 4. Association analysis of genetic variations in *ZmTCP32 and ZmTCP42* with maize drought tolerance. (**A**) Quantile–quantile plots of the estimated −log10(*p*) from the *ZmTCP* gene family-based association analysis, using three methods. The gray line is the expected line under the null distribution. The white square represents the observed *p* values using general linear model (GLM); the gray square represents the GLM model with the first two principal components (PC_2); the black diamond represents the observed *p* values using the mixed linear model (MLM) model incorporating both PC_2 and a Kinship matrix. (**B**,**C**) Schematic diagrams of *ZmTCP32* (**B**) and *ZmTCP42* (**C**), including the UTR (gray), intron (thin black line), and protein coding regions (thick black line), are presented in the x-axis. The *p* value is shown on a −log10 scale.

2.6. ZmTCP32 and ZmTCP42 Are Both Induced by ABA Treatment and Drought Stress

To confirm whether *ZmTCP32* and *ZmTCP42* RNA levels are truly associated with drought tolerance, we used RT-qPCR to directly analyze the RNA expression of the *ZmTCP32* and *ZmTCP42* genes, in response to ABA treatment and drought stress. As illustrated in Figure 5, in the B73 genotype, *ZmTCP32* and *ZmTCP42* are both significantly induced by ABA by roughly four-fold, relative to the controls at 48 h after the ABA treatment. More importantly, *ZmTCP42* was highly induced by drought stress (+7.6-fold at 24 h) and by the PEG6000 treatment (+5.9-fold at 24 h); *ZmTCP42* was significantly more responsive to dehydration stress and PEG treatment than *ZmTCP32*, even though *ZmTCP32* was also responsive to drought stress (+3.9-fold at 24 h) and PEG treatment (+3.0-fold at 24 h) (Figure 5).

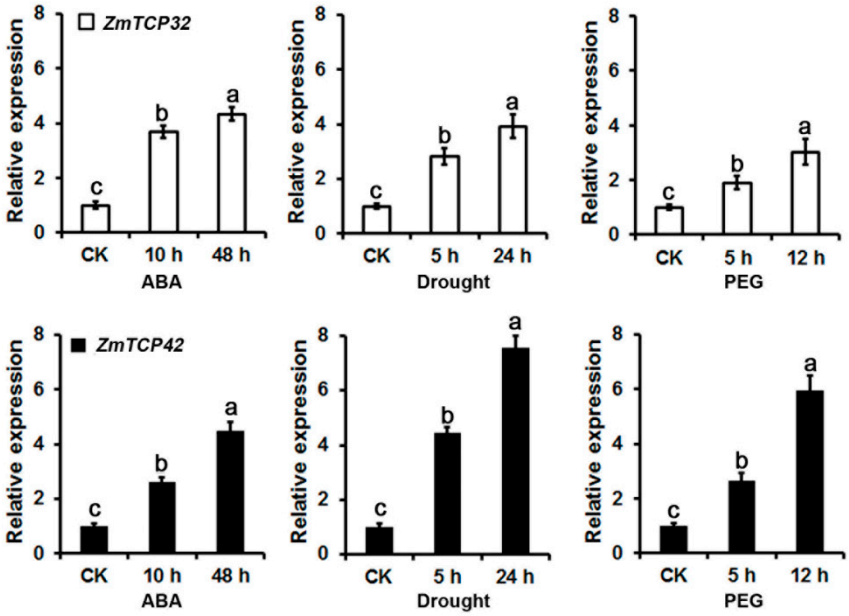


Figure 5. The RNA levels of *ZmTCP32* and *ZmTCP42* in maize B73 leaves measured by RT-qPCR, after treatment with 100 μM ABA, drought stress, or 10% PEG treatment. The drought stress was performed as previously described [45], on hydroponically cultured seedlings, for 5 and 24 h, and their relative leaf water contents (RLWC) were determined to be approximately 70% and 58%, at the corresponding time points, respectively. *ZmUbi-2* transcript levels were used as an internal control for data normalization; the gene-specific primers are listed in Table S2. Each data point represents the mean ± SD (n = 3) of three biological replicates. Significant differences were calculated by one-way ANOVA with Duncan's multiple test (SAS Institute, Inc., Cary, NC, USA). Different letters indicate a significant statistical difference between sample means, at p = 0.05, while means with the same letters were not significantly different.

2.7. Overexpression of ZmTCP42 Enhances Drought Resistance in Transgenic Arabidopsis

Given that the polymorphism in the 5′-UTR of *ZmTCP42* suggests a potential role in drought resistance on maize seedlings, we selected *ZmTCP42* to directly test for its function. To validate *ZmTCP42* function in response to drought stress, we generated transgenic Arabidopsis plants that overexpress the *ZmTCP42* gene, by using the enhanced cauliflower mosaic virus 35S promoter. After screening for the *ZmTCP42* expression levels, we selected two independent transgenic lines, *ZmTCP42-OE16* and *ZmTCP42-OE25*, with enhanced RNA levels, for further experiments (Figure 6A,B). We investigated the ABA sensitivity of *ZmTCP42* transgenic lines, in response to exogenous ABA, during seed germination. In the absence of exogenous ABA, all lines germinated completely, as did the wild-type seeds; on the contrary, in the presence of exogenous ABA, the expanding, greening cotyledons were significantly affected in the *ZmTCP42*-OE lines; the number of seedlings with green cotyledons in *ZmTCP42-OE16* and *ZmTCP42-OE25* were significantly lower than those of the wild-type at 1 μM ABA (Figure 6C,D), suggesting that the *ZmTCP42* overexpression lines are more sensitive to ABA. Furthermore, we analyzed the tolerance levels of *ZmTCP42*-OE lines to drought stress. When the 28-day-old transgenic seedlings were subjected to the drought test, only ~30% of the wide-type plants were able to recover from the stress, whereas ~65 of *ZmTCP42-OE16* and ~87% of *ZmTCP42-OE25* transgenic plants survived (Figure 6E,F). Similar results were obtained in repeated experiments, indicating that overexpression of *ZmTCP42* enhanced drought tolerance in Arabidopsis.

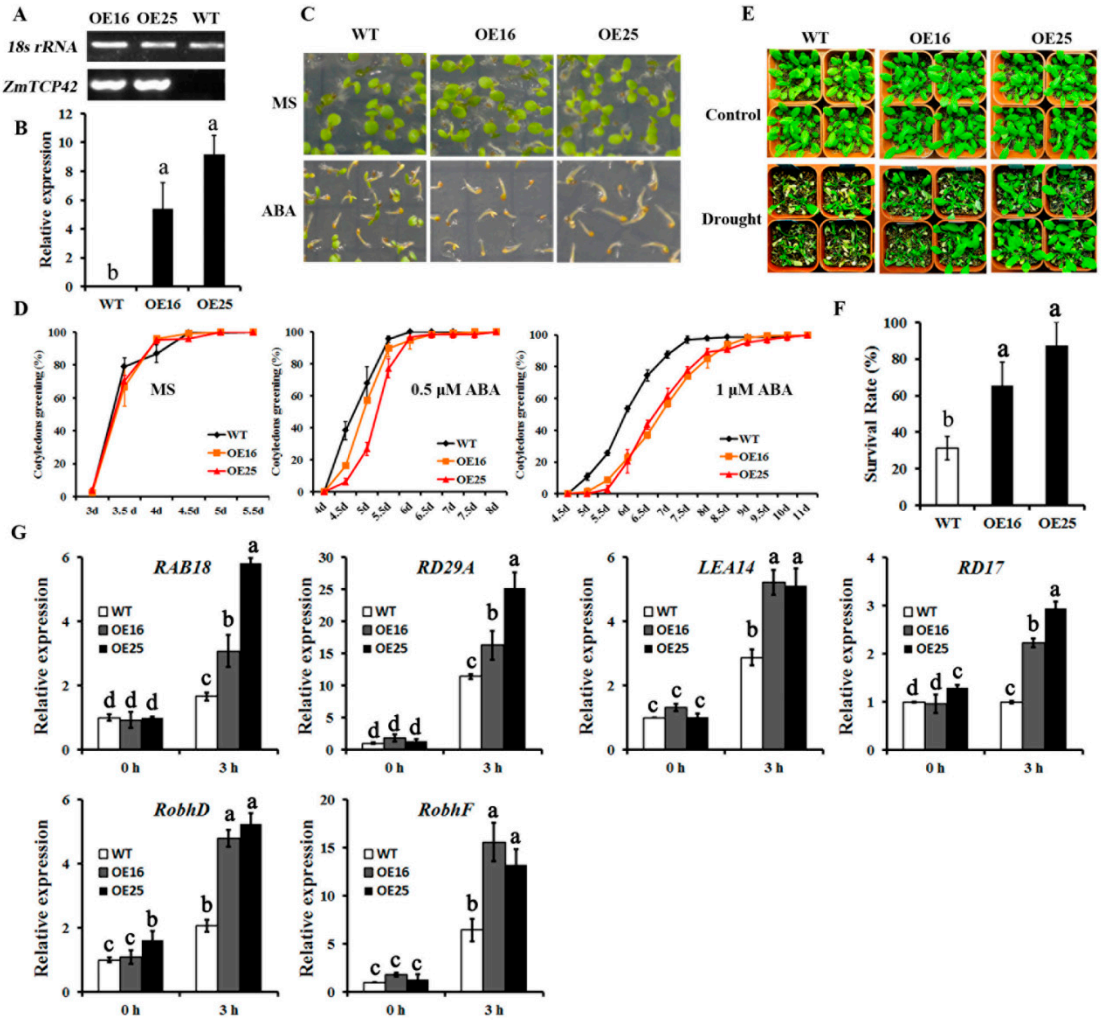


Figure 6. Analysis of wild-type (WT) and *ZmTCP42*-overexpressing transgenic *Arabidopsis* plants. (**A**) RT-PCR analysis of *ZmTCP42* transcript levels in *ZmTCP42* transgenic lines. (**B**) RT-qPCR analysis of *ZmTCP42* transcript levels in *ZmTCP42* transgenic lines. (**C**) Seed germination of the *ZmTCP42* transgenic lines and in the wild-type, in response to abscisic acid (ABA). Germination rates with 1 μM ABA for 5 days, defined by cotyledon greening, in wild-type *ZmTCP42-OE16* and *ZmTCP42-OE25*, compared to the MS medium alone. (**D**) Statistical analysis of a green seedling's rate of *ZmTCP42* transgenic lines and wild-type grown on Murashige and Skoog (MS) medium with 0, 0.5 μM, and 1 μM ABA. (**E**) Drought tolerance of the *ZmTCP42* transgenic Arabidopsis plants. Photographs were taken before and after the drought treatment, followed by 3 days of rewatering. (**F**) Statistical analysis of survival rates after the drought-stress treatment. The average survival rates and standard errors were calculated from three independent experiments. (**G**) RNA levels of stress-responsive genes in the *ZmTCP42* transgenic lines and in the wild-type, in response to drought stress. Total RNA was obtained from 3 weeks-old seedlings treated by 3 h of dehydration stress and was analyzed by RT-qPCR, using the gene-specific primers listed in Table S2. The mean value and standard error were calculated from three replicates after normalization to *ACTIN2*. The RNA levels in the wild-type grown under non-stress conditions was taken as 1.0. Significant differences were calculated by one-way ANOVA with Duncan's multiple test (SAS Institute, Inc., Cary, NC, USA). Different letters indicate a significant statistical difference between the sample means at $p = 0.05$, while means with the same letters were not significantly different.

We further assessed the expression levels of ABA- inducible or drought-inducible genes by using real-time RT-qPCR to analyze their responses in *Arabidopsis*. Upon dehydration stress, the RNA levels of *RAB18*, *RD29A*, *LEA14*, and *RD17*, which are well-known drought-responsive, positive regulator

genes of drought tolerance, were up-regulated, relative to wild-type *Arabidopsis* plants (Figure 6G). Similar changes were observed for the RNA levels of the *RbohD* and *RbohF* genes, encoding two NADPH oxidases, which were directly responsible for the reactive oxygen species (ROS) production in leaves, in response to stress (Figure 6G). These results clearly showed that *ZmTCP42* overexpression improved the inducibility of *Arabidopsis* drought-tolerance-associated genes upon drought stress, leading to an enhanced drought tolerance.

3. Discussion

To date, the TCP family members have been described in various species; for instance, 24 *TCP* genes in *Arabidopsis*, 23 in rice, 20 in sorghum, and 66 in wheat (*Triticum aestivum* L.) [11,12,16,17]. Chai et al. (2017)reported 29 *ZmTCP* genes in maize, including their chromosomal location, structure and domain conservation analysis, and gene duplication analysis. They also analyzed the phylogenetic inference and expression profiling of these 29 *ZmTCP* genes, and then speculated that the *ZmTCP* genes influence stem and ear growth. In our study, we retrieved a total of 46 *ZmTCP* genes, by searching against the updated maize genome B73_RefGen_v3, including the 29 *ZmTCPs* previously reported by Chai et al. and the 17 newly identified *ZmTCPs* containing a canonical TCP domain. Subsequently, we further systematically determined their phylogenetic relationship, synteny with rice, sorghum, and Arabidopsis *TCPs*, pattern of drought-responsiveness, association analysis of their natural variations with drought tolerance, and functional analysis of *ZmTCP42* in drought tolerance. Collectively, our data demonstrated that a few *ZmTCP* family proteins are likely involved in plant drought tolerance; in particular, ZmTCP42 functions as an important positive regulator for drought tolerance.

Recently, genome-wide identification revealed that segmental duplication might be the main contributor to the expansion of the *TCP* gene family in some monocots, including rice, sorghum and wheat, as well as in some dicot species, such as cotton and soybean [14,17,48–52]. The report showed that there were about 2.6–2.8-fold duplication of the *TCP* gene family in allotetraploid upland cotton genome (*G. hirsutum*), compared to *Arabidopsis* [48]. Similarly, we found that the duplication ratio of the *ZmTCP* family was about 2-fold in the allopolyploid maize genome, compared with *Arabidopsis*, rice, and sorghum. The change in the ratio of gene numbers suggests that the *ZmTCP* family has undergone lineage-specific expansion and functional divergence during the course of evolution. Furthermore, as shown in the phylogenetic analysis, monocot ZmTCP proteins clustered independently from those in *Arabidopsis*, suggesting a potential functional divegence between the dicot and monocot TCPs. We also found that *ZmTCP* gene expansion was not uniform, and occurred mainly in the class II CYC/TB1 clade. However, the biological significance of the *ZmTCP* gene duplication in the maize genome remains to be determined. Our collinearity analysis on *TCP* genes within several monocot plants showed that 20 *ZmTCP* genes had syntenic members or collinear genes in rice and sorghum, and that the chromosomal segments containing these genes were also duplicated (Figure 2), supporting the concept that the maize genome might have arisen from an ancestral allotetraploid, half of which share a common ancestor with sorghum, which in turn probably represents a lineage split from rice [53,54]. Taken together, our results might provide some useful clues for future studies on a homologous gene function.

A growing body of research suggest that TCP TFs play important roles in plant development, as well as in response to abiotic stress [8,19,33,34]. However, the role of *TCP* genes in plant response to drought stress in maize was still obscure, without answers to which, the *ZmTCP* genes were directly associated with the levels of drought tolerance in maize. We were, therefore, prompted to perform this study to address this question. The answer would not only help facilitate the genetic improvement of drought tolerance but would also increase our knowledge of the biological function of this gene family. Generally, *ZmTCP* genes exhibit great differential expression in response to drought stress in maize seedlings, not only among subgroups but members within the same subgroup, suggesting that these *ZmTCP* genes might function very diversely. Our current results showed that 14 *ZmTCP* genes were significantly up-regulated and 19 genes were clearly down-regulated, indicating that these genes might function as key mediators of drought stress responses in maize (Figure 3). Notably, three genes,

ZmTCP1, *ZmTCP9*, and *ZmTCP24*, belonging to the class II CYC/TB1 clade, showed similar inducible expression patterns (Figure 3), implying that they might play a redundant role in regulating drought stress response in maize. Moreover, the orthologs of *ZmTCP9* and *ZmTCP24* in rice and *Arabidopsis* are *OsPCF5* and *AtTCP3*/*AtTCP4*, respectively, which are regulated by the conserved miR319. Previous reports have shown that knockdown of miR319-dependent TCPs (by constitutively overexpressing miR319)increased drought and salinity tolerance [34,55]. The rice homolog of *ZmTCP1* and *ZmTCP13* was named *OsTB1*, which plays an important role in stress response, especially in regulation of cold tolerance [55,56]. Additionally, *ZmTCP37* and *ZmTCP44*, orthologs of *AtTCP20*, also showed a slightly inducible expression pattern under drought stress (Figure 3). In *Arabidopsis*, *AtTCP20* represses the transcription of *LIPOXYGENASE2* (*AtLOX2*) gene, which is involved in jasmonic acid synthesis and promotes leaf senescence [57].

To explore the association of *TCPs* with stress tolerance in plants, we have investigated homologous *ZmTCP* genes in a small number of inbred lines and proposed predictive conclusions. Among the 46 *ZmTCP* genes analyzed, the genetic polymorphisms of *ZmTCP32* and *ZmTCP42* were significantly associated with the phenotypic variations of drought tolerance (p-value ≤ 0.01, PC_2) (Table 2). While *ZmTCP42* was the most significantly (p-value ≤ 0.001, MLM) associated with drought tolerance in this natural variation panel, the natural variation in the *ZmTCP42* promoter might contribute to maize drought tolerance (Figure 4). In the B73 genotype of maize, *ZmTCP32* was constitutively highly expressed in various tissues (Figure 3A). In comparison with *ZmTCP32*, *ZmTCP42* RNA was detected at a low level in various tissues, except in leaves and endosperm, and showed a relatively high level at 10 h, upon drought stress in a micrroarry analysis (Figure 3B). Both *ZmTCP32* and *ZmTCP42* were significantly induced by ABA, dehydration and PEG treatment; more notably, *ZmTCP42* was induced higher than *ZmTCP32* in response to dehydration and PEG treatment. These results suggested that *ZmTCP42* was involved in plant drought response. Transgenic *Arabidopsis* overexpressing *ZmTCP42* exhibited a higher ABA sensitivity, improved drought stress tolerance, and enhanced the induction of Arabidopsis ABA- or drought-inducible genes, which strongly suggests that ZmTCP42 might regulate these ABA/drought-response genes. In summary, we have identified ZmTCP42 as an important positive regulator of drought tolerance, through analyses of gene expression and natural variations.

4. Materials and Methods

4.1. Plant Growth Conditions and Stress Treatments

Maize (Inbred line B73) growth conditions and stress treatments were performed according to Wang et al. [45]. The hydroponically cultured 3-leaf stage seedlings were used. For drought treatments, the seedlings were placed on a clean bench and subjected to dehydration (at 28 °C, with a relative humidity of 40–60%). Samples were exposed for 0, 5 h, and 24 h, at which the relative leaf water contents were measured (RLWC) to be 98%, 70%, and 60%, respectively, in the drought-treated leaf samples. ABA and PEG treatments were applied by immersing the seedlings in 100 µM ABA or 10% PEG 6000. All leaf samples from a minimum of 3 seedlings were immediately frozen in liquid nitrogen, then stored at −80 °C, prior to RNA isolation. The *Arabidopsis thaliana* ecotype Col-0 was used in this study. Seeds were grown on the Murashige and Skoog (MS) medium with an addition of 3% sucrose and 0.6% agar (pH 5.8). Drought treatment was applied to the 3-week-old seedlings by removing them from the MS medium plates and desiccating them on Whatman 3MM paper, on a clean bench, for 0 h and 3 h, respectively.

4.2. Reverse Transcription PCR and RT-qPCR Analysis

Total RNA was isolated from the collected leaf samples using the TRIZOL reagent (Biotopped, China) according to the manufacturer's instructions. One microgram of the total RNA from each sample was used in reverse transcription. RT-qPCR analyses were performed in the optical 48-well plates, using the ABI7300 Thermo-cycler (Applied Biosystems, USA). Reactions were carried out in a

10 µL volume containing 1 µL diluted cDNA, 200 nM gene-specific primers, and 5 µL SYBR Premix Ex Taq II (Takara, China) with the following conditions—10 min at 95 °C, 40 cycles of 15 s at 95 °C, and 30 s at 60 °C. Each experiment was performed using at least three independent biological replicates. *ZmTCP32* and *ZmTCP42* RNA levels in response to different stresses were calculated according to a standard curve calibration, based on amplification of a dilution series of *ZmTCP32* and *ZmTCP42* plasmid, respectively, according to the manufacturer's protocol. *ZmUbi-2* (UniProtKB/TrEMBL; ACC: Q42415) was used as the internal control. The expression levels of different stress-responsive genes were compared, based on the delta Ct method with normalization to *ACTIN2*.

4.3. Generation of Transgenic Plants

The coding region of the *ZmTCP42* cDNA of the maize B73 inbred line was inserted into the pGreenII vector [58]. The constructed plasmid carrying the desired gene was transformed into *Agrobacterium tumefaciens* GV3101 + pSoup and transformed into ecotype Col-0, as described previously [58]. Using the kanamycin-based selection, several independent T_2 transgenic lines were obtained, and expression of the *ZmTCP42* transgene was confirmed in these lines by RT-PCR and RT-qPCR. Two independent overexpression lines *ZmTCP42-OE16* and *ZmTCP42-OE25* (with a single inserted copy) were selected, based on the level of transgene expression, and the T3 homozygous seeds were subjected to further analyses.

4.4. Seed Germination and Drought Phenotype Analysis

To study the effect of ABA on germination and cotyledon greening, the seeds were planted on MS medium plates, with an addition of plus 1% sucrose and different concentrations of ABA (Sigma-Aldrich)in a growth chamber, at 22 °C under a 16-h-light/8-h-dark photo-period, with a 60% relative humidity, after 2 days of vernalization in darkness at 4 °C. For the drought stress tolerance test, plants were grown for 28 days under normal conditions and subjected to water stress by withholding watering for 14 days [58]. Four plants were planted in each small cup with 100 g soil (1:1 of black soil/vermiculite)and grown under the condition of 16-h-light/8-h-dark. Four independent experiments were performed and at least 16 plants each were observed, for the *ZmTCP42-OE16* and *ZmTCP42-OE25* lines and the WT, in each experiment. The plants were rewatered when significant differences in wilting were observed. Three days after rewatering, the surviving plants were counted.

4.5. Identification of the TCP Proteins in the Maize Genome

To identify TCP proteins in the B73 maize genome, some putative TCP protein sequences from maize (genome assembly: B73_RefGen_v3) were first retrieved from the Plant Transcription Factor Database 3.0 (Jin et al. 2014). Conserved TCP DNA-binding domain (PF03634) from the Pfam database [59] was used to search and retrieve sequences from the Phytozome database v10.0 (available online: http://www.phytozome.net/eucalyptus.php). In addition, BLASTP searches were also performed against the maize genome, to identify any additional TCP members using *Arabidopsis* and rice TCP proteins sequences (*E-value* ≤ 0.00001). The overlapped genes were removed. The presence of a TCP domain in all family proteins was evaluated using the CDD database searches (https://www.ncbi.nlm.nih.gov/Structure/cdd). Arabidopsis TCP proteins were downloaded from TAIR 10 (available online: http://www.arabidopsis.org), which contained 24 members. TCP proteins in rice (Genome assembly: Rice Genome Annotation Project Database release 7) and sorghum (Genome assembly: V3) were both downloaded from the Plant Transcription Factor Database. The ExPASy program (http://www.expasy.org/tools/) was used to predict the molecular weight (kDa) and the isoelectric point (PI) of each protein.

4.6. Gene Structure and Phylogenetic Analysis

Genomic sequences of maize TCP genes were downloaded from Phytozome V10, and the untranslated regions were removed. To show the exon/intron organization for individual TCP genes,

coding sequences were aligned to genomic sequences and schematics generated using GSDS 2.0 (Gene Structure Display Server) (available online: http://gsds.cbi.pku.edu.cn) [60]. To determine the phylogenetic relationships of the TCP proteins, full-length amino acid sequences of TCPs identified in maize, rice, *Arabidopsis*, and sorghum were aligned by the ClustalW program [61]. The Phylogenetic tree was constructed by the neighbor-joining method, with 1000 bootstrap replicates.

4.7. Association Analysis

Association analysis for *ZmTCPs* was performed by using a maize association mapping population containing 367 inbred lines and the corresponding drought tolerance phenotype data from a previous study [45]. Among the 556,945 high-quality single nucleotide polymorphism (SNP) data, with minor allele frequency (MAF)≥ 0.05, 317 SNPs were found in the genic region of the 46 *ZmTCPs*. The general linear model (GLM) model, the GLM model with the first 2 principal components (PC_2), and the mixed linear model (MLM) model were chosen to detect the SNPs significantly associated with drought tolerance, by using the TASSEL5.0 program [62].

4.8. Tissue Expression Profile and Microarray Analysis

Expression patterns of 46 *ZmTCPs* in different maize tissues were analyzed using the genome-wide gene expression atlas of the inbred B73 line of maize that was reported previously [43]. Expression data for the 15 tissues were combined from 60 growth stages. Normalized expression values of each gene in different tissues were averaged. The gene expression level was presented as a log value. To analyze the gene expression patterns of *ZmTCPs* in maize during drought stress, the microarray data and analysis methods were employed as described previously [63].

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/11/2762/s1.

Author Contributions: S.D. and Z.C. designed the experiments, performed the experiments, analyzed the data, and prepared figures and tables. H.D. gave good advice on the work. H.W. wrote the manuscript. All authors read and approved the final manuscript.

Funding: This work was supported by the National Natural Science Foundation (31701062 and 31701439), the Hubei Province Educational Commission Foundation (Q20171307), the Plan in Scientific and Technological Innovation Team of Outstanding Young, Hubei Provincial Department of Education (T2017004).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

TCP	the teosinte branched1/cycloidea/proliferating
TF	transcription factor
CYC/TB1	cycloidea/teosinte branched 1
PCF	proliferating cell nuclear antigen factor
CIN	CINCINNATA of Antirrhinum
RT-qPCR	reverse transcription quantitative polymerase chain reaction
GLM	general linear model
SNP	single nucleotide polymorphism
MLM	mixed linear model
ABA	abscisic acid
PEG	polyethylene glycol

References

1. Barnabas, B.; Jager, K.; Feher, A. The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* **2007**, *31*, 11–38. [CrossRef] [PubMed]
2. Yordanov, I.; Velikova, V.; Tsonev, T. Plant Responses to Drought, Acclimation, and Stress Tolerance. *Photosynthetica* **2000**, *38*, 171–186. [CrossRef]

3. Daryanto, S.; Wang, L.; Jacinthe, P. Global Synthesis of Drought Effects on Maize and Wheat Production. *PLoS ONE* **2016**, *11*, e0156362. [CrossRef] [PubMed]
4. Lobell, D.B.; Roberts, M.J.; Schlenker, W.; Braun, N.; Little, B.B.; Rejesus, R.M.; Hammer, G.L. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science* **2014**, *344*, 516–519. [CrossRef] [PubMed]
5. Akhtar, M.; Jaiswal, A.; Taj, G.; Jaiswal, J.P.; Qureshi, M.I.; Singh, N.K. DREB1/CBF transcription factors: Their structure, function and role in abiotic stress tolerance in plants. *J. Genet.* **2012**, *91*, 385–395. [CrossRef] [PubMed]
6. Baldoni, E.; Genga, A.; Cominelli, E. Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. *Int. J. Mol. Sci.* **2015**, *16*, 15811–15851. [CrossRef]
7. Yamaguchi-Shinozaki, K.; Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* **2006**, *57*, 781–803. [CrossRef]
8. Martin-Trillo, M.; Cubas, P. TCP genes: A family snapshot ten years later. *Trends Plant Sci.* **2010**, *15*, 31–39. [CrossRef]
9. Kosugi, S.; Ohashi, Y. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* **2002**, *30*, 337–348. [CrossRef]
10. Navaud, O.; Dabos, P.; Carnus, E.; Tremousaygue, D.; Herve, C. TCP Transcription Factors Predate the Emergence of Land Plants. *J. Mol. Evol.* **2007**, *65*, 23–33. [CrossRef]
11. Riechmann, J.L.; Heard, J.E.; Martin, G.; Reuber, L.; Jiang, C.; Keddie, J.; Adam, L.; Pineda, O.; Ratcliffe, O.J.; Samaha, R. Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science* **2000**, *290*, 2105–2110. [CrossRef] [PubMed]
12. Yao, X.; Ma, H.; Wang, J.; Zhang, D. Genome-Wide Comparative Analysis and Expression Pattern of TCP Gene Families in Arabidopsis thaliana and *Oryza sativa*. *J. Integr. Plant Biol.* **2007**, *49*, 885–897. [CrossRef]
13. Parapunova, V.; Busscher, M.; Busscher-Lange, J.; Lammers, M.; Karlova, R.; Bovy, A.G.; Angenent, G.C.; de Maagd, R.A. Identification, cloning and characterization of the tomato TCP transcription factor family. *BMC Plant Biol.* **2014**, *14*, 157. [CrossRef] [PubMed]
14. Ma, J.; Wang, Q.; Sun, R.; Xie, F.; Jones, D.C.; Zhang, B. Genome-wide identification and expression analysis of TCP transcription factors in Gossypium raimondii. *Sci. Rep.* **2014**, *4*, 6645. [CrossRef] [PubMed]
15. Kosugi, S.; Ohashi, Y. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* **1997**, *9*, 1607–1619. [CrossRef] [PubMed]
16. Zhao, J.; Zhai, Z.; Li, Y.; Geng, S.; Song, G.; Guan, J.; Jia, M.; Wang, F.; Sun, G.; Feng, N.; et al. Genome-Wide Identification and Expression Profiling of the TCP Family Genes in Spike and Grain Development of Wheat (*Triticum aestivum* L.). *Front. Plant Sci.* **2018**, *9*, 1282. [CrossRef] [PubMed]
17. Francis, A.; Dhaka, N.; Bakshi, M.; Jung, K.H.; Sharma, M.K.; Sharma, R. Comparative phylogenomic analysis provides insights into TCP gene functions in Sorghum. *Sci. Rep.* **2016**, *6*, 38488. [CrossRef] [PubMed]
18. Cubas, P.; Lauter, N.; Doebley, J.; Coen, E. The TCP domain: A motif found in proteins regulating plant growth and development. *Plant J.* **1999**, *18*, 215–222. [CrossRef]
19. Manassero, N.G.U.; Viola, I.L.; Welchen, E.; Gonzalez, D.H. TCP transcription factors: Architectures of plant form. *Biomol. Concepts* **2013**, *4*, 111–127. [CrossRef]
20. Li, S. The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development. *Plant Signal. Behav.* **2015**, *10*, e1044192.
21. Danisman, S. TCP Transcription Factors at the Interface between Environmental Challenges and the Plant's Growth Responses. *Front. Plant Sci.* **2016**, *7*, 1930. [CrossRef] [PubMed]
22. Dhaka, N.; Bhardwaj, V.; Sharma, M.K.; Sharma, R. Evolving Tale of TCPs: New Paradigms and Old Lacunae. *Front. Plant Sci.* **2017**, *8*, 479. [CrossRef]
23. Nicolas, M.; Cubas, P. The Role of TCP Transcription Factors in Shaping Flower Structure, Leaf Morphology, and Plant Architecture. In *Plant Transcription Factors*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 249–267.
24. Resentini, F.; Felipo-Benavent, A.; Colombo, L.; Blázquez, M.A.; Alabadí, D.; Masiero, S. TCP14 and TCP15 Mediate the Promotion of Seed Germination by Gibberellins in Arabidopsis thaliana. *Mol. Plant* **2015**, *8*, 482–485. [CrossRef] [PubMed]
25. Yanai, O.; Shani, E.; Russ, D.; Ori, N. Gibberellin partly mediates LANCEOLATE activity in tomato. *Plant J.* **2011**, *68*, 571–582. [CrossRef] [PubMed]

26. Schommer, C.; Palatnik, J.F.; Aggarwal, P.; Chételat, A.; Cubas, P.; Farmer, E.E.; Nath, U.; Weigel, D. Control of Jasmonate Biosynthesis and Senescence by miR319 Targets. *PLoS Biol.* **2008**, *6*, e230. [CrossRef] [PubMed]
27. Koyama, T.; Mitsuda, N.; Seki, M.; Shinozaki, K.; Ohme-Takagi, M. TCP Transcription Factors Regulate the Activities of ASYMMETRIC LEAVES1 and miR164, as Well as the Auxin Response, during Differentiation of Leaves in Arabidopsis. *Plant Cell* **2010**, *22*, 3574–3588. [CrossRef] [PubMed]
28. Efroni, I.; Han, S.-K.; Kim, H.; Wu, M.-F.; Steiner, E.; Birnbaum, K.; Hong, J.; Eshed, Y.; Wagner, D. Regulation of Leaf Maturation by Chromatin-Mediated Modulation of Cytokinin Responses. *Dev. Cell* **2013**, *24*, 438–445. [CrossRef] [PubMed]
29. Koornneef, M.; Hanhart, C.J.; Hilhorst, H.W.; Karssen, C.M. In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana. *Plant Physiol.* **1989**, *90*, 463–469. [CrossRef]
30. Finkelstein, R.R.; Gampala, S.S.; Rock, C.D. Abscisic acid signaling in seeds and seedlings. *Plant Cell* **2002**, *14*, S15–S45. [CrossRef]
31. Osakabe, Y.; Osakabe, K.; Shinozaki, K.; Tran, L.P. Response of plants to water stress. *Front. Plant Sci.* **2014**, *5*, 86. [CrossRef] [PubMed]
32. Raghavendra, A.S.; Gonugunta, V.K.; Christmann, A.; Grill, E. ABA perception and signalling. *Trends Plant Sci.* **2010**, *15*, 395–401. [CrossRef] [PubMed]
33. Mukhopadhyay, P.; Tyagi, A.K. OsTCP19 influences developmental and abiotic stress signaling by modulating ABI4-mediated pathways. *Sci. Rep.* **2015**, *5*, 9998. [CrossRef] [PubMed]
34. Zhou, M.; Li, D.; Li, Z.; Hu, Q.; Yang, C.; Zhu, L.; Luo, H. Constitutive Expression of a miR319 Gene Alters Plant Development and Enhances Salt and Drought Tolerance in Transgenic Creeping Bentgrass. *Plant Physiol.* **2013**, *161*, 1375–1391. [CrossRef] [PubMed]
35. Rueda-Romero, P.; Barrero-Sicilia, C.; Gomez-Cadenas, A.; Carbonero, P.; Onate-Sanchez, L. Arabidopsis thaliana DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. *J. Exp. Bot.* **2012**, *63*, 1937–1949. [CrossRef]
36. Gonzalez-Grandio, E.; Poza-Carrion, C.; Sorzano, C.O.S.; Cubas, P. BRANCHED1 Promotes Axillary Bud Dormancy in Response to Shade in Arabidopsis. *Plant Cell* **2013**, *25*, 834–850. [CrossRef] [PubMed]
37. Gonzalez-Grandio, E.; Pajoro, A.; Franco-Zorrilla, J.M.; Tarancon, C.; Immink, R.G.; Cubas, P. Abscisic acid signaling is controlled by a BRANCHED1/HD-ZIP I cascade in Arabidopsis axillary buds. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E245–E254. [CrossRef] [PubMed]
38. Li, H.; Peng, Z.; Yang, X.; Wang, W.; Fu, J.; Wang, J.; Han, Y.; Chai, Y.; Guo, T.; Yang, N.; et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **2013**, *45*, 43–50. [CrossRef]
39. Yan, J.; Kandianis, C.B.; Harjes, C.E.; Bai, L.; Kim, E.-H.; Yang, X.; Skinner, D.J.; Fu, Z.; Mitchell, S.; Li, Q.; et al. Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. *Nat. Genet.* **2010**, *42*, 322. [CrossRef]
40. Yang, Q.; Li, Z.; Li, W.; Ku, L.; Wang, C.; Ye, J.; Li, K.; Yang, N.; Li, Y.; Zhong, T.; et al. CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16969–16974. [CrossRef]
41. Wang, H.; Qin, F. Genome-Wide Association Study Reveals Natural Variations Contributing to Drought Resistance in Crops. *Front. Plant Sci.* **2017**, *8*, 1110. [CrossRef]
42. Chai, W.; Jiang, P.; Huang, G.; Jiang, H.; Li, X. Identification and expression profiling analysis of TCP family genes involved in growth and development in maize. *Physiol. Mol. Biol. Plants* **2017**, *23*, 779–791. [CrossRef] [PubMed]
43. Sekhon, R.S.; Lin, H.; Childs, K.L.; Hansey, C.N.; Buell, C.R.; de Leon, N.; Kaeppler, S.M. Genome-wide atlas of transcription during maize development. *Plant J.* **2011**, *66*, 553–563. [CrossRef] [PubMed]
44. Liu, S.; Wang, X.; Wang, H.; Xin, H.; Yang, X.; Yan, J.; Li, J.; Tran, L.S.; Shinozaki, K.; Yamaguchi-Shinozaki, K.; et al. Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. *PLoS Genet.* **2013**, *9*, e1003790. [CrossRef] [PubMed]
45. Wang, X.; Wang, H.; Liu, S.; Ferjani, A.; Li, J.; Yan, J.; Yang, X.; Qin, F. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* **2016**, *48*, 1233–1241. [CrossRef] [PubMed]
46. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006**, *17*, 155–160. [CrossRef] [PubMed]

47. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [CrossRef] [PubMed]
48. Li, W.; Li, D.D.; Han, L.H.; Tao, M.; Hu, Q.Q.; Wu, W.Y.; Zhang, J.B.; Li, X.B.; Huang, G.Q. Genome-wide identification and characterization of TCP transcription factor genes in upland cotton (*Gossypium hirsutum*). *Sci. Rep.* **2017**, *7*, 10118. [CrossRef]
49. Ma, X.; Ma, J.; Fan, D.; Li, C.; Jiang, Y.; Luo, K. Genome-wide Identification of TCP Family Transcription Factors from Populus euphratica and Their Involvement in Leaf Shape Regulation. *Sci. Rep.* **2016**, *6*, 32795. [CrossRef]
50. Mondragón-Palomino, M.; Trontin, C. High time for a roll call: Gene duplication and phylogenetic relationships of TCP-like genes in monocots. *Ann. Bot.* **2011**, *107*, 1533–1544. [CrossRef]
51. Sharma, R.; Kapoor, M.; KTyagi, A.; Kapoor, S. Comparative transcript profiling of TCP family genes provide insight into gene functions and diversification in rice and Arabidopsis. *J. Plant Mol. Biol. Biotechnol.* **2010**, *1*, 24–38.
52. Feng, W.; Lindner, H.; Robbins, N.E., 2nd; Dinneny, J.R. Growing Out of Stress: The Role of Cell- and Organ-Scale Growth Control in Plant Water-Stress Responses. *Plant Cell* **2016**, *28*, 1769–1782. [CrossRef] [PubMed]
53. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The Sorghum bicolor genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556. [CrossRef] [PubMed]
54. Gaut, B.S.; Doebley, J. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6809–6814. [CrossRef] [PubMed]
55. Yang, C.; Li, D.; Mao, D.; Liu, X.; Ji, C.; Li, X.; Zhao, X.; Cheng, Z.; Chen, C.; Zhu, L. Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (*Oryza sativa* L.). *Plant Cell Environ.* **2013**, *36*, 2207–2218. [CrossRef] [PubMed]
56. Chen, L.; Zhao, Y.; Xu, S.; Zhang, Z.; Xu, Y.; Zhang, J.; Chong, K. OsMADS57 together with OsTB1 coordinates transcription of its target OsWRKY94 and D14 to switch its organogenesis to defense for cold adaptation in rice. *New Phytol.* **2018**, *218*, 219–231. [CrossRef] [PubMed]
57. Danisman, S.; van der Wal, F.; Dhondt, S.; Waites, R.; de Folter, S.; Bimbo, A.; van Dijk, A.D.; Muino, J.M.; Cutri, L.; Dornelas, M.C.; et al. Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiol.* **2012**, *159*, 1511–1523. [CrossRef]
58. Ding, S.; Zhang, B.; Qin, F. Arabidopsis RZFP34/CHYR1, a Ubiquitin E3 Ligase, Regulates Stomatal Movement and Drought Tolerance via SnRK2.6-Mediated Phosphorylation. *Plant Cell* **2015**, *27*, 3228–3244. [CrossRef] [PubMed]
59. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]
60. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
61. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]
62. Bradbury, P.J.; Zhang, Z.W.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
63. Mao, H.; Yu, L.; Li, Z.; Liu, H.; Han, R. Molecular evolution and gene expression differences within the HD-Zip transcription factor family of *Zea mays* L. *Genetica* **2016**, *144*, 243–257. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Review

Genetic and Molecular Control of Floral Organ Identity in Cereals

Zulfiqar Ali [1,†,*], Qasim Raza [2,3,†], Rana Muhammad Atif [2,4,†], Usman Aslam [2], Muhammad Ajmal [2] and Gyuhwa Chung [5,*]

1 Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Sharif University of Agriculture, Multan 66000, Pakistan
2 Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan; qasimnazami@gmail.com (Q.R.); dratif@uaf.edu.pk (R.M.A.); wellusman@hotmail.com (U.A.); ajmalrana100@gmail.com (M.A.)
3 Molecular Breeding Laboratory, Division of Plant Breeding and Genetics, Rice Research Institute, Kala Shah Kaku 39020, Pakistan
4 Centre for Advanced Studies in Agriculture and Food Security, University of Agriculture, Faisalabad 38000, Pakistan
5 Department of Biotechnology, Chonnam National University, Chonnam 59626, Korea
* Correspondence: zulfiqar.ali@mnsuam.edu.pk (Z.A.); chung@chonnam.ac.kr (G.C.); Tel.: +92-061-920-1684 (Z.A.); +83-10-36303-1260 (G.C.)
† These authors contributed equally to this work.

Received: 2 May 2019; Accepted: 28 May 2019; Published: 4 June 2019

Abstract: Grasses represent a major family of monocots comprising mostly cereals. When compared to their eudicot counterparts, cereals show a remarkable morphological diversity. Understanding the molecular basis of floral organ identity and inflorescence development is crucial to gain insight into the grain development for yield improvement purposes in cereals, however, the exact genetic mechanism of floral organogenesis remains elusive due to their complex inflorescence architecture. Extensive molecular analyses of Arabidopsis and other plant genera and species have established the ABCDE floral organ identity model. According to this model, hierarchical combinatorial activities of A, B, C, D, and E classes of homeotic genes regulate the identity of different floral organs with partial conservation and partial diversification between eudicots and cereals. Here, we review the developmental role of A, B, C, D, and E gene classes and explore the recent advances in understanding the floral development and subsequent organ specification in major cereals with reference to model plants. Furthermore, we discuss the evolutionary relationships among known floral organ identity genes. This comparative overview of floral developmental genes and associated regulatory factors, within and between species, will provide a thorough understanding of underlying complex genetic and molecular control of flower development and floral organ identity, which can be helpful to devise innovative strategies for grain yield improvement in cereals.

Keywords: ABCDE model; cereals; evolutionary relationships; flower organ identity; floral speciation; MADS-box genes

1. Introduction

Cereals are clearly critical for global food security. They provide approximately 60% of human caloric requirement and this figure can even exceed 80% in resource-poor countries [1]. However, the exponential increase in the world population, soaring food prices and constant depletion of arable land resources due to climate change have made it inevitable to develop cereal crops with increased grain yield [2]. Cereals belong to the grass family Poaceae, which is one of the largest groups of

monocotyledonous plants, with almost 12,000 species [3]. The grass family is monophyletic and diverged from eudicots approximately 125–150 million years ago [4,5]. Grasses show remarkable diversity in overall plant morphology, physiology, genetics, and ecology compared to their eudicot counterparts [6]. For example, spikelets are characteristic structural units of grass inflorescence, which (depending upon species) show determinate or indeterminate growth. Spikelets are composed of one to several florets but unlike eudicot flowers, these florets possess bract-like structures called lemma, palea, and lodicules, instead of sepals and petals [7,8].

The Poaceae family has two important model crop species; rice (*Oryza sativa)* and maize (*Zea mays).* Each has been used to study flower development processes at the molecular level. The genome of rice is exceptionally small compared to other grass species and has been fully sequenced [9,10]. In addition, the rice genome is conducive for effective positional cloning and genetic transformation; making it ideal for developmental biology studies [11–13]. Similarly, the genome of maize has also been fully sequenced [14], is amenable to positional cloning, and the species has simple reproductive biology [7,15]. Both of these species show synteny [9], thus the progress in one species has been facilitating the progress in the other species.

In addition to rice and maize, the Poaceae family also contains *Brachypodium distachyon*, a promising model plant that is anatomically similar to the majority of forage grasses and temperate cereals including wheat (*Triticum aestivum*), the "king of cereals". *B. distachyon* has a short life cycle and is readily cultivatable. The genome of *B. distachyon* has already been sequenced [16] and it offers a highly efficient genetic transformation system. These qualities make *B. distachyon* suitable for functional genomic studies of grass related traits [17–19]. In comparison, wheat is the most important staple crop in temperate zones and a major source of starch, energy and dietary fiber. As an example, bread wheat alone provides 20% of the daily calorie intake in the UK [20]. However, wheat functional genomic studies were limited due to the lack of a quality reference genome sequence and hexaploid nature of the species [21,22]. More recently, a high quality, fully annotated reference genome of hexaploid wheat has been delivered which can accelerate research in wheat developmental biology and genomics assisted breeding [23]. In recent years, significant progress has been made towards understanding the genetic regulation of spike development in Brachypodium, wheat, and barley [24–31]. These studies revealed striking similarities between Brachypodium, wheat, and barley, with highly conserved genetic regulation of inflorescence development in these species. Thus, understanding the molecular control of inflorescence development and floral organ identity in model species will expand our knowledge about the genetic architecture of the spike development in all economically important grasses.

Floral organs control grain development. Previously, a simple yet elegant ABC model of floral organ identity was devised to demonstrate the molecular control of floral development in model plants [32]. This model proposed that combinatorial activities of three homeotic gene classes specify four floral organs i.e., sepal, petal, stamen, and carpel. Class A genes, when expressed alone, produce sepals. The expression of classes A and B together directs petal identity. The expression of classes B and C together regulates stamen identity and the expression of Class C genes alone determines carpel identity. Subsequently, two other floral identity gene classes were identified. Class D genes in Petunia [33] and the redundant class E genes (*SEP1*–4) in Arabidopsis [34,35]. The current model consists of these five classes of floral-homeotic, MADS-box genes (A, B, C, D, and E). The hierarchical combination of these five gene classes thus determines floral organ identity [36].

In higher model plants, especially Arabidopsis and rice, the ABCDE model has helped explain the molecular control of floral organ identity to some extent. This is largely due to their relatively small genome size and the extensive research associated with each of these model species. Analyses of the floral homeotic genes of these species suggest that the same flower organ identity model can be applied to other cereals [37], including Brachypodium, maize, and wheat. This review explores recent advances in rice, maize, Brachypodium, and wheat floral development and subsequent organ specification, with reference to the model plant Arabidopsis. Plethora of studies revealed novel regulatory factors and pathways that contribute to the unique morphology of the grasses. However, the vast array of

functions performed by floral homeotic genes and the large body of literature devoted to this subject makes it difficult to comprehensively review all aspects of the genetic control of floral development. Here, we tried to review the comparisons of floral development genes, within and between species that will expand our understanding of the complex molecular genetic control of floral development and flower organ identity especially in grasses.

2. Inflorescence Morphology and Development

The grass family includes several agriculturally and economically important species including rice, wheat, maize, sorghum, and barley. Developmental and genetic pathways controlling the shape of inflorescence architecture and development in these important crops have been reviewed briefly [27,28,38–40]. All grass inflorescences have a characteristic basal structural unit, the spikelet, composed of one to several florets depending upon the species [6]. These florets are surrounded by bract-like structures known as glumes. Most grass species possess unique inflorescence organization and structure distinct from eudicots and even from other monocots [41]. For example, Arabidopsis bears indeterminate inflorescence with several branched flowers. The grasses like Brachypodium, Hordeum, Secale, and Triticum inflorescences carry sessile spikelets on the rachis. In contrast, Avena, Echinochloa, Oryza, Panicum, Setaria, and Sorghum bear long branched inflorescence where spikelets are pedunculate [42] (Figure 1A). Moreover, the Arabidopsis inflorescence meristem normally differentiates only into branch meristem and floral meristem whereas several specialized axillary meristems are formed in grass spikes [40] (Figure 1B). Unlike their eudicot counterparts, the grass florets possess bract-like structures; lemma, palea, and lodicules in place of sepals and petals [8].

Among cereals, rice exhibits distinct inflorescence morphology compared to that of Brachypodium and wheat [24,38,40,43]. The spikelet is the basal structural unit in these three grass species. The rice inflorescence is relatively complex and comprised of long stalked panicles in which primary branches are directly attached to the main axis (rachis) that further produce secondary branches, lateral spikelets and terminal spikelets [38]. By contrast, there exists only one rachis in Brachypodium and wheat that directly bears the spikelets in an alternating configuration [24,28]. Spikelets in these species also bear rudimentary glumes and floret primordia. In rice, the single spikelet can produce only a single floret [44], whereas the wheat spikelet contains several florets and normally four or five of these reach anthesis [45]. Unlike wheat and rice, the inflorescence of Brachypodium carries only two or three lateral spikelets and a single terminal spikelet [24]. Each spikelet contains ~11 florets, arranged in a distichous phyllotaxy around central axis. Overall, the organization and structure of floral organs are conserved among rice, Brachypodium, and wheat, with the exception of three additional stamens within a floret in rice [24,28,38] (Figure 1C). The grass floret contains lemma, palea, lodicules, stamens, and pistils. The pistil is comprised of three fused carpels which surround a single ovule. The apical region of the pistil bifurcates with feathery stigmas. Morphological analysis suggested that lodicules are homologues of petals [46], which together with the lemma and palea are unique to grasses.

Inflorescence development is regulated by several types of meristems [44,47] and starts with the transition of the shoot apical meristem (SAM) into an inflorescence meristem (IM). In Brachypodium and wheat, the IM directly generates the spikelet meristem (SM) [24,28,43], while in rice the IM generates the primary branch meristem (pBM) followed by the secondary branch meristem (sBM) which then finally configure the spikelet meristems (SMs) [40] (Figure 1B). The SMs generate floral meristems (FMs), which subsequently determine floral organ identity. All grasses show indeterminate growth starting from SAM to just before SM determinacy. However, the SM to FM transition is determinate and critical [48] as it is the final phase at which the meristem activity stops. By this stage, stem cells are believed to exhaust all their energy due to continuous formation of floral organs and floret primordia [47].

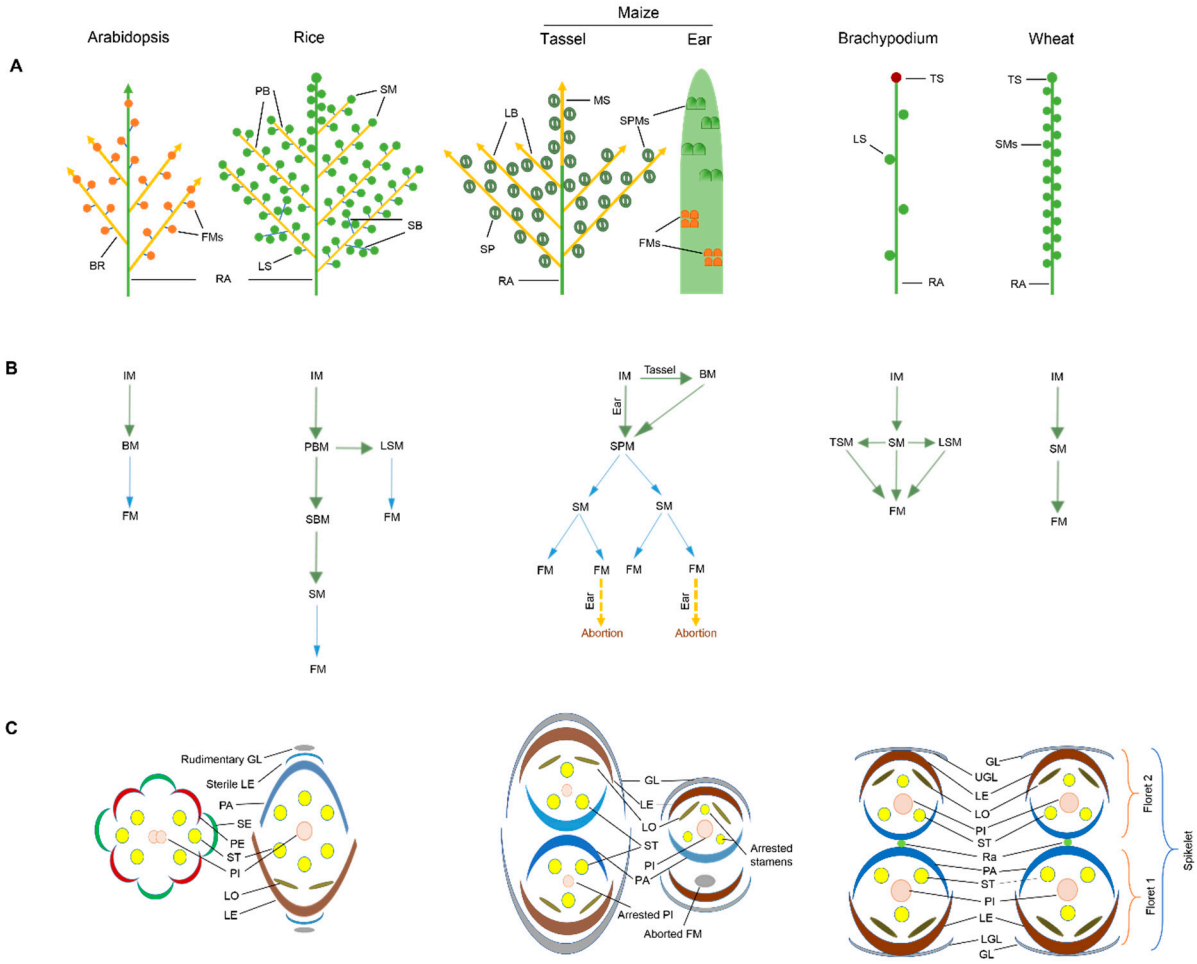


Figure 1. Graphical representation of inflorescences, phase transition and transverse flowers. (**A**) Structural configuration of inflorescence in Arabidopsis, rice, maize, Brachypodium and wheat. Color codes. Green line: rachis; yellow line: primary branch; blue line: secondary branch; green circles: spikelet/spikelet pair meristems; maroon circle/oval: terminal spikelet; orange circle: floral meristems. (**B**) Regulation of meristem transition in Arabidopsis, rice, maize, Brachypodium, and wheat. Green arrow: multiple meristems formation; blue arrow: single meristem formation; orange dashed arrow: abortion of floral meristems. (**C**) Schematic representation of transverse spikelet/flower. Color codes: Blue: palea; dark orange: lemma; gold: lodicules; green: sepal; green circle: rachis; gray: glume; pink: pistil; red: petal; yellow: stamen. Abbreviations: BR: branch; BM: branch meristem; FM: floral meristem; GL: glume; IM: inflorescence meristem; LE: lemma; LO: lodicule; LS: lateral spikelet; LSM: lateral spikelet meristem; PA: palea; PB: primary branch; PE: petal; PI: pistil; PBM: primary branch meristem; RA: rachis; Ra: rachilla; SB: secondary branch; SE: sepal; SBM: secondary branch meristem; SM: spikelet meristem: SPM: spikelet pair meristem; ST: stamen; TS: terminal spikelet.

In contrast to Brachypodium, wheat, and rice, maize is a monoecious crop in which male and female organs occur separately on the same plant. The male inflorescence at the shoot apex is known as tassel that bears paired spikelets while the female inflorescence occurs in the leaf axil which is termed as ear [40]. Male IM produce long indeterminate branches which further differentiate into short secondary branches that bear spikelet pair meristems (SPMs). Each SPM initiate two SMs, which in turn produce two FMs each (Figure 1B). The female inflorescence (ear) is produced on the main stalk, hence SPMs are directly attached to the main stem. SPMs are transient and bear a pair of SMs. SMs are also transient which in turn produce two FMs. Each spikelet bears two staminate flowers called florets and only one of these florets produces a fertile flower. Flowers further develop into different floral organs such as lemma, palea, lodicules, stamens, and carpels. Apart from shapes and

position of male and female inflorescences, the arrest of stamen formation in ear florets and of pistil formation in tassel florets makes it easy to distinguish male and female inflorescences [49]. Over the last two decades, several genetic factors involved in flower development have been identified which mainly function as trans-regulatory elements. Here, we will discuss the latest knowledge about the association of MADS-box- and non-MADS-box-related gene families with inflorescence development and floral organ identity in grasses.

3. Role of MADS-Box Transcription Factors in Floral Organ Identity

MADS-box transcription factors are involved in various biological processes and have been identified in almost all groups of eukaryotes. The name MADS was derived from combining the names of *MINICHROMOSOME MAINTENANCE 1* of *Saccharomyces cerevisiae*, *AGAMOUS* of *Arabidopsis thaliana*, *DEFICIENS* of *Antirrhinum majus*, and *SERUM RESPONSE FACTOR* of *Homo sapiens* [50]. All MADS-box TFs have a highly conserved ~60 amino acids long DNA binding MADS domain at the N-terminal region which binds to CArG boxes on DNA [51]. Flowering plant genomes contain approximately 100 MADS-box genes, which are further categorized into M-type and MIKC-type MADS genes [52]. Only a few M-type MADS are functionally characterized so far [53], however, plant MIKC-type MADS-box genes have been extensively studied [54]. In plants, the diversification of MADS-box genes is closely linked to the evolution of important organs, such as seeds, flowers, and fruits [55]. Moreover, morphological variations in inflorescence of grass family are closely associated with changes in copy number, expression patterns, and interactions between MIKC-type MADS-box genes [56]. In flowering plants, combinatorial activities of the five classes of MIKC-type MADS-domain genes define floral organ identity. According to the Arabidopsis "floral quartet model", sepals are specified by class A and E genes in the first whorl; petals by class A, B, and E genes in the second whorl; stamens by class B, C, and E genes in the third whorl; and carpels by class C, and E genes in the fourth whorl [54]. The ovule identity gene *FLORAL BINDING PROTEIN 11* (*FBP11*) was first identified and functionally characterized in Petunia and classified as D class gene [33]. In Arabidopsis, ovule identity is controlled by *AGAMOUS* subfamily member *SEEDSTICK* (*STK*) [57]. Functional divergence, duplication, and evolutionary relationships among these five classes of homeotic genes, identified in Arabidopsis and major cereals, are summarized in Table 1. Modified ABCDE models showing the complex genetic interaction of MADS-box TFs' and other important regulators in Arabidopsis and cereals are illustrated in Figure 2.

Table 1. Genes controlling floral organ identity in cereals.

Arabidopsis Gene	Subfamily	Class	Putative Role	Rice	Maize	Wheat	Brachypodium	References
AP1	*AP1*	A	Promote floral meristems, sepals, petals or lemma/palea identities	*OsMADS14* *OsMADS15* *OsMADS18* *OsMADS20*	*ZMM4, 15* *ZAP1* - -	*WFUL1, TaAGL25* *WFUL2* *WFUL3, TaAGL10* -	*BrMADS33* *BrMADS10* *BrMADS3* *BrMADS31*	[31,58–66]
AP2	*AP2*	A	Spikelet/floral meristem identity and lodicule identity	*IDS1* *SNB* *MFS1* *FZP*	*ids1* *sid1* - *BD1*	*TaAP2* *TaQ* - *WFZP*	- - - *MOS1*	[24,25,67–75]
AP3	*AP3*	B	Lodicule and stamen identity	*OsMADS16/SPW1*	*Silky1*	*WAP3*	*BrMADS5*	[31,76–80]
PI	*PI*	B		*OsMADS2*	*ZMM16*	*WPI2*	*BrMADS20*	[31,77,81–86]
				OsMADS4	*ZMM18, 29*	*WPI1*	*BrMADS16*	
ABS/TT16, GOA	B_{SISTER}	-	Integuments and seed development	*OsMADS29, 30*	*ZMM17*	WB_{sis}, *TaAGL35*	*BrMADS17, 23, 38*	[31,84,87–91]
AG, SHP1, SHP2	*AG*	C	Stamen and carpel identity	*OsMADS3*	*ZMM2*	*WAG-2*	*BrMADS14*	[31,57,92–98]
				OsMADS58	*ZAG1*	*WAG-1*	*BrMADS18*	
STK	*AG*	D	Ovule identity	*OsMADS13*	*ZMM1, ZAG2*	*WSTK, TaAGL9, 31*	*BrMADS2*	[30,31,57,62,90,99–103]
				OsMADS21	-	-	*BrMADS4*	
SEP (1-4)	*SEP*	E	FM determinacy and floral organ identity	*OsMADS7/45*	*ZMM6*	*WSEP, TaAGL16, 28,30*	*BrMADS26*	[31,34,104–108]
				OsMADS8/24	*ZMM7/27*	*TaMADS1*	*BrMADS32*	
-	*LOFSEP*	E		*OsMADS1/LHS1*	*ZMM8/14*	*WLHS1, TaAGL24*	*BrMADS11*	[31,104,107,109–112]
				OsMADS5	-	*TaAGL3, 5, 8, 34, 40*	*BrMADS7*	
				OsMADS34/PAP2	*ZMM24, 31*	*TaAGL27*	*BrMADS1*	
-	*AGL6*	E		*OsMADS6/MFO1*	*ZAG3, 5*	*TaAGL6, 37*	*BrMADS28*	[31,82,105,113–117]
				OsMADS17	-	-	-	
CRC	*YABBY*-like	-	Carpel identity	*DL*	*Drl1, 2*	*TaDL*	-	[77,80,118–120]
UFO	F-box	-	Positive regulation of B & C class MADS	*APO1*	-	-	-	[121–123]
-	DoF	-	Lemma and Palea identity?	*CFO1, G1/ELE, ld-1*	-	-	-	[124–126]

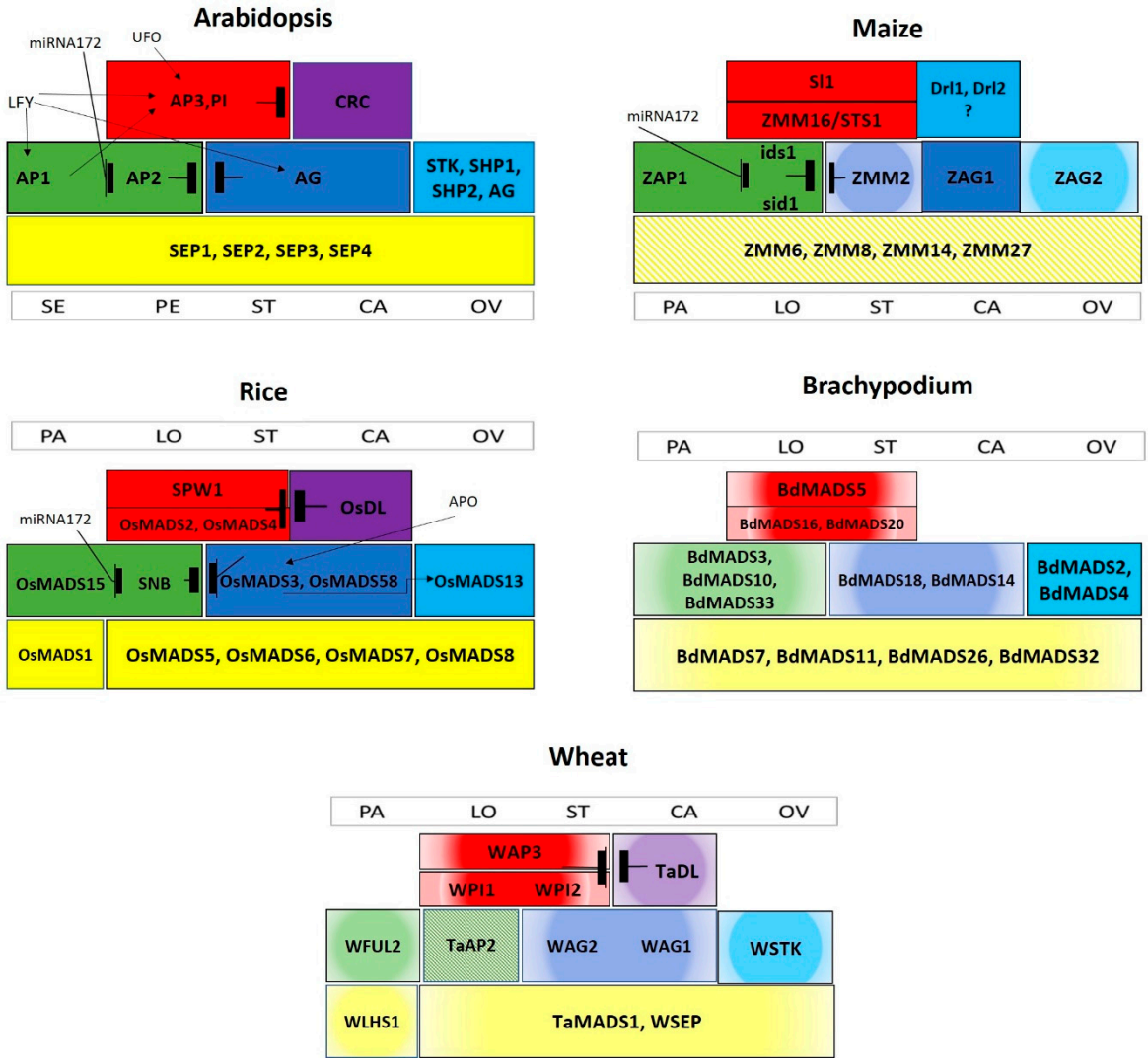


Figure 2. ABCDE models of floral organ identity. Revised floral organ identity models in Arabidopsis, rice, maize, Brachypodium, and wheat. Class (A)-genes indicated in green, class B in red, class C in dark blue, class D in light blue, class E yellow, and non-MADS in purple. Solid colors show functional data, color gradients represent expression analysis data, while color patterns indicate hypothesized functions. Antagonistic interactions are indicated with barred lines, black arrows illustrate positive regulation of the corresponding genes, and a comma symbolizes duplicated gene interaction. Abbreviations: CA: Carpel; LO: Lodicule; OV: Ovule; PA: Palea; PE: Petal; SE: Sepal; ST: Stamen. For gene abbreviations see text.

3.1. Class A Homeotic Genes

In Arabidopsis, class A genes include *APETALA 1* and *2* (*AP1* and *AP2*), of which only *AP1* encodes a MADS-box TF. *AP1* is expressed only in sepals and petals (two outer whorls) and has an additional role in floral meristem determinacy [61]. Similar expression and functional patterns have been reported in Antirrhinum class-(A) ortholog [127]. In eudicots, class (A)-functions are defined by *AP1*, *CAULIFLOWER (CAL)*, and *FRUITFULL (FUL)* genes, whereas in monocots only *FUL*-like genes are present which are associated with class (A)-function [128]. Recently, Wu et al. [65] demonstrated that grass-specific *FUL*-like genes are required to specify palea and lodicule identities in addition to their function of specifying meristem identity. Similar results were reported for rice and wheat, wherein *AP1* clade genes together with class E *SEPALLATA* (*SEP*) genes were shown to participate in the transition from SAM to IM [112,129].

The *AP1* homologs identified in rice include *OsMADS14*, *OsMADS15*, *OsMADS18*, and *OsMADS20*, all of which belong to the *FRUITFULL (FUL)* lineage [59]. Ectopic expression of *OsMADS14* in rice suggests its involvement in floral meristem control to promote flowering. On the other hand, loss-of-function loss-of-function mutations in *Osmads15* indicate the role of *AP1* in palea formation with no effect on lodicule development [130]. A more recent study in rice employing both single and double mutants of *OsMADS14* and *OsMADS15* [65] provided strong evidence that rice *AP1/FUL*-like genes are essential for specifying lemma/palea and lodicule identities during the floral development process. Because lemma and palea are considered homologous structures to sepals and petals of eudicots, respectively, therefore, it is possible that *AP1/FUL*-like genes are independently recruited to fulfil the function of class A genes in grass species. Maize orthologs of *AP1* include *Zea mays APETALA1* (*ZAP1*), *Zea mays MADS4*, and *15* (*ZMM4*, *ZMM15*) [62]. Phylogenetic analysis showed that *ZAP1* is an ortholog of *OsMADS15* [131]. Northern blot analysis demonstrated that *ZAP1* was expressed in the lemma/palea and lodicules, but not in stamens and pistils [62]. These results suggest that *ZAP1* is a putative class-(A) gene with a possible repressive interaction with class C genes. *ZMM4* and *ZMM15* are orthologs of *OsMADS14* and *ZMM4* and have been reported to be involved in inflorescence development and floral induction [58], which is consistent with the function of *AP1* homologs from Arabidopsis and rice.

The Brachypodium genome contains at least four (A)-class genes, *BdMADS3*, *10*, *31*, and *33*, which are orthologs of *OsMADS18*, *OsMADS15*, *OsMADS20*, and *OsMADS14*, respectively. *BdMADS3*, *10*, and *33* were observed to be strongly expressed in the lemma and palea, but not in lodicules and stamens with the exception of *BdMADS3* that also strongly expressed in stamens [31]. *BdMADS31* was absent in all floral organs but was weakly expressed in leaves similar to the expression pattern of Arabidopsis and rice orthologs [132,133]. These expression pattern studies suggest the involvement of *BdMADS3*, *10*, and *33* in (A)-class performance; however, further functional analyses are required to confirm their regulatory roles in floral organ identity.

Wheat has five *FUL*-like paralogs including *WFUL1/VERNALIZATION1* (*VRN1*), *WFUL2*, *WFUL3*, *TaAGL10*, and *TaAGL25* [117,128]. Phylogenetic analysis showed that these are the orthologs of *OsMADS14*, *OsMADS15*, and *OsMADS18* [60], an observation consistent with the current phylogenetic tree (Figure 3). Previously it was thought that *WFUL1* had no (A)-class function and was only involved in the transition from the vegetative to reproductive phase [63,134], but recent studies suggest that *VRN1/WFUL1* is expressed in leaves and the shoot apex, where it is required for the long-day flowering response and inflorescence meristem identity [64,134,135]. *ODDSOC2* is a MADS-box TF and downstream target of *VRN1* that functions to repress flowering and has been observed to be downregulated in plants with active *VRN1* alleles and vernalization [136]. Another study reported that *WFUL1* and *WFUL3* are expressed in all floral organs with limited or no expression of *WFUL2* in stamens and pistils [60], suggesting that *WFUL2* has diversified functions in outer (palea and lodicule) and inner (stamen and pistil) floral whorls. Yeast two-hybrid and yeast three-hybrid analyses demonstrated that *WFUL2* interacts with the B and E classes of MADS-box genes [60]. These findings in combination with the expression pattern analysis illustrate that *WFUL2* has a major role in lemma/palea and lodicule identities in wheat florets. It is noteworthy that *WFUL1/VRN1* has a more important role in leaf development indicating functional diversification between wheat *FUL*-like genes. Similarly, functional diversification between rice *FUL1* (*OsMADS14*) and *FUL2* (*OsMADS15*) has been observed. Single mutant of *OsMADS14* showed lower seed setting, but no floret-specific mutant phenotype could be observed when grown under natural field conditions. However, under greenhouse conditions the mutant plants had small paleae and showed the homeotic transformation from lodicules to stamen-like organs. Whereas paddy field-grown *osmads15* plants showed 45% smaller paleae, without affecting the organ identity. However, greenhouse-grown *osmads15* plants had elongated empty glumes and 100% reduced paleae. Additionally, *osmads15* plants showed no homeotic transformation of inner three floral organs under both growing conditions [65].

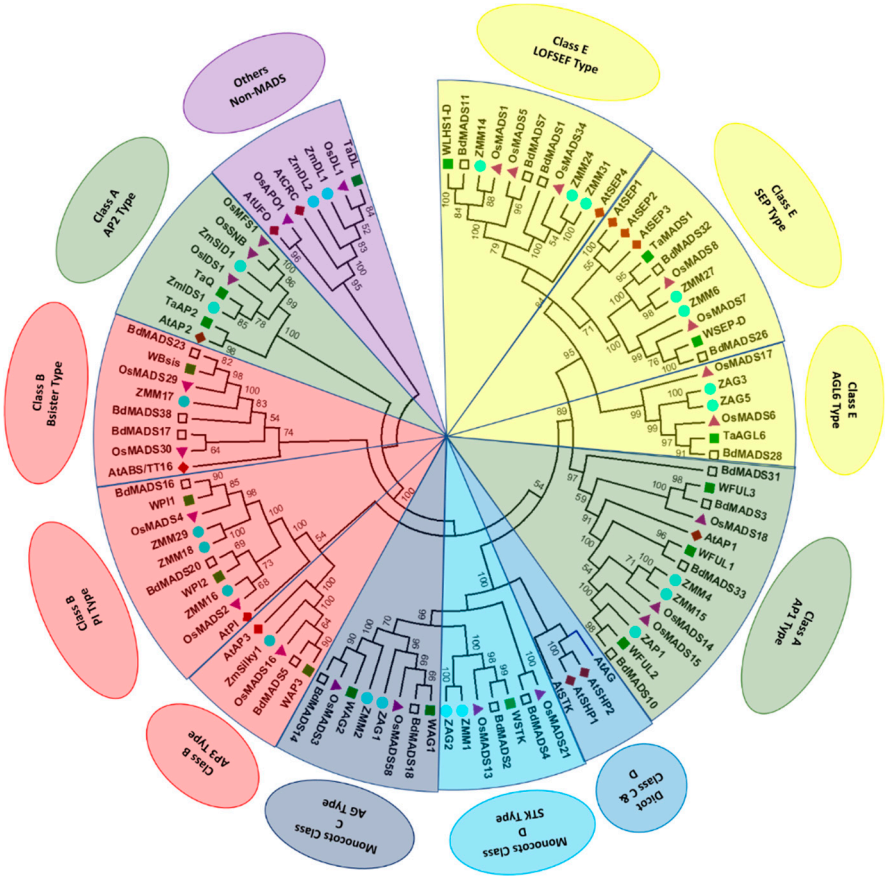


Figure 3. Evolutionary relationships among MADS-box genes. Phylogenetic tree constructed from the deduced amino acid sequences of Arabidopsis, Brachypodium, maize, rice, and wheat genes obtained from NCBI database. (Sequence ID information can be seen in supplementary file) The tree was inferred after amino acid sequence alignment by Clustal Omega [137], using the neighbor-joining method [138] and visualized in topology-only mode. Only bootstrap values >50%, as calculated from 100 replicates, are shown. Phylogenetic analysis was conducted in MEGA version 6 [139]. Markers: Diamond: Arabidopsis; triangle: rice; circle: maize; green filled square: wheat; hollow square: Brachypodium genes.

All angiosperms contain *AP2* TFs, which, in addition to their role in the regulation of floral development, are implicated in primary and secondary metabolism, growth and development, and response to stress [140]. In Arabidopsis, *AP2* is required for the establishment of floral meristems, floral organ identity, and regulation of floral homeotic gene expression [70]. In rice, two *AP2*-like genes—*INDETERMINATE SPIKELET1* (*IDS1*) and *SUPERNUMERARY BRACT* (*SNB*)—synergistically control lodicule development [72]. Another *AP2*-like gene, named *FRIZZLE PANICLE* (*FZP*), prevents the formation of axillary meristem in rice but controls the spikelet meristem identity [71]. *FZP* has also been observed to regulate the transition from panicle branching to spikelet formation in rice by repressing *RICE FLORICAULA LEAFY* (*RFL*)/*ABERRANT PANICLE ORGANIZATION2* (*APO2*). In addition, *FZP* overexpression positively regulate B and E class MADS-box genes in floral meristem suggesting its role in floral organ identity [67]. *MULTI-FLORET SPIKELET1* (*MFS1*) is another *AP2*-type gene that positively regulates rice *IDS1* and *SNB* genes [74]. Rice *IDS1* and *SNB* regulate the transition from spikelet meristem to floral meristem [141]. Both of these genes display strong functional resemblance to maize indeterminate spikelet1 (*ids1*) and sister of indeterminate spikelet1 (*sid1*), respectively, which are also required to initiate floral meristems and to control spikelet meristem determinacy [68]. Similar to the function of *AP2* in Arabidopsis, *ids1* and *sid1* negatively regulate class C gene function within the lateral organs of the spikelet. Likewise, maize *BRANCHED SILKLESS1*

(*BD1*) encodes an ethylene responsive factor (*ERF/AP2*) that regulates the spikelet meristem identity and mutation in *BD1* produces indeterminate floral branching [69]. Like rice *FZP* and maize *BD1*, Brachypodium *MORE SPIKELETS1* (*MOS1*) determines spikelet meristem identity as the *mos1* mutant showed increased number of axillary meristems compared with the wild type [24]. In wheat, Wheat *FZP (WFZP)* controls spikelet meristem identity that drives the formation of supernumerary spikelets by repressing floral meristem formation and differentiation [25]. The regulation of spikelet meristem identity by *AP2*-like genes in rice, maize, Brachypodium, and wheat indicates that their function is conserved among distantly related grass species including agriculturally important crops. In addition, wheat genome also contain *TaQ* and *TaAP2*. The wheat domestication gene (*TaQ*) has a role in inflorescence shape, glume shape, glume tenacity, and spike length [27,75]. Phylogenetic analysis and transcriptional pattern of wheat *TaAP2* revealed its orthologous relationship with barley *HvAP2/Cly1*, which is involved in lodicule identity [73,142]. This observation demonstrates that like rice *AP2*-like orthologs, wheat *TaAP2* might also associated with lodicule identity [72,74], suggesting their functional similarities in grasses.

In recent years, evolutionary conserved micro-RNAs (miRNAs) have been identified and played a crucial role in plant organogenesis. miR172 appears with the evolution in angiosperms and has been identified in Arabidopsis, rice, maize, barley, and wheat. The level of miR172 increases with plant age and its expression is under photoperiodic control [143]. It is an active repressor of all *AP2*-like TFs, which are thought to participate in floral patterning. *AP2* has been demonstrated to bind and repress the expression of miR172b [144]. Early studies reported *AP2* transcripts in all floral organs [70], however recent observations show that *AP2* expression is restricted to sepals and petals compared to that of miR172 that predominantly expressed in inner floral whorls (stamen, carpel, and ovule) [145]. These findings suggest an antagonistic interaction of *AP2* and miR172 in plant developmental transitions.

In cereals, functionally characterized targets of miR172 include *Zea mays* indeterminate spikelet1 (*ids1*) and sister of indeterminate spikelet1 (*sid1*) [68], *Oryza sativa SUPERNUMERARY BRACT (OsSNB)* [146], and *Hordeum vulgare Cleistogamy1 (Cly1)* [142]. Wheat domestication gene *TaQ* is also a target of miR172, however it is not clear if miR172 mediated regulation has a role in domestication [147]. These investigations provide new insights into the ancient role of miRNAs about floral organ regulation in cereals.

3.2. Class B Homeotic Genes

Arabidopsis class B homeotic genes include *AP3* and *PISTILLATA* (*PI*) that are required for petal and stamen identities. Single mutants of these genes caused conversion of petals to sepals in the second floral whorl and stamens to carpels in the third floral whorl [78,81]. Rice has two orthologs of *PI*: *OsMADS2* and *OsMADS4* [148]. RNAi suppression of *OsMADS2* showed homeotic changes in lodicules with no effect on stamens [83], whereas RNAi suppression of *OsMADS4* showed no alteration in these floral organs [86]. Interestingly, simultaneous mutations in both genes caused the conversion of lodicules and stamens into palea and carpel-like structures respectively. These observations suggest an equal role for both genes in stamen development, with *OsMADS2* more important in lodicule identity. Similarly, maize contains three orthologs of Arabidopsis *PI*; *Zea mays MADS16*, *18*, and *29* (*ZMM16*, *ZMM18*, and *ZMM29*). Mutation in *ZMM16* produced a *Sterile Tassel Silky Ear1* (*STS1*) phenotype in which lodicules transformed into palea-like and stamens into carpel-like structures [85]. Phylogenetic analysis showed that *ZMM16* as an ortholog of *OsMADS2* while *ZMM18* and *ZMM29* are orthologous to *OsMADS4*. Recently, a study reported that *ZMM16/STS1* (together with its paralogs *ZMM18* and *ZMM29*) forms obligate heterodimers with maize *SILKY1 (Sl1)* and specifies organ identity in second and third floral whorls [149]. Interestingly, RNAi knockdowns of *ZMM18* and *ZMM29* showed no detectable floret phenotype, indicating that *STS1* can compensate *ZMM18/29* reduction, but *ZMM18*/29 cannot compensate for *STS1* reduction. With this evidence, it is possible to speculate a role for maize *AP3/PI*-like genes.

The sole ortholog of *AP3* in rice; *OsMADS16/SUPERWOMAN1* (*SPW1)* has been observed to interact with rice *PI*-like genes. *OsMADS16* knockdown mutant showed homeotic conversion of lodicules and stamens to palea and carpel-like structures, respectively, similar to *PI*-types [80]. Similarly, the loss-of-function mutant of maize *SILKY1* (*AP3* ortholog) showed alterations in lodicules and stamens [76]. As lodicules represent second whorl (petals), their transformation into palea-like structures support the hypothesis that petals of eudicots are likely to be modified into lodicules in grasses. Furthermore, Arabidopsis and maize class B genes showed similar biochemical activities in vivo and in vitro [85]. Collectively, these findings suggest that the function of class B genes is somewhat conserved between grasses and eudicots.

The Brachypodium genome contains three B class genes: *BdMADS5*, *16*, and *20*. *BdMADS5* is an ortholog of Arabidopsis *AP3* and rice *OsMADS16*. Similarly, *BdMADS16* and *BdMADS20* are orthologs of *OsMADS4* and *OsMADS2*, respectively, and are clustered with Arabidopsis *PI* [31]. Strong expression of Brachypodium B class genes was detected in lodicules and stamens, with *BdMADS16* expressed in carpels as well. However, transcript abundance of all B class genes was very low in the lemma and palea in Brachypodium similar to those of Arabidopsis and rice B class genes [132,133]. Although their expression patterns suggest that *BdMADS5*, *16*, and *20* have conserved roles in lodicule and stamen identity, functional analyses of these genes remain to be conducted to confirm these hypotheses.

WAP3, also called *TaAP3* is a wheat ortholog of Arabidopsis *AP3*, which is encoded by two highly homeologous genes: *TaMADS#51* and *TaMADS#82* [82,131]. Northern blot analysis revealed that *WAP3* expression was restricted to young spikes during floral development and possibly associated with the induction of pistillody (homeotic conversion of stamens into carpel-like structures) [79]. *WAP3* is also involved in the homeotic transformation of lodicules and stamens into palea and pistil-like structures, respectively [77]. Wheat genome also contains two *PI*-like genes: *WPI1* and *WPI2*. Phylogenetic analysis revealed close orthologous relationships of *WPI1* with *OsMADS4*, and that of *WPI2* with *OsMADS2*. Similar to *WAP3*, wheat *PI*-type genes were reported to be involved in lodicule and stamen development and their homeotic transformation into palea and pistil-like structures. Hama et al. [77] reported that *WAP3* and *WPI* were highly expressed in the primordia of lodicules and stamens. Low expression patterns of wheat B class genes were detected in pistil-like stamens of an alloplasmic wheat line having the *Aegilops crassa* Boiss. cytoplasm and lacking the *Rfd1* gene, indicating that these genes gradually disappear from the fourth whorl (carpel/pistil) just like Arabidopsis *PI* [81]. These observations strongly suggest that wheat class B genes are associated with the induction of pistillody, a direct consequence of changes in copy number and expression of *WAP3* and *WPI's* in third and fourth whorls confirming that *WAP3* and *WPI's* exhibit class B functions.

BSISTER genes, closely related to class B MADS-box genes, have been identified through phylogenetic studies. Members of this subfamily regulate female reproductive organs and seed development [150]. All *BSISTER* MADS-box genes investigated to date are expressed during early ovule development indicating that these genes may be required for ovule identity. Arabidopsis has two *BSISTER* genes—*ARABIDOPSIS BSISTER (ABS)/TRANSPARENT TESTA16* (*TT16*) and *GORDITA* (*GOA*)—both expressed in mature ovules [88,89]. Yang et al. [91] has functionally characterized the rice *BSISTER* MADS-box gene; *OsMADS29*. His findings demonstrate that *OsMADS29* expressed only in floral but not vegetative organs. Another study involving RT-PCR revealed that *OsMADS29* expressed in ovules, consistent with previously reported patterns for wheat *BSISTER* (WB_{sis}) and maize *ZMM17* [87,90]. However, knock-down of *OsMADS29* by double-stranded RNA-mediated interference (RNAi) resulted in shriveled and/or aborted seeds [91], suggesting that *OsMADS29* also has important functions in seed development of rice by regulating cell degeneration of maternal tissues. Furthermore, Arabidopsis and rice *BSISTER* and D-class genes show overlapping expression patterns [151]. More recently, Schilling et al. [84] investigated another *BSISTER* gene (*OsMADS30*) in rice. This gene was weakly expressed in ovules. Further, the plants carrying a T-DNA insertion in *OsMADS30* showed no aberrant phenotype, indicating that this gene is either not required for ovule specification or its function is obscured by another class D gene (*OsMADS21*). Brachypodium also

contains three *BSISTER* genes: *BdMADS17*, *23*, and *38*. Weak expression of *BdMADS17* and *BdMADS23* was detected in palea but absent in ovules. However, *BdMADS38* was weakly expressed in stamens only [31]. Altogether, these results suggest that *BSISTER* genes do not possess a strict function, instead of play overlapping roles in whole reproductive ontogeny.

3.3. Class C and D homeotic genes

It is believed that during the divergence of angiosperm and gymnosperm lineages, an ancient duplication resulted in the class C origin, including all stamen and carpel identity genes and class D or ovule specification genes [98,99,152]. This type of classification is reported in several phylogenetic studies [30,31,131,153], and has therefore been adopted in this review. Arabidopsis has three class C homeotic genes; *AGAMOUS* (*AG*) and *SHATTERPROOF1* and *2* (*SHP1* and *SHP2*). Arabidopsis typical class C gene *AG* specifies stamen (third whorl) and carpel (fourth whorl) identities and has an additional role in floral meristem determinacy [93]. In the absence of *AG* activity, class (A)-gene function expands to the 3rd and 4th whorls [32,154], which suggests antagonistic interaction between these two classes of homeotic genes. The additional C class genes of Arabidopsis, *SHP1* and *SHP2*, are required for carpel and fruit dehiscence zone specifications [57,155]. Like grass class B genes, the function of class C genes are also diversified in grasses due to events of duplication and subfunctionalization of these genes during evolution. Rice has two duplicated class C genes; *OsMADS3* and *OsMADS58*. Yamaguchi et al. [98] investigated single mutants of rice class C genes and reported interaction with the class D gene *OsMADS13*, regulating floral meristem determinacy with redundant mediation of class C gene functions. Mutant and transgenic analyses showed that *OsMADS58* regulates floral determinacy with minor effects on carpel identity, while *OsMADS3* predominately regulates stamen identity and prevents lodicule development with minor effects on floral determinacy. As floral determinacy is defined by class (A)-genes, *OsMADS58* probably has reduced *AG* activity in the third and fourth whorls compared to *OsMADS3*. Furthermore, the rice class B gene *OsMADS16* interacts with class C genes to suppress indeterminate growth within floral meristems [156]. These findings indicate that class C genes play a dominant role in stamen and carpel identity, with a minor role in floral meristem determinacy and possible antagonistic interaction between A and C class genes. A study conducted by Dreni et al. [92] demonstrated redundant mediation of the class C associated functions by *OsMADS3* and *OsMADS58*. He also observed strong defects in stamens and carpels of *osmads3* flowers, whereas most of the *osmads58* flowers were indistinguishable from wild type flowers. The contribution of *OsMADS3* in specifying C-function seems to be more important when compared with *OsMADS58*, consistent with the reports of Yamaguchi et al. [98]. The double mutants of *osmads3* and *osmads58* were corresponding to the *ag* mutant of Arabidopsis with some differences between their phenotypes. The *osmads3* and *osmads58* mutants showed homeotic conversion of stamens and carpels into lodicule and palea-like structures, respectively. Dreni et al. [92] also reported FM determinacy by *AG* subfamily genes. Out of four *AG* subfamily genes in rice, three (*OsMADS3*, *OsMADS13*, and *OsMADS58*) redundantly regulated the FM determinacy. All the three possible double mutant combinations (*osmads3* and *osmads58*, *osmads3* and *osmads13*, and *osmads13* and *osmads58*) resulted in an enhanced FM indeterminacy.

Maize has three class C genes: *Zea mays AGAMOUS1* (*ZAG1*), *ZMMS2*, and *ZMM23* [152,153]. Like rice class C genes, *ZAG1* and *ZMM2* both have functional diversification as these are orthologs of *OsMADS58* and *OsMADS3*, respectively. Expression analysis detected *ZAG1* transcript abundance in early stamen and carpel primordia with partial floral meristem determinacy [97]. However, a later study with *ZAG1* mutants demonstrated a loss of floral meristem determinacy with little change in stamen and carpel identity [96]. *ZMM2* transcripts were expressed in stamens and carpels, while stronger expression patterns were detected in stamens only, suggesting an involement in stamen and carpel development. Although, *ZMM2* mutants have not been identified, these observations indicate overlapping but nonidentical activities for both maize C class genes.

Like rice, Brachypodium also has two C class genes—*BdMADS14* and *18*—that show high sequence similarity with *OsMADS3* and *OsMADS58*, respectively [31]. Strong expression of *BdMADS18*

was detected in stamens and carpels, whereas *BdMADS14* was weakly expressed in stamens only. In contrast to their rice homologs, where *OsMADS3* and *OsMADS58* have important roles in floral organ identity [98]; the gene *BdMADS18* appears to have a more dominant role in stamen and carpel identity. Similarly, wheat also has two orthologs of *AG*; wheat *AGAMOUS-1* and *2* (*WAG-1* and *WAG-2*) [94]. However, unlike rice and maize orthologs, these have possible roles in ectopic ovule formation and the conversion of stamens into pistil-like structures. Meguro et al. [95] reported that *WAG* transcription levels were low in young spikes but increased during later stages of spike development and were highest between the booting and spike emergence stages. *WAG* was expressed in both reproductive and non-reproductive parts of the flower with an extra transcript of *WAG* detected in the pistillody line. These observations suggest that *WAG* is associated with pistillody induction. Loss-of-function analysis of *WAG* genes would further elucidate their role in stamen and carpel identity. Other names for *WAG-1* and *WAG-2* are *TaAG1* and *TaAG2*/*TaAGL39*, respectively [117,131]. Phylogenetic analysis showed that rice class C genes, *OsMADS58* and *OsMADS3*, are orthologous to *WAG-1* and *WAG-2*, respectively (Figure 3). In conclusion, both class B and C genes in wheat appear to have a role in the induction of pistillody [77,79,95].

Previous studies demonstrated that class D is a more specialized version of class C and define ovule identity [37,153]. Class D genes were first identified in Petunia as *FLORAL BINDINGPROTEIN 7* and *11* (*FBP7, FBP11*). Their cosuppression transforms ovules into carpelloid structures [157]. Overexpression of *FBP11* results in ectopic ovules on sepals and petals [33] indicating its function in ovule identity. In Arabidopsis, class D gene functions are specified by *SEEDSTICK* (*STK*). Biochemically, STK protein interacts with class C (AG, SHP1 and SHP2) and class E proteins to define ovule identity [101]. Triple mutants of *STK*, *SHP1*, and *SHP2* transform ovules into carpelloid structures [57] confirming that Class D genes specify ovule identity in Arabidopsis. Phylogenetic analysis clustered *STK* and *SHP*s into single clade (Figure 3). The functional divergence between *STK* and *SHP* paralogous genes may arise due to diversification in their DNA binding site motifs or through alterations in their tissue-specific expression levels [158]. Rice contains two orthologs of *STK*: *OsMADS13* and *OsMADS*21 [99]. Expression analysis, loss-of-function and protein–protein interaction studies suggest that *OsMADS13* is involved in ovule identity [99,100,102,103]. Moreover, *OsMADS13* acts synergistically with *OsMADS3* (a class C gene) to regulate ovule development and floral meristem termination [159]. Loss-of-function in *OsMADS*21 showed no ovule defects suggesting a loss of ovule specification by this gene [92,99].

Maize has three class D genes: *ZMM1*, *ZAG2*, and *ZMM*25 [62]. Phylogenetic analysis showed *ZMM1* and *ZAG2* to be closely related to rice *OsMADS13*, whereas *ZMM*25 had a close relationship with rice *OsMADS21* [131]. Similar to Arabidopsis *STK*, the expression of *ZAG2* was primarily identified in carpels and ovules [97,160] indicating a possible role in ovule specification.

Like rice, Brachypodium also has two D class genes—*BdMADS2* and *4*—orthologous to *OsMADS13* and *OsMADS21*, respectively. Quantitative RT-PCR revealed comparable expressions of both genes in all floral organs, with the exception of carpels and ovules, where the expression of *BdMADS2* was more than 5 times to that of *BdMADS4* [30]. Ectopic expression of both genes in Arabidopsis demonstrated that overexpression of *BdMADS4* produced more significant phenotypic changes than transgenic Arabidopsis carrying *BdMADS2*. Interestingly, in contrast to Arabidopsis and rice D class genes, overexpression of Brachypodium D-lineage genes did not directly affect carpel and ovule development in transgenic Arabidopsis. Further studies involving loss-of-function mutants would be required to confirm their role in ovule identity.

Wheat *SEEDSTICK* (*WSTK*) is an ortholog of Arabidopsis *STK* and rice *OsMADS13*. In wheat, its homologous genes are identified as *TaAGL9* and *TaAGL31* [117]. *WSTK* expression was observed in young to mature spikes, although its transcription was only restricted to pistils. During ovule development, the highest expression of *WSTK* was observed in the developing ovule of the pistils, suggesting an involvement in ovule specification and development [90]. Moreover, *WSTK* was expressed not only in true pistils but also in pistil-like stamens of an alloplasmic wheat line having

the *Aegilops crassa* Boiss. cytoplasm, which arose due to the homeotic transformation of stamens into pistil-like structures. During the homeotic transformation of stamens into pistil-like structures in an alloplasmic wheat line no significant difference were recorded in the expression of wheat class C gene homologs *WAG-1* and *WAG-2* [161]. Furthermore, yeast two-hybrid analysis demonstrated that the WSTK protein formed a complex with a class E protein (WSEP) [90]. These observations suggest that similar to Arabidopsis STK, wheat STK protein interacts with the E class protein to specify ovule identity providing an evidence functional conservation of class D genes in Arabidopsis and wheat.

3.4. Class E Homeotic Genes

Class E genes work in all floral organs and act as cofactors for A, B, C, and D class proteins to form higher order MADS-box protein complexes, which regulate the floral organ identity (floral quartet model) [162]. In Arabidopsis, four *SEPALLATA* genes (*SEP1-4*) have been reported and these specify sepal, petal, stamen, carpel, and ovule identity [34,35]. Knockdown of all of these genes results in the transformation of floral organs into bract-like structures and sepals.

In grasses, *SEP*-like genes are further classified into *SEP* and *LOFSEP* clades and *AGL6*-like genes [114,115,163]. The *SEP* clade in rice include *OsMADS7* and *OsMADS8.* Cosuppression of both genes results in severe homeotic and meristematic changes in all floral organs, especially in lodicules [104]. *OsMADS1*/*LEAFY HULL STERILE1* (*LHS1*), *OsMADS5*, and *OsMADS34*/*PANICLE PHYTOMER 2* (*PAP2*) are placed into the *LOFSEP* clade [110]. Mutations in *OsMADS1* produced an abnormal phenotype, which is described by the presence of lemma/palea-like leaves and lodicules [59]. Loss of *OsMADS1* transforms the lemma into glume-like structures [164]. Simultaneous knockdown of *OsMADS1*, *OsMADS5*, *OsMADS7*, and *OsMADS8* transforms the inner floral whorls into bract-like structures with no effect on the lemma [104] suggesting that *OsMADS1* is associated with lemma and palea differentiation. A more recent study confirms that *OsMADS1* is involved in floral meristem identity and activity because defective floral organs were observed in the outer two whorls of *osmads-1* flowers [165]. Yeast two-hybrid analysis showed OsMADS1 protein to form heterodimers with B, C and, D class proteins that modulate floral meristem determinacy and organ identity. *OsMADS1* and *OsMADS13* regulate meristem determinacy in partially independent pathways while *OsMADS17* is a direct target of *OsMADS1* during floral development. *OsMADS1* interacts physically and genetically with *OsMADS3* and *OsMADS58* to specify stamen identity and suppression of spikelet meristem reversion [165]. These findings suggest that *OsMADS1*, through physical and genetic interaction with floral homeotic regulators, has diversified functions in floral meristem maintenance and specification of organ identity. In contrast, mutations in *OsMADS34* disturb inflorescence morphology and interfere with primary and secondary branches [111]. *OsMADS34* has been shown to interact with rice class (A)-genes to define inflorescence meristem identity [112]. These observations demonstrate that *OsMADS34* plays an important role in inflorescence and spikelet meristem determination. Ren et al. [166] recently demonstrated that a new mutant allele (*m34-z*) of *OsMADS34* homeotically converted empty glumes into lemma-like organs, suggesting that *OsMADS34* is required for glume specification.

The rice *AGL6* clade contains two genes: *OsMADS6*/*MOSAIC FLORAL ORGANS 1* (*MFO1*) and *OsMADS17* [114,115]. Both genes specify floral organ identity, although the predominant role is played by *MFO1*. Mutant analysis showed that *MFO1* determines floral organs by synergistically interacting with all classes of homeotic genes, except class-(A) [113]. Interestingly, a null allele of *MFO1* converts all floral organs into lemma like structures, except the lemma, suggesting a critical role for *MFO1* in floral organ specification [167].

At least ten putative E class genes have been identified in maize, which can be further subcategorized into *SEP*, *LOFSEP*, and *AGL6* clades [105,109]. The *SEP* clade contains three genes (*ZMM6*, *7*, and *27*). The *LOFSEP* clade contains five genes (*ZMM3*, *8*, *14*, *24*, and *31*) and the *AGL6* clade contains two genes (*ZAG3* and *5*). Sequence and phylogenetic analysis showed that *ZMM3* was orthologous to *OsMADS5*, *ZMM6* was orthologous to *OsMADS7*, *ZMM7* and *ZMM27* were orthologous

to *OsMADS8*, *ZMM24* and *ZMM31* were orthologous to *PAP2*/*OsMADS34*, and *ZMM8* and *ZMM14* were orthologous to *LHS1*/*OsMADS1* (Figure 3). During spikelet development, *ZMM8* and *ZMM14* were expressed in upper florets, although not in the lower florets of floral organs [109], indicating a possible role in floral meristem determinacy. In situ hybridization showed that *ZMM6* and *ZMM27* were strongly expressed during the maize kernel development, with lower expression during inflorescence development and no expression at all during vegetative growth [105]. Furthermore, neither single nor double knockdown mutants of *ZMM6* and/or *ZMM27* resulted in kernel abnormalities or alterations in flower development indicating functional redundancy of class E genes in maize.

Similar to rice and maize, Brachypodium also contains six cass-E genes—*BdMADS1*, *7*, *11*, *26*, *28*, and *32*—which are further classified into *SEP*, *LOFSEP*, and *AGL6* clades [31]. The *SEP* clade contains two genes—*BdMADS26* and *32*—both of which expressed highly in all floral organs except for the lemma and palea. The *LOFSEP* clade contains *BdMADS1*, *7*, and *11*. Strong expression of *BdMADS7* and *BdMADS11* was detected in all floral organs. On the other hand, the *AGL6* clade contains only one gene—*BdMADS28*—which was weakly expressed in lodicules and stamens only [31]. These diversified expression patterns are consistent with those of rice and maize homologs and indicate functional divergence among different class E clades. To date, none of the *SEP* encoding genes has yet been functionally characterized in Brachypodium.

Wheat E class genes are also subcategorized into *SEP*, *LOFSEP*, and *AGL6* clades. The *SEP* clade contains wheat *SEP* (*WSEP*), *TaMADS1*, *TaAGL16*, *TaAGL28*, and *TaAGL30* genes [107,117]. In situ expression analysis showed *WSEP* in lodicules, stamens, and carpels during floral organ differentiation. Stronger expression of *WSEP* was observed in the palea after determination of floral organ identity, supporting the concept that in addition to organ differentiation, *WSEP* has a role in subsequent development [107]. Rice *AGL6* clade gene *OsMADS6* also showed palea specific expression [116]. Similar to Arabidopsis *SEP3*, *WSEP* also interacts with class B and C homeotic genes, suggesting a conserved role for grass E genes. *TaMADS1* is another *SEP* encoding gene that is characterized as an E class gene [108] and orthologous to the rice class E gene *OsMADS8/24*. *TaMADS1* is functionally similar to *WSEP* in that overexpression of both genes in transgenic Arabidopsis caused early flowering and terminal flower formation [107].

The *LOFSEP* clade in wheat contains eight genes: *WLHS1*, *TaAGL3*, *5*, *8*, *24*, *27*, *34*, and *40*. *LEAFY HULL STERILE 1* (*WLHS1*) is an ortholog of *OsMADS1/LHS1* [107,117]. High transcript levels of *WLHS1* accumulate in glumes, lemma, palea, and lodicules while stamens and pistils exhibiting low levels. This differential expression behavior of *LHS1* was also reported in other grass species including *Avena sativa*, *Chasmanthium latifolium*, *Pennisetum glaucum*, and *Sorghum bicolor* [168], which may be due to differences in corresponding inflorescence structures. The wheat *AGL6* clade contains two genes: *TaAGL6* and *TaMADS37* [117]. Mutants of the wheat *LOFSEP* and *AGL6* clades have not been identified and thus their function is unknown.

4. Non MADS-Box Genes Involved In Floral Organ Identity

Some non-MADS-box genes are also reported to regulate floral development. Mutants of rice *aberrant panicle organization1* (*apo1*) showed phenotypic resemblance with class C gene mutants. *APO1* mutants convert stamens into lodicules with extra carpels, suggesting that *APO1* positively regulates class C gene functions [121,122]. Moreover, expression of the rice class C gene (*OsMADS3*) was reduced in *apo1* mutants indicating that *APO1* positively regulates *OsMADS3* expression [169]. The Arabidopsis genes *UNUSUAL FLORAL ORGANS (UFO)* and *APO1* are orthologs and both encode F-box proteins. *UFO* activates class B genes [123], suggesting a distinct role for both genes irrespective of their similar biochemical functions. Arabidopsis *FLORICAULA* (*FLO*)/*LEAFY* (*LFY*) and its rice ortholog *RICE FLORICAULA/LEAFY* (*RFL*)/*APO2* displayed different yet overlapping functions. For example, *RFL/APO2* specifies inflorescence meristem identity through interaction with *APO1* [170], whereas *FLO/LFY* specifies floral meristem identity and activates class A, B, and C genes [171].

As described above, carpel identity is defined by class C MADS-box genes, however *YABBY* TFs are also reported to play a major role in carpel identity. The rice mutant *drooping leaf (dl)* has some functional similarity to class C MADS TFs in specifying the carpel identity and mutation in *DL* converts carpels into stamens [80]. These findings support the notion that candidate carpel identity genes in rice (class C and *DL*) redundantly regulate class C gene functions. *DL* and the class B gene *OsMADS16/SPW1* antagonize each other and this antagonism is critical to setting boundaries between stamen and carpel identity [80]. The Arabidopsis gene *CRABSCLAW (CRC)* and rice *DL* both encode *YABBY* TFs, although in addition to its function in carpel identity; *CRC* also has a role in nectary development [118]. The expression pattern of *CRC* in homeotic mutants suggests negative regulation by class-(A) and B genes. The wheat *DL* ortholog (*TaDL*), which was identified by homology screening [119] and expression in alloplasmic wheat, was found in true pistils as well as pistil-like stamens, suggesting its role in carpel specification. Moreover, class B genes were not expressed in pistil like stamens indicating that *TaDL* and class B genes are mutually antagonistic [77]. Like rice and wheat, the maize *DL* mutants (*drl1* and *drl2*) have been characterized and cloned. The *drl* mutants displayed ectopic inner-whorl organs in pistillate and staminate florets [120]. Although meristem activity was influenced by the expression of the *Drl* loci, *Drl* transcripts were absent in floral meristems suggesting that *Drl* genes may function autonomously. Sang et al. [125] characterised *CHIMERIC FLORAL ORGANS 1 (CFO1)*, a MADS-box gene, which regulates floral organ identity in rice. Mutants of *CFO1* showed disrupted marginal palea with ectopic but chimeric floral organs. Expression pattern analysis revealed that rice *DL* was ectopically expressed in defective floral organs of *cfo1* flowers, suggesting negative regulation between *CFO1* and *DL* [125].

More recently, Liu et al. [124] reported that *LONG STERILE LEMMA1 (G1)/ELONGATED EMPTY GLUME (ELE)* and *OsMADS34/PAP2* were associated with rice lemma development and determination of empty glume identity. Mutants of *G1/ELE* showed homeotic transformation of empty glumes into lemma like organs. Single and double mutants of *G1/ELE* and *OsMADS1/OsLHS1* showed redundant roles for both genes in controlling empty glume identity and lemma development. Expression analysis of *G1/ELE* in *osmads-1* flowers and *OsMADS1/LHS1* in *g-1* flowers indicated that both genes are regulated through independent pathways and do not interact at the transcription level. In *G1/ELE* mutant plants, downregulation of empty glume identity genes and ectopic expression of lemma identity genes provides strong evidence that empty glumes are in-fact sterile lemmas. Moreover, Yang et al. [126] identified a single recessive gene *lemma-distortion1 (ld1)* associated with lemma development in rice. This gene encodes a zinc finger protein. Overall these reports suggest that a plethora of non-MADS-box genes are also involved in floral organ identity in eudicots and cereals.

5. Functional Conservation and Diversification between Distinct Floral Specification Systems

If it is assumed that all angiosperms have homologous reproductive organs, then divergent angiosperm groups may have a single ancestral flower specifying genetic mechanism. The modified floral organ identity model (ABCDE) in Arabidopsis suggests occurrence of genetic interactions among floral homeotic genes. The same model can be used to interpret molecular control of inflorescence identity in other crop plants including cereals [37]. High-throughput forward and reverse genetic approaches have led to the identification, cloning, and functional characterization of several genes involved in the regulation of floral development especially in grasses. Interestingly, most of these genes exhibit highest sequence similarites and share expression patterns and functional properties with those of eudicot A, B, C, D, and E floral homeotic genes. However, some grass-specific floral regulators have also been identified that do not have eudicot homologs and perform distinct functions in grass floral development. This review integrates current knowledge of floral organ identity genes in an attempt to adopt the eudicot floral organ identity model to other crop species. Considering grass organ-identity models illustrated in Figure 2, it is apparent that further research is needed to functionally characterize maize, Brachypodium, and wheat MADS-box genes to manipulate for crop grain yield improvement.

Prior to the discovery of loss-of-function mutants, gene function was usually examined through sequence conservation and expression pattern comparisons with already characterised genes. Due to genome complexities and difficulties in the use of modern genetic approaches in grasses, very little was known about the role of MADS-box genes in controlling spikelet and floret development. However, recent studies have provided new insights into the conservation of class (A)-gene function among eudicots and cereals [65]. The mutational analyses of *AP1/FUL* like genes in rice and demonstrated that in addition to their role in floral meristem identity, they also influenced the specification of palea and lodicule identities. As these grass-specific organs are thought to be homologous to eudicot's sepals and petals, these divergent groups may share a conserved nonreproductive floral organ specification system.

Comparison of all the proposed models indicates a partially conservative, partially-diverse floral regulation among grasses and higher eudicots. Based on studies in model eudicots, gene expression patterns and mutant phenotypes appear to be consistent with functional predictions. This is essentially true for class B, C, and D MADS-box genes. Like eudicots, the functions of class B and C genes have diverged in grasses due to duplication and subfunctionalization of separate genes. For example, rice *PI*-like class B genes show unequal redundancy in their function. Individual mutant analysis of *OsMADS2* indicates the homeotic conversion of lodicules without affecting stamens, whereas *OsMADS4* shows no alteration in lodicules and stamens. Additionally, double mutants of both genes show the conversion of lodicule and stamen into palea and carpel-like structures, suggesting an equal role of both genes in stamen identity. However, *OsMADS2* is more important than *OsMADS4* in lodicule identity. Mutant analysis of maize *PI* orthologs also indicated that *ZMM16/STS1* can compensate *zmm18/29* reduction while specifying class B gene functions. However, *ZMM18/29* cannot compensate for *zmm16/sts1* reduction. Similar roles may be speculated for Brachypodium and wheat *PI* orthologs as *BdMADS16* and *WPI1* show sequence similarity with *OsMADS4* and *ZMM18*, respectively, while *BdMADS20* and *WPI2* have greater similarity with respect to *OsMADS2* and *ZMM16* (Figure 3).

Subfunctionalization of duplicated genes was also observed among grass related class C genes. According to Yamaguchi et al. [98], *OsMADS58* plays a major role in floral meristem determinacy with minor effects on floral organ identity; whereas *OsMADS3* has a dominant role in stamen identity with a minor role in meristem determinacy. Similar findings were reported by Dreni et al. [92], in which *OsMADS3* showed to regulate stamen identity compared with *OsMADS58*. Severe defects were obsereved in *osmads3* mutant flowers, whereas most of the *osmads58* mutant flowers were indistinguishable from the wild type flowers. Likewise, the maize orthologs *ZAG1* and *ZMM2* exhibit functional diversification and are homologs to *OsMADS58* and *OsMADS3*, respectively. *ZAG1* mutant showed loss of floral meristem determinacy and have a minor role in stamen and carpel identity. By contrast, *ZMM2* has yet to be characterized for floral organ identity. In Brachypodium, *BdMADS14* and *BdMADS18* also indicated overlapping but diverse expression patterns as *BdMADS14* highly expressed in stamens only compared to *BdMADS18* that strongly expressed both in stamens and carpels. Based on orthologous relationships, Brachypodium and maize class C genes have overlapping but nonidentical functions, similar to rice genes. In contrast, the wheat class C genes *WAG-1* and *WAG-2* are involved in ectopic ovule formation and homeotic conversion of stamens into pistil- like structures and are orthologous to *OsMADS58* and *OsMADS3,* respectively. Knockdown mutants of *WAG* genes will help elucidate their function in floral organ identity. Mutations in rice and maize class C genes result in homeotic conversion of stamens and carpels into lodicules and paleae-like organs, respectively. Similarly, class C gene mutations in Arabidopsis caused homeotic conversion of both reproductive organs, with few exceptions. Mutations of these genes in wheat result in ectopic ovule formation and homeotic conversion of stamens into pistil-like structures. Thus, diverse carpel specification systems operate in these two divergent groups.

Genetic interactions between rice class C and D MADS-box and non-MADS-box (*DL*) genes provided new insights into the partially conservative and partially diversified mechanisms regulating floral development in eudicots and grasses. Li et al. [159] studied double mutants of *OsMADS3*, *OsMADS13*, and *DROOPING LEAF* (*DL*) to investigate their role in floral development. Termination

of floral meristem determinacy and several carpelloid structures were observed in *osmads3/osmads13* double mutants, while noteably their single mutants lacked these alterations at all. Furthermore, gene expression and protein–protein interaction analyses revealed that both C and D class genes neither regulate nor interact at the transcription or protein levels, suggesting that class E genes would mediate their interaction to synergistically control the termination of floral meristem and ovule identity. These obervations support the notion that grasses have retained their class C and D gene functions, despite undergone duplication and subfunctionalization. Dramatically, double mutants of *osmads3/dl* showed no *AG* activity, with production of lodicule-like structures within the fourth whorl and the termination of floral meristem determinacy. Mutual suppression was also absent and normal expression patterns were observed for *OsMADS3* and *DL* genes in ectopic stamens of *dl* flowers and *osmads3* mutants, respectively, suggesting a redundant role for both genes in floral meristem termination. In contrast, single and double mutants of *OsMADS13* and *DL* suggest that *DL* is epistatic to *OsMADS3*, and both have identical roles in ovule identity. More recently, the role of ovule in cereal grain development has been briefly reviewd [172]. The phenotype of single *dl* mutants was identical to that of double *osmads13/dl* mutants. The *dl* mutants lacked *OsMADS13* expression, whereas single mutant of *osmads13* showed abundance of *DL* transcripts, indicating direct or indirect *OsMADS13* regulation by *DL*. Dreni et al. [92] investigated genetic interactions between rice class C and D genes (*OsMADS13* and *OsMADS58*) through single- and multiple-gene mutants. As expected, *osmads13/58* double mutants showed accumulation of lodicule and palea-like organs in the third and fourth whorls accompanied by loss of floral organ identity and the triggering of floral meristem indeterminacy due to reduced *AG* activity. These observations suggest that a highly conserved class C gene functional mechanism exists in grasses despite partial subfunctionalization among duplicated genes. These results led to the proposition that *DL* lacks class C gene functional activity and cannot specify carpel identity alone, requiring both *OsMADS3* and *OsMADS58*. Although these findings provide new insights into the floral development process, further examination of single and multiple mutants of class B, C, D, and *DL* genes will be required to elucidate the roles of MADS-box and *DL* genes in floral organ specification.

The function of class E genes, particularly the *SEP*-like, are somewhat conserved among eudicots and grasses. In rice, detailed spatial and temporal mRNA expression studies, protein interaction patterns, and mutant analysis indicated a consistent role for *SEP*-like genes in floral meristem and organ identity specification [104], since divergence from eudicots. Mutant phenotypes of Arabidopsis *SEP* redundant genes (*SEP1/2/3/4*) and rice *OsMADS1*, *OsMADS7*, and *OsMADS8* genes indicated a redundant but interdependent role for both groups, suggesting a partial overlap but subfunctionalization among class E genes. Furthermore, characterization of *AGL6* clade mutants of rice and maize indicated a similar functional role in floral meristem determinacy and organ identity [114,115,159]. Similar expression patterns of monocot *AGL6* and *SEP* clade genes and their complex interactions with class B, C, and D genes indicated conserved floral specification systems. The functional similarity of *SEP* and *AGL6* clades is provided by Petunia floral development genes [173].

6. Future Perspectives

Understanding of grass inflorescence morphogenesis has expanded significantly over the last two decades. Extensive studies in model plants have demonstrated common genetic factors regulating eudicot and grass floral development including MADS-box and non-MADS-box genes and epigenetic regulators. For Arabidopsis and rice, the genetic and molecular mechanisms of transition from vegetative to reproductive phase and the role of MADS-box genes in floral organ identity are well understood. However, this is less well defined for Brachypodium, maize, and wheat because loss-of-function mutant analysis is rare in these species. Currently, advances in genetics analysis has made mutant development and characterization easy in grasses which is being used to define grass floral developmental biology. Deciphering the molecular control of transition from shoot apical meristem to floral meristem development and the determination of floral organ identity will provide

new insights to devise innovative strategies for the development of cereals with enhanced grain yield and adaptation to multiple environments [174].

Biological research in general and plant evolutionary biology have been revolutionized by advances in next generation sequencing. Enormous amounts of genomic and transcriptomic sequence data have also been generated through the 1000 Plants Project (1KP) and the Floral Genome Project [175,176]. These gene resources provide an unprecedented opportunity to bridge the evolutionary gap between floral morphogenesis in model plants and economically important cereals by characterizing floral genetic components of ABCD model. Here, it is important to note that most of the cereal orthologs are merely retrieved by phylogenetic analysis from resource genome databases; therefore, specified experimental studies will be required to support genetic framework of underlying mechanisms of floral organ identity in cereals. In this perspective, genome editing tools such as virus-induced gene silencing (VIGS) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system has been widely useful in several eudicots and monocot species to investigate gene function in floral organ identity and symmetry between basal and core eudicots [177–181]. Effective application of these systems could herald a new generation of multidisciplinary evo–devo research that better describes the evolutionary changes in gene regulatory networks underlying floral development. Moreover, these approaches along with TILLING resources can provide new avenues for grain yield improvement in cereals through translational research.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/11/2743/s1.

Author Contributions: Z.A. conceived the idea. Z.A., Q.R., and R.M.A. retrieved relevant literature and drafted the manuscript. Q.R., R.M.A., and M.A. performed phylogenetic analysis, interpreted results, and drew figures. Z.A., U.A., and G.C. read, reviewed, and edited the manuscript. All authors listed have made substantial, direct, and intellectual contribution to the work and approved the final manuscript.

Funding: This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ014344)" Rural Development Administration, Korea.

Acknowledgments: The authors gratefully acknowledge the review effort of Richard Trethowan (University of Sydney, Australia) for improving the manuscript and financial support from the USPCAS-AFS, University of Agriculture Faisalabad, Pakistan and Higher Education Commission of Pakistan.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

ABS	*ARABIDOPSIS BSISTER*
AG	*AGAMOUS*
AGL6	*AGAMOUS-LIKE 6*
AP1	*APETALA1*
AP2	*APETALA2*
AP3	*APETALA3*
APO1	*Aberrant panicle organization-1*
BD1	*BRANCHED SILKLESS1*
BdMADS	*Brachypodium MADS*
BM	Branch meristem
BR	Branch
CA	Carpel
CAL	*CAULIFLOWER*
CFO1	*CHIMERIC FLORAL ORGANS 1*
CRC	*CRABSCLAW*
DL	*Drooping leaf*
ELE	*ELONGATED EMPTY GLUME*
FBP11	*FLORAL BINDINGPROTEIN 11*
FBP7	*FLORAL BINDINGPROTEIN 7*

FLO/LFY	*FLORICAULA/ LEAFY*
FM	Floral meristem
FUL1	*FRUITFULL 1*
FUL2	*FRUITFULL 2*
FUL3	*FRUITFULL 3*
FZP	*FRIZZY PANICLE*
G1	*LONG STERILE LEMMA1*
GL	Glume
GOA	*GORDITA*
HvAP2/Cly1	*Hordeum vulgare APETALA2/Cleistogamy1*
IDS1	*INDETERMINATE SPIKELET1*
IM	Inflorescence meristem
ld-1	*lemma-distortion 1*
Le	Lemma
LHS1	*LEAFY HULL STERILE 1*
LO	Lodicule
LS	Lateral spikelet
LSM	Lateral spikelet meristem
MADS	*MINICHROMOSOME MAINTENANCE 1, AGAMOUS, DEFICIENS, and SERUM RESPONSE FACTOR*
MEGA6	Molecular Evolutionary Genetic Analysis version 6
MFO1	*MOSAIC FLORAL ORGANS 1*
MFS1	*MULTI-FLORET SPIKELET 1*
miRNA	micro-RNA
MOS1	*MORE SPIKELET 1*
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
OsIDS1	*Oryza sativa INDETERMINATE SPIKELET 1*
OsMADS	*Oryza sative MADS*
OV	Ovule
PA	Palea
PAP2	*PANICLE PHYTOMER2*
PB	Primary branch
PBM/pBM	Primary branch meristem
PE	Petal
PI	*PISTILLATA*
PI	Pistil
RA	Rachis
Ra	Rachilla
RFL	*RICE FLORICAULA*
RNAi	RNA interference
RT-PCR	Reverse transcriptase-polymerase chain reaction
SAM	Shoot apical meristem
SB	Secondary branch
SBM/sBM	Secondary branch meristem
SBP/SPL	*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE*
SE	Sepal
SEP	*SEPALLATA*
SHP	*SHATTERPROOF*
SID1	*SISTER OF INDETERMINATE SPIKELET1*
Sl1	*SILKY1*
SM	Spikelet meristem
SNB	*SUPERNUMERARY BRACT*
SPM	Spikelet pair meristem
SPW1	*SUPERWOMAN1*
ST	Stamen

STK	SEEDSTICK
STS1	*Sterile Tassel Silky Ear1*
TaAG3	*Triticum aestivum AGAMOUS3*
TaAP3	*Triticum aestivum AP3*
TaDL	*Triticum aestivum Drooping Leaf*
TaMADS	*Triticum aestivum MADS*
TFs	Transcription factors
TS	Terminal spikelet
TSM	Terminal spikelet meristem
TT16	*TRANSPARENT TESTA 16*
UFO	*UNUSUAL FLORAL ORGANS*
VRN1	*VERNALIZATION 1*
WAG-1	*Wheat AGAMOUS-1*
WAG-2	*Wheat AGAMOUS-2*
WAP1	*Wheat APETALA1*
WAP3	*Wheat APETALA3*
WBsis	*Wheat BSISTER*
WFUL1	*Wheat FRUITFULL 1*
WFUL2	*Wheat FRUITFULL 2*
WFUL3	*Wheat FRUITFULL 3*
WLHS1	*Wheat LEAFY HULL STERILE 1*
WPl1	*Wheat PISTILLATA 1*
WPI2	*Wheat PISTILLATA 2*
WSEP	*Wheat SEPALLATA*
WSTK	*Wheat SEEDSTICK*
ZAG1	*Zea mays AGAMOUS1*
ZAG2	*Zea mays AGAMOUS2*
ZAP1	*Zea mays APETALA1*
ZmIDS1	*Zea mays INDETERMINATE SPIKELET1*
ZMM	*Zea mays MADS*

References

1. Awika, J.M. Major cereal grains production and use around the world. In *Advances in Cereal Science: Implications to Food Processing and Health Promotion*; ACS Publications: Washington, DC, USA, 2011; pp. 1–13.
2. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food security: The challenge of feeding 9 billion people. *Science* **2010**, *327*, 812–818. [CrossRef]
3. Soreng, R.J.; Peterson, P.M.; Romaschenko, K.; Davidse, G.; Zuloaga, F.O.; Judziewicz, E.J.; Filgueiras, T.S.; Davis, J.I.; Morrone, O. A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* **2015**, *53*, 117–137. [CrossRef]
4. Bell, C.D.; Soltis, D.E.; Soltis, P.S. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **2010**, *97*, 1296–1303. [CrossRef] [PubMed]
5. Soltis, D.E.; Bell, C.D.; Kim, S.; Soltis, P.S. Origin and early evolution of angiosperms. *Ann. NY Acad. Sci.* **2008**, *1133*, 3–25. [CrossRef] [PubMed]
6. Schmidt, R.J.; Ambrose, B.A. The blooming of grass flower development. *Curr. Opin. Plant Biol.* **1998**, *1*, 60–67. [CrossRef]
7. Bommert, P.; Satoh-Nagasawa, N.; Jackson, D.; Hirano, H.-Y. Genetics and evolution of inflorescence and flower development in grasses. *Plant Cell Physiol.* **2005**, *46*, 69–78. [CrossRef] [PubMed]
8. Lombardo, F.; Yoshida, H. Interpreting lemma and palea homologies: A point of view from rice floral mutants. *Front. Plant Sci.* **2015**, *6*, 61. [CrossRef]
9. Goff, S.A.; Ricke, D.; Lan, T.-H.; Presting, G.; Wang, R.; Dunn, M.; Glazebrook, J.; Sessions, A.; Oeller, P.; Varma, H. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* **2002**, *296*, 92–100. [CrossRef] [PubMed]

10. Yu, J.; Hu, S.; Wang, J.; Wong, G.K.-S.; Li, S.; Liu, B.; Deng, Y.; Dai, L.; Zhou, Y.; Zhang, X. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* **2002**, *296*, 79–92. [CrossRef]
11. Itoh, J.-I.; Nonomura, K.-I.; Ikeda, K.; Yamaki, S.; Inukai, Y.; Yamagishi, H.; Kitano, H.; Nagato, Y. Rice plant development: From zygote to spikelet. *Plant Cell Physiol.* **2005**, *46*, 23–47. [CrossRef]
12. Kurata, N.; Miyoshi, K.; Nonomura, K.-I.; Yamazaki, Y.; Ito, Y. Rice mutants and genes related to organ development, morphogenesis and physiological traits. *Plant Cell Physiol.* **2005**, *46*, 48–62. [CrossRef] [PubMed]
13. Matsumoto, T.; Wu, J.; Kanamori, H.; Katayose, Y. The map-based sequence of the rice genome. *Nature* **2005**, *436*, 793.
14. Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.-S. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524. [CrossRef] [PubMed]
15. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [CrossRef] [PubMed]
16. Initiative, I.B. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **2010**, *463*, 763.
17. Brkljacic, J.; Grotewold, E.; Scholl, R.; Mockler, T.; Garvin, D.F.; Vain, P.; Brutnell, T.; Sibout, R.; Bevan, M.; Budak, H. Brachypodium as a model for the grasses: Today and the future. *Plant Physiol.* **2011**, *157*, 3–13. [CrossRef] [PubMed]
18. Draper, J.; Mur, L.A.; Jenkins, G.; Ghosh-Biswas, G.C.; Bablak, P.; Hasterok, R.; Routledge, A.P. Brachypodium distachyon. A new model system for functional genomics in grasses. *Plant Physiol.* **2001**, *127*, 1539–1555. [CrossRef] [PubMed]
19. Vogel, J.P.; Garvin, D.F.; Mockler, T.C.; Schmutz, J.; Rokhsar, D.; Bevan, M.W.; Barry, K.; Lucas, S.; Harmon-Smith, M.; Lail, K. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **2010**, *463*, 763–768.
20. Shewry, P.R.; Hey, S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* **2015**, *4*, 178–202. [CrossRef]
21. Consortium, I.W.G.S. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **2014**, *345*, 1251788.
22. Jia, J.; Zhao, S.; Kong, X.; Li, Y.; Zhao, G.; He, W.; Appels, R.; Pfeifer, M.; Tao, Y.; Zhang, X. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **2013**, *496*, 7443. [CrossRef]
23. IWGSC, I. Shifting the limits in wheat research and breeding using a fully annotated reference genome by the international wheat genome sequencing consortium (iwgsc). *Science* **2018**, *361*.
24. Derbyshire, P.; Byrne, M.E. MORE SPIKELETS1 is required for spikelet fate in the inflorescence of Brachypodium. *Plant Physiol.* **2013**, *161*, 1291–1302. [CrossRef] [PubMed]
25. Dobrovolskaya, O.; Pont, C.; Sibout, R.; Martinek, P.; Badaeva, E.; Murat, F.; Chosson, A.; Watanabe, N.; Prat, E.; Gautier, N. FRIZZY PANICLE drives supernumerary spikelets in bread wheat. *Plant Physiol.* **2015**, *167*, 189–199. [CrossRef] [PubMed]
26. Feng, N.; Song, G.; Guan, J.; Chen, K.; Jia, M.; Huang, D.; Wu, J.; Zhang, L.; Kong, X.; Geng, S. Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. *Plant Physiol.* **2017**, *174*, 1779–1794. [CrossRef] [PubMed]
27. Gauley, A.; Boden, S.A. Genetic pathways controlling inflorescence architecture and development in wheat and barley. *J. Integr. Plant Biol.* **2019**. [CrossRef] [PubMed]
28. Koppolu, R.; Schnurbusch, T. Developmental pathways for shaping spike inflorescence architecture in barley and wheat. *J. Integr. Plant Biol.* **2019**, *61*, 278–295. [CrossRef] [PubMed]
29. Peng, F.Y.; Hu, Z.; Yang, R.-C. Genome-wide comparative analysis of flowering-related genes in Arabidopsis, wheat, and barley. *Int. J. Plant Genom.* **2015**, 874361. [CrossRef]
30. Wei, B.; Liu, D.; Guo, J.; Leseberg, C.H.; Zhang, X.; Mao, L. Functional divergence of two duplicated D-lineage MADS-box genes BdMADS2 and BdMADS4 from Brachypodium distachyon. *J. Plant Physiol.* **2013**, *170*, 424–431. [CrossRef]
31. Wei, B.; Zhang, R.-Z.; Guo, J.-J.; Liu, D.-M.; Li, A.-L.; Fan, R.-C.; Mao, L.; Zhang, X.-Q. Genome-wide analysis of the MADS-box gene family in Brachypodium distachyon. *PLoS ONE* **2014**, *9*, e84781. [CrossRef]

32. Coen, E.S.; Meyerowitz, E.M. The war of the whorls: Genetic interactions controlling flower development. *Nature* **1991**, *353*, 31–37. [CrossRef]
33. Colombo, L.; Franken, J.; Koetje, E.; van Went, J.; Dons, H.; Angenent, G.C.; van Tunen, A.J. The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell* **1995**, *7*, 1859–1868. [CrossRef]
34. Ditta, G.; Pinyopich, A.; Robles, P.; Pelaz, S.; Yanofsky, M.F. The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr. Biol.* **2004**, *14*, 1935–1940. [CrossRef]
35. Pelaz, S.; Ditta, G.S.; Baumann, E.; Wisman, E.; Yanofsky, M.F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **2000**, *405*, 200–203. [CrossRef]
36. Theißen, G. Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **2001**, *4*, 75–85. [CrossRef]
37. Ciaffi, M.; Paolacci, A.R.; Tanzarella, O.A.; Porceddu, E. Molecular aspects of flower development in grasses. *Sex. Plant Reprod.* **2011**, *24*, 247–282. [CrossRef]
38. Chongloi, G.L.; Prakash, S.; Vijayraghavan, U. Rice shoot and floral meristem development: An overview of developmental regulators of meristem maintenance and organ identity. *J. Exp. Bot.* **2019**, *70*, 1719–1736. [CrossRef]
39. Schnurbusch, T. Wheat and Barley Biology: Towards new frontiers. *J. Integr. Plant Biol.* **2019**, *61*, 198–203. [CrossRef]
40. Zhang, D.; Yuan, Z. Molecular control of grass inflorescence development. *Ann. Rev. Plant Biol.* **2014**, *65*, 553–578. [CrossRef]
41. Benlloch, R.; Berbel, A.; Serrano-Mislata, A.; Madueño, F. Floral initiation and inflorescence architecture: A comparative view. *Ann. Bot.* **2007**, *100*, 659–676. [CrossRef]
42. Doust, A. Architectural evolution and its implications for domestication in grasses. *Ann. Bot.* **2007**, *100*, 941–950. [CrossRef]
43. Kirby, E. Botany of the wheat plant. In *Bread Wheat. Improvement and Production*; Food and Agriculture Organization of the United Nation: Rome, Italy, 2002; pp. 19–37.
44. Hirano, H.-Y.; Tanaka, W.; Toriba, T. Grass flower development. In *Flower Development*; Springer: Berlin, Germany, 2014; pp. 57–84.
45. Langer, R.H.M.; Hanif, M. A study of floret development in wheat (*Triticum aestivum* L.). *Ann. Bot.* **1973**, *37*, 743–751. [CrossRef]
46. Hoshikawa, K. The growing rice plant: An anatomical monograph. *Nosan Gyoson Bunka* **1989**, 199–205.
47. Tanaka, W.; Pautler, M.; Jackson, D.; Hirano, H.-Y. Grass meristems II: Inflorescence architecture, flower development and meristem fate. *Plant Cell Physiol.* **2013**, *54*, 313–324. [CrossRef]
48. Kyozuka, J. Grass Inflorescence: Basic Structure and Diversity. In *Advances in Botanical Research*; Academic Press: Cambridge, MA, USA, 2014; pp. 191–219.
49. Liu, C.; Thong, Z.; Yu, H. Coming into bloom: The specification of floral meristems. *Development* **2009**, *136*, 3379–3391. [CrossRef]
50. Gramzow, L.; Theißen, G. Phylogenomics of MADS-box genes in plants—two opposing life styles in one gene family. *Biology* **2013**, *2*, 1150–1164. [CrossRef]
51. Shore, P.; Sharrocks, A.D. The MADS-box family of transcription factors. *Eur. J. Bioche.* **1995**, *229*, 1–13. [CrossRef]
52. Gramzow, L.; Ritz, M.S.; Theißen, G. On the origin of MADS-domain transcription factors. *Trends Genet.* **2010**, *26*, 149–153. [CrossRef]
53. Masiero, S.; Colombo, L.; Grini, P.E.; Schnittger, A.; Kater, M.M. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **2011**, *23*, 865–872. [CrossRef]
54. Smaczniak, C.; Immink, R.G.; Muiño, J.M.; Blanvillain, R.; Busscher, M.; Busscher-Lange, J.; Dinh, Q.P.; Liu, S.; Westphal, A.H.; Boeren, S. Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Nat. Acad. Sci.* **2012**, *109*, 1560–1565. [CrossRef] [PubMed]
55. Theissen, G.; Becker, A.; Di Rosa, A.; Kanno, A.; Kim, J.T.; Münster, T.; Winter, K.-U.; Saedler, H. A short history of MADS-box genes in plants. *Plant Mol. Biol.* **2000**, *42*, 115–149. [CrossRef]
56. Malcomber, S.T.; Preston, J.C.; Reinheimer, R.; Kossuth, J.; Kellogg, E.A. Developmental gene evolution and the origin of grass inflorescence diversity. *Adv. Bot. Res.* **2006**, *44*, 425–481.
57. Pinyopich, A.; Ditta, G.S.; Savidge, B.; Liljegren, S.J.; Baumann, E.; Wisman, E.; Yanofsky, M.F. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **2003**, *424*, 85–88. [CrossRef]

58. Danilevskaya, O.N.; Meng, X.; Selinger, D.A.; Deschamps, S.; Hermon, P.; Vansant, G.; Gupta, R.; Ananiev, E.V.; Muszynski, M.G. Involvement of the MADS-box gene ZMM4 in floral induction and inflorescence development in maize. *Plant Physiol.* **2008**, *147*, 2054–2069. [CrossRef]
59. Jeon, J.-S.; Lee, S.; Jung, K.-H.; Yang, W.-S.; Yi, G.-H.; Oh, B.-G.; An, G. Production of transgenic rice plants showing reduced heading date and plant height by ectopic expression of rice MADS-box genes. *Mol. Breed.* **2000**, *6*, 581–592. [CrossRef]
60. Kinjo, H.; Shitsukawa, N.; Takumi, S.; Murai, K. Diversification of three APETALA1/FRUITFULL-like genes in wheat. *Mol. Genet. Genom.* **2012**, *287*, 283–294. [CrossRef]
61. Mandel, M.A.; Gustafson-Brown, C.; Savidge, B.; Yanofsky, M.F. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. *Nature* **1992**, *360*, 273–277. [CrossRef]
62. Münster, T.; Deleu, W.; Wingen, L.; Cacharrón, N.; Ouzunova, M.; Faigl, W. Maize MADS-box genes galore. *Maydica* **2002**, *47*, 287–301.
63. Murai, K.; Miyamae, M.; Kato, H.; Takumi, S.; Ogihara, Y. WAP1, a wheat APETALA1 homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant Cell Physiol.* **2003**, *44*, 1255–1265. [CrossRef]
64. Preston, J.C.; Kellogg, E.A. Discrete developmental roles for temperate cereal grass VERNALIZATION1/FRUITFULL-like genes in flowering competency and the transition to flowering. *Plant Physiol.* **2008**, *146*, 265–276. [CrossRef]
65. Wu, F.; Shi, X.; Lin, X.; Liu, Y.; Chong, K.; Theißen, G.; Meng, Z. The ABCs of flower development: Mutational analysis of AP1/FUL-like genes in rice provides evidence for a homeotic (A)-function in grasses. *Plant J.* **2017**, *89*, 310–324. [CrossRef]
66. Yan, L.; Loukoianov, A.; Tranquilli, G.; Helguera, M.; Fahima, T.; Dubcovsky, J. Positional cloning of the wheat vernalization gene VRN1. *Proc. Nat. Acad. Sci.* **2003**, *100*, 6263–6268. [CrossRef]
67. Bai, X.; Huang, Y.; Mao, D.; Wen, M.; Zhang, L.; Xing, Y. Regulatory role of FZP in the determination of panicle branching and spikelet formation in rice. *Sci. Rep.* **2016**, *6*, 19022. [CrossRef]
68. Chuck, G.; Meeley, R.; Hake, S. Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes ids1 and sid1. *Development* **2008**, *135*, 3013–3019. [CrossRef]
69. Chuck, G.; Muszynski, M.; Kellogg, E.; Hake, S.; Schmidt, R.J. The control of spikelet meristem identity by the branched silkless1 gene in maize. *Science* **2002**, *298*, 1238–1241. [CrossRef]
70. Jofuku, K.D.; Den Boer, B.; Van Montagu, M.; Okamuro, J.K. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* **1994**, *6*, 1211–1225. [CrossRef]
71. Komatsu, M.; Chujo, A.; Nagato, Y.; Shimamoto, K.; Kyozuka, J. FRIZZY PANICLE is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. *Development* **2003**, *130*, 3841–3850. [CrossRef]
72. Lee, D.Y.; An, G. Two AP2 family genes, supernumerary bract (SNB) and Osindeterminate spikelet 1 (OsIDS1), synergistically control inflorescence architecture and floral meristem establishment in rice. *Plant J.* **2012**, *69*, 445–461. [CrossRef] [PubMed]
73. Ning, S.; Wang, N.; Sakuma, S.; Pourkheirandish, M.; Wu, J.; Matsumoto, T.; Koba, T.; Komatsuda, T. Structure, transcription and post-transcriptional regulation of the bread wheat orthologs of the barley cleistogamy gene Cly1. *Theor. Appl. Genet.* **2013**, *126*, 1273–1283. [CrossRef] [PubMed]
74. Ren, D.; Li, Y.; Zhao, F.; Sang, X.; Shi, J.; Wang, N.; Guo, S.; Ling, Y.; Zhang, C.; Yang, Z. MULTI-FLORET SPIKELET1, which encodes an AP2/ERF protein, determines spikelet meristem fate and sterile lemma identity in rice. *Plant Physiol.* **2013**, *162*, 872–884. [CrossRef]
75. Simons, K.J.; Fellers, J.P.; Trick, H.N.; Zhang, Z.; Tai, Y.-S.; Gill, B.S.; Faris, J.D. Molecular characterization of the major wheat domestication gene Q. *Genetics* **2006**, *172*, 547–555. [CrossRef]
76. Ambrose, B.A.; Lerner, D.R.; Ciceri, P.; Padilla, C.M.; Yanofsky, M.F.; Schmidt, R.J. Molecular and genetic analyses of the silky1 gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* **2000**, *5*, 569–579. [CrossRef]
77. Hama, E.; Takumi, S.; Ogihara, Y.; Murai, K. Pistillody is caused by alterations to the class-B MADS-box gene expression pattern in alloplasmic wheats. *Planta* **2004**, *218*, 712–720.
78. Jack, T.; Brockman, L.L.; Meyerowitz, E.M. The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. *Cell* **1992**, *68*, 683–697. [CrossRef]

79. Murai, K.; Takumi, S.; Koga, H.; Ogihara, Y. Pistillody, homeotic transformation of stamens into pistil-like structures, caused by nuclear–cytoplasm interaction in wheat. *Plant J.* **2002**, *29*, 169–181. [CrossRef]
80. Nagasawa, N.; Miyoshi, M.; Sano, Y.; Satoh, H.; Hirano, H.; Sakai, H.; Nagato, Y. SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. *Development* **2003**, *130*, 705–718. [CrossRef]
81. Goto, K.; Meyerowitz, E.M. Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. *Genes Dev.* **1994**, *8*, 1548–1560. [CrossRef]
82. Murai, K.; Murai, R.; Takumi, S.; Ogihara, Y. Cloning and characterization of cDNAs corresponding to the wheat MADS box genes. In Proceedings of the 9th Int Wheat Genet Symp, Saskatoon, SK, Canada, 2–7 August 1998; pp. 89–94.
83. Prasad, K.; Vijayraghavan, U. Double-stranded RNA interference of a rice PI/GLO paralog, OsMADS2, uncovers its second-whorl-specific function in floral organ patterning. *Genetics* **2003**, *165*, 2301–2305.
84. Schilling, S.; Gramzow, L.; Lobbes, D.; Kirbis, A.; Weilandt, L.; Hoffmeier, A.; Junker, A.; Weigelt-Fischer, K.; Klukas, C.; Wu, F. Non-canonical structure, function and phylogeny of the Bsister MADS-box gene OsMADS30 of rice (Oryza sativa). *Plant J.* **2015**, *84*, 1059–1072. [CrossRef]
85. Whipple, C.J.; Ciceri, P.; Padilla, C.M.; Ambrose, B.A.; Bandong, S.L.; Schmidt, R.J. Conservation of B-class floral homeotic gene function between maize and Arabidopsis. *Development* **2004**, *131*, 6083–6091. [CrossRef]
86. Yao, S.-G.; Ohmori, S.; Kimizu, M.; Yoshida, H. Unequal genetic redundancy of rice PISTILLATA orthologs, OsMADS2 and OsMADS4, in lodicule and stamen development. *Plant Cell Physiol.* **2008**, *49*, 853–857. [CrossRef]
87. Becker, A.; Kaufmann, K.; Freialdenhoven, A.; Vincent, C.; Li, M.-A.; Saedler, H.; Theissen, G. A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genom.* **2002**, *266*, 942–950.
88. Nesi, N.; Debeaujon, I.; Jond, C.; Stewart, A.J.; Jenkins, G.I.; Caboche, M.; Lepiniec, L. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **2002**, *14*, 2463–2479. [CrossRef]
89. Prasad, K.; Zhang, X.; Tobón, E.; Ambrose, B.A. The Arabidopsis B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant J.* **2010**, *62*, 203–214. [CrossRef]
90. Yamada, K.; Saraike, T.; Shitsukawa, N.; Hirabayashi, C.; Takumi, S.; Murai, K. Class D and Bsister MADS-box genes are associated with ectopic ovule formation in the pistil-like stamens of alloplasmic wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* **2009**, *71*, 1–14. [CrossRef]
91. Yang, X.; Wu, F.; Lin, X.; Du, X.; Chong, K.; Gramzow, L.; Schilling, S.; Becker, A.; Theißen, G.; Meng, Z. Live and let die-The B sister MADS-box gene OsMADS29 controls the degeneration of cells in maternal tissues during seed development of rice (Oryza sativa). *PLoS ONE* **2012**, *7*, e51435. [CrossRef]
92. Dreni, L.; Pilatone, A.; Yun, D.; Erreni, S.; Pajoro, A.; Caporali, E.; Zhang, D.; Kater, M.M. Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. *Plant Cell* **2011**, *23*, 2850–2863. [CrossRef]
93. Gómez-Mena, C.; de Folter, S.; Costa, M.M.R.; Angenent, G.C.; Sablowski, R. Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development* **2005**, *132*, 429–438. [CrossRef]
94. Hirabayashi, C.; Murai, K. Class C MADS-box gene AGAMOUS was duplicated in the wheat genome. *Wheat Inf. Serv.* **2009**, *107*, 13–16.
95. Meguro, A.; Takumi, S.; Ogihara, Y.; Murai, K. WAG, a wheat AGAMOUS homolog, is associated with development of pistil-like stamens in alloplasmic wheats. *Sex. Plant Reprod.* **2003**, *15*, 221–230.
96. Mena, M.; Ambrose, B.A.; Meeley, R.B.; Briggs, S.P. Diversification of C-function activity in maize flower development. *Science* **1996**, *274*, 1537. [CrossRef] [PubMed]
97. Schmidt, R.J.; Veit, B.; Mandel, M.A.; Mena, M.; Hake, S.; Yanofsky, M.F. Identification and molecular characterization of ZAG1, the maize homolog of the Arabidopsis floral homeotic gene AGAMOUS. *Plant Cell* **1993**, *5*, 729–737. [CrossRef] [PubMed]
98. Yamaguchi, T.; Lee, D.Y.; Miyao, A.; Hirochika, H.; An, G.; Hirano, H.-Y. Functional diversification of the two C-class MADS box genes OSMADS3 and OSMADS58 in Oryza sativa. *Plant Cell* **2006**, *18*, 15–28. [CrossRef]
99. Dreni, L.; Jacchia, S.; Fornara, F.; Fornari, M.; Ouwerkerk, P.B.; An, G.; Colombo, L.; Kater, M.M. The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. *Plant J.* **2007**, *52*, 690–699. [CrossRef]

100. Favaro, R.; Immink, R.; Ferioli, V.; Bernasconi, B.; Byzova, M.; Angenent, G.; Kater, M.; Colombo, L. Ovule-specific MADS-box proteins have conserved protein-protein interactions in monocot and dicot plants. *Mol. Genet. Genom.* **2002**, *268*, 152–159. [CrossRef]
101. Favaro, R.; Pinyopich, A.; Battaglia, R.; Kooiker, M.; Borghi, L.; Ditta, G.; Yanofsky, M.F.; Kater, M.M.; Colombo, L. MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **2003**, *15*, 2603–2611. [CrossRef]
102. Lopez-Dee, Z.P.; Wittich, P.; Pe, M.E.; Rigola, D.; Del Buono, I.; Gorla, M.S.; Kater, M.M.; Colombo, L. OsMADS13, a novel rice MADS-box gene expressed during ovule development. *Dev. Genet.* **1999**, *25*, 237–244. [CrossRef]
103. Yamaki, S.; Nagato, Y.; Kurata, N.; Nonomura, K.-I. Ovule is a lateral organ finally differentiated from the terminating floral meristem in rice. *Dev. Biol.* **2011**, *351*, 208–216. [CrossRef]
104. Cui, R.; Han, J.; Zhao, S.; Su, K.; Wu, F.; Du, X.; Xu, Q.; Chong, K.; Theißen, G.; Meng, Z. Functional conservation and diversification of class E floral homeotic genes in rice (Oryza sativa). *Plant J.* **2010**, *61*, 767–781. [CrossRef] [PubMed]
105. Lid, S.E.; Meeley, R.B.; Min, Z.; Nichols, S.; Olsen, O.-A. Knock-out mutants of two members of the AGL2 subfamily of MADS-box genes expressed during maize kernel development. *Plant Sci.* **2004**, *167*, 575–582. [CrossRef]
106. Li, Q.; Liu, B. Genetic regulation of maize flower development and sex determination. *Planta* **2017**, *245*, 1–14. [CrossRef] [PubMed]
107. Shitsukawa, N.; Tahira, C.; Kassai, K.-i.; Hirabayashi, C.; Shimizu, T.; Takumi, S.; Mochida, K.; Kawaura, K.; Ogihara, Y.; Murai, K. Genetic and epigenetic alteration among three homoeologous genes of a class E MADS box gene in hexaploid wheat. *Plant Cell* **2007**, *19*, 1723–1737. [CrossRef] [PubMed]
108. Zhao, X.Y.; Cheng, Z.J.; Zhang, X.S. Overexpression of TaMADS1, a SEPALLATA-like gene in wheat, causes early flowering and the abnormal development of floral organs in Arabidopsis. *Planta* **2006**, *223*, 698–707. [CrossRef]
109. Cacharrón, J.; Saedler, H.; Theißen, G. Expression of MADS box genes ZMM8 and ZMM14 during inflorescence development of Zea mays discriminates between the upper and the lower floret of each spikelet. *Dev. Genes Evo.* **1999**, *209*, 411–420. [CrossRef]
110. Christensen, A.R.; Malcomber, S.T. Duplication and diversification of the LEAFY HULL STERILE1 and Oryza sativa MADS5 SEPALLATA lineages in graminoid Poales. *EvoDevo* **2012**, *3*, 4. [CrossRef]
111. Gao, X.; Liang, W.; Yin, C.; Ji, S.; Wang, H.; Su, X.; Guo, C.; Kong, H.; Xue, H.; Zhang, D. The SEPALLATA-like gene OsMADS34 is required for rice inflorescence and spikelet development. *Plant Physiol.* **2010**, *153*, 728–740. [CrossRef] [PubMed]
112. Kobayashi, K.; Yasuno, N.; Sato, Y.; Yoda, M.; Yamazaki, R.; Kimizu, M.; Yoshida, H.; Nagamura, Y.; Kyozuka, J. Inflorescence meristem identity in rice is specified by overlapping functions of three AP1/FUL-like MADS box genes and PAP2, a SEPALLATA MADS box gene. *Plant Cell* **2012**, *24*, 1848–1859. [CrossRef] [PubMed]
113. Li, H.; Liang, W.; Hu, Y.; Zhu, L.; Yin, C.; Xu, J.; Dreni, L.; Kater, M.M.; Zhang, D. Rice MADS6 interacts with the floral homeotic genes SUPERWOMAN1, MADS3, MADS58, MADS13, and DROOPING LEAF in specifying floral organ identities and meristem fate. *Plant Cell* **2011**, *23*, 2536–2552. [CrossRef]
114. Li, H.; Liang, W.; Jia, R.; Yin, C.; Zong, J.; Kong, H.; Zhang, D. The AGL6-like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Res.* **2010**, *20*, 299–313. [CrossRef]
115. Ohmori, S.; Kimizu, M.; Sugita, M.; Miyao, A.; Hirochika, H.; Uchida, E.; Nagato, Y.; Yoshida, H. MOSAIC FLORAL ORGANS1, an AGL6-like MADS box gene, regulates floral organ identity and meristem fate in rice. *Plant Cell* **2009**, *21*, 3008–3025. [CrossRef]
116. Reinheimer, R.; Kellogg, E.A. Evolution of AGL6-like MADS box genes in grasses (Poaceae): Ovule expression is ancient and palea expression is new. *Plant Cell* **2009**, *21*, 2591–2605. [CrossRef]
117. Zhao, T.; Ni, Z.; Dai, Y.; Yao, Y.; Nie, X.; Sun, Q. Characterization and expression of 42 MADS-box genes in wheat (*Triticum aestivum* L.). *Mol. Genet. Genom.* **2006**, *276*, 334–350. [CrossRef]
118. Bowman, J.L.; Smyth, D.R. CRABS CLAW, a gene that regulates carpel and nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **1999**, *126*, 2387–2396.

119. Ishikawa, M.; Ohmori, Y.; Tanaka, W.; Hirabayashi, C.; Murai, K.; Ogihara, Y.; Yamaguchi, T.; Hirano, H.-Y. The spatial expression patterns of DROOPING LEAF orthologs suggest a conserved function in grasses. *Genes Genet. Sys.* **2009**, *84*, 137–146. [CrossRef]
120. Strable, J. Functional and Genomic Analyses of the Maize Yabby Transcription Factors Drooping Leaf1 and Drooping Leaf2. Ph.D. Thesis, Low State University, Uganda, Africa, 2015.
121. Ikeda, K.; Ito, M.; Nagasawa, N.; Kyozuka, J.; Nagato, Y. Rice ABERRANT PANICLE ORGANIZATION 1, encoding an F-box protein, regulates meristem fate. *Plant J.* **2007**, *51*, 1030–1040. [CrossRef] [PubMed]
122. Ikeda, K.; Nagasawa, N.; Nagato, Y. ABERRANT PANICLE ORGANIZATION 1 temporally regulates meristem identity in rice. *Dev. Biol.* **2005**, *282*, 349–360. [CrossRef] [PubMed]
123. Lee, I.; Wolfe, D.S.; Nilsson, O.; Weigel, D. A LEAFY co-regulator encoded by unusual floral organs. *Curr. Biol.* **1997**, *7*, 95–104. [CrossRef]
124. Liu, M.; Li, H.; Su, Y.; Li, W.; Shi, C. G1/ELE Functions in the Development of Rice Lemmas in Addition to Determining Identities of Empty Glumes. *Front. Plant Sci.* **2016**, *7*. [CrossRef] [PubMed]
125. Sang, X.; Li, Y.; Luo, Z.; Ren, D.; Fang, L.; Wang, N.; Zhao, F.; Ling, Y.; Yang, Z.; Liu, Y. CHIMERIC FLORAL ORGANS1, encoding a monocot-specific MADS box protein, regulates floral organ identity in rice. *Plant Physiol.* **2012**, *160*, 788–807. [CrossRef]
126. Yang, D.; Ye, X.; Zheng, X.; Cheng, C.; Ye, N.; Lu, L.; Huang, F.; Li, Q. Identification and fine mapping of lemma-distortion1, a single recessive gene playing an essential role in the development of lemma in rice. *J. Agri. Sci.* **2016**, *154*, 989–1001. [CrossRef]
127. Müller, B.M.; Saedler, H.; Zachgo, S. The MADS-box gene DEFH28 from Antirrhinum is involved in the regulation of floral meristem identity and fruit development. *Plant J.* **2001**, *28*, 169–179. [CrossRef]
128. Litt, A.; Irish, V.F. Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: Implications for the evolution of floral development. *Genetics* **2003**, *165*, 821–833.
129. Chen, A.; Dubcovsky, J. Wheat TILLING mutants show that the vernalization gene VRN1 down-regulates the flowering repressor VRN2 in leaves but is not essential for flowering. *PLOS Genet.* **2012**, *8*, e1003134. [CrossRef]
130. Wang, K.; Tang, D.; Hong, L.; Xu, W.; Huang, J.; Li, M.; Gu, M.; Xue, Y.; Cheng, Z. DEP and AFO regulate reproductive habit in rice. *PLOS Genet.* **2010**, *6*, e1000818. [CrossRef]
131. Paolacci, A.R.; Tanzarella, O.A.; Porceddu, E.; Varotto, S.; Ciaffi, M. Molecular and phylogenetic analysis of MADS-box genes of MIKC type and chromosome location of SEP-like genes in wheat (*Triticum aestivum* L.). *Mol. Genet. Genom.* **2007**, *278*, 689–708. [CrossRef]
132. Arora, R.; Agarwal, P.; Ray, S.; Singh, A.K.; Singh, V.P.; Tyagi, A.K.; Kapoor, S. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genom.* **2007**, *8*, 242. [CrossRef]
133. Pařenicová, L.; de Folter, S.; Kieffer, M.; Horner, D.S.; Favalli, C.; Busscher, J.; Cook, H.E.; Ingram, R.M.; Kater, M.M.; Davies, B. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis new openings to the MADS world. *Plant Cell* **2003**, *15*, 1538–1551. [CrossRef]
134. Trevaskis, B.; Bagnall, D.J.; Ellis, M.H.; Peacock, W.J.; Dennis, E.S. MADS box genes control vernalization-induced flowering in cereals. *Proc. Nat. Acad. Sci.* **2003**, *100*, 13099–13104. [CrossRef]
135. Deng, W.; Casao, M.C.; Wang, P.; Sato, K.; Hayes, P.M.; Finnegan, E.J.; Trevaskis, B. Direct links between the vernalization response and other key traits of cereal crops. *Nature Commun.* **2015**, *6*. [CrossRef]
136. Greenup, A.G.; Sasani, S.; Oliver, S.N.; Talbot, M.J.; Dennis, E.S.; Hemming, M.N.; Trevaskis, B. ODDSOC2 is a MADS box floral repressor that is down-regulated by vernalization in temperate cereals. *Plant Physiol.* **2010**, *153*, 1062–1073. [CrossRef] [PubMed]
137. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Sys. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]
138. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evo.* **1987**, *4*, 406–425.
139. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evo.* **2013**, *30*, 2725–2729. [CrossRef] [PubMed]

140. Licausi, F.; Ohme-Takagi, M.; Perata, P. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytol.* **2013**, *199*, 639–649. [CrossRef] [PubMed]
141. Lee, D.Y.; Lee, J.; Moon, S.; Park, S.Y.; An, G. The rice heterochronic gene SUPERNUMERARY BRACT regulates the transition from spikelet meristem to floral meristem. *Plant J.* **2007**, *49*, 64–78. [CrossRef]
142. Nair, S.K.; Wang, N.; Turuspekov, Y.; Pourkheirandish, M.; Sinsuwongwat, S.; Chen, G.; Sameri, M.; Tagiri, A.; Honda, I.; Watanabe, Y. Cleistogamous flowering in barley arises from the suppression of microRNA-guided HvAP2 mRNA cleavage. *Proc. Nat. Acad. Sci.* **2010**, *107*, 490–495. [CrossRef]
143. Jung, J.-H.; Seo, Y.-H.; Seo, P.J.; Reyes, J.L.; Yun, J.; Chua, N.-H.; Park, C.-M. The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. *Plant Cell* **2007**, *19*, 2736–2748. [CrossRef]
144. Yant, L.; Mathieu, J.; Dinh, T.T.; Ott, F.; Lanz, C.; Wollmann, H.; Chen, X.; Schmid, M. Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *Plant Cell* **2010**, *22*, 2156–2170. [CrossRef]
145. Wollmann, H.; Mica, E.; Todesco, M.; Long, J.A.; Weigel, D. On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Development* **2010**, *137*, 3633–3642. [CrossRef]
146. Zhu, Q.-H.; Upadhyaya, N.M.; Gubler, F.; Helliwell, C.A. Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (Oryza sativa). *BMC Plant Biol.* **2009**, *9*, 149. [CrossRef] [PubMed]
147. Zhu, Q.-H.; Helliwell, C.A. Regulation of flowering time and floral patterning by miR172. *J. Exp. Bot.* **2011**, erq295. [CrossRef]
148. Chung, Y.-Y.; Kim, S.-R.; Kang, H.-G.; Noh, Y.-S.; Park, M.C.; Finkel, D.; An, G. Characterization of two rice MADS box genes homologous to GLOBOSA. *Plant Sci.* **1995**, *109*, 45–56. [CrossRef]
149. Bartlett, M.E.; Williams, S.K.; Taylor, Z.; DeBlasio, S.; Goldshmidt, A.; Hall, D.H.; Schmidt, R.J.; Jackson, D.P.; Whipple, C.J. The Maize PI/GLO Ortholog Zmm16/sterile tassel silky ear1 interacts with the zygomorphy and sex determination pathways in flower development. *Plant Cell* **2015**, *27*, 3081–3098. [CrossRef] [PubMed]
150. De Folter, S.; Shchennikova, A.V.; Franken, J.; Busscher, M.; Baskar, R.; Grossniklaus, U.; Angenent, G.C.; Immink, R.G. A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. *Plant J.* **2006**, *47*, 934–946. [CrossRef]
151. Mizzotti, C.; Mendes, M.A.; Caporali, E.; Schnittger, A.; Kater, M.M.; Battaglia, R.; Colombo, L. The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J.* **2012**, *70*, 409–420. [CrossRef]
152. Kramer, E.M.; Jaramillo, M.A.; Di Stilio, V.S. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **2004**, *166*, 1011–1023. [CrossRef]
153. Zahn, L.M.; Leebens-Mack, J.H.; Arrington, J.M.; Hu, Y.; Landherr, L.L.; Depamphilis, C.W.; Becker, A.; Theissen, G.; Ma, H. Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: Evidence of independent sub-and neofunctionalization events. *Evo. Dev.* **2006**, *8*, 30–45. [CrossRef] [PubMed]
154. Yanofsky, M.F.; Ma, H.; Bowman, J.L.; Drews, G.N.; Feldmann, K.A.; Meyerowitz, E.M. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **1990**, *346*, 35. [CrossRef]
155. Liljegren, S.J.; Ditta, G.S.; Eshed, Y.; Savidge, B. SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* **2000**, *404*, 766. [CrossRef]
156. Yun, D.; Liang, W.; Dreni, L.; Yin, C.; Zhou, Z.; Kater, M.M.; Zhang, D. OsMADS16 genetically interacts with OsMADS3 and OsMADS58 in specifying floral patterning in rice. *Mol. Plant* **2013**, *6*, 743–756. [CrossRef]
157. Angenent, G.C.; Franken, J.; Busscher, M.; van Dijken, A.; van Went, J.L.; Dons, H.; van Tunen, A.J. A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* **1995**, *7*, 1569–1582. [CrossRef]
158. Singh, L.N.; Hannenhalli, S. Functional diversification of paralogous transcription factors via divergence in DNA binding site motif and in expression. *PLoS ONE* **2008**, *3*, e2345. [CrossRef]

159. Li, H.; Liang, W.; Yin, C.; Zhu, L.; Zhang, D. Genetic interaction of OsMADS3, DROOPING LEAF, and OsMADS13 in specifying rice floral organ identities and meristem determinacy. *Plant Physiol.* **2011**, *156*, 263–274. [CrossRef]
160. Theissen, G.; Strater, T.; Fischer, A.; Saedler, H. Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of AGAMOUS-like MADS-box genes from maize. *Gene* **1995**, *156*, 155–166.
161. Mizumoto, K.; Hatano, H.; Hirabayashi, C.; Murai, K.; Takumi, S. Altered expression of wheat AINTEGUMENTA homolog, WANT-1, in pistil and pistil-like transformed stamen of an alloplasmic line with Aegilops crassa cytoplasm. *Dev. Genes Evo.* **2009**, *219*, 175–187. [CrossRef]
162. Theissen, G.; Saedler, H. Plant biology: Floral quartets. *Nature* **2001**, *409*, 469–472. [CrossRef]
163. Malcomber, S.T.; Kellogg, E.A. SEPALLATA gene diversification: Brave new whorls. *Trends Plant Sci.* **2005**, *10*, 427–435. [CrossRef]
164. Prasad, K.; Parameswaran, S.; Vijayraghavan, U. OsMADS1, a rice MADS-box factor, controls differentiation of specific cell types in the lemma and palea and is an early-acting regulator of inner floral organs. *Plant J.* **2005**, *43*, 915–928. [CrossRef]
165. Hu, Y.; Liang, W.; Yin, C.; Yang, X.; Ping, B.; Li, A.; Jia, R.; Chen, M.; Luo, Z.; Cai, Q. Interactions of OsMADS1 with floral homeotic genes in rice flower development. *Mol. Plant* **2015**, *8*, 1366–1384. [CrossRef]
166. Ren, D.; Rao, Y.; Leng, Y.; Li, Z.; Xu, Q.; Wu, L.; Qiu, Z.; Xue, D.; Zeng, D.; Hu, J. Regulatory Role of OsMADS34 in the Determination of Glumes Fate, Grain Yield, and Quality in Rice. *Front. Plant Sci.* **2016**, *7*. [CrossRef]
167. Duan, Y.; Xing, Z.; Diao, Z.; Xu, W.; Li, S.; Du, X.; Wu, G.; Wang, C.; Lan, T.; Meng, Z. Characterization of Osmads6-5, a null allele, reveals that OsMADS6 is a critical regulator for early flower development in rice (*Oryza sativa* L.). *Plant Mol. Biol.* **2012**, *80*, 429–442. [CrossRef] [PubMed]
168. Malcomber, S.T.; Kellogg, E.A. Heterogeneous expression patterns and separate roles of the SEPALLATA gene LEAFY HULL STERILE1 in grasses. *Plant Cell* **2004**, *16*, 1692–1706. [CrossRef] [PubMed]
169. Tanaka, W.; Toriba, T.; Hirano, H.-Y. Flower development in rice. The Molecular Genetics of Floral Transition and Flower Development, (ed. Fornara, F.) 2014, 221–262.
170. Ikeda-Kawakatsu, K.; Maekawa, M.; Izawa, T.; Itoh, J.I.; Nagato, Y. ABERRANT PANICLE ORGANIZATION 2/RFL, the rice ortholog of Arabidopsis LEAFY, suppresses the transition from inflorescence meristem to floral meristem through interaction with APO1. *Plant J.* **2012**, *69*, 168–180. [CrossRef]
171. Rao, N.N.; Prasad, K.; Kumar, P.R.; Vijayraghavan, U. Distinct regulatory role for RFL, the rice LFY homolog, in determining flowering time and plant architecture. *Proc. Nat. Acad. Sci.* **2008**, *105*, 3646–3651. [CrossRef] [PubMed]
172. Wilkinson, L.G.; Dayton, C.B.; Matthew, R.T. Exploring the role of the ovule in cereal grain development and reproductive stress tolerance. *Ann. Plant Rev. Online* **2018**, *1*, 1–35.
173. Rijpkema, A.S.; Zethof, J.; Gerats, T.; Vandenbussche, M. The petunia AGL6 gene has a SEPALLATA-like function in floral patterning. *Plant J.* **2009**, *60*, 1–9. [CrossRef]
174. Raza, Q.; Ali, Z.; Karim, I.; Ajmal, M.; Khan, M.U. Genetic analysis of triple pistil wheat derived two F2 populations to enhance genetic yield potential. *Res. Plant Biol.* **2019**, *9*, 1–8.
175. Albert, V.A.; Soltis, D.E.; Carlson, J.E.; Farmerie, W.G.; Wall, P.K.; Ilut, D.C.; Solow, T.M.; Mueller, L.A.; Landherr, L.L.; Hu, Y. Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biol.* **2005**, *5*, 5. [CrossRef]
176. Matasci, N.; Hung, L.-H.; Yan, Z.; Carpenter, E.J.; Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Ayyampalayam, S.; Barker, M. Data access for the 1,000 Plants (1KP) project. *GigaScience* **2014**, *3*, 17. [CrossRef]
177. Bortesi, L.; Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotec. Adv.* **2015**, *33*, 41–52. [CrossRef]
178. Dinesh-Kumar, S.; Anandalakshmi, R.; Marathe, R.; Schiff, M.; Liu, Y. Virus-induced gene silencing. *Plant Funct. Genom.* **2003**, 287–293.
179. Gould, B.; Kramer, E.M. Virus-induced gene silencing as a tool for functional analyses in the emerging model plant Aquilegia (columbine, Ranunculaceae). *Plant Methods* **2007**, *3*, 6. [CrossRef]

180. Hidalgo, O.; Bartholmes, C.; Gleissberg, S. Virus-induced gene silencing (VIGS) in Cysticapnos vesicaria, a zygomorphic-flowered Papaveraceae (Ranunculales, basal eudicots). *Ann. Bot.* **2012**, *109*, 911–920. [CrossRef]
181. Sharma, B.; Guo, C.; Kong, H.; Kramer, E.M. Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytol.* **2011**, *191*, 870–883. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

The Class III Peroxidase (POD) Gene Family in Cassava: Identification, Phylogeny, Duplication, and Expression

Chunlai Wu [1,2,†], Xupo Ding [1,†], Zehong Ding [1], Weiwei Tie [1], Yan Yan [1], Yu Wang [3], Hai Yang [4,*] and Wei Hu [1,*]

1 Key Laboratory of Biology and Genetic Resources of Tropical Crops of Ministry of Agriculture, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China; wuchunlai19900109@126.com (C.W.); dingxupo@itbb.org.cn (X.D.); dingzehong@itbb.org.cn (Z.D.); tieweiwei@itbb.org.cn (W.T.); yanyan@itbb.org.cn (Y.Y.)
2 The Genetic Engineering International Cooperation Base of Chinese Ministry of Science and Technology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
3 Beijing Commerce and Trade School, Beijing 100162, China; wangyu7071@126.com
4 National Engineering Research Center for Nanomedicine, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
* Correspondence: YangHai@hust.edu.cn (H.Y.); huwei2013@itbb.org.cn (W.H.)
† These authors contributed equally to this work.

Received: 15 April 2019; Accepted: 31 May 2019; Published: 3 June 2019

Abstract: The class III peroxidase (POD) enzymes participate in plant development, hormone signaling, and stress responses. However, little is known about the POD family in cassava. Here, we identified 91 cassava *POD* genes (*MePODs*) and classified them into six subgroups using phylogenetic analysis. Conserved motif analysis demonstrated that all MePOD proteins have typical peroxidase domains, and gene structure analysis showed that *MePOD* genes have between one and nine exons. Duplication pattern analysis suggests that tandem duplication has played a role in *MePOD* gene expansion. Comprehensive transcriptomic analysis revealed that *MePOD* genes in cassava are involved in the drought response and postharvest physiological deterioration. Several *MePODs* underwent transcriptional changes after various stresses and related signaling treatments were applied. In sum, we characterized the POD family in cassava and uncovered the transcriptional control of *POD* genes in response to various stresses and postharvest physiological deterioration conditions. These results can be used to identify potential target genes for improving the stress tolerance of cassava crops.

Keywords: expression; genome-wide; identification of peroxidase genes; duplication pattern; stress; cassava

1. Introduction

Peroxidases (EC 1.11.1.X) form a large family of enzymes that are widely distributed in living organisms and catalyze the oxidoreduction reaction between hydrogen peroxide (H_2O_2) as an electron acceptor and diverse electron donors, such as auxin, phenolic compounds, or secondary metabolites [1,2]. According to their protein sequences and structure, peroxidases are classified as either non-heme peroxidases or heme peroxidases [3]. The majority of heme peroxidase members are divided into animal and non-animal groups [4]. On the basis of their amino acid sequences and catalytic properties, non-animal heme peroxidases are assigned to one of three large families: class I, II, or III [3,5]. The class

III peroxidases (EC 1.11.1.7) are plant-specific oxidoreductases and have various abbreviations (POX, POD, Px, PER, and Prx) [2].

There are many multigenic class III peroxidasesin land plants, which is commonly secreted into the vacuole and cell wall [4–7]. The structures and weights of Prx proteins are highly conserved between paralogs and orthologs [1,3]. The class III plant peroxidases contain 10–12 conserved α-helices and two short β-strands [2,7–10], and they mainly participate in the peroxidative cycle and hydroxylic cycle to reduce the production of hydrogen peroxide and the formation of reactive oxygen species (ROS) [4,5,11,12]. Prx proteins are involved in a variety of physiological processes, such as the cross-linking of cell wall components, salt tolerance, defense against pathogen attack, the oxidation of toxic reductants, and the metabolism of phytohormones [2,3,13–17].

Some genetic evidence supports Prx proteins' role in the plant response to biotic and abiotic stresses. Overexpression of *AtPrx64* was able to enhance tolerance to aluminum stress in transgenic tobacco plants [18]. *AtPrx3* was shown to positively regulate plant tolerance to drought and salt stresses in Arabidopsis [19]. Overexpression of the *Catharanthus roseus* genes *CrPrx* and *CrPrx1* in tobacco led to enhanced chilling resistance and increased germination rates under dehydration and salt treatments, respectively [20]. Repressing the expression of *Ep5C* in tomato resulted in reduced susceptibility to bacterial speck caused by the pathogen *Pseudomonas syringae* pv *tomato* [21]. *CaPO2* gene-silenced pepper plants were shown to be susceptible to infection by *Xanthomonas campestris* pv *vesicatoria*, whereas overexpression of *CaPO2* in transgenic *Arabidopsis thaliana* conferred bacterial disease resistance [22]. Transgenic carrot plants overexpressing *OsPrx114* exhibited enhanced resistance to necrotrophic fungal pathogens [23]. Together, these previous studies reveal the positive role of class III plant peroxidases in the response to biotic and abiotic stresses.

To date, the peroxidase (POD) family members have been characterized by whole-genome analyses in several plants, including 73 PODs in Arabidopsis [24–26], 138 PODs in rice [9], 93 PODs in *Populus trichocarpa* [27], 102 PODs in *Medicago sativa* [28], 119 PODs in maize [29], and 94 PODs in *Pyrus bretschneideri* [30]. However, there is less known about the POD family in cassava, a major tropical crop. Cassava is the third most valuable crop after maize and rice in Africa, Latin America, and Asia, supplying a carbohydrate source to 600 million people in tropical and subtropical regions [31]. Cassava can efficiently use water, heat, and light resources, and it is resistant to dehydration stress and lower-fertility soils [32,33]. Unfortunately, the potential of cassava as a food and industrial crop is restricted because its storage roots deteriorate within 72 h of its harvest [34]. ROS production is an early event that leads to the postharvest physiological deterioration (PPD) of cassava storage roots [35]. The mechanisms underlying cassava's resistance to drought and sensitivity to PPD are not well understood. POD proteins function by reducing the production of H_2O_2 and formation of ROS, which are involved in various physiological processes. Systematic investigations of the cassava POD family would provide novel insights into the POD-mediated stress response and regulation of root deterioration.

2. Results

2.1. Genome-Wide Identification of PODs in Cassava

According to the 211 POD protein sequences from Arabidopsis and rice genome databases, 91 POD members were predicted from the cassava genome using BLAST and HMMER methods. After conserved domain detection was confirmed, these cassava POD proteins (MePODs) were named MePOD01 to MePOD91. The full length of these putative cassava POD proteins ranges from 153 (MePOD63) to 422 (MePOD46) amino acid residues, and their relative molecular weight varies from 16.64 (MePOD63) to 46.12 kDa (MePOD46), with isoelectric points ranging from 4.43 (MePOD26) to 9.63 (MePOD39) (Table S1).

2.2. Phylogenetic and Comparative Analyses of PODs in Cassava

The homology and similarity of the *POD* genes in cassava were determined by performing multiple sequence alignments. Then, the radiation phylogenetic tree of the 91 *POD* genes was constructed using the neighbor-joining (NJ) method with a bootstrap value of 1000 using MEGA 5.1 (University College Dublin, Dublin, Ireland). The phylogenetic analyses indicated that *MePOD* genes can be divided into six subgroups on the basis of the observed genetic distance and bootstrap support (Figure 1). The large subgroups A and D consist of 23 and 24 MePOD members, respectively, whereas the small subgroups C and F contain 9 and 8 MePOD members, respectively. Subgroups B and E are composed of 15 and 12 MePODs members, respectively. These results show that a diversified POD family exists in cassava.

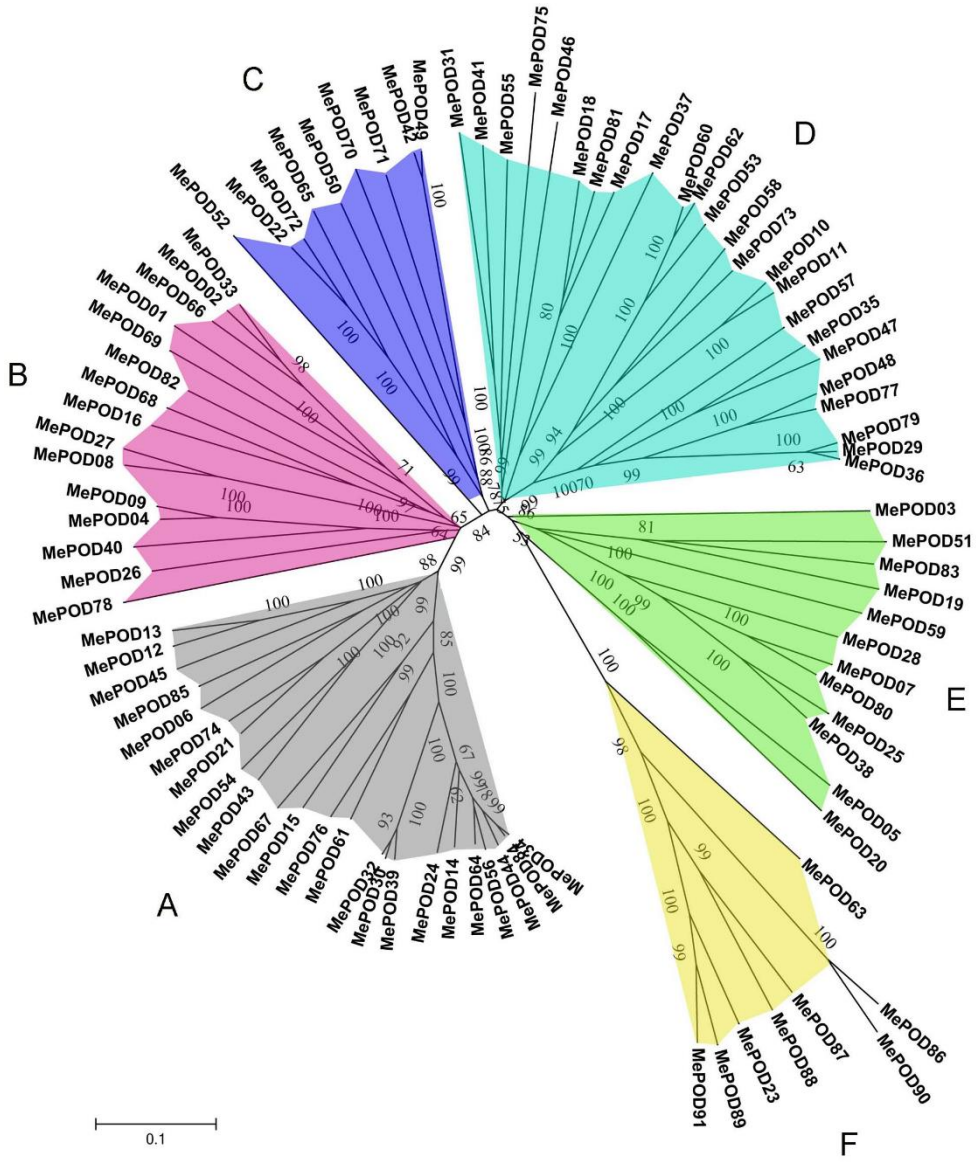


Figure 1. Phylogenetic analyses of PODs from cassava. A total of 91 PODs from cassava were used to create the neighbor-joining (NJ) tree with 1000 bootstraps.

2.3. Conserved Motif and Gene Structure Analysis of POD Families in Cassava

The structural features of MePODs were investigated by identifying 10 conserved motifs using the MEME database in accordance with the phylogenetic relationship. Then, the conserved motifs were submitted in their entirety to the InterProScan database for annotation. Eight domains (domains 1, 2, 3, 4, 5, 6, 7, and 9) were noted as POD protein motifs, which are an essential feature of the peroxidase

family. On the basis of the motif analyses, 83 MePODs were assigned to one of five subgroups (A–E). Each of these 83 MePODs contains at least nine POD motifs, except for MePOD57 (in subgroup D) and MePOD69 (in subgroup B), which have seven and five motifs, respectively. The presence of these motifs suggests that the identified proteins are characteristic of the POD family (Figure 2). Subgroup F is distinct from the others: its members contain domains 2, 4, 5, 7, and 8. These results indicate that the proteins assigned to the same subfamilies share similar POD motif characteristics, further supporting their phylogenetic classification as PODs in cassava.

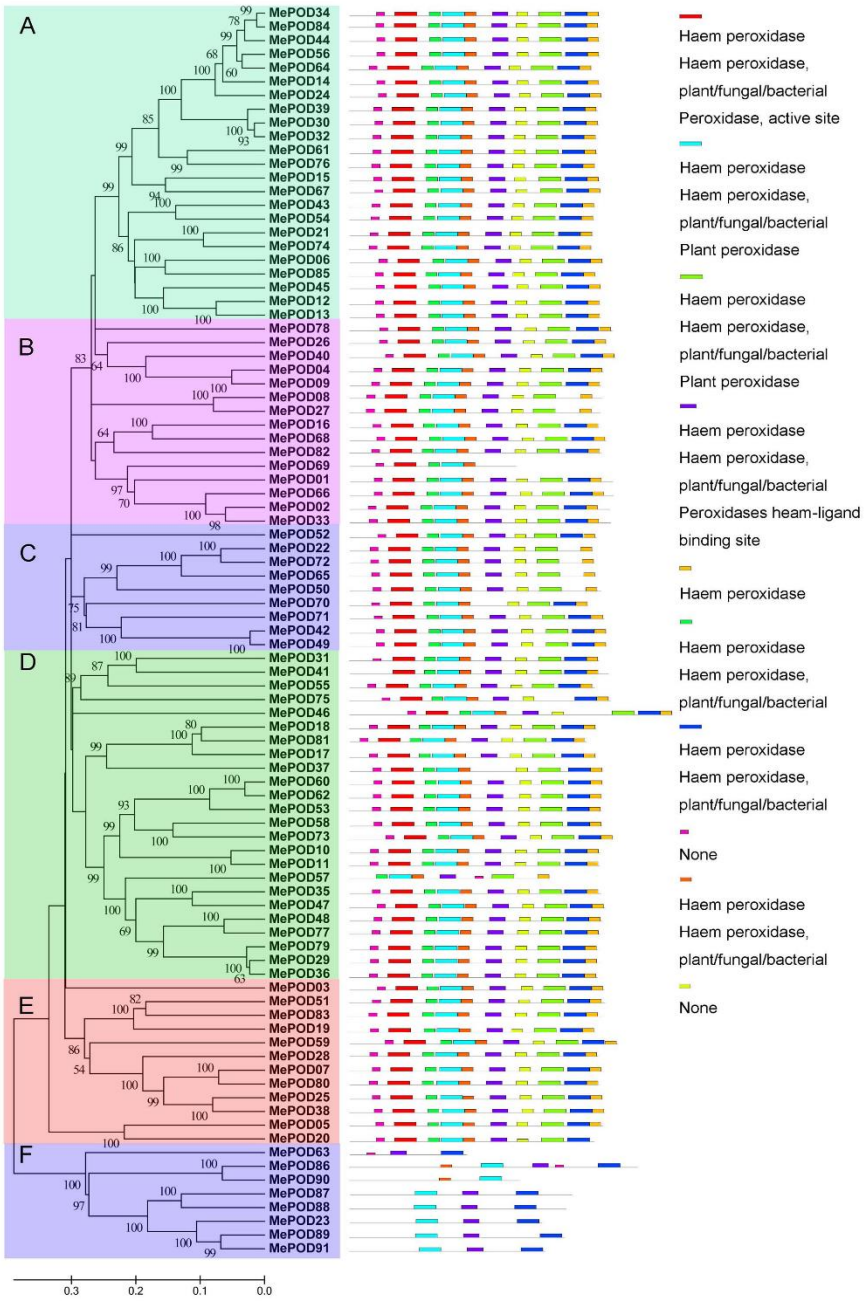


Figure 2. The motif analyses of POD family members in cassava according to their evolutionary relationship. The POD motifs were identified by the MEME database. The 10 different colors of the boxes on the right represent diverse conserved motifs, while the gray lines indicate non-conserved sequences.

Next, the exon–intron structures of cassava *POD* genes were analyzed. Subgroup F is exon-rich, with five to nine exons, whereas other subfamilies have fewer (between one and four) exons, except for

MePOD02 (in subgroup B), which has five exons (Figure 3). High proportions of *POD* genes contain four exons in subgroups A, B, C, and D, with four-exon *POD* genes forming 84%, 67%, 57%, and 76% of the genes in these subgroups, respectively, whereas only 50% of the subgroup E genes have four exons. Generally, *POD* genes in the same subgroup show similar exon–intron features, providing further evidence of their phylogenetic relationship.

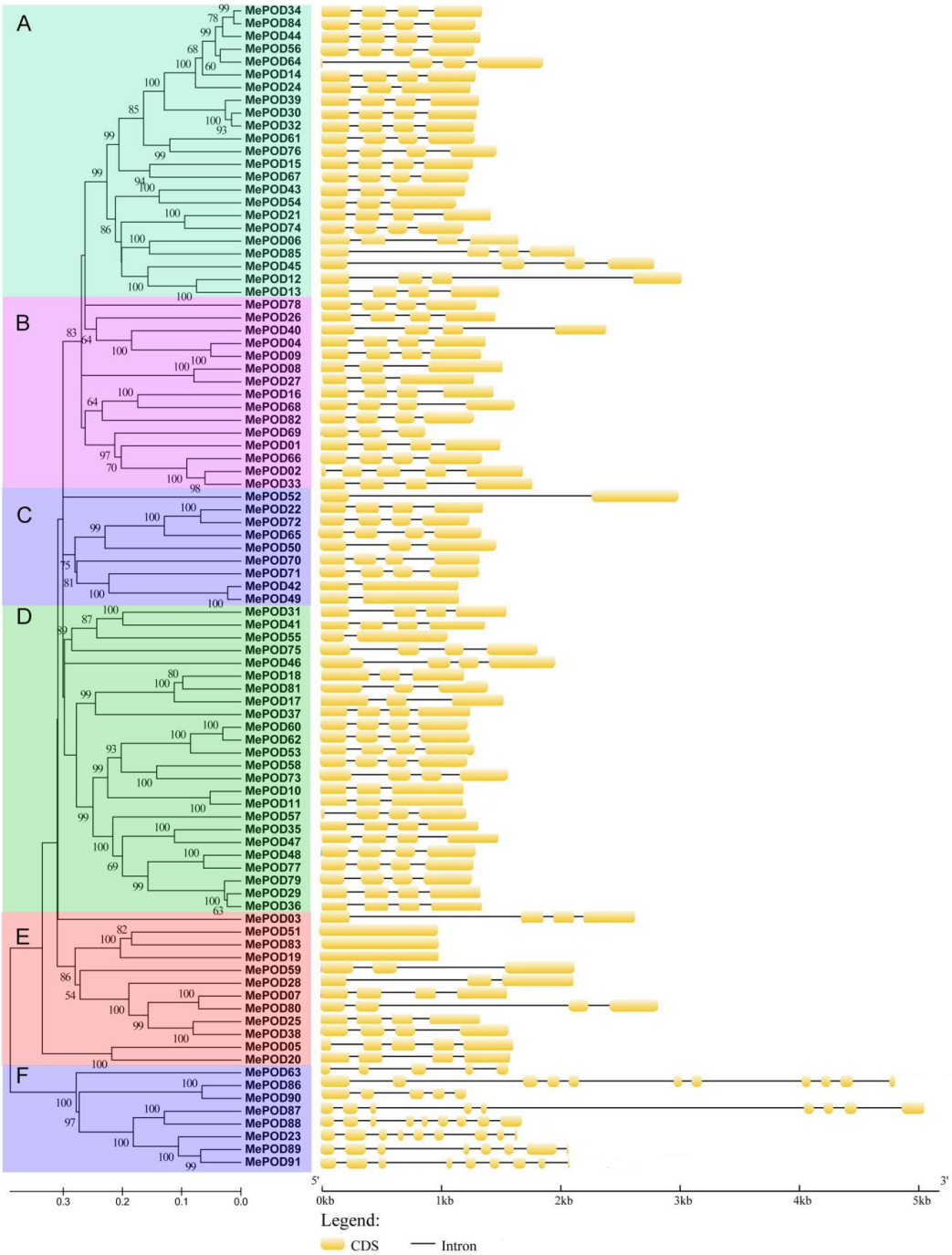


Figure 3. The exon–intron organization analyses of cassava PODs on the basis of the phylogenetic relationship. The exon–intron distribution was established using the GSDS database. The yellow boxes and the black lines represent exons and introns, respectively.

2.4. Analyses of Chromosomal Distribution and Duplication Events of the Cassava POD Genes

The locations of the cassava *POD* genes were determined by analyzing their chromosomal distribution (Figure 4). The 91 MePODs were mapped to chr1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, and 18, and scaffold01119. The 24 *POD* genes in subgroup D were distributed among chr1, 2, 3, 5, 6, 8, 11, 12, 13, 15, 16, and 17, making subgroup D the most widely distributed subgroup. Subgroup F contains eight *POD* genes, which are located on chr1, 2, 4, 8, 11, 16, and 18; thus, of the six subgroups identified, subgroup F is dispersed among the fewest chromosomes. The 23 members of subgroup A are found on chr1, 2, 7, 8, 9, 10, 15, 17, and 18, among which chromosomes 7, 9, and 10 only contain subgroup A genes. Generally, the cassava *POD* genes are widely distributed among chromosomes.

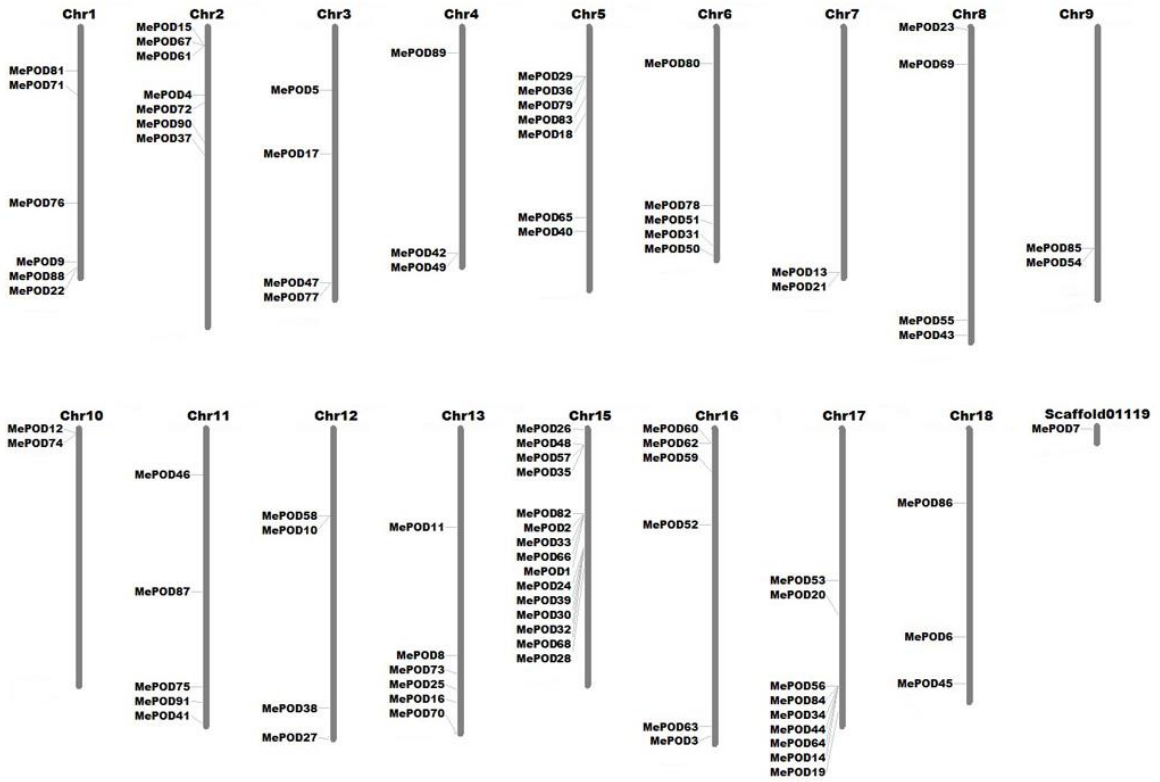


Figure 4. Chromosome distribution analyses of *POD* gene subgroups in cassava. The chromosomal information of 91 MePODs was collected from the Phytozome 12.0 cassava database, and the genes were then mapped to 17 chromosomes and one scaffold. MapInspect software (Wageningen University, Wageningen, Netherlands) was used to draw this figure.

To further investigate the expansion of *POD* genes in cassava, we aligned the total nucleotide sequences of the 91 *MePOD* genes to identify duplication events. We identified 15 events involving 16 paralogs (*MePOD2/MePOD33*, *MePOD29/36/79*, *MePOD30/32/39*, *MePOD34/44/56/84*, *MePOD42/MePOD49*, *MePOD60/MePOD62*), suggesting that tandem duplication played a significant role in POD family expansion in the cassava genome (Figure 5).

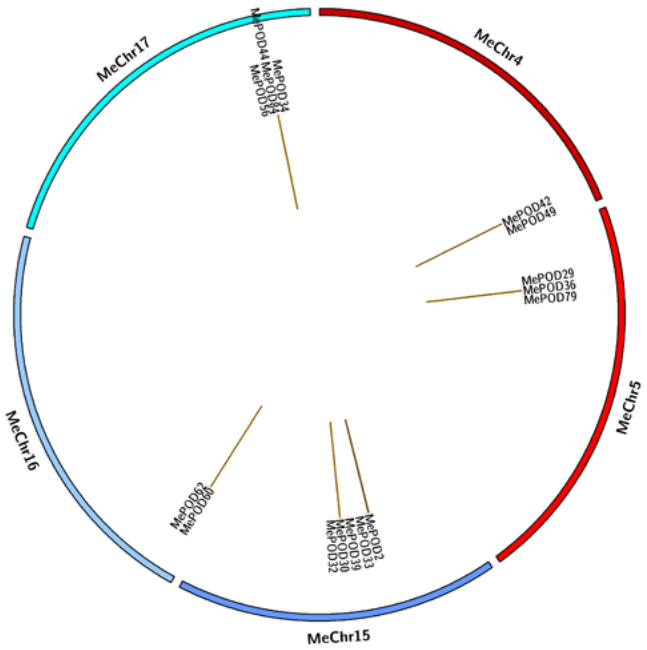


Figure 5. Analyses of *POD* gene duplication in cassava. The program Circos (Canada's Michael Smith Genome Sciences Center, Vancouver, Canada) was used to draw different chromosomes in a circular distribution. The brown connection lines represent tandem duplication events of *POD* genes in cassava.

Next, we calculated nonsynonymous (Ka) and synonymous (Ks) ratios to understand the modes of evolutionary selection for the duplicated *MePOD* genes. We found that the Ka/Ks ratios of the paralogous genes are between 0.05 and 0.29, indicating that these genes underwent purifying selection during evolution (Table S2).

2.5. Expression Profiles of POD Genes in Different Tissues of Two Cassava Genotypes

The expression levels of *MePOD* genes in different tissues were investigated by performing RNA-Seq analysis on the storage roots, stems, and leaves of a cultivated variety (Arg7) and wild subspecies (W14). The resulting expression data covered 59 and 56 *MePOD* genes in the transcriptome dataset of Arg7 and W14, respectively (Figure 6A; Table S3). Of these genes in Arg7, 15 (25%), 9 (15%), and 8 (14%) *MePODs* had high transcriptional levels (log2-based > 4) in stems, leaves, and storage roots, respectively. The number of *MePODs* with high expression (log2-based > 4) in the stems, leaves, and storage roots of W14 was 9 (16%), 7 (13%), and 10 (18%), respectively. Notably, *MePOD5* in subgroup E and *MePOD89* in subgroup F were strongly expressed (log2-based fold change > 4) in the three diverse tissues of Arg7 and W14. These *POD* genes may play a molecular role in the development and function of different cassava tissues.

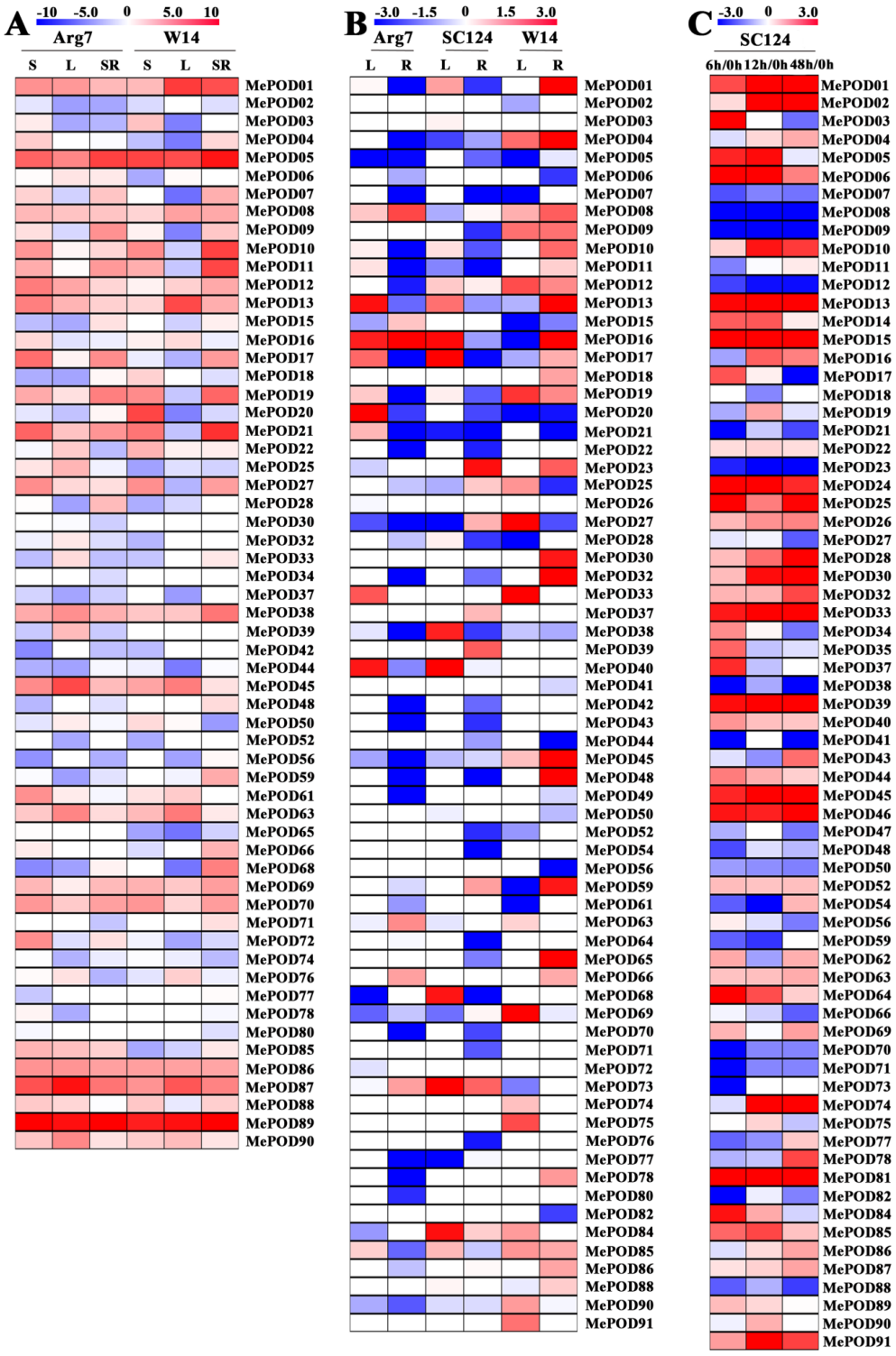


Figure 6. Transcriptomic analysis of cassava *POD* genes. (**A**) Expression of MePODs in the stem (S), leaf (L), and storage root (SR) of W14 and Arg7. The log2-based FPKM value was applied to build the heat map using Mev4.9.0 software (CCCB, Boston, USA). (**B**) Expression of MePODs in the leaf (L) and root (R) of Arg7, SC124, and W14 after drought treatment relative to under normal conditions. Log2-based fold changes (L/control; R/control) were applied to build the heat map using Mev4.9.0 software. (**C**) Expression of MePODs in the storage root at 6, 12, and 48 h relative to 0 h after harvest. Log2-based fold changes were applied to build the heat map using Mev4.9.0 software.

2.6. Expression Profiles of POD Genes After Drought Treatment

To study the possible role of MePODs in the cassava response to drought stress, water was withheld from a wild subspecies (W14) and two cultivated varieties (Arg7 and SC124) for 12 days. The leaves and roots of these samples were then collected to perform RNA-Seq. Of the transcriptome data, the expression data were obtained for 71 out of the 91 cassava *POD* genes (Figure 6B; Table S4). After drought treatment, 6 (8%) and 5 (7%) *MePOD* genes in Arg7 were transcriptionally upregulated (log2-based fold change > 1), whereas 7 (10%) and 29 (41%) were down-regulated (log2 based fold change < −1) in the leaves and roots, respectively. After SC124 was subjected to drought stress, 9 (13%) and 4 (6%) *MePOD* genes were upregulated (log2-based fold change > 1) but 6 (8%) and 29 (41%) were downregulated (log2-based fold change <−1) in the leaves and roots, respectively. After the W14 subspecies was exposed to drought, 13 (18%) and 21 (30%) *MePOD* genes were induced (log2-based fold change > 1), whereas 11 (15%) and 9 (13%) were depressed (log2-based fold change < −1) in the leaves and roots, respectively. *MePOD13* (subgroup A) and *MePOD16* (subgroup B) were upregulated (log2-based fold change > 1) by drought stress in the leaves of all three genotypes. The above data reveal that more *MePOD* genes were upregulated in response to drought treatment in W14 than in Arg7 and SC124.

2.7. Expression Profiles of Cassava PODs During PPD

Transcriptome analyses were conducted at different postharvest periods of the storage roots of SC124 to examine the possible function of *MePOD* genes during PPD (Figure 6C; Table S5). Transcriptional data were obtained for 71 of the 91 MePODs. Of these MePODs, 38 genes were induced (log2-based fold change > 1) at a minimum of one time point. Notably, *MePOD-1, -6, -13, -15, -24, -25, -33, -39, -45, -46, -81,* and *-91* were continuously upregulated (log2-based fold change > 1) at all points. Conversely, *MePOD-7, -8, -9, -12, -23, -38, -50, -70, -71,* and *-88* were downregulated (log2-based fold change < −1) at all tested times. These results suggest the possibility that MePODs play a role during PPD in cassava.

2.8. Expression Analysis of MePOD Genes in Response to Various Abiotic and Biotic Stresses and Related Signals

To test the transcription of *MePOD* genes upon exposure to methyl jasmonate (MeJA), salicylic acid (SA), abscisic acid (ABA), H_2O_2, salt, osmotic stress (by mannitol treatment), cold stress, and *Xanthomonas axonopodis* pv *manihotis* (*Xam*), nine genes (*MePOD-13, -16, -17, -19, -23, -68, -74, -85, -86*) that were induced by drought stress in at least two tissues in all three varieties were selected for quantitative real-time PCR (qRT-PCR) analysis (Figure 7; Table S6). We define upregulation as: log2-based fold change > 1. With MeJA treatment, the expression of eight genes (*MePOD-13, -17, -19, -23, -68, -74, -85,* and *-86)* was induced, with particularly high levels expressed 24 h after treatment. ABA treatment resulted in increased transcript levels of *MePOD-17* and *-85*. With SA treatment, the expression levels of *MePOD-13, -16, -19, -23, -68, -85,* and *-86* were amplified. H_2O_2 treatment led to the induction of *MePOD-13, -17, -23,* and *-85*. Salt treatment induced the expression of *MePOD17* after two and six hours, and three days of treatment, but the gene was repressed after 14 days of treatment. Under osmotic stress, *MePOD-13, -17, -19, -23, -85,* and *-86* were upregulated, among which *MePOD-85* was induced throughout the entire treatment time. In response to cold treatment, *MePOD17* was upregulated at 2 and 15 h. Exposure to the pathogen *Xam* led to the upregulated expression of six *POD* genes (*MePOD-13, -16, -17, -74, -85,* and *-86*) at a minimum of two time points. Together, these results demonstrate that the *POD* genes of cassava respond to multiple stresses and related signals (Figure 7).

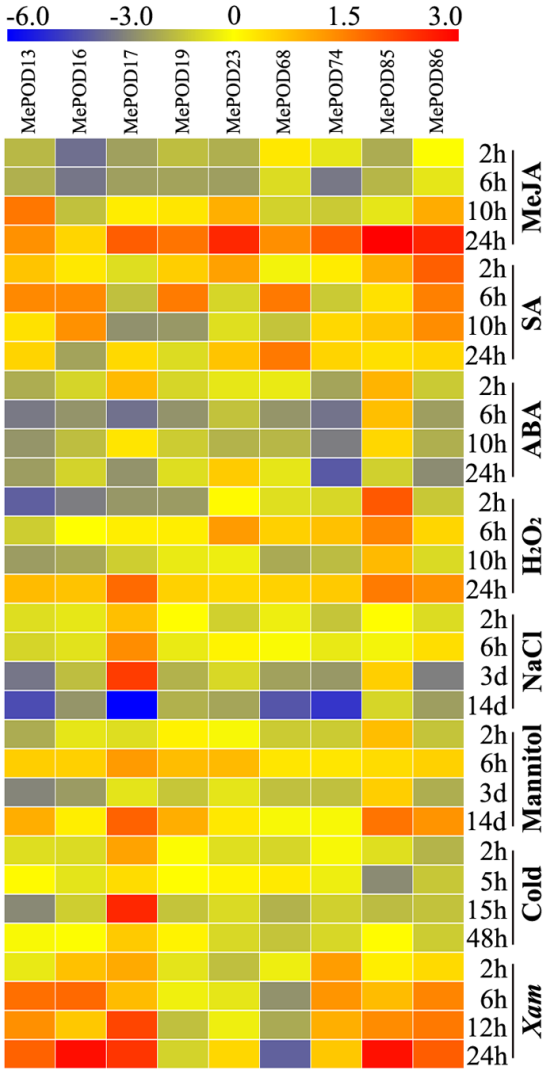


Figure 7. Expression profiles of cassava *POD* genes in the leaves of Arg7 after exposure to MeJA, SA, ABA, H_2O_2, salt, osmotic stress (mannitol treatment), cold stress, and *Xam*. Log2-based qRT-PCR fold changes were used to build the heat map with Mev4.9.0 software. The changes in color represent the relative gene expression level.

3. Discussion

Given the significant role of PODs in various physiological processes, including responses to biotic and abiotic stresses, it was necessary to scientifically investigate the potential functions of *POD* genes in cassava, which is an important crop. In this study, we identified 91 PODs in the cassava genome (Figure 1); thus, cassava has more POD members than Arabidopsis but fewer than rice, *Populus trichocarpa*, *Medicago sativa*, maize, and *Pyrus bretschneideri* [9,24–30]. We found that 92% (84/91) of the MePODs have a molecular mass in the range of 30 to 45 kDa, which is in accordance with previous studies [2,7]. Most of the *POD* genes (89/91) in cassava harbor more than one exon (Figure 3), which is similar to the proportion of single-exon *POD* genes in *Pyrus bretschneideri* and *Zea mays* (*PbPRX* (90/94) and *ZmPRX* (89/107), respectively) [29,30]. The similarities in gene structure and motif composition among the members in each MePOD subgroup support the phylogenetic classification presented here.

The expansion of a gene family primarily occurs via three kinds of modes: segmental duplication of multiple genes, tandem duplication of individual genes, and whole-genome duplication [36,37]. To analyze the duplication modes of the *POD* genes in cassava, we first identified the chromosomal

locations of the *MePOD* genes. Chromosomal mapping revealed that these genes are widely distributed among 17 chromosomes and one scaffold (cassava has 18 chromosomes in total) (Figure 4), which is in accordance with the wide chromosomal distribution of PODs in Arabidopsis, rice, *Populus trichocarpa*, maize, and *Pyrus bretschneideri* [9,24,27,29,30]. Secondly, 16 paralogous *POD* genes were characterized in the cassava genome, indicating that tandem duplication contributed to MePOD expansion. Accumulated evidence has demonstrated that duplication events have been important for gene expansion in the POD family. A total of 37 PRX genes in *Populus* and 24 *POD* genes in maize were identified as tandem duplications, further supporting that tandem duplication has been a significant means of *POD* gene expansion [27,29]. Almost all these paralogous MePODs had low or no expression after drought treatment, but 63% (10 out of 16) from the postharvest transcriptome were expressed (Table S6), of which *MePOD-2, -30, -32, -33, -39,* and *-44* were significantly upregulated and *MePOD-34, -56,* and *-62* were repressed at some time point during the PPD process (Figure 6C). These results indicate that most of the *MePOD* genes resulting from tandem duplication-driven expansion are involved in the PPD process of cassava storage roots. Paralogous PRX genes were also found to be involved in other biological processes. In maize, paralogous genes *ZmPRX-26, -42,* and *-75* were induced after NaCl, PEG, SA, or H_2O_2 treatment [29]. In Chinese pear, the expression of paralogous genes *PbPRX-42* and *-64* increased during fruit development [30].

The POD family is positively related to the reduced production of hydrogen peroxide and the decreased formation of reactive oxygen species, and the suppression of these species increases plant resistance to stresses [4,5,11,12,19,20]. In this study, the total number of *POD* genes responding to drought (log2-based fold change > 1) was greater in both the roots and leaves of W14 than that in Arg7 and SC124, suggesting the comprehensive activation of PODs in response to drought in W14 (Figure 6B). The wild ancestor W14 has been previously confirmed to be more resistant to drought than the two cultivars SC124 and Arg7 [38,39]. Accumulated evidence suggests that the overexpression of *POD* genes results in increased plant tolerance to drought and osmotic stresses [19,20,40,41]. The activity of POD enzyme was significantly enhanced under drought stress [42]. Consequently, we conclude that the high ratio of MePODs induced by drought in W14 might contribute to its strong drought tolerance.

Previous studies have suggested that ROS production results in the deterioration process in cassava during the postharvest period, and a reduction in ROS accumulation could delay the PPD process [35,43]. The POD family mainly participates in the peroxidative cycle and hydroxylic cycle, resulting in the reduced production of H_2O_2 and the decreased formation of ROS [4,5,11,12]. Some PRXs have been shown to change in expression during the fruit storage process [44,45]. The activity of POD enzyme significantly increased during cassava PPD process, suggesting their possible role during the postharvest period of cassava [35,44]. In this study, we found that 78% (71 out of 91) of PODs (log2-based fold change > 1) were upregulated in the storage roots of SC124 (Figure 6C). Interestingly, 13% (12 out of 91) of PODs (log2-based fold change > 1) were induced at all points. Collectively, these results indicate that *MePOD* genes are involved in the PPD process in cassava storage roots.

Previous research has indicated that PODs can extensively participate in plants' responsse to biotic and abiotic stresses [18–23]. Here, we selected nine genes (*MePOD-13*, *-16*, *-17*, *-19*, *-23*, *-68*, *-74*, *-85*, and *-86*) to further examine their expression levels after various treatments (Figure 7; Table S6). These genes are located on different regions of chr7, 13, 3, 17, 8, 15, 10, 9, and 18, respectively (Figure 4). Phylogenetic analysis indicates that *MePOD-16*, *-68*, *-74*, and *-85* belong to subgroup A; *MePOD-17* belongs to subgroup D; and *MePOD-13*, *-19*, *-23*, and *-86* belong to subgroup E (Figure 1). The results show that all nine of the analyzed MePODs were upregulated in response to at least two types of treatments. *MePOD17* and *MePOD85* (log2-based fold change > 1) were induced by six treatments (MeJA, salt, cold stress, osmotic stress, ABA, and *Xam* and MeJA, osmotic stress, SA, ABA, H_2O_2, and *Xam*, respectively); *MePOD13* was upregulated by five treatments (MeJA, osmotic stress, ABA, SA, H_2O_2, and *Xam*); and *MePOD23* and *MePOD86* were upregulated by four treatments (MeJA, osmotic stress, SA, and H_2O_2 and MeJA, SA, H_2O_2, and *Xam*, respectively). Of these, *MePOD13* and

MePOD23 were induced after H_2O_2 treatment in cassava leaves (Figure 7) but exhibited the opposite trend of expression during the PPD process in storage roots (Figure 6C), suggesting their differential roles in diverse tissues. *MePOD-13, -19, -23, -68,* and *-86* (belonging to subgroup E, except for *MePOD68*) were upregulated by MeJA and SA treatments but downregulated by ABA treatment (Figures 1 and 7). The expression of some PODs has been induced by MeJA and SA treatments in other plant species [2,46]. The opposite direction of expression of these *POD* genes between MeJA and SA treatments and ABA treatment may be due to the antagonism between MeJA/SA and ABA [47,48]. Whereas ABA plays a prominent role in plants' tolerance to drought stress [38], MeJA- and SA-mediated signaling pathways are also activated under drought stress [49,50]. The induction of these genes by MeJA, SA, and drought suggests their possible involvement in MeJA- and SA-mediated drought responses in cassava. The responses of *POD* genes to multiple treatments have been observed in other plants. In Arabidopsis, *AtPrx33* and *AtPrx34* were upregulated after H_2O_2 and flg22 treatments [51]. In maize, *ZmPRX-26*, *-42*, and *-71* were induced by H_2O_2, salt, and PEG treatments [29]. Phylogenetic analysis of MePODs with AtPrx-33 and -34 and ZmPRX-26, -42, and -71 found that MePOD86 shares a close phylogenetic relationship with ZmPRX71 (Figure S1), suggesting their functional conservation in multiple treatments. Multiple stresses, such as cold, salt, or PEG, induced the activity of POD enzyme, demonstrating the response of *POD* genes to environmental stress [52–54]. These results suggest that MePODs participate in the response to multiple stresses or related signals and are candidate targets for the genetic improvement of cassava.

4. Materials and Methods

4.1. Plant Materials and Treatments

Three cassava genotypes, W14, SC124, and Arg7, were planted in the greenhouse of the Chinese Academy of Tropical Agricultural Science (Haikou, China). Their characteristics were described in our previous studies [38,55]. Stem segments containing three nodes were cut from eight-month-old cassava plants and planted in pots, as described in Hu's study [51]. The transcripts of W14 and Arg7 *MePOD* genes in stems and leaves after being planted for 90 days and middle storage roots after being planted for 150 days were examined by RNA-Seq. After W14, SC124, and Arg7 were cultured for 90 days, they were subjected to drought stress by withholding water for 12 days, after which their leaves and roots were sampled to study the transcriptional responses by RNA-Seq. To examine the expression profiles of *MePOD* genes after the plants were exposed to stress and related signaling treatments, the 60-day-old Arg7 variety was treated with 100 μM MeJA for 0, 2, 6, 10, and 24 h; 300 mM NaCl for 0 h, 2 h, 6 h, 3 days, and 14 days; a low temperature (4 °C) for 0, 2, 5, 15, and 48 h; 100 μM SA for 0, 2, 6, 10, and 24 h; 200 mM mannitol (to induce osmotic stress) for 0 h, 2 h, 6 h, 3 days, and 14 days; 100 μM ABA for 0, 2, 6, 10, and 24 h; 10% H_2O_2 for 0, 2, 6, 10, and 24 h; or *Xam* for 0, 2, 6, 12, and 24 h. Ten-month-old cassava storage roots (CSR) were cut into 5-mm-thick slices and placed into Petri dishes containing wet filter paper for 0, 6, 12, and 48 h to study the expression changes in *MePOD* genes via RNA-Seq during CSR deterioration. All samples were frozen immediately in liquid nitrogen and stored at −80 °C for RNA-Seq and qRT-PCR.

4.2. Identification and Phylogenetic Analysis of PODs in Cassava

MePOD genes were identified in cassava on the basis of homology with 73 POD protein sequences from the *Arabidopsis* genome database (available online: http://www.arabidopsis.org/index.jsp) and 138 POD protein sequences from the rice genome database (available online: http://rice.plantbiology.msu.edu/index.shtml) [56,57]. The Hidden Markov Model-based search (HMMER: http://hmmer.wustl.edu/) profile of these confirmed POD proteins was constructed to search the cassava genome hub (available online: http://www.phytozome.net/cassava.php) [58]. Finally, all predicted POD protein sequences were further examined by CDD (available online: http://www.ncbi.nlm.nih.gov/cdd/) and PFAM (available online: http://pfam.sanger.ac.uk/) after being checked by BLAST analyses [59,60].

All the predicted cassava POD genes identified from HMMER and BLAST were confirmed only if they included the POD special domains examined by SMART (available online: http://smart.embl-heidelberg.de/) [61]. Multiple sequence alignment of all predicted MePOD protein sequences was performed with Clustal W in BioEdit software [62]. The phylogenetic tree of the full-length MePOD protein sequences was created using MEGA 5.0 (University College Dublin, Dublin, Ireland) with the neighbor-joining (NJ) method, and bootstrap analysis was conducted with 1000 replicates [63].

4.3. Protein Properties and Structure Analyses of PODs in Cassava

The ProtParam database (available online: http://web.expasy.org/protparam/) was used to predict the properties, including amino acid numbers, molecular weights (MW), and isoelectric points (pI), of all presumed POD proteins [64]. The motifs were analyzed using the MEME program (available online: http://meme-suite.org/tools/meme), in which the maximum number of motifs was set to 10, the optimum width of motifs was set to 15–50 amino acid residues, and the other settings were kept at default values [65]. Subsequently, these 10 motifs were annotated in InterProScan (available online: http://www.ebi.ac.uk/Tools/pfa/iprscan/) [66]. The gene structures of each MePOD were investigated using GSDS software (available online: http://gsds.cbi.pku.edu.cn/) using each MePOD's genomic DNA sequence and its corresponding CDS sequence, which were retrieved from the cassava genome database [67].

4.4. Chromosomal Location and Duplication Pattern Analyses

According to the results of BLASTN in the Phytozome 12.0 cassava database, *MePOD* genes were mapped to different chromosomes. On the basis of the calculated value of nucleotide sequence similarity and the phylogenetic relationship of cassava *POD* genes, paralogous genes were identified. The gene duplication pattern of paralogous *MePOD* genes was determined by the following two criteria: (1) the identity of the aligned region was >90% and (2) the alignment covered >90% of the longer gene. Circos software (Canada's Michael Smith Genome Sciences Center, Vancouver, Canada) was used to draw the duplication events of *MePOD* genes [68,69]. The values of nonsynonymous substitution (Ka) and synonymous substitution (Ks) were calculated suing DnaSP 5.0 software [70]. Ka/Ks rate > 1 indicates positive evolution, Ka/Ks rate = 1 indicates neutral evolution, and Ka/Ks rate < 1 indicates negative evolution [71].

4.5. Transcriptome Analyses of PODs in Cassava

RNA-Seq was used to determine the expression of cassava *MePOD* genes. Total RNA was isolated from frozen stems, leaves, roots, and storage roots using the plant RNeasy extraction kit (TIANGEN, Beijing, China) and quantified with a NanoDrop 2000c (Thermo Scientific Inc., Waltham, MA, USA). Total RNA (3 μg) of each sample was used to construct the cDNA library according to the Illumina instructions and then sequenced using an Illumina GA II (Illumina Inc., San Diego, USA). The original data processing and analysis methods were described in our previous study [72].

4.6. Quantitative Real-Time PCR Analyses

Leaf samples of the Arg7 variety subjected to MeJA, SA, ABA, NaCl, low temperature, mannitol, H_2O_2, or *Xam* treatments were collected to perform qRT-PCR analysis. Total RNA (1 μg) from each sample was used to synthesize the first-strand cDNA using oligo-dT primer by SuperScript reverse transcriptase (Takara, Dalian, China). The cDNA product was diluted to 50 ng·μL^{-1}, and 1 μL was used for qRT-PCR. The qRT-PCR reaction mixtures (20 μL) contained 0.6 μL of each gene-specific primer (300 nmol·μL^{-1}), 10 μL of 2× FastFire qPCR PreMix (Tiangen, Beijing, China), and 7.8 μL of RNase-free water. The qRT-qPCR thermal cycling included cDNA denaturation at 95 °C for 1 min, with 40 cycles of 95 °C for 5 s and 60 °C for 15 s in the Mx3005P Real-Time PCR System (Agilent Inc., Palo Alto, CA, USA) with the SYBR green method. The β-tubulin gene of cassava was chosen as an internal control. All qRT-qPCR experiments were performed in triplicate, and the gene-specific primers

used in expression analysis are listed in Table S7. The data obtained from the qRT-qPCR were analyzed with Tukey's post-hoc ANOVA in SPSS 22.0 (SPSS Inc., Chicago, USA) ($P < 0.05$) after fold treatment with the $2^{-\triangle\triangle Ct}$ method.

5. Conclusions

In this study, we identified 91 PODs from the cassava genome and studied their basic classification, protein motif, gene structure, chromosomal distribution, and duplication pattern. Comprehensive transcriptional level analyses revealed the involvement of MePODs in biotic and abiotic stress responses, hormone responses, and storage root deterioration. Several *MePOD* genes (*MePOD-13*, *-17*, *-85*, and *-86*) were found to be transcriptionally upregulated after multiple different treatments, suggesting that these genes are good candidates to target for cassava improvement. These findings increase our understanding of POD-mediated stress and hormone responses and storage root deterioration in cassava, laying a foundation for the genetic improvement of cassava.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/11/2730/s1. Table S1. Characteristics of PODs in cassava. Table S2. The Ka/Ks ratios of duplicated *POD* genes in cassava. Table S3. The expression profiles (log2-based values) of the cassava *POD* genes in different tissues. Table S4. The expression profiles (log2-based fold changes) of the cassava *POD* genes after drought treatment. Table S5. The expression profiles (log2-based fold changes) of the cassava *POD* genes after harvest. Table S6. The expression profiles (log2-based values) of the cassava *POD* genes after various stress treatments. Table S7. Primers used in qRT-PCR analysis. Figure S1. Phylogenetic analyses of 91 MePODs, 2 AtPRXs, 3 OsPRXs, and 3 ZmPRXs.

Author Contributions: W.H., X.D., and H.Y. designed the research; C.W., X.D., Z.D., and Y.Y. performed the research; X.D., W.T., C.W., and Y.W. analyzed the data; X.D. and C.W. wrote the paper; W.H., Z.D., and H.Y. reviewed the paper. All authors have read and approved the final manuscript.

Funding: This research was funded by the Natural Science Foundation of Hainan Province (318MS092), the National Natural Science Foundation of China (31771859), the Central Public-Interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630052016005, 1630052016006, 1630052017021, 1630012019009, 1630052019023), the Central Public-Interest Scientific Institution Basal Research Fund for Innovative Research Team Program of CATAS (17CXTD-28, 1630052017017), and the earmarked fund for Modern Agro-industry Technology Research System (CARS-11).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Passardi, F.; Cosio, C.; Penel, C.; Dunand, C. Peroxidases have more functions than a Swiss army knife. *Plant Cell Rep.* **2005**, *24*, 255–265. [CrossRef]
2. Almagro, L.; Gómez Ros, L.V.; Belchi-Navarro, S.; Bru, R.; Ros Barceló, A.; Pedreno, M.A. Class III peroxidases in plant defence reactions. *J. Exp. Bot.* **2009**, *60*, 377–390. [CrossRef]
3. Welinder, K.G. Superfamily of plant, fungal and bacterial peroxidases. *Curr. Opin. Struct. Biol.* **1992**, *2*, 388–393. [CrossRef]
4. Mathé, C.; Barre, A.; Jourda, C.; Dunand, C. Evolution and expression of class III peroxidases. *Arch. Biochem. Biophys.* **2010**, *500*, 58–65. [CrossRef]
5. Hiraga, S.; Sasaki, K.; Ito, H.; Ohashi, Y.; Matsui, H. A large family of Class III plant peroxidases. *Plant Cell Physiol.* **2001**, *42*, 462–468. [CrossRef]
6. Zheng, X.H.; Van Huystee, R.B. Oxidation of tyrosine by peroxidase isozymes derived from peanut suspension culture medium and by isolated cell walls. *Plant Cell Tiss. Org.* **1991**, *25*, 35–43. [CrossRef]
7. Barceló, A.R.; Ros, L.V.; Carrasco, A.E. Looking for syringyl peroxidases. *Trends Plant Sci.* **2007**, *12*, 486–491. [CrossRef]
8. Henriksen, A.; Mirza, O.; Indiani, C.; Teilum, K.; Smulevich, G.; Welinder, K.G.; Gajhede, M. Structure of soybean seed coat peroxidase: A plant peroxidase with unusual stability and haem-apoprotein interactions. *Protein Sci.* **2001**, *10*, 108–115. [CrossRef]
9. Passardi, F.; Longet, D.; Penel, C.; Dunand, C. The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* **2004**, *65*, 1879–1893. [CrossRef]

10. Cosio, C.; Ranocha, P.; Francoz, E.; Burlat, V.; Zheng, Y.; Perry, S.E.; Ripoll, J.J.; Yanofsky, M.; Dunand, C. The class III peroxidase PRX17 is a direct target of the MADS-box transcription factor AGAMOUS-LIKE15 (AGL15) and participates in lignified tissue formation. *New Phytol.* **2016**, *213*, 250–263. [CrossRef]
11. Passardi, F.; Penel, C.; Dunand, C. Performing the paradoxical: How plant peroxidases modify the cell wall. *Trends Plant Sci.* **2004**, *9*, 534–540. [CrossRef]
12. Ostergaard, L.; Teilum, K.; Mirza, O.; Mattsson, O.; Petersen, M.; Welinder, K.G.; Mundy, J.; Gajhede, M.; Henriksen, A. Arabidopsis ATP A2 peroxidase. Expression and high-resolution structure of a plant peroxidase with implications for lignification. *Plant Mol. Biol.* **2000**, *44*, 231–243. [CrossRef]
13. Hiraga, S.; Yamamoto, K.; Ito, H.; Sasaki, K.; Matsui, H.; Honma, M.; Nagamura, Y.; Sasaki, T.; Ohashi, Y. Diverse expression proćles of 21 rice peroxidase genes. *Febs Lett.* **2000**, *471*, 245–250. [CrossRef]
14. Chen, D.; Ding, Y.; Guo, W.; Zhang, T. Molecular cloning and characterization of a flower-specific class III peroxidase gene in *G. Hirsutum*. *Mol. Biol. Rep.* **2009**, *36*, 461–469. [CrossRef]
15. González, A.M.; Marcel, T.C.; Kohutova, Z.; Stam, P.; van der Linden, C.G.; Niks, R.E. Peroxidase profiling reveals genetic linkage between peroxidase gene clusters and basal host and non-host esistance to rusts and mildew in barley. *PLoS ONE* **2010**, *5*, e10495. [CrossRef]
16. Jin, J.; Hewezi, T.; Baum, T.J. Arabidopsis peroxidase AtPRX53 influences cell elongation and susceptibility to Heterodera schachtii. *Plant Signal. Behav.* **2011**, *6*, 1778–1786. [CrossRef]
17. Shigeto, J.; Tsutsumi, Y. Diverse functions and reactions of class III peroxidases. *New Phytol.* **2016**, *209*, 1395–1402. [CrossRef]
18. Wu, Y.; Yang, Z.; How, J.; Xu, H.; Chen, L.; Li, K. Overexpression of a peroxidase gene (*AtPrx64*) of *Arabidopsis thaliana* in tobacco improves plant's tolerance to aluminum stress. *Plant Mol. Biol.* **2017**, *95*, 157–168. [CrossRef]
19. Llorente, F.; López-Cobollo, R.M.; Catalá, R.; Martínez-Zapater, J.M.; Salinas, J. A novel cold-inducible gene from Arabidopsis, RCI3, encodes a peroxidase that constitutes a component for stress tolerance. *Plant J.* **2002**, *32*, 13–24. [CrossRef]
20. Kumar, S.; Jaggi, M.; Sinha, A.K. Ectopic overexpression of vacuolar and apoplastic *Catharanthus roseus* peroxidases confers differential tolerance to salt and dehydration stress in transgenic tobacco. *Protoplasma* **2012**, *249*, 423–432. [CrossRef]
21. Coego, A.; Ramirez, V.; Ellul, P.; Mayda, E.; Vera, P. The H_2O_2-regulated *Ep5C* gene encodes a peroxidase required for bacterial speck susceptibility in tomato. *Plant J.* **2005**, *42*, 283–293. [CrossRef] [PubMed]
22. Choi, H.W.; Kim, Y.J.; Lee, S.C.; Hong, J.K.; Hwang, B.K. Hydrogen peroxide generation by the pepper extracellular peroxidase CaPO2 activates local and systemic cell death and defense response to bacterial pathogens. *Plant Physiol.* **2007**, *145*, 890–904. [CrossRef] [PubMed]
23. Wally, O.; Punja, Z.K. Enhanced disease resistance in transgenic carrot (*Daucus carota L.*) plants over-expressing a rice cationic peroxidase. *Planta* **2010**, *232*, 1229–1239. [CrossRef]
24. Tognolli, M.; Penel, C.; Greppin, H.; Simon, P. Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene* **2002**, *288*, 129–138. [CrossRef]
25. Valério, L.; De Meyer, M.; Penel, C.; Dunand, C. Expression analysis of the Arabidopsis peroxidase multigenic family. *Phytochemistry* **2004**, *65*, 1331–1342. [CrossRef]
26. Duroux, L.; Welinder, K.G. The peroxidase gene family in plants: A phylogenetic overview. *J. Mol. Evol.* **2003**, *57*, 397–407. [CrossRef]
27. Ren, L.L.; Liu, Y.J.; Liu, H.J.; Qian, T.T.; Qi, L.W.; Wang, X.R.; Zeng, Q.Y. Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of populus Class III peroxidase family. *Plant Cell* **2014**, *26*, 2404–2419. [CrossRef]
28. Behr, M.; Legay, S.; Hausman, J.F.; Guerriero, G. Analysis of cell wall-related genes in organs of *Medicago sativa* L. under different abiotic stresses. *Int. J. Mol. Sci.* **2015**, *16*, 16104–16124. [CrossRef]
29. Wang, Y.; Wang, Q.; Zhao, Y.; Han, G.; Zhu, S. Systematic analysis of maize class III peroxidase gene family reveals a conserved subgroup involved in abiotic stress response. *Gene* **2015**, *566*, 95–108. [CrossRef]
30. Cao, Y.; Han, Y.; Meng, D.; Li, D.; Jin, Q.; Lin, Y.; Cai, Y. Structural, evolutionary, and functional analysis of the Class III peroxidase gene family in Chinese pear (*Pyrus bretschneideri*). *Front Plant Sci.* **2016**, *7*, 1874. [CrossRef]
31. Oliveira, E.J.; Santana, F.A.; Oliveira, L.A.; Santos, V.S. Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. *Genet Mol. Res.* **2014**, *13*, 6683–6700. [CrossRef]

32. International Cassava Genetic Map Consortium. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta Crantz*) from 10 populations. *G3 (BethesdaMd.)* **2014**, *5*, 133–144. [CrossRef]
33. Maran, J.P.; Sivakumar, V.; Thirugnanasambandham, K.; Sridhar, R. Degradation behavior of biocomposites based on cassava starch buried under indoor soil conditions. *Carbohydr. Polym.* **2014**, *30*, 20–28. [CrossRef]
34. Vanderschuren, H.; Nyaboga, E.; Poon, J.S.; Baerenfaller, K.; Grossmann, J.; Hirsch-Hoffmann, M.; Kirchgessner, N.; Nanni, P.; Gruissem, W. Large-Scale proteomics of the cassava storage root and identification of a target gene to reduce postharvest deterioration. *Plant Cell* **2014**, *26*, 1913–1924. [CrossRef]
35. Hu, W.; Kong, H.; Guo, Y.; Zhang, Y.; Ding, Z.; Tie, W.; Yan, Y.; Huang, Q.; Peng, M.; Shi, H.; et al. Comparative physiological and transcriptomic analyses reveal the actions of melatonin in the delay of postharvest physiological deterioration of cassava. *Front Plant Sci.* **2016**, *7*, 736. [CrossRef]
36. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu Rev. Plant Biol.* **2009**, *60*, 433–453. [CrossRef]
37. Cao, J.; Shi, F. Evolution of the RALF gene family in plants: Gene duplication and selection patterns. *Evol. Bioinf.* **2012**, *8*, 271–292. [CrossRef]
38. Hu, W.; Wei, Y.; Xia, Z.; Yan, Y.; Hou, X.; Zou, M.; Lu, C.; Wang, W.; Peng, M. Genome-wide identification and expression analysis of the NAC transcription factor family in cassava. *PLoS ONE* **2015**, *10*, e0136993. [CrossRef]
39. Shang, S.; Wu, C.; Huang, C.; Tie, W.; Yan, Y.; Ding, Z.; Xia, Z.; Wang, W.; Peng, M.; Tian, L.; et al. Genome-wide analysis of the GRF family reveals their involvement in abiotic stress response in cassava. *Genes* **2018**, *9*, 110. [CrossRef]
40. Guo, X.H.; Jiang, J.; Wang, B.C.; Li, H.Y.; Wang, Y.C.; Yang, C.P.; Liu, G.F. ThPOD3, a truncated polypeptide from *Tamarix hispida*, conferred drought tolerance in *Escherichia coli*. *Mol. Biol. Rep.* **2010**, *37*, 1183–1190. [CrossRef]
41. Choi, H.W.; Hwang, B.K. The pepper extracellular peroxidase CaPO2 is required for salt, drought and oxidative stress tolerance as well as resistance to fungal pathogens. *Planta* **2012**, *235*, 1369–1382. [CrossRef]
42. Hu, W.; Huang, C.; Deng, X.; Zhou, S.; Chen, L.; Li, Y.; Wang, C.; Ma, Z.; Yuan, Q.; Wang, Y.; et al. *Ta ASR1*, a transcription factor gene in wheat, confers drought stress tolerance in transgenic tobacco. *Plant Cell Env.* **2013**, *36*, 1449–1464. [CrossRef]
43. Iyer, S.; Mattinson, D.S.; Fellman, J.K. Study of the early events leading to cassava root postharvest deterioration. *Trop. Plant Biol.* **2010**, *3*, 151–165. [CrossRef]
44. Park, M.H. Sucrose delays senescence and preserves functional compounds in *Asparagus officinalis* L. *Biochem. Biophys. Res. Commun.* **2016**, *480*, 241–247. [CrossRef]
45. Xi, Y.; Jiao, W.; Cao, J.; Jiang, W. Effects of chlorogenic acid on capacity of free radicals scavenging and proteomic changes in postharvest fruit of nectarine. *PLoS ONE* **2017**, *12*, e0182494. [CrossRef]
46. Wang, J.E.; Liu, K.K.; Li, D.W.; Zhang, Y.L.; Zhao, Q.; Gong, Z.H. A novel peroxidase *CanPOD* gene of pepper is involved in defense responses to *Phytophtora capsici* infection as well as abiotic stress tolerance. *Int. J. Mol. Sci.* **2013**, *14*, 3158–3177. [CrossRef]
47. Moons, A. Antagonistic effects of Abscisic Acid and Jasmonates on salt stress-inducible transcripts in rice roots. *Plant Cell* **1997**, *9*, 2243–2259. [CrossRef]
48. Yasuda, M.; Ishikawa, A.; Jikumaru, Y.; Seki, M.; Nakashita, H. Antagonistic interaction between systemic acquired resistance and the Abscisic Acid-mediated abiotic stress response in *Arabidopsis*. *Plant Cell* **2008**, *20*, 1678–1692. [CrossRef]
49. Li, H.L.; Deng, H.P.; Sun, Z.Y.; Zhao, L.J.; Han, L.; Ju, G.S.; Qian, Y.Q. Effect of methyl jasmonate on physiological indexes of chrysanthemum cuttage under natural drought stress. *For. Res.* **2010**, *23*, 733–737. [CrossRef]
50. Pandey, S.; Chakraborty, D. Salicylic acid and drought stress response: Biochemical to molecular crosstalk: Stress responses in plants. *Stress Responses Plants* **2015**, *3*, 247–265. [CrossRef]
51. Daudi, A.; Cheng, Z.; O'Brien, J.A.; Mammarella, N.; Khan, S.; Ausubel, F.M.; Bolwell, G.P. The apoplastic oxidative burst peroxidase in *Arabidopsis* is a major component of pattern-triggered immunity. *Plant Cell* **2012**, *24*, 275–287. [CrossRef]
52. Huang, Q.; Wang, Y.; Li, B.; Chang, J.; Chen, M.; Li, K.; Yang, G.; He, G. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* **2015**, *15*, 268. [CrossRef]

53. Meng, D.; Yu, X.; Ma, L.; Hu, J.; Liang, Y.; Liu, X.; Yin, H.; Liu, H.; He, X.; Li, D. Transcriptomic Response of Chinese Yew (*Taxus chinensis*) to cold stress. *Front. Plant Sci.* **2017**, *8*, 868. [CrossRef]
54. Fu, L.; Ding, Z.; Han, B.; Hu, W.; Li, Y.; Zhang, J. Physiological investigation and transcriptome analysis of polyethylene glycol (PEG)-induced dehydration stress in cassava. *Int. J. Mol. Sci.* **2016**, *17*, 283. [CrossRef]
55. Hu, W.; Yang, H.; Yan, Y.; Wei, Y.; Tie, W.; Ding, Z.; Zuo, J.; Peng, M.; Li, K. Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Sci. Rep.* **2016**, *6*, 22783. [CrossRef]
56. Lamesch, P.; Berardini, T.Z.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; et al. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **2012**, *40*, D1202–D1210. [CrossRef]
57. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 461. [CrossRef]
58. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [CrossRef]
59. Marchler-Bauer, A.; Anderson, J.B.; Chitsaz, F.; Derbyshire, M.K.; DeWeese-Scott, C.; Fong, J.H.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; et al. CDD: Specific functional annotation with the conserved domain database. *Nucleic Acids Res.* **2008**, *37*, D205–D210. [CrossRef]
60. Finn, R.D.; Tate, J.; Mistry, J.; Coggill, P.C.; Sammut, S.J.; Hotz, H.R.; Ceric, G.; Forslund, K.; Eddy, S.R.; Sonnhammer, E.L.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2007**, *36*, D281–D288. [CrossRef]
61. Letunic, I.; Copley, R.R.; Schmidt, S.; Ciccarelli, F.D.; Doerks, T.; Schultz, J.; Chris Ponting, C.P.; Bork, P. SMART 4.0: Towards genomic data integration. *Nucleic Acids Res.* **2004**, *32*, D142–D144. [CrossRef]
62. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
63. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [CrossRef]
64. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]
65. Brown, P.; Baxter, L.; Hickman, R.; Beynon, J.; Moore, J.D.; Ott, S. MEME-LaB: Motif analysis in clusters. *Bioinformatics* **2013**, *29*, 1696–1697. [CrossRef]
66. McDowall, J.; Hunter, S. InterPro protein classification. *Methods Mol. Biol.* **2011**, *694*, 37–47. [CrossRef]
67. Hu, B.; Jin, J.P.; Guo, A.Y.; Zhang, H.; Luo, J.C.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
68. Wang, M.; Yue, H.; Feng, K.; Deng, P.; Song, W.; Nie, X. Genome-wide identification, phylogeny and expressional profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (*Triticum aestivum L.*). *BMC Genom.* **2016**, *17*, 668. [CrossRef]
69. Jin, X.; Zhu, L.; Yao, Q.; Meng, X.; Ding, G.; Wang, D.; Xie, Q.; Tong, Z.; Tao, C.; Yu, L.; et al. Expression profiling of mitogen-activated protein kinase genes reveals their evolutionary and functional diversity in different rubber tree (*Hevea brasiliensis*) cultivars. *Genes* **2017**, *8*, 261. [CrossRef]
70. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef]
71. Liang, Y.; Xiong, Z.; Zheng, J.; Xu, D.; Zhu, Z.; Xiang, J.; Gan, J.; Raboanatahiry, N.; Yin, Y.; Li, M. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci. Rep.* **2016**, *6*, 24265. [CrossRef]
72. Wu, C.; Hu, W.; Yan, Y.; Tie, W.; Ding, Z.; Guo, J.; He, G. The late embryogenesis abundant protein family in cassava (*Manihot esculenta crantz*): Genome-wide characterization and expression during abiotic stress. *Molecules* **2018**, *23*, 1196. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Salt Tolerance Improvement in Rice through Efficient SNP Marker-Assisted Selection Coupled with Speed-Breeding

Md Masud Rana [1,2], Takeshi Takamatsu [1,3], Marouane Baslam [3], Kentaro Kaneko [1], Kimiko Itoh [1,3], Naoki Harada [1,3], Toshie Sugiyama [1,3], Takayuki Ohnishi [4], Tetsu Kinoshita [5], Hiroki Takagi [6] and Toshiaki Mitsui [1,3,*]

1 Department of Life and Food Sciences, Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan; f16m502g@mail.cc.niigata-u.ac.jp (M.M.R.); takamatsutakeshi@yahoo.co.jp (T.T.); k-neko@gs.niigata-u.ac.jp (K.K.); kimi@agr.niigata-u.ac.jp (K.I.); naharada@agr.niigata-u.ac.jp (N.H.); sugiyama@agr.niigata-u.ac.jp (T.S.)
2 Agronomy Division, Bangladesh Rice Research Institute, Gazipur-1701, Bangladesh
3 Faculty of Agriculture, Niigata University, Niigata 950-2181, Japan; mbaslam@gs.niigata-u.ac.jp
4 Center for Education and Research of Community Collaboration, Utsunomiya University, Utsunomiya 321-8505, Japan; ohnishi@cc.utsunomiya-u.ac.jp
5 Kihara Institute for Biological Research, Yokohama City University, Yokohama 244-0813, Japan; tkinoshi@yokohama-cu.ac.jp
6 Faculty of Bioresources and Environmental Sciences, Ishikawa Prefectural University, Ishikawa 921-8836, Japan; h-takagi@ishikawa-pu.ac.jp
* Correspondence: t.mitsui@agr.niigata-u.ac.jp; Tel.: +81-25-262-6641

Received: 19 April 2019; Accepted: 22 May 2019; Published: 26 May 2019

Abstract: Salinity critically limits rice metabolism, growth, and productivity worldwide. Improvement of the salt resistance of locally grown high-yielding cultivars is a slow process. The objective of this study was to develop a new salt-tolerant rice germplasm using speed-breeding. Here, we precisely introgressed the *hst1* gene, transferring salinity tolerance from "Kaijin" into high-yielding "Yukinko-mai" (WT) rice through single nucleotide polymorphism (SNP) marker-assisted selection. Using a biotron speed-breeding technique, we developed a BC_3F_3 population, named "YNU31-2-4", in six generations and 17 months. High-resolution genotyping by whole-genome sequencing revealed that the BC_3F_2 genome had 93.5% similarity to the WT and fixed only 2.7% of donor parent alleles. Functional annotation of BC_3F_2 variants along with field assessment data indicated that "YNU31-2-4" plants carrying the *hst1* gene had similar agronomic traits to the WT under normal growth condition. "YNU31-2-4" seedlings subjected to salt stress (125 mM NaCl) had a significantly higher survival rate and increased shoot and root biomasses than the WT. At the tissue level, quantitative and electron probe microanalyzer studies indicated that "YNU31-2-4" seedlings avoided Na^+ accumulation in shoots under salt stress. The "YNU31-2-4" plants showed an improved phenotype with significantly higher net CO_2 assimilation and lower yield decline than WT under salt stress at the reproductive stage. "YNU31-2-4" is a potential candidate for a new rice cultivar that is highly tolerant to salt stress at the seedling and reproductive stages, and which might maintain yields under a changing global climate.

Keywords: *hst1*; Na^+ accumulation; SNP; rapid generation advance; salt tolerant; variant annotation; whole-genome sequencing

1. Introduction

Projected climate change will aggravate a variety of abiotic stresses of rice plants, including salinity, heat, drought, and submergence, thus reducing world rice production [1–4]. At the same time, we must increase global rice production by at least 70% to feed the anticipated 9.6×10^9 people by 2050 [5,6]. Under these conditions, the improvement of the salinity tolerance of locally grown high-yielding rice cultivars is one of the most promising breeding objectives by which to meet global food demand.

Rice is considered the most salt-sensitive cereal crop [7], with a threshold EC_e (electrical conductivity of saturated extract) of 3 dSm^{-1}, above which yield starts to decline [8–10]. Salinity imposes osmotic effects, ion toxicity, and nutritional imbalance and substantially affects almost all phases of growth [7,11]. Possible salt tolerance mechanisms in rice involve ion homeostasis and compartmentalization, ion transport and uptake, biosynthesis and accumulation of osmoprotectants, osmolytes, and compatible solutes, activation of antioxidant enzymes for ROS detoxification, and hormone modulation [12–17]. The *Saltol* [18,19] and *SHOOT* K^+ *CONCENTRATION 1* [20,21] genes have been identified from major quantitative trait loci (QTLs) of salt-tolerant landraces Pakkali and Nona Bokra, respectively. These QTLs have been introgressed into some widely grown, high-yielding rice cultivars to improve salt tolerance [22–26], but the rate of improvement is slow.

Rice biotechnology has made advances in identifying single nucleotide polymorphisms (SNPs) controlling salinity tolerance [27–31]. Mutant lines of "Hitomebore" were generated by treatment with ethyl methanesulfonate (EMS), an inducer of nucleotide substitutions, and isolated a salt-tolerant line carrying the *hitomebore salt tolerant 1* (*hst1*) gene. The causative SNP conferring the high salinity tolerance of the *hst1* mutant line corresponded to the third exon of the *Os06g0183100* gene, which is predicted to encode a B-type response regulator designated *OsRR22*. We backcrossed the *hst1* line with "Hitomebore" to breed the salt-tolerant cultivar "Kaijin", with a yield ability of 5.88 t ha^{-1} [32,33]. "Yukinko-mai" is an early-maturing standard cultivar derived from a cross between "Yukino-sei" and "Domannaka" at the Niigata Agricultural Research Institute's Crop Research Center; it has a high yield potential of 6.84 t ha^{-1} [34] and is tolerant to high temperatures during grain filling [35].

To combat earthquake- and tsunami-induced soil salinity in Japan, it is crucial to improve the salt resistance of locally grown popular rice cultivars, most of which are salt sensitive [32,36–38]. In addition, developing Japanese cultivars for international appeal and fine-tuning their yield performance under various ecosystems around the world are time-demanding tasks. To generate new rice cultivars quickly, in response to evolving consumer preferences and crises, we crossed the salt-tolerant "Kaijin" with "Yukinko-mai" to develop a salt-tolerant line with elite agronomic traits through the use of marker-assisted selection (MAS). MAS is the most advanced tool yet developed for the precise introgression of genes of interest into elite rice cultivars [39,40], and allows breeders to recover most of the recurrent parent genome in only two or three generations [41]. Salt, heat, and drought stress-responsive genes or QTLs revealed by recent advances in genomics and biotechnology are being used for MAS of rice all over the world [42–44].

In recent years, rapid generation-advance technology called "speed-breeding" has been used to shorten the generation cycle, accelerating the progress of genomics and breeding studies in multiple crops [45–48]. This technique has been used for the genetical improvement of rice, such as recombinant inbred lines, backcrossed inbred lines, and isogenic cultivars [46,47]. The speed breeding method has been reported for six major crops such as spring wheat (*Triticum aestivum*), durum wheat (*Triticum durum*), barley (*Hordeum vulgare*), chickpea (*Cicer arietinum*), pea (*Pisum sativum*), and canola (*Brassica napus*), that uses a prolonged photoperiod to reduce the generation time [48]. Nagatoshi and Fujita [49] developed a breeding technique for short-day soybean plant applying supplemental CO_2 in combination with long-day and appropriate temperature cycles. Using speed-breeding, which combines temperature, light duration, and humidity control, tiller removal, and embryo rescue, breeders can obtain four to five advanced generations in a year [45,46].

In this study, we precisely transferred the *hst1* (*OsRR22*) gene, which confers salinity tolerance, from "Kaijin" into high-yielding "Yukinko-mai" (WT) rice through SNP MAS coupled with speed-breeding. We sequenced the whole genome of a BC_3F_2 *hst1* homozygous line and determined the genome recovery rate. We also examined important physiological and biochemical parameters of the BC_3F_3 generation that confer salt tolerance and evaluated the phenotype under salt stress and normal field conditions.

2. Results

2.1. Breeding Scheme for Development of Advanced Plant Material

We introgressed the salt-tolerance *hst1* gene from "Kaijin" into the genetic background of the high-yielding "Yukinko-mai" by three backcrosses followed by two rounds of self-fertilization (Figure 1A). To accelerate the breeding cycle, we used a biotron speed-breeding system, with controlled temperature and daylength, restriction of tillers, and embryo rescue (Figure S1). At each cross, the plants produced a good quantity of fertilized seed; the cross success rate ranged between 54% and 69% (Table S1), and seeds from three or four plants were sufficient to develop new progeny. Each advanced generation took approximately 70 days from germination to flowering and 10 days from pollination to embryo rescue (Figure S1). The total duration of each generation varied according to days to flowering. Using this speed-breeding technique, we developed the BC_3F_3 population, carrying our desired allele in the homozygous state, in six generations and 17 months.

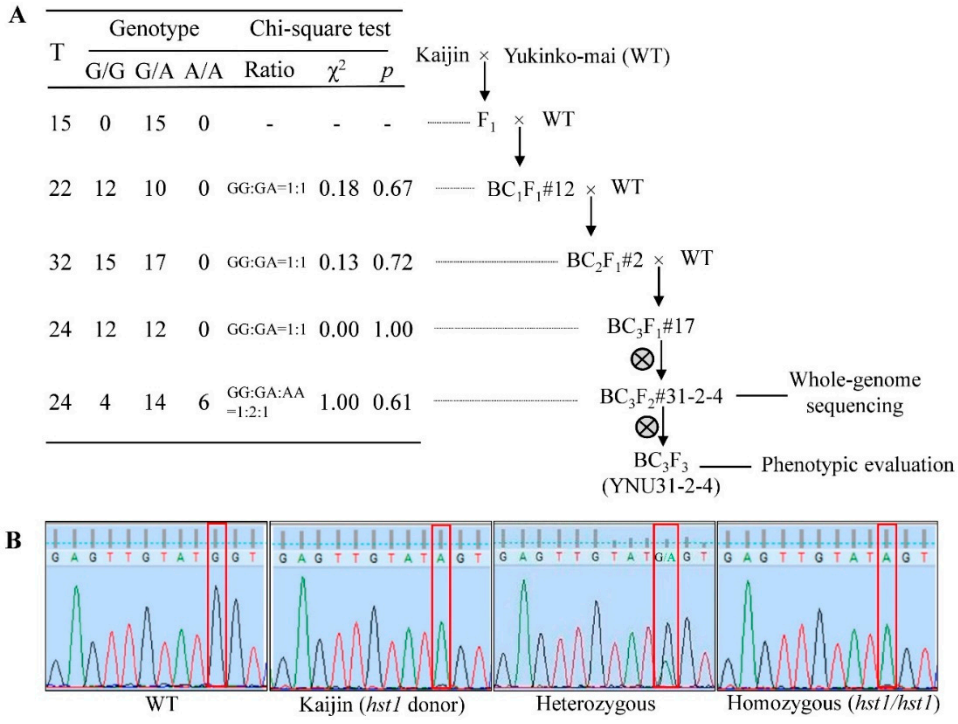


Figure 1. Single nucleotide polymorphism (SNP) marker-aided introgression of *hst1* from "Kaijin" into "Yukinko-mai". (**A**) The *hst1* gene was transferred from highly salt-tolerant "Kaijin" into "Yukinko-mai". "Kaijin" was backcrossed to "Yukinko-mai" (WT) 3 times followed by 2 rounds of self-pollination. The table shows the selection results at each generation: T, total number of tested plants; G/G, number of plants not carrying donor allele; G/A, number of plants carrying donor allele in heterozygous state; A/A, plants carrying donor allele in homozygous state. The number after "#" is the individual plant number used in backcrossing or self-pollination. (**B**) Advanced breeding individuals were genotyped by direct sequencing. The red box represents nucleotide 1975 of the *OsRR22* locus, which is responsible for salt tolerance.

2.2. Introgression of hst1 into "Yukinko-mai" (WT) and Genotyping of Advanced Progeny Using SNP Marker

We used SNP-based genotyping by Sanger sequencing to identify plants harboring the donor allele in each breeding round. A SNP in the *OsRR22* (*hst1*) gene confers the salinity tolerance of the donor parent "Kaijin". The breeding lines were selected on the basis of target peaks of G/A heterozygosity (nucleotide 1975 of the *OsRR22* locus) in the F_1, BC_1F_1, BC_2F_1, and BC_3F_1 generations and of A/A homozygosity at the same locus in the BC_3F_2 generation (Figure 1B). In the F_1 to BC_3F_1 generations, we obtained two genotypes: either homozygous, lacking the donor allele (G/G), or heterozygous (G/A) (Figure 1A). The heterozygous BC_3F_1 population was self-pollinated to develop BC_3F_2 lines that carried the donor allele in the homozygous state (A/A). BC_3F_2 plants (A/A) morphologically similar to the recurrent parent were self-pollinated to develop the BC_3F_3 generation. We sequenced the whole genome of the BC_3F_2 line #31-2-4 to compare it with the parental genome, and characterized it.

2.3. Recovery Rate and Characterization of BC_3F_2 #31-2-4 Genome

To investigate the genetic similarities between our advanced line and the parents, we analyzed BC_3F_2 #31-2-4 using whole-genome sequencing. After we filtered out low-reliability SNPs/indels, #31-2-4 showed 118 454 SNPs/indels, comprising 106 288 "Yukinko-mai"-type homozygous alleles, 3157 "Kaijin"-type homozygous alleles, and 9009 heterozygous alleles (Table 1). These SNPs/indels lie across the genome with deep coverage (Figure 2, dots), indicating high resolution and successful genome-wide genotyping. Allele types formed dense blocks on chromosomes (Figure 2, vertical bars), clearly showing recovered regions ("Yukinko-mai" homozygous blocks), "Kaijin" genome segments ("Kaijin" homozygous blocks), and unfixed segments (heterozygous blocks) (Figure 2, horizontal bar). "Yukinko-mai" chromosomes (Chrs.) 5, 11, and 12 were recovered almost completely. There were small "Kaijin" segments in Chrs. 1, 4, 7, 9, and 10, large "Kaijin" segments in Chrs. 2, 3, and 6, small heterozygous segments in Chr. 4, and large heterozygous segments in Chrs. 3, 6, 8, and 9. Interestingly, we identified some genotype blocks overlapping other genotype blocks (Figure 2, chr08, 5–10 Mb; chr09, 10–12 Mb), resulting from continuous recombination events in these extremely short regions [50,51]. We calculated the genome recovery rate from the number of "Yukinko-mai" alleles out of the total number; the BC_3F_2 genome recovered 93.5% of the "Yukinko-mai" genome, from 89.7% homozygous alleles and 7.6% heterozygous alleles (Table 1; Section 4.7). This score is close to the theoretical value of 93.7% following three backcrosses and one self-fertilization. In addition, 2.7% of the BC_3F_2 genome was "Kaijin" homozygous and 7.6% remained unfixed as heterozygous (Table 1).

Table 1. Single nucleotide polymorphism (SNP)/indel detection in BC_3F_2 #31-2-4. SNPs/indels in BC_3F_2 #31-2-4 are classified into "Yukinko-mai"-type (homo[zygous]), "Kaijin"-type (homo(zygous)), and Hetero(zygous) alleles.

Genotype	Number of Variants			Genome %
	SNPs	Indels	Total	
'Yukinko-mai" (homo)	84 927	21 361	106 288	89.7
"Kaijin" (homo)	2329	828	3157	2.7
Hetero	7556	1453	9009	7.6
Total	94 812	23 642	118 454	100.0

To estimate the effects of variants on phenotypes, we listed variants causing protein sequence alterations (e.g., frameshift and in-frame indels, non-synonymous SNPs, SNPs/indels at splice donor/acceptor sites etc.) in SnpEff software for Sequence Ontology and effect prediction (Figure 3). In BC_3F_2, 207 "Kaijin" homozygous and 536 heterozygous variants were found in exon and splice sites. Only 71 and 230, respectively, of those caused protein sequence alterations (Figure 3; Table S2). These non-synonymous changes occurred in only four genes for agronomic traits (Table 2).

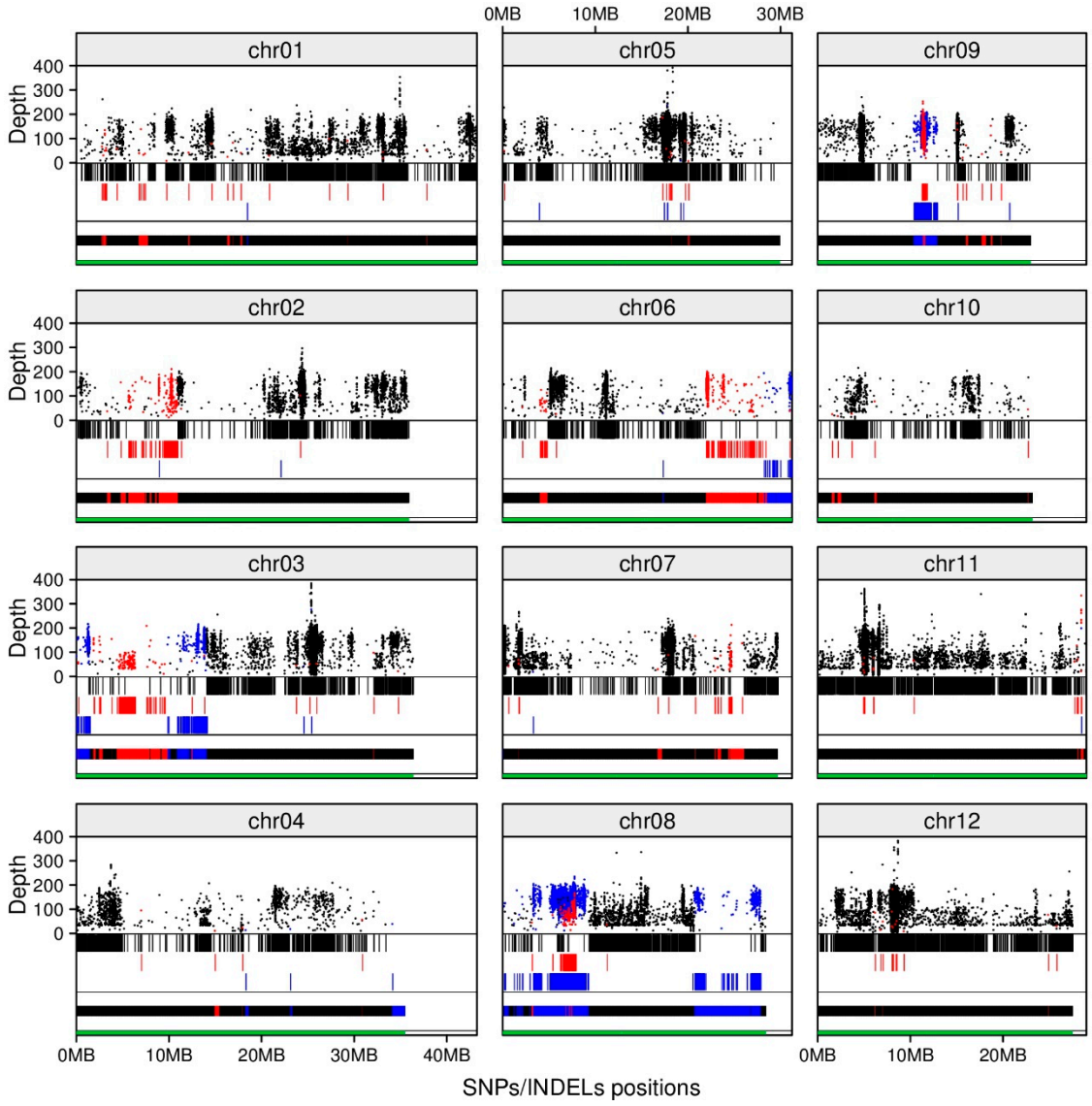


Figure 2. Positions and depths of SNPs/indels and genotype blocks on BC_3F_2 #31-2-4 genome. Dots show coverage depth of SNPs/indels. Vertical bars show their positions by genotype (color). Horizontal bar shows densities of SNPs/indels densities in 10,000-nt sliding window by genotype (colour). Black, "Yukinko-mai"-type homozygous; red, "Kaijin"-type homozygous; blue, heterozygous. Green bars represent chromosome length.

2.4. Field Evaluation of the "YNU31-2-4" for Main Agronomic Traits

We evaluated "YNU31-2-4" in the field to compare its major agronomic traits with those of the parents. We found a significant difference in days-to-heading between "YNU31-2-4" and WT: some plants headed significantly earlier than WT (Table 3; Figure S2B). Heading date was associated negatively with the *Os09g0356200* genotype and positively with the *hst1* genotype. These results support the accuracy of gene-based GWAS [53], and also suggest the potential early flowering function of *hst1* or the effect of unknown flowering-related genes. Grain width, grain thickness, and 1000-grain weight were significantly higher in "YNU31-2-4" plants than WT (Table 3). The *hst1* mutant also had a wider grain than WT [32]. Together these results, at least partly, suggest that *hst1* might be involved in the increase of grain width. In contrast, other important morphological traits of "YNU31-2-4" were highly similar to those of WT, particularly flag leaf color, length, and width, plant height, tiller number

per plant, panicle number per plant, panicle length, spikelet number per panicle, grain yield per plant, aboveground biomass per plant, and grain length (Table 3; Figure S3A–C).

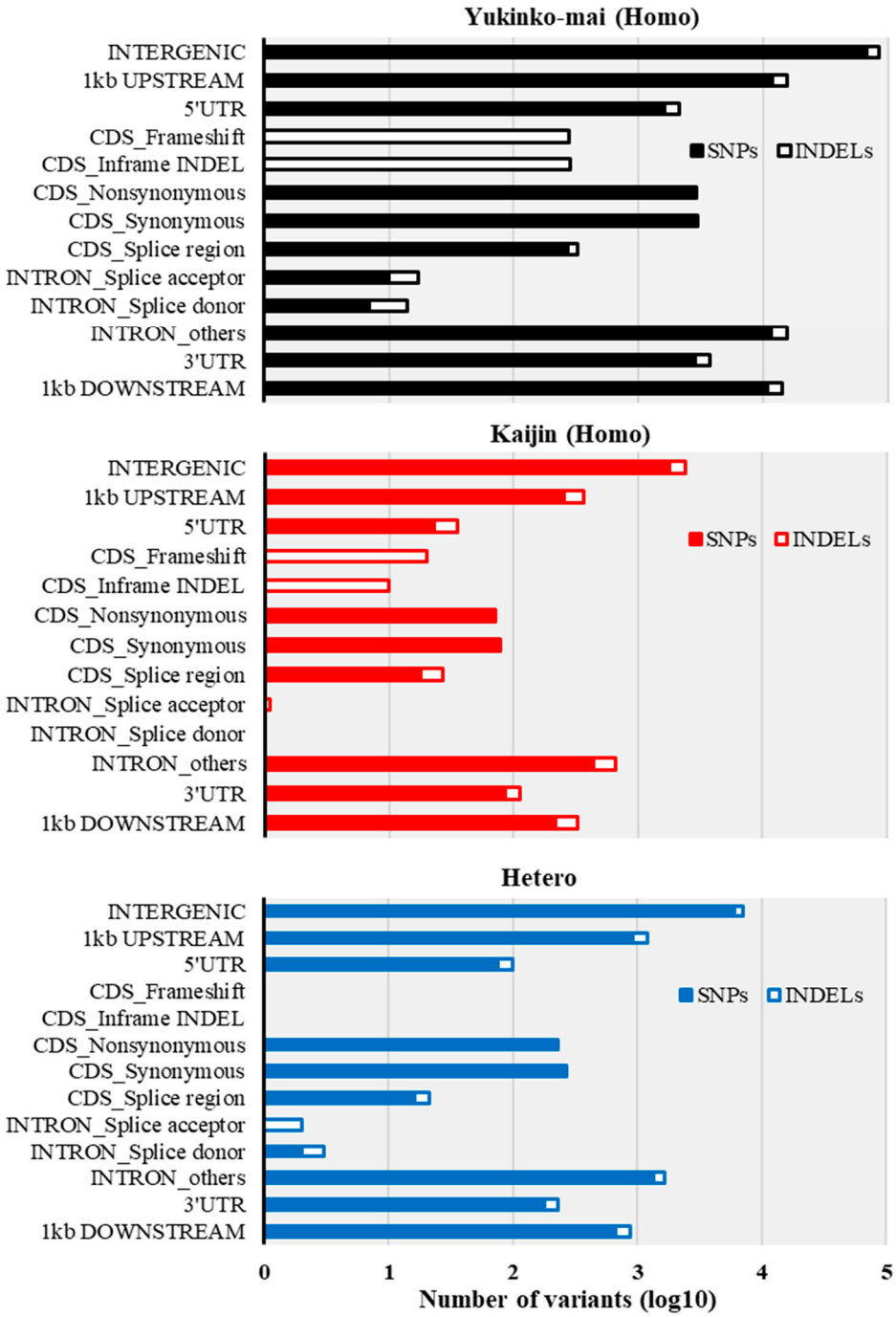


Figure 3. Sequence Ontology analysis of BC_3F_2 SNPs/indels. Sequence Ontology is based on SnpEff terms with minor modifications. Closed bars, SNPs; open bars, indels. Black, "Yukinko-mai"-type homozygous, red, "Kaijin"-type homozygous; blue, heterozygous.

Table 2. List of agronomic-trait-related genes harboring alternative protein sequences in BC_3F_2 #31-2-4. Chr., chromosome; Ref., Nipponbare reference allele; Hetero, heterozygous. Dataset 1, Overview of Functionally Characterized Genes in Rice Online database (OGRO) [52]; Dataset 2, potential agronomic functional gene set by gene-based GWAS of Japanese rice population [53].

Chr./Position	Ref. SNP/Indel	Genotype			Alteration Type	RAP ID	Protein Encoded	Dataset 1	Dataset 2
		BC_3F_2	"Kaijin"	"Yukinko-mai"					
Chr06/26277010	G	A	A	G	missense SNP	Os06g0644200	Vacuolar HD-translocating inorganic pyrophosphatase 1	Cold tolerance	–
Chr08/6268486	GCACGGCCACGGC	Hetero	GCACGGCCACGGC	G	in-frame deletion	Os08g0207500	Zn-regulated transporter, iron (Fe)-regulated transporter–like protein 4	Other soil stress tolerance	–
Chr08/26913261	G	Hetero	G	A	missense SNP	Os08g0538300	Chitin elicitor receptor kinase 1	Blast resistance	–
Chr09/11449688	G	Hetero	G	A	missense SNP	Os09g0356200	Malectin-like carbohydrate-binding domain–containing protein	–	Days to heading
Chr09/11449800	AGG	Hetero	AGG	AC	frameshift deletion			–	

Table 3. Growth and yield performance of "Kaijin", WT, and "YNU31-2-4" genotypes under normal field conditions. Grain yield is weight of filled spikelets at 14% moisture content. Data are mean ± SD (n = 10–12). Values with the same letter within a row are not statistically different (Duncan's multiple range test, $p < 0.05$).

Agronomic Traits	Genotype		
	"Kaijin"	WT	"YNU31-2-4"
Flag leaf greenness (SPAD value)	42.93 ± 2.63 a	43.49 ± 1.82 a	43.24 ± 2.42 a
Flag leaf length (cm)	27.21 ± 2.53 b	29.85 ± 2.13 a	29.12 ± 1.85 a
Flag leaf width (cm)	1.11 ± 0.06 b	1.23 ± 0.10 a	1.21 ± 0.04 a
Days-to-heading (day)	105.17 ± 1.40 a	105.08 ± 1.38 a	103.75 ± 1.76 b
Plant height (cm)	98.42 ± 3.56 a	94.75 ± 3.49 b	95.83 ± 2.41 ab
Tiller number per plant	20.40 ± 3.31 b	23.80 ± 3.22 a	24.40 ± 3.57 a
Panicle number per plant	20.40 ± 3.31 b	23.80 ± 3.22 a	24.20 ± 3.12 a
Panicle length (cm)	18.71 ± 0.69 b	19.79 ± 0.52 a	19.11 ± 0.97 ab
Spikelet number per panicle	71.99 ± 8.91 a	73.55 ± 6.27 a	72.66 ± 5.80 a
Grain number per panicle	63.61 ± 8.20 a	67.36 ± 5.81 a	60.59 ± 7.29 a
1000-grain weight (g)	22.42 ± 0.14 b	21.56 ± 0.22 c	23.29 ± 0.43 a
Seed setting rate (%)	88.34 ± 3.71 a	91.61 ± 2.52 a	83.43 ± 7.58 b
Grain yield (g per plant)	33.77 ± 4.24 b	40.07 ± 4.59 a	37.84 ± 4.46 a
Aboveground biomass (g per plant)	61.80 ± 8.50 b	72.30 ± 7.63 a	67.13 ± 9.14 ab
Grain length (mm)	5.05 ± 0.03 b	5.17 ± 0.01 a	5.21 ± 0.06 a
Grain width (mm)	2.78 ± 0.01 b	2.66 ± 0.01 c	2.86 ± 0.04 a
Grain thickness (mm)	2.14 ± 0.01 a	2.07 ± 0.01 c	2.10 ± 0.01 b

2.5. Salinity Tolerance of the "YNU31-2-4" Line at Seedling Stage

We evaluated the salinity tolerance of "Kaijin", "Yukinko-mai" (WT), and "YNU31-2-4" seedlings during three weeks at 0, 75, and 125 mM NaCl. At 0 mM NaCl, there was no phenotypic difference between "YNU31-2-4" and WT (Figure 4A). Under salt stress, however, the WT leaves were rolled, whereas those of "Kaijin" and "YNU31-2-4" remained flat and stayed green even at 125 mM NaCl. Further, 125 mM NaCl reduced the survival of WT seedlings to 52.5%, whereas all seedlings of "YNU31-2-4" survived (Figure 4B). Moreover, "Kaijin" and "YNU31-2-4" had better shoot and root growth under salinity vs. WT (Figure 4C–F); at 125 mM NaCl, "YNU31-2-4" had 30% and 38% better shoot and root dry weight, respectively than WT. In a separate experiment (Figure S4), at 0 mM NaCl, there were no significant differences in leaf relative water content or chlorophyll content between WT and "YNU31-2-4" plants. Under salt stress, the "YNU31-2-4" plants were able to maintain significantly higher relative water content and chlorophyll levels than WT (Figure S4A,B). Under the control condition, "YNU31-2-4" had significantly higher proline content than the parents. Exposure to salinity led to a considerable increase in proline levels in all genotypes, and to 1.6× the WT level in "YNU31-2-4" (Figure S4C). These results clearly indicate that "YNU31-2-4" has stronger salt tolerance than "Yukinko-mai" (WT) at the seedling stage.

2.6. Na^+ and K^+ Content in Shoot and Roots of the Tested Genotypes under Salt Stress

We assayed the K^+ and Na^+ contents in shoots and roots of seedlings under salinity, since the degree of stress depends on their uptake and translocation. Under control condition, shoot K^+ was significantly lower in "YNU31-2-4" than in its parent (Figure 5A), whereas under salt stress, it was 1.4× the WT level in shoot and 2.6× in root (Figure 5A,D). Under control condition, Na^+ levels in leaves (Figure 5B) and roots (Figure 5E) of all genotypes remained similarly low. Salinity stress increased Na^+ concentration in WT shoots relative to the other two genotypes, reaching 6.5× that in "YNU31-2-4" plants (Figure 5B). Under control conditions, the Na^+/K^+ ratio did not differ significantly among the tested genotypes in shoots (Figure 5C) and roots (Figure 5F). Under salinity, it was 9.2× the "YNU31-2-4" level in WT shoots and 2.9× in WT roots (Figure 5C,F).

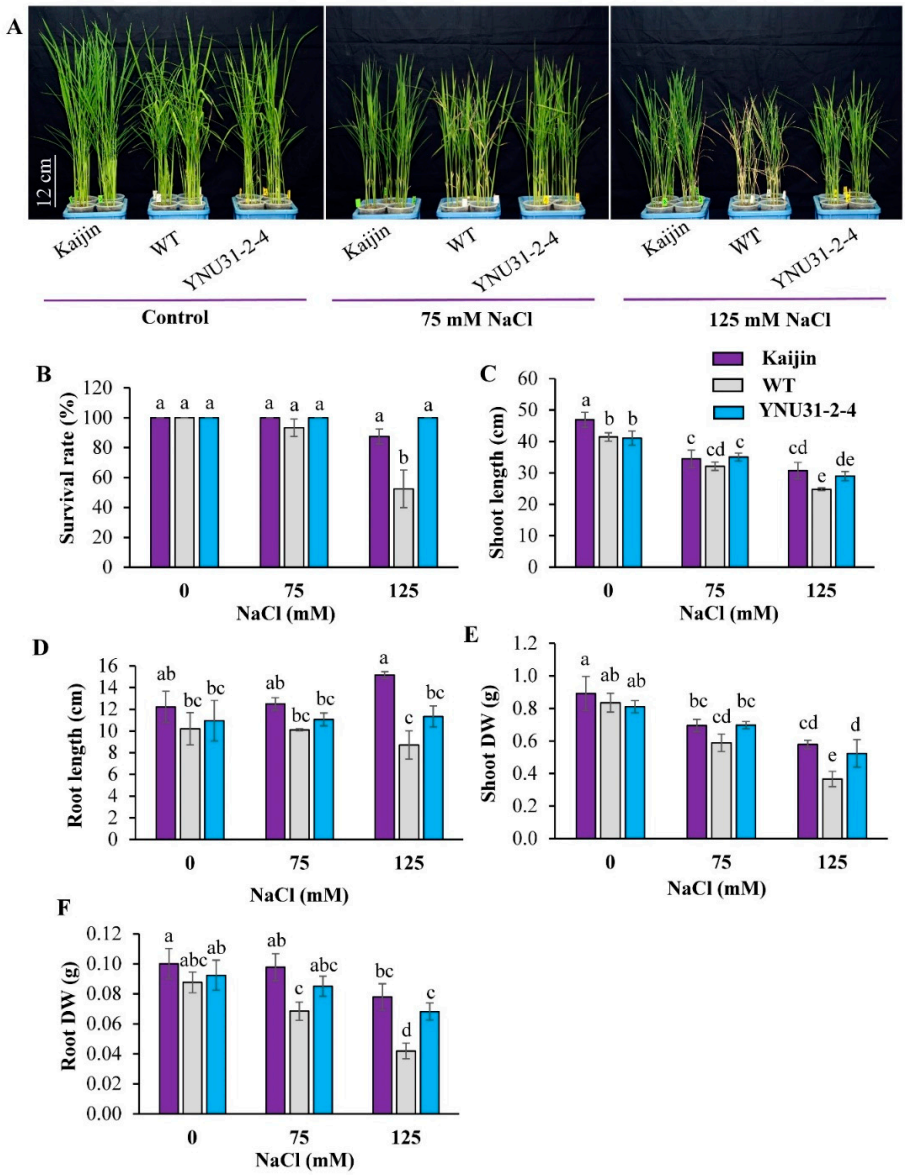


Figure 4. Salinity tolerance of BC_3F_3 line "YNU31-2-4" at seedling stage. (**A**) Phenotypic comparison of "Kaijin", WT, and "YNU31-2-4" seedlings grown in 0, 75, or 125 mM NaCl for two weeks. (**B**) Survival rates, (**C**) shoot length, (**D**) root length, (**E**) shoot dry weight, and (**F**) root dry weight of seedlings shown in **A**. Data in **B**–**F** are mean ± SD of four independent biological replicates; data in **E** and **F** are dry weight (DW) of 10 plants in each treatment. Bars labeled with the same letter are not statistically different (Tukey's test, $p < 0.05$).

Under the control condition, electron probe microanalysis revealed a dense distribution of K^+ in the basal portion of the shoot of all genotypes, but only a very sparse distribution of Na^+ (Figure 6). Under salt stress, the Na^+ distribution in cells was increased in all genotypes, and the salt-sensitive WT accumulated significantly more Na^+ than the other two genotypes (Figure 6).

2.7. Salinity Tolerance and Yield Assessment at Reproductive Stage

Under control condition at the reproductive stage, there was no obvious phenotypic difference between WT and "YNU31-2-4" plants (Figure 7A). Salt stress for five weeks caused severe burning and wilting symptoms in WT, but "YNU31-2-4" and "Kaijin" plants maintained green leaves. Under control condition, the penultimate leaves of "YNU31-2-4" plants maintained a slightly higher net CO_2

assimilation rate than the parents (Figure 7B). Salt stress for four weeks significantly reduced the net CO_2 assimilation rate of WT plants relative to "YNU31-2-4" and "Kaijin". Under the control condition, WT and "YNU31-2-4" plants had similar phenotypic characters and yield attributes except for a higher 1000-spikelet weight than the parents (Figure 7C–G). Under salt stress, in contrast, "YNU31-2-4" had higher plant height, yield characters, and aboveground biomass than WT (Figure 7C–G,I). Relative to control condition, salt stress reduced grain yield by 68% in WT but by 38% in "YNU31-2-4" (Figure 7H). As a result, the grain yield of "YNU31-2-4" was 10% higher than that of the donor parent "Kaijin" under control condition and 45% higher than that of WT under saline condition.

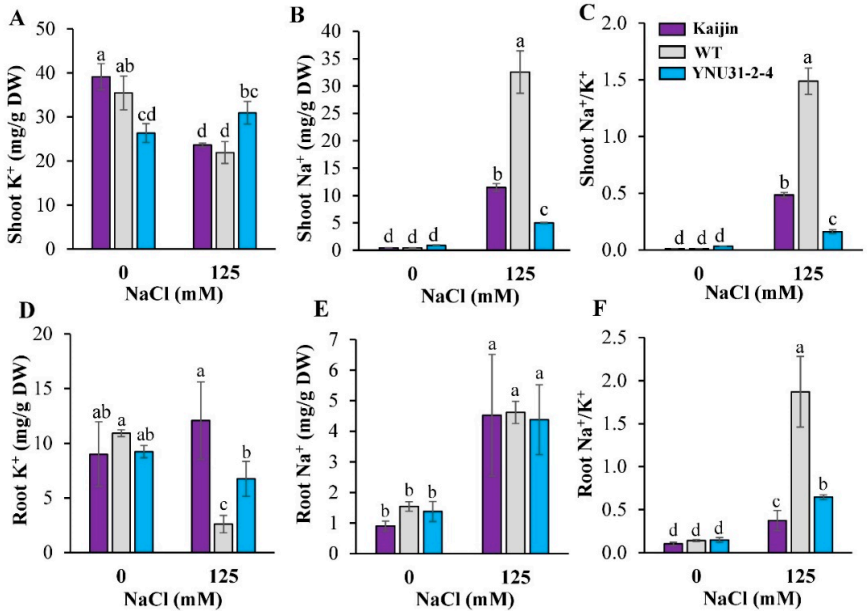


Figure 5. Na^+ and K^+ content in shoot and roots of the tested genotypes under salt stress. (**A**) Shoot K^+, (**B**) shoot Na^+, (**C**) shoot Na^+/K^+, (**D**) root K^+, (**E**) root Na^+, and (**F**) root Na^+/K^+ of 20-day-old seedlings. Data are mean ± SD of three independent biological replicates. Bars with the same letter are not statistically different (Duncan's multiple range test, $p < 0.05$).

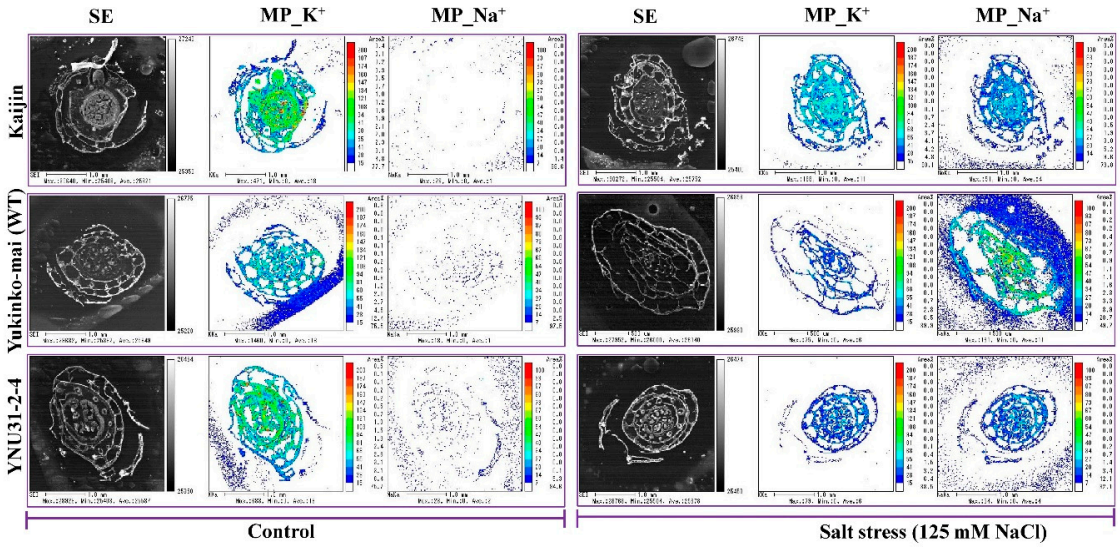


Figure 6. Accumulation and distribution of K^+ and Na^+ in cell clusters of 20-day-old seedlings. Relative amounts of K^+ and Na^+ are indicated by color coding. SE, secondary electron image; MP_K^+, mapping pattern of K^+; MP_Na^+, mapping pattern of Na^+.

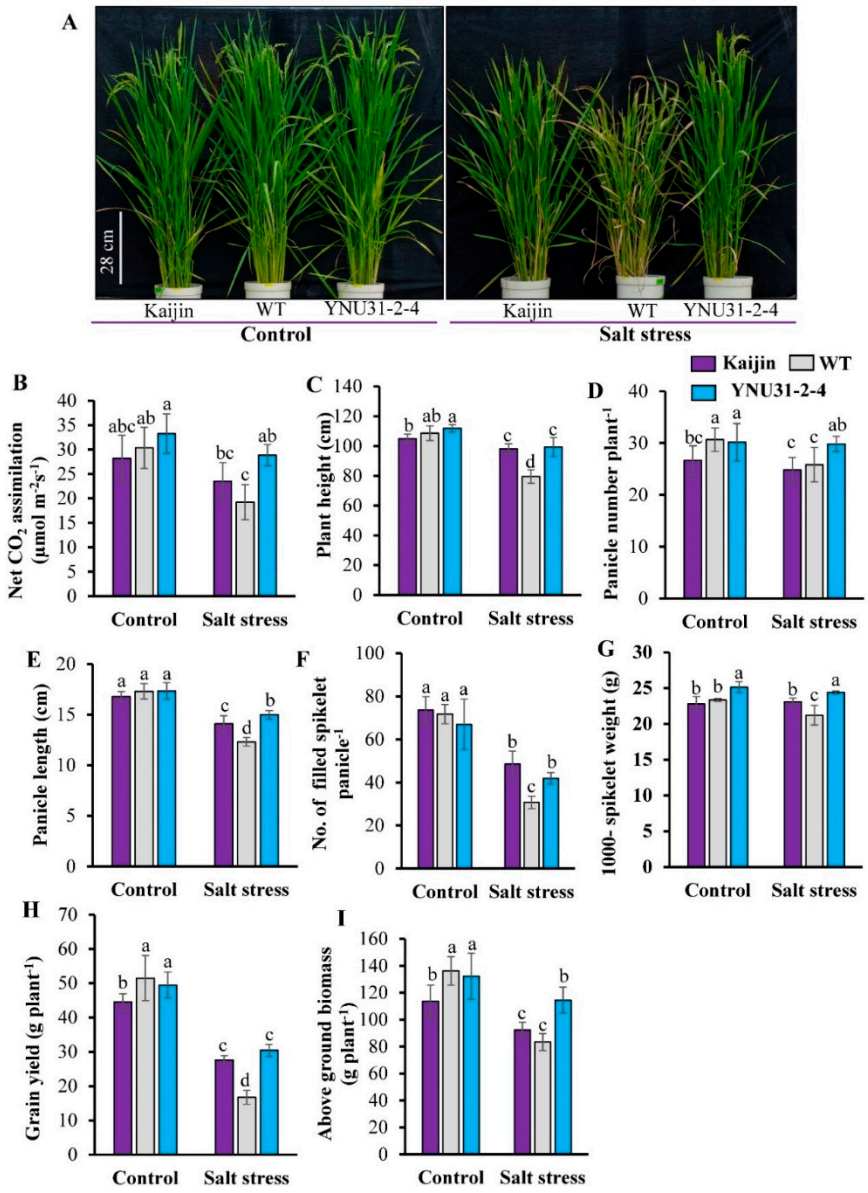


Figure 7. Salinity tolerance of BC_3F_3 line "YNU31-2-4" at heading. (**A**) Phenotypic comparison of "Kaijin", WT, and "YNU31-2-4" plants grown with or without salt stress. Salt-stressed plants were grown in 50 mM NaCl from 60 days after germination (DAG) and then in 75 mM from 74 DAG until 95 DAG (booting stage), and then in fresh water until 110 DAG (heading). (**B**) Net CO_2 assimilation rate of the penultimate leaf four weeks after imposition of salt treatment. (**C**) Comparison of plant height at 110 DAG. (**D–I**) Comparisons of (**D**) number of panicles/plant, (**E**) panicle length, (**F**) number of filled spikelets/panicle, (**G**) weight of 1000 filled spikelets, (**H**) grain yield/plant, and (**I**) dry weight of aboveground biomass at harvest. Data are mean ± SD of 6 individuals. Bars with the same letter are not statistically different (Duncan's multiple range test, $p < 0.05$).

3. Discussion

Soil salinity is a major threat to the future food production, affecting more than 6% of the total land area [54]. The rapid global warming and sea level rise pose threats to rice yield and quality in South Asian rice-growing countries. In addition, a tsunami contaminated paddy field in Miyagi prefecture, Japan, with salt in 2011 [32,37,55]. Therefore, it is important to introgress genes/QTLs/SNPs conferring salt tolerance in locally grown popular rice cultivars, focusing on higher grain yield, to ensure food security under changing climatic conditions. Cultivar improvement through conventional breeding

is feasible, but it takes a long time to minimize linkage drag through phenotypic screening [56,57]. For these reasons and to achieve breeding goals, we introgressed the *hst1* gene from "Kaijin" into "Yukinko-mai", which has excellent yield stability. We developed the BC_3F_3 generation, named "YNU31-2-4", through SNP marker-assisted selection (Figure 1A).

To accelerate the breeding cycle, we used a biotron speed-breeding system (Figure S1) without a CO_2 supply, since the application of 475 ppm CO_2 in growth chambers did not greatly change the breeding cycle [45], and many rice breeders do not have CO_2 regulation facilities owing to the high cost. By using a longer daylength (14/10 h light/dark) for first 30 days to accelerate the vegetative growth followed by a shorter daylength (10/14 h light/dark) to induce reproduction, tiller removal, and embryo rescue to decrease the period before seed maturity (Figure S1), we were able to achieve four breeding generations within 11 months (Table S1). Developing four to five generations a year is the ultimate objective [46–48]. This simplified, faster, efficient method for reducing the duration and number of breeding cycles will contribute significantly to genomic studies and the deployment of superior rice.

We used whole-genome sequencing (WGS) to characterize the advanced breeding lines and revealed genome recovery rate, genotype blocks, and putative phenotypes (Figures 2 and 3; Tables 1 and 2). WGS identified 118 454 SNPs/indels as markers. As WGS provides higher resolution of genome blocks than conventional SSR marker methods [58], it can thus be used for advancing generations from parents with low genetic variation, as here. Furthermore, in subsequent selection, knowledge of these genotype blocks helped us to rapidly fix heterozygous regions into the recipient allele through the MABC selection method (Figure S2A). Our high-resolution analysis with massive numbers of SNP/indel markers not only enabled accurate genome-wide genotyping, but also highlighted potential recombination hotspots.

Functional annotation was able to predict five SNP/indels (Table 2) that may affect agronomic traits and thus phenotype in our advanced line "YNU31-2-4". One of two non-synonymous changes in *Os09g0356200*, a putative heading-date-associated gene identified in a gene-based genome-wide association study (GWAS) [53], caused a frameshift deletion probably causing loss-of-function. BC_3F_3 progeny of BC_3F_2 plants harboring the *Os09g0356200* gene in the heterozygous state likely shows a wide range of heading date. Indeed, the field assessment results demonstrate the difference in the distribution of heading dates between "YNU31-2-4" and WT ("Yukinko-mai") (Table 3; Figure S2B). Additionally, the heading date of the BC_3F_3 population was associated with the genotype of the putative heading-date-associated gene *Os09g0356200* (Table 2; Figure S2). Our finding indicates the potential value and reliability of gene-based GWAS among the Japanese rice population data set [53] in breeding by incorporating the genetic variations into those cultivars, and its practical value in predicting genes governing complex traits of agronomic phenotypes. The results of whole-genome sequencing have demonstrated the success in developing improved cultivars using our rapid breeding system. They also reveal undesirable genome regions and genes from the donor parent "Kaijin", valuable information for estimating agronomic properties without further phenotyping study. Identifying heterozygous genes can provide mechanistic insights toward homogenizing phenotypes for ongoing breeding. It is important to note that the heterozygote disadvantage has been overcome by taking advantage of the SNP-based selection of the homozygous WT allele from the YNU31-2-4 population (Figure S2). Furthermore, other morphological traits of "YNU31-2-4" were similar to those of WT (Table 3; Figure S3A–C). The field assessment results along with the high-resolution genotyping data indicate no apparent grain yield reduction in "YNU31-2-4" in relation to the presence of the target *hst1* gene. In fact, "YNU31-2-4" increased ca. 11% yield than the donor parent "Kaijin" owing to the higher number of panicles per plant and 1000-grain weight.

Rice is very sensitive to salt stress at the seedling stage [7], and its sensitivity varies with the developmental stage [59]. To assess the practical utility of *hst1* in our introgression line, we exposed seedlings to moderate (75 mM) and high (125 mM) salt stresses. Salt tolerance at this stage is of importance in saline environments, as crop establishment is fundamentally determined during the earliest stages of development. Our findings revealed that under high salt stress, the "YNU31-2-4"

plants had a significantly higher survival rate, shoot, and root biomass than WT (Figure 4A–F), which suggest strong tolerance similar to that of the donor parent. The "YNU31-2-4" plants maintained significantly higher plant growth, proline content, and plant water status under salinity, which could indicate a physiological and biochemical tolerance mechanism [54,60]. In fact, previous studies showed that salinity might reduce the fertility of the spike and the translocation of assimilates to the grain in bread wheat and rice. Physiologically, the "YNU31-2-4" maintain its fully hydrated state under saline condition, which could at least partially have rapid and large effects on cell expansion, cell division, stomatal opening, maintain normal rates of transpiration, abscisic acid (ABA) accumulation, etc. Assay of soluble proline levels is a useful way to monitor physiological status and to assess stress tolerance, since plants under salt stress accumulate this osmoprotectant against ion-dependent protein degradation [61]. Proline is accumulated in taxonomically diverse sets of plants [62], providing stress tolerance by protecting the cell membrane and maintaining osmotic balance within the cell, and also serves as an organic nitrogen reserve during stress recovery [61,63,64].

We also assessed yield traits of "YNU31-2-4" under salt stress at the early reproductive to booting stages, when salt stress reduces panicle and spikelet numbers per plant, leading to significant yield losses [65,66]. The improvement of rice grain yield under salt stress is the focus of breeding [67]. Owing to the significant increases in panicle number per plant, spikelet number per panicle, and 1000-spikelet weight, the final grain yield of "YNU31-2-4" plants was 45% higher than WT under salt stress at the reproductive stage (Figure 7D,F–H). Under control condition, there was no yield difference between "YNU31-2-4" and WT (Figure 7H). Interestingly, "YNU31-2-4" has higher yield potential than the donor parent under control condition owing to the higher number of panicles per plant and 1000-spikelet weight, comparable to the field evaluation results (Figure 7D, G). The higher number of panicles can be attributable to the WT background. The higher seed weight of "YNU31-2-4" could be due, at least partly, to the improved photosynthetic efficiency (Figure 7B) due to coordination of leaf morphological (Table 3) and physiological (Figure 7B; Figure S4) traits, which has great potential for use in breeding for higher yield. Accordingly, "YNU31-2-4" showed larger flag leaf than the donor "Kaijin", which could play an important role to grain filling and hence determining yield potential. The superior tiller growth with higher leaf size rendered the source, sink, and flow stronger and more harmonized and consequently increased the cereal yield [68–72]. Thus, this study clearly shows that the introgression of *hst1* to the WT significantly increased salt resistance without any reduction in grain yield. Thus, "YNU31-2-4" has significant breeding value without a noticeable yield penalty under normal and salt stress conditions.

Roots absorb minerals and water from the soil and play a key role in transporting them to leaves. In the context of salt tolerance, roots are sensitive to NaCl and are the first site of defense, directly limiting or excluding sodium uptake [54,73]. Roots are often used as a biomarker of salt stress. Root architecture differed between WT and "YNU31-2-4" plants after two weeks of normal hydroponic culture (Figure S5). Roots of "YNU31-2-4" and "Kaijin" exposed to high salt stress elongated more than WT roots (Figure 4D). The better morphophysiological and biochemical characters of "YNU31-2-4" under salt stress demonstrate the success of introgression of *hst1* into "Yukinko-mai".

The genes involved in conferring salt tolerance, which is likely a complex trait controlled by a combination of multiple genes, are yet to be elucidated. Recent research advances have identified major genes conferring salinity tolerance in rice, including *OsHKT1;1*, *OsHKT2;1*, *OsSOS1*, *OsNHX1*, *OsCAX1*, *OsAKT1*, *OsKCO1*, *OsNRT1;2*, *OsCLC1*, *OsADS31* and *OsTPC1*; however, their functional pathways during salt stress are not coordinately linked for explaining the very complex phenomenon of salt tolerance [16,74,75]. The *hst1* (loss-of-function in *OsRR22*) gene primarily led to the upregulation of *OsHKT1;1* (encoding a high-affinity K^+ transporter) that functions as a Na^+ transporter contributing salt resistance of the *hst1* mutant and "Kaijin" [32]. In our experiment, Na^+ content and Na^+/K^+ ratio in leaf and root were divergent between WT and "YNU31-2-4" and therefore may represent the effects of "basic" strategies related to salt tolerance or susceptibility. The quantification and localization results demonstrate that like "Kaijin", "YNU31-2-4" plants maintained a very low Na^+/K^+ ratio in both

shoot and root under salt stress (Figure 5C,F), which is one of the most important mechanisms used by plants to withstand salt stress [76,77]. Under salt stress, the susceptible WT plants had more Na^+ densely localized in shoot tissue (Figure 6). An overload of Na^+ can dramatically depolarize the plasma membrane, leading to K^+ efflux via depolarization-activated outward-rectifying K^+ channels [78]. It is notable that *hst1*-regulated salt stress resistance involved K^+ homeostasis. These results suggest that the accumulation of more K^+ with less Na^+ in "YNU31-2-4" plants would be mediated by a mechanism of K^+ influx and Na^+ efflux. The possible roles of the high-affinity K^+ transporter *OsHKT1;1*, upregulated by *hst1*, could mediate salt stress resistance in "YNU31-2-4". Further investigation will be needed to elucidate the molecular mechanisms mediating K^+ and Na^+ homeostasis in "YNU31-2-4".

In summary, our results demonstrate that the modified biotron breeding system coupled with SNP MAS offers a rapid and effective way to improve single traits in rice. The precise introgression of *hst1*, combined with suitable genetic resources and phenotyping results, resulted in the selection of a line, "YNU31-2-4", adapted to salt stress at the vegetative and reproductive stages with improved yield due to improved water relations, photosynthesis, ion homeostasis, regulation of Na^+ uptake, and xylem loading of Na^+ to shoot. In order to corroborate the obtained salt stress data, the future perspective of this study is to evaluate the phenotype of the promising line under large-scale field trials. "YNU31-2-4" is a potential candidate for new rice cultivar with markedly improved salinity tolerance, which might sustain grain yield and food security in a changing climate.

4. Materials and Methods

4.1. Planting Materials

Seeds of "Yukinko-mai" (elite cultivar) and "Kaijin" (salt tolerant) were obtained from the Niigata Agricultural Research Institute's Crop Research Center (Nagaoka city, Niigata, Japan) and the Iwate Biotechnology Research Center (Kitakami city, Iwate, Japan), respectively.

4.2. Speed-Breeding–Modified Controlled-Biotron Breeding Conditions

We developed advanced generations using the protocol described by Ohnishi et al. [45] with some modifications. Plants were grown in a growth chamber (CFH-415; Tomy Seiko, Tokyo, Japan) equipped with temperature, light, and humidity controls. Seeds were sterilized in 2.5% sodium hypochlorite and incubated at 30 °C in the dark for 2 days. They were then placed on seedling nursery trays and cultured. Ten-day-old seedlings were transplanted (1 per pot) into 230-mL plastic pots filled (4/5) with granulated rice nursery culture soil. Plants were grown under a long daylength (14/10 h light/dark) for 30 days to accelerate vegetative growth and then under a short daylength (10/14 h light/dark) to accelerate reproductive development. The temperature was maintained at 30/25 °C light/dark. Relative humidity was set to 70% and light intensity was set to 350 μmol m^{-2} s^{-1} (Figure S1). Each plant was restricted to the main culm by removing tillers. The flowers of the female parent were emasculated and pollinated according to Ohnishi et al. [45]. At 10 days after pollination, we rescued embryos from developing seeds and cultured them for 10 days according to the protocol. Healthy rice seedlings were then transplanted and raised to the next breeding step.

4.3. Developing Salt-Tolerant Line by Backcrossing "Kaijin" to "Yukinko-mai"

We performed backcrossing to develop an advanced line for salinity tolerance due to the *hst1* gene derived from "Kaijin" using the recurrent parent "Yukinko-mai" (Figure 1A). F_1 plants were confirmed as heterozygous at the *hst1* (*OsRR22*) locus by Sanger sequencing, and were backcrossed to "Yukinko-mai" to produce BC_1F_1 plants. We followed the same strategy of selecting plants heterozygous at *hst1* and backcrossing to develop BC_2F_1 and BC_3F_1 generations. Selected BC_3F_1 heterozygous plants were self-pollinated to generate BC_3F_2 lines with the donor allele in the homozygous state. We sequenced the genome of BC_3F_2 line #31-2-4 to compare with the genomes of the parents. Self-pollinated seeds of line #31-2-4 were named "YNU31-2-4" (BC_3F_3 generation) and used for phenotypic evaluation.

4.4. Confirmation of Genotypes by Sanger Sequencing

We used a PCR primer set to amplify a 545-bp region around the selected SNP (nucleotide 1975 of the *OsRR22* locus) [32] from genomic DNA extracted from young leaves of 20-day-old plants using the CTAB method [79]. Well defined PCR product was gel-purified with a High Pure PCR Product Purification Kit (Roche Applied Science, Tokyo, Japan). Sanger sequencing was performed using a BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) on a Prism 3130 Genetic Analyzer (Applied Biosystems). Sequence chromatogram data were visualized in FinchTV software (Geospiza, Inc., Seattle, WA, USA) to determine the genotype at the SNP position.

4.5. DNA Library Construction and Whole-Genome Sequencing

Total genomic DNA was extracted from leaves of "Yukinko-mai" and BC_3F_2 line #31-2-4 according to the protocol of Walbot and Warren [80] with some modifications. The quantity of genomic DNA was tested with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and the quality was tested by 0.8% agarose gel electrophoresis. The DNA was sent to Macrogen Japan Corp. (Sakyo-ku, Kyoto, Japan) for Illumina HiSeq X Ten sequencing with NGS libraries prepared by the TruSeq DNA PCR-Free Library Prep Kit (Illumina, Inc., San Diego, CA, USA). The sequence data have been deposited in the DDBJ Sequence Read Archive: DRR151851 (BC_3F_2) and DRR151852 ("Yukinko-mai").

4.6. Read Mapping, Variant Calling, and Variant Annotation

"Kaijin" whole-genome sequencing reads (DRR021949, DRR021950, DRR021951, DRR021952) were downloaded from public databases. The raw paired-end reads from "Yukinko-mai", BC_3F_2 #31-2-4, and "Kaijin" sequences were trimmed in Trimmomatic v. 0.33 software [81] with the following parameters: SLIDINGWINDOW, 8:20; TRAILING, 30; MINLEN, 70. The processed reads were mapped to the Nipponbare reference genome (IRGSP-1.0) by using the BWA-MEM v. 0.7.15 algorithm [82]. PCR duplicates in the binary alignment map (BAM) file of "Kaijin" were marked in Picard Tools v. 1.68 software (http://broadinstitute.github.io/picard/). Then indel realignment and base recalculation were done in Genome Analysis Toolkit (GATK) v. 3.6 software [83]. For multi-sample variant calling, we used GATK HaplotypeCaller in gVCF mode followed by GATK GenotypeGVCFs. We filtered out variants with missing data, multi-allelic sites, heterozygous sites in "Kaijin" and "Yukinko-mai", low coverage depth (DP < 6), and low quality (QUAL < 20). We further filtered out heterozygous variants in BC_3F_2 #31-2-4 outside the range of 40%–60% allele frequency by a custom script, and then visualized the genotype map of BC_3F_2 #31-2-4 in the gtrellis package of R software [84]. We annotated variants in SnpEff v. 4.0e software [85] and summarized the results in the Python programming language. Sequentially, we extracted "HIGH"- and "MODERATE"-impact variants flagged by SnpEff and performed functional annotation analysis based on two agronomic data sets: data set 1, the Overview of Functionally Characterized Genes in Rice Online (OGRO) database [52]; and data set 2, the potential agronomic functional gene set selected by gene-based GWAS of the Japanese rice population [53]. One of non-synonymous variant in *Os09g0356200* was sequenced by Sanger sequencing using forward primer: 5′-cactggaggtcgaaactgct-3′ and reverse primer: 5′-tccggtcccagaaatgaagc-3′. The analyses were all based on gene annotation information and genome sequences from the Rice Annotation Project Database (RAP-DB: http://rapdb.dna.affrc.go.jp/).

4.7. Estimation of Genome Recovery Rate

We estimated the genome recovery rate of BC_3F_2 #31-2-4 by calculating the "Yukinko-mai"-type allele frequency out of total variants as:

$$\text{Genome recovery rate} = \frac{\text{YY} + \text{YK}/2}{\text{YY} + \text{YK} + \text{KK}}$$

where YY = number of "Yukinko-mai" homozygous variants, YK = number of heterozygous variants, and KK = number of "Kaijin" homozygous variants.

4.8. Phenotypic Evaluation under Field Condition

We grew "YNU31-2-4" plants in paddy fields of the Crop Research Center, Niigata University, Japan (37°51'20.75"N 138°57'37.9"E), during May-September in 2018, to evaluate the major agro-morphological traits. The experiment was laid out in a randomized complete block design with three replications. We used "Kaijin" and "Yukinko-mai" as salt tolerant and high yielding check cultivars, respectively. Thirty-day-old seedlings were transplanted at a spacing of 20 cm × 15 cm. All agronomic practices were performed uniformly for all the genotypes following the local cultural practices. Four uniform looking plants of each genotype from the central row of each replication were selected to determine the phenotype. The major agronomic traits such as (1) flag leaf color, (2) flag leaf length (3) flag leaf width, (4) heading date, (5) plant height, (6) tiller number, (7) panicle number, (8) panicle length, (9) spikelet number, (10) grain number, (11) 1000-grain weight, (12) seed setting rate, (13) grain yield, and (14) above ground biomass were determined. Grain length, width, and thickness were determined with a rice grain grader (RGQI20A; Satake, Hiroshima, Japan).

4.9. Growth Conditions and Evaluation of Salinity Tolerance at Seedling Stage

We evaluated "YNU31-2-4" and the parents for seedling-stage salt tolerance in the growth chamber at 26/23 °C (12/12 h) and a relative humidity of 70%. Pre-germinated seeds were placed in 230-mL plastic pots filled with rice nursery culture soil containing 0.5 g N, 0.9 g P, and 0.5 g K/kg. The experiment consisted of four treatments: 0 (control), 50, 75, and 125 mM NaCl (pH 5) The salt stress was imposed ten days after germination. The experiment used four biological replicates, each with 10 seedlings. Phenotype was evaluated 2 weeks after salt was imposed.

4.10. Determination of Leaf Relative Water Content, Chlorophyll, and Proline

A separate experiment was conducted to measure biochemical and physiological traits related to salinity tolerance. "YNU31-2-4", "Kaijin", and "Yukinko-mai" seedlings were cultured hydroponically [86]. Ten-day-old rice seedlings were subjected to 0 or 125 mM NaCl (pH 5.0). Samples were collected 10 days after salt was imposed.

The relative water content (RWC%) of control and salt-treated leaves was determined according to Sade et al. [87] as:

$$\%RWC = (\text{fresh weight} - \text{dry weight})/(\text{turgid weight} - \text{dry weight}) \times 100$$

Fully expanded leaves of plants cultured in the absence or presence of salt stress were harvested at the end of the light period, snap-frozen, and ground to a fine powder in liquid nitrogen with a pestle and mortar. Total chlorophyll content was determined according to Lichtenthaler [88]. Free proline content was measured by a colorimetric assay as described by Bates et al. [89].

4.11. Measurement of Na^+ and K^+ Concentrations

Sodium and potassium ions in shoots and roots were quantified by a wet digestion method [90]. Dried, finely powered plant samples (50 mg) were digested in HNO_3/H_2O_2 solution (2:1) in a microwave oven for 4–5 min until the solution became clear. The digested solution was shaken gently and filtered through 0.2-µm filters (Whatman, Maidstone, England), and the solid fraction was discarded. The contents of Na^+ and K^+ in the extract were quantified by atomic absorption spectrophotometry (Z-6100, Hitachi, Tokyo, Japan).

For localization of sodium and potassium ions, we prepared samples according to the protocol of Mitsui et al. [91]. Harvested basal portions of shoots were immediately frozen and embedded in OCT compound medium (Sakura Finetek USA, Inc., Torrance, CA, USA), which contained 10.24%

w/w polyvinyl alcohol, 4.26% *w/w* polyethylene glycol, and 85.50% *w/w* of a nonreactive ingredient. Then 5-μm sections were scanned with an electron probe microanalyzer (EPMA-1605; Shimadzu, Kyoto, Japan).

4.12. Evaluation of Salt-Stress Tolerance at Reproductive Stage

"YNU31-2-4", salt-tolerant "Kaijin", and susceptible "Yukinko-mai" plants were evaluated for salt-stress tolerance at the reproductive stage in a semi-controlled greenhouse. Thirty-day-old seedlings were transplanted (1 per bucket) into 2.5-L plastic bucket and each treatment had six replicates. Plants were subjected to 0 or 50 mM NaCl in irrigation water (pH 5) at 60 days after germination (DAG). After 2 weeks, the salt concentration was increased to 75 mM until booting stage (95 DAG), and then plants were recovered by irrigating with fresh water. The net assimilation rate of penultimate leaves was measured with an LI-6400 gas exchange system (LI-COR Inc., Lincoln, NE, USA) 4 weeks after salt was imposed. Gas exchange was determined at 25 °C at a photosynthetic photon flux density of 350 μmol m^{-2} s^{-1}. Yield and its attributes, particularly panicle number, spikelet number, and 1000-spikelet weight were determined at harvest.

4.13. Statistical Analysis

Values are presented as mean ± standard deviation (SD). Means were tested by analysis of variance (ANOVA) followed by Tukey's or Duncan's multiple range test at $p < 0.05$ in SPSS software (SPSS Inc., Chicago, IL, USA).

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/10/2585/s1. Table S1: Cross-efficiency rate and advanced generation duration in biotron breeding system. Table S2: Annotation of SNPs and indels in BC_3F_2 #31-2-4. Figure S1: Schematic representation of biotron speed-breeding system. Figure S2: Frameshift deletion in *Os09g0356200* affects days-to-heading of YNU31-2-4 plants. Figure S3: Agronomic trait assessment of "Kaijin", WT, and "YNU31-2-4" plants under control field condition. Figure S4: "YNU31-2-4" maintains higher relative water, chlorophyll, and proline contents at the seedling stage under salt stress. Figure S5: "YNU31-2-4" plants have an enlarged root system.

Author Contributions: M.M.R., T.T., M.B., K.K. and T.M. designed the experiments and analyzed the data; M.M.R. performed most of the experiments; T.T. performed the computational analysis of N.G.S. data; K.K., N.H., T.S. and T.O. contributed analytic tools; K.I., T.O., T.K. and H.T. jointly conceived the work; M.M.R., M.B. and T.M. wrote the article with contributions from all authors; and T.M. conceived the project and research plans and supervised the research.

Funding: This research was supported by KAKENHI Grants-in-Aid for Scientific Research (A) (15H02486) from Japan Society for the Promotion of Sciences, Strategic International Collaborative Research Program by the Japan Science and Technology Agency (JST SICORP), and Grant for Promotion of KAAB Projects (Niigata University) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

Acknowledgments: Computations were performed on the NIG supercomputer at the ROIS National Institute of Genetics.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wassmann, R.; Jagadish, S.V.K.; Heuer, S.; Ismail, A.; Redona, E.; Serraj, R.; Singh, R.K.; Howell, G.; Pathak, H.; Sumfleth, K. Chapter 2 Climate Change Affecting Rice Production. The Physiological and Agronomic Basis for Possible Adaptation Strategies. *Adv. Agron.* **2009**, *101*, 59–122.
2. Welch, J.R.; Vincent, J.R.; Auffhammer, M.; Moya, P.F.; Dobermann, A.; Dawe, D. Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14562–14567. [CrossRef]
3. Sreenivasulu, N.; Butardo, V.M.; Misra, G.; Cuevas, R.P.; Anacleto, R.; Kavi Kishor, P.B. Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. *J. Exp. Bot.* **2015**, *66*, 1737–1748. [CrossRef] [PubMed]

4. Calanca, P.P. Effects of Abiotic Stress in Crop Production. In *Quantification of Climate Variability, Adaptation and Mitigation for Agricultural Sustainability*; Springer International Publishing: Cham, Switzerland, 2017; pp. 165–180.
5. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food security: The challenge of feeding 9 billion people. *Science* **2010**, *327*, 812–818. [CrossRef]
6. United Nations. World Population Prospects: The 2012 Revision. Highlights and Advance Tables. *Popul. Dev. Rev.* **2013**, *36*, 775–801. [CrossRef]
7. Munns, R.; Tester, M. Mechanisms of Salinity Tolerance. *Annu. Rev. Plant Biol.* **2008**, *59*, 651–681. [CrossRef] [PubMed]
8. Rao, P.S.; Mishra, B.; Gupta, S.R.; Rathore, A. Reproductive stage tolerance to salinity and alkalinity stresses in rice genotypes. *Plant Breed.* **2008**, *127*, 256–261. [CrossRef]
9. Negrão, S.; Courtois, B.; Ahmadi, N.; Abreu, I.; Saibo, N.; Oliveira, M.M. Recent updates on salinity stress in rice: From physiological to molecular responses. *CRC Crit. Rev. Plant Sci.* **2011**, *30*, 329–377. [CrossRef]
10. Marcos, M.; Sharifi, H.; Grattan, S.R.; Linquist, B.A. Spatio-temporal salinity dynamics and yield response of rice in water-seeded rice fields. *Agric. Water Manag.* **2018**, *195*, 37–46. [CrossRef]
11. Todaka, D.; Nakashima, K.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. *Rice* **2012**, *5*, 6. [CrossRef]
12. Horie, T.; Karahara, I.; Katsuhara, M. Salinity tolerance mechanisms in glycophytes: An overview with the central focus on rice plants. *Rice* **2012**, *5*, 11. [CrossRef] [PubMed]
13. Roy, S.J.; Negrão, S.; Tester, M. Salt resistant crop plants. *Curr. Opin. Biotechnol.* **2014**, *26*, 115–124. [CrossRef] [PubMed]
14. Deinlein, U.; Stephan, A.B.; Horie, T.; Luo, W.; Xu, G.; Schroeder, J.I. Plant salt-tolerance mechanisms. *Trends Plant Sci.* **2014**, *19*, 371–378. [CrossRef] [PubMed]
15. Hanin, M.; Ebel, C.; Ngom, M.; Laplaze, L.; Masmoudi, K. New Insights on Plant Salt Tolerance Mechanisms and Their Potential Use for Breeding. *Front. Plant Sci.* **2016**, *7*, 1787. [CrossRef] [PubMed]
16. Reddy, I.N.B.L.; Kim, B.-K.; Yoon, I.-S.; Kim, K.-H.; Kwon, T.-R. Salt Tolerance in Rice: Focus on Mechanisms and Approaches. *Rice Sci.* **2017**, *24*, 123–144. [CrossRef]
17. Chen, Z.C.; Yamaji, N.; Horie, T.; Che, J.; Li, J.; An, G.; Ma, J.F. A Magnesium Transporter OsMGT1 Plays a Critical Role in Salt Tolerance in Rice. *Plant Physiol.* **2017**, *174*, 1837–1849. [CrossRef]
18. Bonilla, D.; Mackill, D.; Deal, K.; Gregorio, G. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (*Oryza sativa* L.) using recombinant inbred lines. *J. Agric. Food Chem.* **2002**, *85*, 68–76.
19. Thomson, M.J.; de Ocampo, M.; Egdane, J.; Rahman, M.A.; Sajise, A.G.; Adorada, D.L.; Tumimbang-Raiz, E.; Blumwald, E.; Seraj, Z.I.; Singh, R.K.; et al. Characterizing the Saltol quantitative trait locus for salinity tolerance in rice. *Rice* **2010**, *3*, 148–160. [CrossRef]
20. Lin, H.X.; Zhu, M.Z.; Yano, M.; Gao, J.P.; Liang, Z.W.; Su, W.A.; Hu, X.H.; Ren, Z.H.; Chao, D.Y. QTLs for Na^+ and K^+ uptake of the shoots and roots controlling rice salt tolerance. *Theor. Appl. Genet.* **2004**, *108*, 253–260. [CrossRef]
21. Ren, Z.-H.; Gao, J.-P.; Li, L.-G.; Cai, X.-L.; Huang, W.; Chao, D.-Y.; Zhu, M.-Z.; Wang, Z.-Y.; Luan, S.; Lin, H.-X. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat. Genet.* **2005**, *37*, 1141–1146. [CrossRef] [PubMed]
22. Vu, H.T.T.; Le, D.D.; Ismail, A.M.; Le, H.H. Marker-assisted backcrossing (MABC) for improved salinity tolerance in rice (*Oryza sativa* L.) to cope with climate change in Vietnam. *Aust. J. Crop Sci.* **2012**, *6*, 1649–1654.
23. Linh, L.H.; Linh, T.H.; Xuan, T.D.; Ham, L.H.; Ismail, A.M.; Khanh, T.D. Molecular Breeding to Improve Salt Tolerance of Rice (Oryza sativa L.) in the Red River Delta of Vietnam. *Int. J. Plant Genom.* **2012**, *2012*, 949038. [CrossRef]
24. Gregorio, G.B.; Islam, M.R.; Vergara, G.V.; Thirumeni, S. Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. *Sabrao J. Breed. Genet.* **2013**, *45*, 31–41.
25. Babu, N.N.; Krishnan, S.G.; Vinod, K.K.; Krishnamurthy, S.L.; Singh, V.K.; Singh, M.P.; Singh, R.; Ellur, R.K.; Rai, V.; Bollinedi, H.; et al. Marker Aided Incorporation of Saltol, a Major QTL Associated with Seedling Stage Salt Tolerance, into Oryza sativa "Pusa Basmati 1121". *Front. Plant Sci.* **2017**, *8*, 41. [CrossRef] [PubMed]
26. Quan, R.; Wang, J.; Hui, J.; Bai, H.; Lyu, X.; Zhu, Y.; Zhang, H.; Zhang, Z.; Li, S.; Huang, R. Improvement of Salt Tolerance Using Wild Rice Genes. *Front. Plant Sci.* **2018**, *8*, 2269. [CrossRef]

27. Negrão, S.; Cecília Almadanim, M.; Pires, I.S.; Abreu, I.A.; Maroco, J.; Courtois, B.; Gregorio, G.B.; McNally, K.L.; Margarida Oliveira, M. New allelic variants found in key rice salt-tolerance genes: An association study. *Plant Biotechnol. J.* **2013**, *11*, 87–100. [CrossRef]
28. Jain, M.; Moharana, K.C.; Shankar, R.; Kumari, R.; Garg, R. Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnol. J.* **2014**, *12*, 253–264. [CrossRef] [PubMed]
29. Rahman, M.A.; Thomson, M.J.; Shah-E-Alam, M.; De Ocampo, M.; Egdane, J.; Ismail, A.M. Exploring novel genetic sources of salinity tolerance in rice through molecular and physiological characterization. *Ann. Bot.* **2016**, *117*, 1083–1097. [CrossRef] [PubMed]
30. Tiwari, S.; Sl, K.; Kumar, V.; Singh, B.; Rao, A.R.; Sv, A.M.; Rai, V.; Singh, A.K.; Singh, N.K. Mapping QTLs for Salt Tolerance in Rice (*Oryza sativa* L.) by Bulked Segregant Analysis of Recombinant Inbred Lines Using 50K SNP Chip. *PLoS ONE* **2016**, *11*, e0153610. [CrossRef]
31. Mishra, S.; Singh, B.; Panda, K.; Singh, B.P.; Singh, N.; Misra, P.; Rai, V.; Singh, N.K. Association of SNP Haplotypes of HKT Family Genes with Salt Tolerance in Indian Wild Rice Germplasm. *Rice* **2016**, *9*, 8. [CrossRef]
32. Takagi, H.; Tamiru, M.; Abe, A.; Yoshida, K.; Uemura, A.; Yaegashi, H.; Obara, T.; Oikawa, K.; Utsushi, H.; Kanzaki, E.; et al. MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* **2015**, *33*, 445–449. [CrossRef] [PubMed]
33. Abe, A.; Takagi, H.; Nakajo, S.; Terauchi, R.; Kuroda, E. Yielding ability and physicochemical characteristics related to eating quality in salt-tolerant rice variety "Kaijin" under normal salt-free cultivated condition. *Tohoku J. Crop Sci.* **2016**, *59*, 19–20. [CrossRef]
34. Ishizaki, K.; Matsui, T.; Kaneda, S.; Kobayashi, K.; Kasaneyama, H.; Abe, S.; Azuma, S.; Hoshi, T.; Sasaki, Y.; Hirao, K.; et al. A New Rice Cultivar "Yukinko-mai". *J. Niigata Agric. Res. Inst.* **2008**, *9*, 89–98.
35. Shiraya, T.; Mori, T.; Maruyama, T.; Sasaki, M.; Takamatsu, T.; Oikawa, K.; Itoh, K.; Kaneko, K.; Ichikawa, H.; Mitsui, T. Golgi/plastid-type manganese superoxide dismutase involved in heat-stress tolerance during grain filling of rice. *Plant Biotechnol. J.* **2015**, *13*, 1251–1263. [CrossRef]
36. Lee, K.-S.; Choi, W.-Y.; Ko, J.-C.; Kim, T.-S.; Gregorio, G.B. Salinity tolerance of japonica and indica rice (*Oryza sativa* L.) at the seedling stage. *Planta* **2003**, *216*, 1043–1046. [CrossRef]
37. Roy, K.; Sasada, K.; Kohno, E. Salinity status of the 2011 Tohoku-oki tsunami affected agricultural lands in northeast Japan. *Int. Soil Water Conserv. Res.* **2014**, *2*, 40–50. [CrossRef]
38. Kurotani, K.; Yamanaka, K.; Toda, Y.; Ogawa, D.; Tanaka, M.; Kozawa, H.; Nakamura, H.; Hakata, M.; Ichikawa, H.; Hattori, T.; et al. Stress Tolerance Profiling of a Collection of Extant Salt-Tolerant Rice Varieties and Transgenic Plants Overexpressing Abiotic Stress Tolerance Genes. *Plant Cell Physiol.* **2015**, *56*, 1867–1876. [CrossRef] [PubMed]
39. Collard, B.C.Y.; Mackill, D.J. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 557–572. [CrossRef] [PubMed]
40. Miah, G.; Rafii, M.Y.; Ismail, M.R.; Puteh, A.B.; Rahim, H.A.; Asfaliza, R.; Latif, M.A. Blast resistance in rice: A review of conventional breeding to molecular approaches. *Mol. Biol. Rep.* **2013**, *40*, 2369–2388. [CrossRef] [PubMed]
41. Miah, G.; Rafii, M.Y.; Ismail, M.R.; Puteh, A.B.; Rahim, H.A.; Latif, M.A. Recurrent parent genome recovery analysis in a marker-assisted backcrossing program of rice (*Oryza sativa* L.). *C. R. Biol.* **2015**, *338*, 83–94. [CrossRef]
42. Fukuoka, S.; Ebana, K.; Yamamoto, T.; Yano, M. Integration of Genomics into Rice Breeding. *Rice* **2010**, *3*, 131–137. [CrossRef]
43. Luo, Y.; Ma, T.; Zhang, A.; Ong, K.H.; Li, Z.; Yang, J.; Yin, Z. Marker-assisted breeding of the rice restorer line Wanhui 6725 for disease resistance, submergence tolerance and aromatic fragrance. *Rice* **2016**, *9*, 66. [CrossRef] [PubMed]
44. Das, G.; Patra, J.K.; Baek, K.-H. Insight into MAS: A Molecular Tool for Development of Stress Resistant and Quality of Rice through Gene Stacking. *Front. Plant Sci.* **2017**, *8*, 985. [CrossRef] [PubMed]
45. Ohnishi, T.; Yoshino, M.; Yamakawa, H.; Kinoshita, T. The biotron breeding system: A rapid and reliable procedure for genetic studies and breeding in rice. *Plant Cell Physiol.* **2011**, *52*, 1249–1257. [CrossRef] [PubMed]

46. Tanaka, J.; Hayashi, T.; Iwata, H. A practical, rapid generation-advancement system for rice breeding using simplified biotron breeding system. *Breed. Sci.* **2016**, *66*, 542–551. [CrossRef]
47. Collard, B.C.Y.; Beredo, J.C.; Lenaerts, B.; Mendoza, R.; Santelices, R.; Lopena, V.; Verdeprado, H.; Raghavan, C.; Gregorio, G.B.; Vial, L.; et al. Revisiting rice breeding methods—Evaluating the use of rapid generation advance (RGA) for routine rice breeding. *Plant Prod. Sci.* **2017**, *20*, 337–352. [CrossRef]
48. Watson, A.; Ghosh, S.; Williams, M.J.; Cuddy, W.S.; Simmonds, J.; Rey, M.D.; Asyraf Md Hatta, M.; Hinchliffe, A.; Steed, A.; Reynolds, D.; et al. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* **2018**, *4*, 23–29. [CrossRef]
49. Nagatoshi, Y.; Fujita, Y. Accelerating Soybean Breeding in a CO_2-Supplemented Growth Chamber. *Plant Cell Physiol.* **2019**, *60*, 77–84. [CrossRef] [PubMed]
50. De Haas, L.S.; Koopmans, R.; Lelivelt, C.L.C.; Ursem, R.; Dirks, R.; Velikkakam James, G. Low-coverage resequencing detects meiotic recombination pattern and features in tomato RILs. *DNA Res.* **2017**, *24*, 549–558. [CrossRef] [PubMed]
51. Si, W.; Yuan, Y.; Huang, J.; Zhang, X.; Zhang, Y.; Zhang, Y.; Tian, D.; Wang, C.; Yang, Y.; Yang, S. Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F_2 plants. *New Phytol.* **2015**, *206*, 1491–1502. [CrossRef]
52. Yamamoto, E.; Yonemaru, J.-I.; Yamamoto, T.; Yano, M. OGRO: The Overview of functionally characterized Genes in Rice online database. *Rice* **2012**, *5*, 26. [CrossRef]
53. Yano, K.; Yamamoto, E.; Aya, K.; Takeuchi, H.; Lo, P.C.; Hu, L.; Yamasaki, M.; Yoshida, S.; Kitano, H.; Hirano, K.; et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **2016**, *48*, 927–934. [CrossRef] [PubMed]
54. Munns, R. Genes and Salt Tolerance. *New Phytol.* **2005**, *167*, 645–663. [CrossRef] [PubMed]
55. Inui, T.; Yasutaka, T.; Endo, K.; Katsumi, T. Geo-environmental issues induced by the 2011 off the Pacific Coast of Tohoku Earthquake and tsunami. *Soils Found.* **2012**, *52*, 856–871. [CrossRef]
56. Iftekharuddaula, K.M.; Salam, M.A.; Newaz, M.A.; Ahmed, H.U.; Collard, B.C.Y.; Septiningsih, E.M.; Sanchez, D.L.; Pamplona, A.M.; Mackill, D.J. Comparison of phenotypic versus marker-assisted background selection for the SUB1 QTL during backcrossing in rice. *Breed. Sci.* **2012**, *62*, 216–222. [CrossRef] [PubMed]
57. Hasan, M.M.; Rafii, M.Y.; Ismail, M.R.; Mahmood, M.; Rahim, H.A.; Alam, M.A.; Ashkani, S.; Malek, M.A.; Latif, M.A. Marker-assisted backcrossing: A useful method for rice improvement. *Biotechnol. Biotechnol. Equip.* **2015**, *29*, 237–254. [CrossRef]
58. Tanweer, F.A.; Rafii, M.Y.; Sijam, K.; Rahim, H.A.; Ahmed, F.; Ashkani, S.; Latif, M.A. Introgression of Blast Resistance Genes (Putative Pi-b and Pi-kh) into Elite Rice Cultivar MR219 through Marker-Assisted Selection. *Front. Plant Sci.* **2015**, *6*, 1002. [CrossRef]
59. Jagadish, S.V.K.; Septiningsih, E.M.; Kohli, A.; Thomson, M.J.; Ye, C.; Redoña, E.; Kumar, A.; Gregorio, G.B.; Wassmann, R.; Ismail, A.M.; et al. Genetic Advances in Adapting Rice to a Rapidly Changing Climate. *J. Agron. Crop Sci.* **2012**, *198*, 360–373. [CrossRef]
60. Ma, Y.-C.; Augé, R.M.; Dong, C.; Cheng, Z.-M.M. Increased salt tolerance with overexpression of cation/proton antiporter 1 genes: A meta-analysis. *Plant Biotechnol. J.* **2017**, *15*, 162–173. [CrossRef] [PubMed]
61. Hayat, S.; Hayat, Q.; Alyemeni, M.N.; Wani, A.S.; Pichtel, J.; Ahmad, A. Role of proline under changing environments: A review. *Plant Signal. Behav.* **2012**, *7*, 1456–1466. [CrossRef]
62. Saxena, S.C.; Kaur, H.; Verma, P.; Petla, B.P.; Andugula, V.R.; Majee, M. Osmoprotectants: Potential for Crop Improvement Under Adverse Conditions. In *Plant Acclimation to Environmental Stress*; Springer: New York, NY, USA, 2013; pp. 197–232.
63. Szabados, L.; Savouré, A. Proline: A multifunctional amino acid. *Trends Plant Sci.* **2010**, *15*, 89–97. [CrossRef]
64. Gupta, B.; Huang, B. Mechanism of Salinity Tolerance in Plants: Physiological, Biochemical, and Molecular Characterization. *Int. J. Genom.* **2014**, *2014*, 701596. [CrossRef]
65. Zeng, L.; Shannon, M.C. Salinity Effects on Seedling Growth and Yield Components of Rice. *Crop Sci.* **2000**, *40*, 996. [CrossRef]
66. Walia, H.; Wilson, C.; Zeng, L.; Ismail, A.M.; Condamine, P.; Close, T.J. Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Mol. Biol.* **2007**, *63*, 609–623. [CrossRef] [PubMed]

67. Zhou, Y.-B.; Liu, C.; Tang, D.-Y.; Yan, L.; Wang, D.; Yang, Y.-Z.; Gui, J.-S.; Zhao, X.-Y.; Li, L.-G.; Tang, X.-D.; et al. The Receptor-Like Cytoplasmic Kinase STRK1 Phosphorylates and Activates CatC, Thereby Regulating H_2O_2 Homeostasis and Improving Salt Tolerance in Rice. *Plant Cell* **2018**, *30*, 1100–1118. [CrossRef]
68. Murchie, E.H.; Yang, J.; Hubbart, S.; Horton, P.; Peng, S. Are there associations between grain-filling rate and photosynthesis in the flag leaves of field-grown rice? *J. Exp. Bot.* **2002**, *53*, 2217–2224. [CrossRef]
69. Zhai, H.; Cao, S.; Wan, J.; Zhang, R.; Lu, W.; Li, L.; Kuang, T.; Min, S.; Zhu, D.; Cheng, S. Relationship between leaf photosynthetic function at grain filling stage and yield in super high-yielding hybrid rice (*Oryza sativa* L). *Sci. China Ser. C Life Sci.* **2002**, *45*, 637–646. [CrossRef]
70. Kasai, M. Regulation of leaf photosynthetic rate correlating with leaf carbohydrate status and activation state of Rubisco under a variety of photosynthetic source/sink balances. *Physiol. Plant.* **2008**, *134*, 216–226. [CrossRef]
71. Sanchez-Bragado, R.; Elazab, A.; Zhou, B.; Serret, M.D.; Bort, J.; Nieto-Taladriz, M.T.; Araus, J.L. Contribution of the ear and the flag leaf to grain filling in durum wheat inferred from the carbon isotope signature: Genotypic and growing conditions effects. *J. Integr. Plant Biol.* **2014**, *56*, 444–454. [CrossRef]
72. Zhou, W.; Lv, T.; Yang, Z.; Wang, T.; Fu, Y.; Chen, Y.; Hu, B.; Ren, W. Morphophysiological mechanism of rice yield increase in response to optimized nitrogen management. *Sci. Rep.* **2017**, *7*, 17226. [CrossRef] [PubMed]
73. Liu, W.; Li, R.-J.; Han, T.-T.; Cai, W.; Fu, Z.-W.; Lu, Y.-T. Salt stress reduces root meristem size by nitric oxide-mediated modulation of auxin accumulation and signaling in Arabidopsis. *Plant Physiol.* **2015**, *168*, 343–356. [CrossRef]
74. Yu, J.; Zhao, W.; Tong, W.; He, Q.; Yoon, M.-Y.; Li, F.-P.; Choi, B.; Heo, E.-B.; Kim, K.-W.; Park, Y.-J.; et al. A Genome-Wide Association Study Reveals Candidate Genes Related to Salt Tolerance in Rice (*Oryza sativa*) at the Germination Stage. *Int. J. Mol. Sci.* **2018**, *19*, 3145. [CrossRef] [PubMed]
75. Ali, A.; Maggio, A.; Bressan, R.; Yun, D.-J.; Ali, A.; Maggio, A.; Bressan, R.A.; Yun, D.-J. Role and Functional Differences of HKT1-Type Transporters in Plants under Salt Stress. *Int. J. Mol. Sci.* **2019**, *20*, 1059. [CrossRef]
76. Okada, T.; Nakayama, H.; Shinmyo, A.; Yoshida, K. Expression of OsHAK genes encoding potassium ion transporters in rice. *Plant Biotechnol.* **2008**, *25*, 241–245. [CrossRef]
77. Hauser, F.; Horie, T. A conserved primary salt tolerance mechanism mediated by HKT transporters: A mechanism for sodium exclusion and maintenance of high K^+/Na^+ ratio in leaves during salinity stress. *Plant. Cell Environ.* **2010**, *33*, 552–565. [CrossRef]
78. Chen, Z.; Pottosin, I.I.; Cuin, T.A.; Fuglsang, A.T.; Tester, M.; Jha, D.; Zepeda-Jazo, I.; Zhou, M.; Palmgren, M.G.; Newman, I.A.; et al. Root plasma membrane transporters controlling K+/Na+ homeostasis in salt-stressed barley. *Plant Physiol.* **2007**, *145*, 1714–1725. [CrossRef]
79. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
80. Walbot, V.; Warren, C. Regulation of Mu element copy number in maize lines with an active or inactive Mutator transposable element system. *Mol. Gen. Genet.* **1988**, *211*, 27–34. [CrossRef]
81. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
82. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
83. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]
84. Gu, Z.; Eils, R.; Schlesner, M. Gtrellis: An R/Bioconductor package for making genome-level Trellis graphics. *BMC Bioinform.* **2016**, *17*, 169. [CrossRef] [PubMed]
85. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]
86. Yoshida, S.; Forno, D.A.; Cock, J.H.; Gomez, K.A. *Laboratory Manual for Physiological Studies of Rice*, 3rd ed.; International Rice Research Institute: Manila, Philippines, 1976; pp. 61–66.
87. Sade, N.; Vinocur, B.J.; Diber, A.; Shatil, A.; Ronen, G.; Nissan, H.; Wallach, R.; Karchi, H.; Moshelion, M. Improving plant stress tolerance and yield production: Is the tonoplast aquaporin SlTIP2;2 a key to isohydric to anisohydric conversion? *New Phytol.* **2009**, *181*, 651–661. [CrossRef] [PubMed]

88. Lichtenthaler, H.K. Chlorophylls and Carotenoids: Pigments of Photosynthetic Biomembranes. *Methods Enzymol.* **1987**, *148*, 350–382. [CrossRef]
89. Bates, L.S.; Waldren, R.P.; Teare, I.D. Rapid determination of free proline for water-stress studies. *Plant Soil* **1973**, *39*, 205–207. [CrossRef]
90. Pequerul, A.; Pérez, C.; Madero, P.; Val, J.; Monge, E. A rapid wet digestion method for plant analysis. In *Optimization of Plant Nutrition*; Springer: Dordrecht, The Netherlands, 1993; pp. 3–6.
91. Mitsui, T.; Loboda, T.; Itoh, A.; Ikarashi, T. Sugar-Controlled Ca^{2+} Uptake and α-Amylase Secretion in Cultured Cells of Rice (*Oryza sativa* L.). *Plant Cell Physiol.* **1999**, *40*, 884–893. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Comparative Analysis of Calcium-Dependent Protein Kinase in Cucurbitaceae and Expression Studies in Watermelon

Chunhua Wei †, Ruimin Zhang †, Xiaozhen Yang, Chunyu Zhu, Hao Li, Yong Zhang, Jianxiang Ma, Jianqiang Yang and Xian Zhang *

State Key Laboratory of Crop Stress Biology in Arid Areas, College of Horticulture, Northwest A&F University, Yangling 712100, China; xjwend020405@nwafu.edu.cn (C.W.); zrm0923@163.com (R.Z.); yxzh5186@126.com (X.Y.); zhucy@nwafu.edu.cn (C.Z.); yuanyilihao123@163.com (H.L.); zhangyong123@nwafu.edu.cn (Y.Z.); majianxiang@126.com (J.M.); yangjq1208@126.com (J.Y.)

* Correspondence: zhangxian@nwsuaf.edu.cn; Tel.: +86-029-8708-2613

† These authors contributed equally to this work.

Received: 21 March 2019; Accepted: 20 May 2019; Published: 23 May 2019

Abstract: Both the calcium-dependent protein kinases (CDPKs) and CDPK-related kinases (CRKs) play numerous roles in plant growth, development, and stress response. Despite genome-wide identification of both families in *Cucumis*, comparative evolutionary and functional analysis of both *CDPKs* and *CRKs* in Cucurbitaceae remain unclear. In this study, we identified 128 *CDPK* and 56 *CRK* genes in total in six Cucurbitaceae species (*C. lanatus*, *C. sativus*, *C. moschata*, *C. maxima*, *C. pepo*, and *L. siceraria*). Dot plot analysis indicated that self-duplication of conserved domains contributed to the structural variations of two *CDPKs* (*CpCDPK19* and *CpCDPK27*) in *C. pepo*. Using watermelon genome as reference, an integrated map containing 25 loci (16 *CDPK* and nine *CRK* loci) was obtained, 16 of which (12 *CDPK* and four *CRK*) were shared by all seven Cucurbitaceae species. Combined with exon-intron organizations, topological analyses indicated an ancient origination of groups CDPK IV and CRK. Moreover, the evolutionary scenario of seven modern Cucurbitaceae species could also be reflected on the phylogenetic trees. Expression patterns of *ClCDPKs* and *ClCRKs* were studied under different abiotic stresses. Some valuable genes were uncovered for future gene function exploration. For instance, both *ClCDPK6* and its ortholog *CsCDPK14* in cucumber could be induced by salinity, while *ClCDPK6* and *ClCDPK16*, as well as their orthologs in *Cucumis*, maintained high expression levels in male flowers. Collectively, these results provide insights into the evolutionary history of two gene families in Cucurbitaceae, and indicate a subset of candidate genes for functional characterizations in the future.

Keywords: calcium-dependent protein kinases; CDPK-related kinases; evolutionary analysis; expression pattern; abiotic stress; cucurbitaceae

1. Introduction

As a ubiquitous second messenger, Calcium (Ca^{2+}) plays an important role in sophisticated signal transduction pathways to survive frequently occurring environmental stresses during plant growth and development [1,2]. Transient changes of the Ca^{2+} concentration in the cytoplasm can be sensed by four types of calcium sensors: calmodulins (CaM), calmodulin-like proteins (CaML), calcineurin B-like proteins (CBL), and the calcium-dependent protein kinase (CDPK) [3,4]. Among these sensors, CDPKs not only sense but also directly translate Ca^{2+} signals into a downstream phosphorylation pathway, thus functioning both as Ca^{2+} sensors and effectors [2,5,6].

The CDPKs (also referred to as CPKs) are ser/thr protein kinases, which consist of four typical domains: a variable N-terminal domain (containing the myristoylation and palmitoylation sites),

a catalytic ser/thr protein kinase domain, an auto-inhibitory domain (acting as a pseudosubstrate combined with the kinase domain to inhibit activity), and a C-terminal regulatory calmodulin-like domain that contains one to four EF-hand motifs for Ca^{2+} binding [7–9]. CDPK-related kinases (CRKs) have similar domain structures than CDPKs (such as the ser/thr kinase domain); however, they do not have EF-hand domains [10,11]. To date, genome-wide identification of *CDPKs* has been widely performed in a large number of plants, e.g., 34 *CDPKs* have been identified in *Arabidopsis* [7], 31 in rice [12], 29 in tomato [10], 18 in melon [11], and 19 in cucumber [9]. Similarly, *CRKs* have also been identified in genomes, e.g., eight *CRKs* in *Arabidopsis* [13], five in rice [14], six in tomato [10], five in pepper [15], and seven in melon [11].

Accumulating evidence indicates that *CDPK* and *CRK* genes are not only involved in plant growth and development, but also in the plant response to abiotic and hormone stresses. For example, AtCPK32 in *Arabidopsis* has been reported to interact with the calcium channel protein CNGC18, thus controlling the polar growth of pollen tubes [16]. In addition, *AtCPK4* and its homolog *AtCPK11* have been reported to function in seed germination and growth, stomatal movement, and response to salt stress [17]. The gene *AtCPK23* acts as a negative regulator and plays important roles in response to drought and salt stresses via controlling K^+ channels; its overexpression can increase stomatal apertures [18]. In grape plants, a large number of *CDPKs* have been reported to be induced in response to various abiotic and biotic stresses, as well as to hormone treatments [19]. A previous study showed that the *CDPK* genes in cucumber are extensively regulated by various stimuli, including salt, cold, heat, waterlogging, and abscisic acid stresses, possibly following different mechanisms [9]. Furthermore, *CmCDPKs* and *CmCRKs* in melon were also shown to be differentially expressed in response to exogenous stresses, such as biotic stress (*Podosphaera xanthii* inoculation), abiotic stresses (salt and cold), and hormone (abscisic acid) treatment [11]. In general, *CDPK* genes are ubiquitously expressed in most of the different plant organs. However, several genes show organ- or tissue-specific expression patterns. For instance, *VvCDPK5* in grape plants can only be detected in pollen [20]. Similarly, *AtCPK17* and *AtCPK34* are both preferentially expressed in mature pollen, regulating the growth of pollen tubes [21].

The Cucurbitaceae family contains several economically important species with already published genomes, including watermelon (*Citrullus lanatus*), melon (*Cucumis melon*), cucumber (*Cucumis sativus*), and bottle gourd (*Lagenaria siceraria*), which belong to the Benincaseae tribe and three *Cucurbita* species (*Cucurbita maxima*, *Cucurbita moschata*, and *Cucurbita pepo*) of the Cucurbiteae tribe [22–27]. Although genome-wide identifications of *CDPKs* and *CRKs* have been performed in the genus *Cucumis* [9,11], comparative evolutionary analysis of both gene families in Cucurbitaceae is lacking. In this study, we identified a total of 128 *CDPK* and 56 *CRK* genes in six Cucurbitaceae species (*C. lanatus*, *C. sativus*, *C. moschata*, *C. maxima*, *C. pepo*, and *L. siceraria*). After mapping these identified genes onto chromosomes of watermelon, we obtained an integrated map including 25 loci (16 *CDPK* and nine *CRK* loci). Evolutionary analyses revealed that four CDPK groups and one CRK group in phylogentic trees could be further divided into loci, consistent with the integrated map. In addition, expression patterns of *ClCDPKs* and *ClCRKs* under different abiotic stresses were also analyzed. Our results provide insights into the evolutionary history of both gene families in Cucurbitaceae, and indicate a subset of candidate genes for future functional analysis.

2. Results

2.1. Genome-Wide Identification of CDPK and CRK Genes

Our previous study verified 18 *CmCDPK* and seven *CmCRK* genes in melon [11], while only 19 *CDPK* homologs were identified genome-wide in cucumber [9]. Hence, to comparatively analyze the *CRK* gene family in Cucurbitaceae, identification of *CRK* genes in cucumber genome was also performed in this study. As a result, a total of 128 *CDPK* (typically containing both STKs_CAMK protein kinase and EF-hand domains) and 56 *CRK* (solely harboring STKs_CAMK protein kinase domain) genes were identified in six Cucurbitaceae species (*C. lanatus*, *C. sativus*, *C. moschata*, *C. maxima*,

C. pepo, and *L. siceraria*). All the identified genes were designated based on their chromosomal locations (Table S2). It is worth noting that two *CDPKs*, referred to as *CsCDPK18* (*Csa018149*) and *CsCDPK15* (*Csa008536*) in the previous study [9], were renamed to *CsCRK6* and *CsCRK7* due to their lack of EF-hand domains. Compared to the similar copy numbers of *CDPK* and *CRK* genes in four species (*C. lanatus*, *C. melon*, *C. sativus*, and *L. siceraria*) of the Benincaseae tribe (Table 1), many more homologs were identified in three Cucurbiteae tribe genomes (*C. moschata*, *C. maxima*, and *C. pepo*). This may have been caused by the whole-genome duplication (WGD) that only occurred in the progenitor of the *Cucurbita* genus [25,26].

Table 1. Numbers and characteristic properties of *CDPKs* and *CRKs* in Cucurbitaceae species.

Species	No. of Genes (CDPK/CRK)	No. of aa (CDPK/CRK)	MW (CDPK/CRK)	pI (CDPK/CRK)	Source
C. lanatus	18/6	491~662/550~696	55.23~74.32/61.88~77.97	5.09~8.84/6.84~9.17	In this study
C. melon	18/7	501~661/561~622	56.25~74.22/63.12~69.44	5.06~8.75/6.91~9.15	Zhang et al, 2017 [11]
C. sativus	17/7	501~661/556~726	56.30~74.30/62.24~81.23	4.99~8.99/7.20~9.33	Xu et al, 2015 [9]; In this study
L. siceraria	17/7	353~1094/555~636	39.92~125.03/62.25~70.80	5.14~8.74/6.82~9.18	In this study
C. moschata	30/12	318~820/517~758	36.26~90.89/57.54~84.44	5.02~9.10/6.49~9.31	In this study
C. maxima	31/12	499~1230/278~687	56.17~138.34/31.62~76.92	5.13~9.03/5.94~9.21	In this study
C. pepo	32/12	439~1393/342~609	49.44~157.82/39.21~68.01	4.84~8.49/5.97~9.28	In this study

Physico-chemical properties of CDPKs and CRKs, including predicted amino acids, molecular weight (MW), and isoelectric point (pI), showed similar ranges in watermelon, melon, and cucumber; however, these characteristics exhibited broader intervals in the remaining four species (Table 1). In this study, six CDPKs were predicted to have long (>1000) amino acid sequences, possibly due to their long CT domains [28]. Compared to CRKs with no EF-hand domain, the majority of CDPKs were predicted to contain four EF-hands, with few exceptions harboring two or three EF-hand motifs (Table S2). Strikingly, two CDPKs (CpCDPK19 and CpCDPK27) in *C. pepo* were confirmed to have more than five EF-hand domains via both online tools ScanProsite and SMART. More detailed information, for example for myristoylation and palmitoylation sites, is also provided in Table S2.

2.2. Structure Variation Analysis of CpCDPK19 and CpCDPK27

As mentioned above, CpCDPK19 and CpCDPK27 in *C. pepo* harbored nine and eight EF-hand motifs, respectively. Previous research indicates that *Cucurbita*, containing two sub-genomes A and B, originated from two progenitors that diverged from one another approximately 30.75 million years ago (Mya) [26]. Sequence analysis showed that CpCDPK5 and CpCDPK13, sharing the highest similarity with CpCDPK19 (amino acid identity: 93.74%) and CpCDPK27 (amino acid identity: 95.84%), contained four and three EF-domains, respectively (Table S2). Dot plot analysis indicated that the EF-hand domains were duplicated in CpCDPK19 compared to CpCDPK5, while both STKs_CAMK protein kinase and EF-hand domains were repeated in CpCDPK27 in contrast to CpCDPK13 (Figure S1). Based on the synteny analysis (Figure 1a), we inferred that a fragment containing four EF-hand motifs in CpCDPK19 was entirely duplicated and inserted into the STKs_CAMK protein kinase domain, and the fifth EF-hand motif resulted from another duplication event. Compared to ClCDPK13, ClCDPK27 harbored three copies of STKs_CAMK protein kinase domain. Moreover, the first EF-hand domain located in the second STKs_CAMK protein kinase, as well as the second and fourth EF-hand domains, were originated from the sixth EF-hand domain, while the third (existed in the third STKs_CAMK protein kinase) and the fifth EF-hand domains were two copies of the seventh motif (Figure 1b). Using the Cucurbit Genomics Database (CuGenDB) [29], both two genes CpCDPK19 and CpCDPK27 could be detected in the fruits of two different materials at transcriptional level (Figure S2). In summary, self-duplication of conserved domains may contribute to the gene structure variations, and result in sub-functionalization or neo-functionalization.

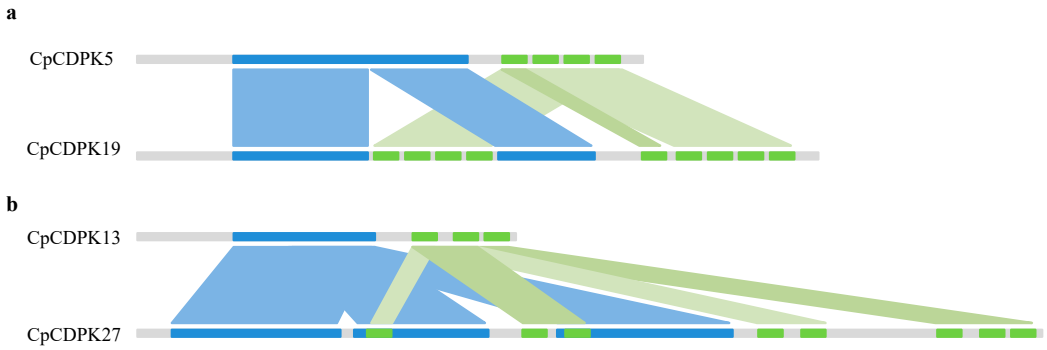


Figure 1. Gene structure variation analysis of CpCDPK19 and CpCDPK27. Syntenic analysis of CpCDPK5 and CpCDPK19 (**a**), as well as CpCDPK12 and CpCDPK27 (**b**). Blue and green rectangles represent STKs_CAMK protein kinase and EF-hand domain, respectively.

2.3. Construction of an Integrated Map for CDPK and CRK Genes

As the genomic distribution of *CDPKs* and *CRKs* in *Cucumis* [9,11], all identified genes in this study were also unevenly distributed on chromosomes in five species, with a few chromosomes harboring none of them (Figure S3). A recent study has confirmed that there are twenty chromosomes in *Cucurbita* species and nineteen of them could be divided into two sub-genomes A and B, except for chromosome 04, consisting of two segments from sub-genome A and one from sub-genome B [26]. Interestingly, almost equal numbers of *CDPKs* and *CRKs* were found to be retained in two sub-genomes (Table S3), consistent with the similar evolved ratios of gene loss or gain of two sub-genomes after polyploidization [26].

The evolutionary scenario of Cucurbitaceae paleohistory hypothesizes that the modern chromosomal structures in cucurbits were derived from an ancestral Cucurbitaceae karyotype (ACK) that consisted of 12 protochromosomes and experienced different times of chromosomal fission and fusion events [27]. Hence, to investigate the conserved loci of *CDPK* and *CRK* in Cucurbitaceae, an integrated map was constructed using watermelon chromosomes as reference, including 16 *CDPK* and nine *CRK* loci (Figure 2). Of these, nine loci (four *CDPK* and five *CRK*) exhibited Presence/Absence polymorphism among genomes, while the remaining 16 loci (12 *CDPK* and four *CRK*) were conserved and shared by all seven species (Figure 2 and Table S4). For example, the first locus on chromosome 01 was lost in *L. siceraria*, while the second locus on chromosome 02 was only absent in watermelon genome.

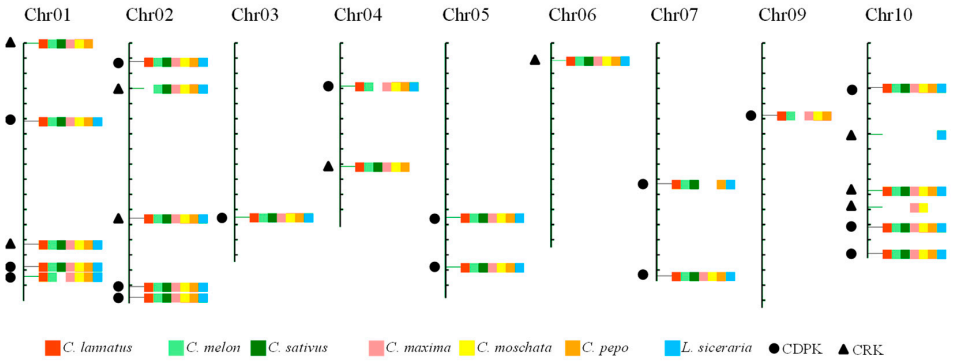


Figure 2. An integrated map of *CDPK* and *CRK* loci in Cucurbitaceae. All *CDPK* and *CRK* genes in Cucurbitaceae were mapped onto 11 chromosomes of watermelon 97103. Loci from *C. lannatus* (red), *C. melon* (light green), *C. sativus* (dark green), *C. maxima* (pink), *C. moschata* (yellow), *C. pepo* (dark yellow), and *L. siceraria* (blue) have been marked with different colors, as indicated in brackets. Black dots and triangles represent *CDPK* and *CRK* loci, respectively.

2.4. Phylogenetic Analysis of CDPK and CRK Gene Families

2.4.1. In the Benincaseae Tribe

In the Benincaseae tribe, watermelon was predicted to diverge from its sister lineage bottle gourd around 10.4–14.6 million years ago (Mya) and from *Cucumis* 17.3–24.3 Mya [27]. To deeply analyze the phylogentic relationships of *CDPK* and *CRK* gene families in these four species, all full-length protein sequences of watermelon (18 CDPKs and six CRKs), melon (18 CDPKs and seven CRKs), cucumber (17 CDPKs and seven CRKs), and bottle gourd (17 CDPKs and seven CRKs) were aligned using the software MUSCLE, and were then used to construct an evolutionary tree using MEGA6.0 (Figure S4). Four CDPK groups (CDPK I, CDPK II, CDPK III, and CDPK IV) and one CRK group (CRK I) were observed in the distance tree. Combined with results from the integrated map, all groups (except for CDPK IV) were found to constitute homologs from at least four loci. Furthermore, most loci contained only one homolog from each individual species. Apparently, homologs from watermelon and bottle gourd were preferentially clustered together, compared to those from *Cucumis* species.

2.4.2. In the Cucurbiteae Tribe

It has been reported that a whole genome duplication (WGD) event has occurred in the progenitor of the *Cucurbita* genus [26,27]. As shown in the phylogenetic tree constructed with all identified full-length protein sequences from *C. moschata* (30 CDPKs and 12 CRKs), *C. maxima* (31 CDPKs and 12 CRKs), and *C. pepo* (32 CDPKs and 12 CRKs), the majority of loci in five groups contained two copies of the homolog gene from each species (Figure S5). Generally, CDPKs or CRKs from the same sub-genome (A or B) of *C. moschata* and *C. maxima* grouped together in the distance tree. For example, both CmoCDPK7 and CmaCDPK8 from sub-genome B of two species clustered together in CDPK IV, while CmoCDPK14 and CmaCDPK15 from sub-genome A formed another sister clade. Based on these observations, origins of CDPKs on chromosome 4 in two *Cucurbita* species could be uncovered, although the boundary sites of three segments (two from sub-genome A and one from B) of chromosome 04 were ambiguous [26], e.g., CmoCDPK26 and CmaCDPK27 in Locus 15 (group CDPK I) were from sub-genome A (Table S3 and Figure S5), suggesting that the members (CmoCDPK9 and CmaCDPK10) in its sister clade originated from sub-genome B. Similarly, CmoCDPK8 and CmaCDPK9 were inferred to originate form sub-genome A, which formed the Locus 20 in group CDPK II. It is worthy to note that both loci 06 and 24 formed two independent clades in the phylogenetic tree, due to having far more CDPK members.

2.4.3. In the Cucurbitaceae Family

To deeply analyze the evolutionary history of *CDPK* and *CRK* gene families in Cucurbitaceae, a total of 226 protein sequences were used to construct a phylogenetic tree (Figure 3). Similar to topological structures mentioned above, four CDPK and one CRK group could be further divided into 25 loci. Notably, six homologs (LsiCRK1, CsCDPK3, CsCDPK19, CpCRK1, CpCDPK1, and CpCDPK2), which failed to be mapped on the integrated map due to lack of enough flanking sequences, were clustered with their orthologs in the phylogenetic trees. CDPKs or CRKs from watermelon and bottle gourd still cluster together in most loci. Compared to sub-genome A, members from sub-genome B usually gathered together with those from the Benincaseae tribe, except for two loci in group CRK I (Figure S6), which is consistent with the evolutionary scenario of modern Cucurbitaceae genomes [27]. CDPK IV, as the smallest group in CDPK lineage, was clustered with CRK I rather than the other three CDPK groups in phylogenetic trees, indicating that groups CDPK IV and CRK I may have originated from a common ancestor [10,11,30].

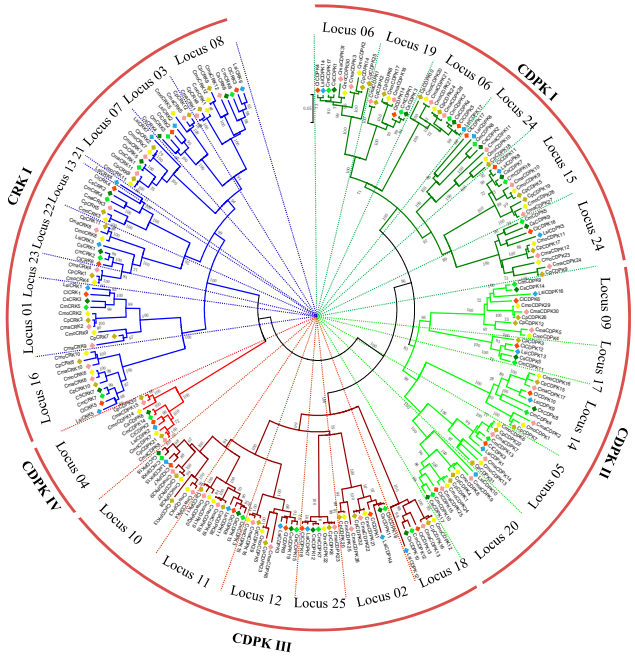


Figure 3. Phylogenetic tree of *CDPK* and *CRK* genes from Cucurbitaceae species. Four CDPK groups and one CRK group can be found in the tree, which were further divided into 25 loci. Numbers on nodes represent bootstrap values, and values <65 are not shown.

For the comparative evolutionary analysis of two gene families in plants, a phylogenetic tree was constructed using the 226 protein sequences in Cucurbitaceae, as well as that from *Arabidopsis* (34 CDPKs and eight CRKs, Cruciferae), tomato (29 CDPKs and six CRKs, Solanaceae), pepper (31 CDPKs and five CRKs, Solanaceae), and rice (31 CDPKs and five CRKs, Poaceae) (Figure S7). There is no doubt that all groups contained homologs from these four families, suggesting that both gene families have conserved basal architectures in their evolutionary process.

2.5. Exon-Intron Organization in CDPK and CRK Groups

The exon-intron organization, as well as the intron numbers, can also provide important evidence to analyze the evolutionary history within gene families [9,11,30]. To obtain further insight into the phylogenetic relationships of two gene families in Cucurbitaceae, gene structures of all 226 CDPKs and CRKs were comparatively depicted, dependent on their gene annotation profiles and genomic sequences (Figure S8). The majority of members in group CDPK I contained six introns with a distinct intron phase pattern 111000, while most CDPKs in group CDPK II had seven introns, sharing a similar intron pattern of 1110020. CDPK III, as a peripheral sister clade of groups CDPK I and II, contained two major intron phases. For instance, 13 out of 17 CDPKs in loci 02 and 25 shared an intron phase 111000, which is identical to that of group CDPK I, while 28 of 37 CDPKs in the other four loci had an intron pattern of 0111000 (Figure S8). CDPK IV, as the smallest group, contained only 10 members with 11 or 12 introns. Among of them, eight CDPKs were constituted of 11 introns with a phase pattern of 02201010000, while the remaining two genes (*CsCDPK6* and *CmoCDPK7*) had 12 introns with an extra intron gain at the 5′ or 3′ end. In group CRK I, 41 out of 63 members contained 10 introns with a phase pattern 0220110000, showing high similarity with that in group CDPK IV. Notably, members in the same loci usually had similar exon or intron lengths, such as Locus 25 in group CDPK III.

2.6. Duplication and Syntenic Analysis of CDPK and CRK Gene Families

In addition to whole genome duplication (WGD) events, both tandem and segmental duplications have also been reported to play vital roles in the expansion and function of a gene family [11,31,32]. To further explore the possible evolutionary relationships of *CDPK* and *CRK* gene families in

Cucurbitaceae, duplication events were investigated in six species (including watermelon, cucumber, bottle gourd, and three *Cucurbita* species). Similar to the melon genome [11], only one or two segmental duplication events were detected in three Benincaceae genomes, with syntenic regions no more than 3.0 Mb (Table S5). However, many more segmental duplication events were observed in *Cucurbita* genus, most of which occurred between sub-genomes.

In the Benincaseae tribe, watermelon, melon, and cucumber are important cucurbit crops widely cultivated throughout the world. Synteny analyses revealed that similar numbers of syntenic regions were detected among three genomes, with average fragment lengths not exceeding 2.5 Mb (Figure 4 and Table S6). As a close sister lineage of *Citrullus*, approximately 15 collinear regions were found between watermelon and bottle gourd, with the largest one spanning about 16.6 Mb. In the Cucurbiteae tribe, many more collinear regions were detected, with the largest one (21.3 Mb) existing between *C. moschata* and *C. pepo* (Table S6).

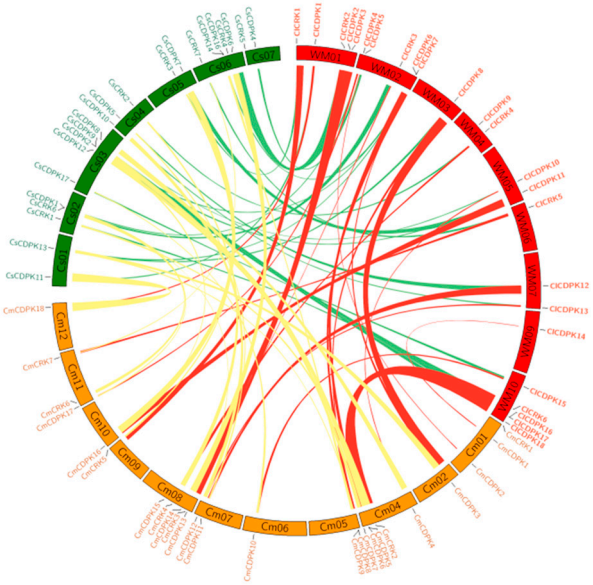


Figure 4. Synteny analysis of *CDPK* and *CRK* genes among watermelon, cucumber, and melon. Chromosomes of three species (watermelon, cucumber, and melon) were depicted in different colors (red, green, and yellow) and in circle form. The approximate distributions of each *CDPK* and *CRK* are presented by short black lines on the circle. Colored curves denote the details of syntenic regions containing *CDPK* and *CRK* genes among genomes.

2.7. Expression Profiles of ClCDPK and ClCRK Genes in Different Tissues

To assess the potential functions of *ClCDPK* and *ClCRK* genes, their expression patterns were investigated in six different tissues, including roots, stems, leaves, tendrils, male flowers, and female flowers. As shown in Figure 5, all identified *ClCDPKs* and *ClCRKs* could be detected in at least one tissue. Some *ClCDPKs* and *ClCRKs*, such as *ClCDPK2*, *ClCDPK3*, *ClCDPK9*, *ClCDPK15*, and *ClCRK1*, showed significantly elevated expression levels in the root, while *ClCDPK6*, *ClCDPK16*, and *ClCDPK17* were strongly expressed in the male flower. Interestingly, *CmCDPK6* in melon, the ortholog of *ClCDPK17* in locus 24, was also reported to have a high transcriptional abundance in the male flower [11]. Moreover, orthologs of *ClCDPK6* and *ClCDPK16* in melon (*CmCDPK9* and *CmCDPK5*) and cucumber (*CsCDPK14* and *CsCDPK9*) were also confirmed to retain high accumulations in male flowers [9,11].

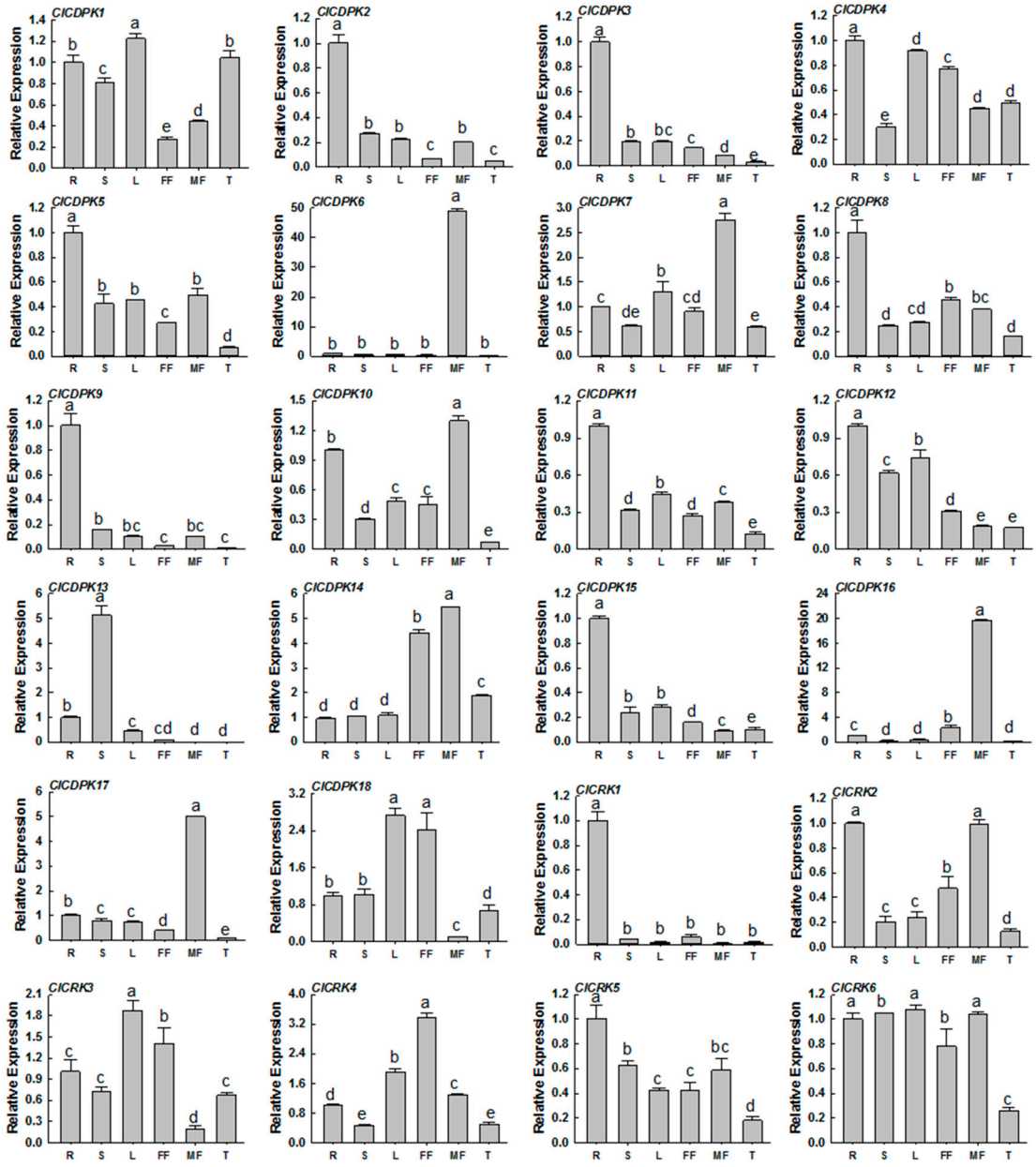


Figure 5. Expression profiles of *ClCDPK* and *ClCRK* genes in different tissues. The transcript levels of the respective genes in roots were used as references and set to a value of 1. The data were showed as means value ± SD. All experiments were performed with three independent replicates. R = roots; S = stems; L = leave; FF = female flowers; MF = male flowers; T = tendrils.

2.8. Expression Patterns of ClCDPK and ClCRK Genes under Abiotic Stresses

Accumulation studies showed that *CDPK* and *CRK* genes are widely involved in the adaptations to environmental stimuli, and that their expression levels are affected by drought, salt, and cold [2,9,11]. To investigate the potential roles of *ClCDPKs* and *ClCRKs* in response to abiotic stresses, their dynamic expressions were analyzed under drought, salt, and cold treatments (Figure 6). Compared to cold stimuli, far more *ClCDPKs* and *ClCRKs* could be induced by drought and NaCl treatments, and seven genes could be up-regulated by both drought and NaCl treatments, including *ClCDPK1*, *ClCDPK5*, *ClCDPK6*, *ClCDPK9*, *ClCDPK10*, *ClCDPK12*, and *ClCDPK14*. Following cold treatment, the transcription levels of four genes (*ClCDPK1*, *ClCDPK5*, *ClCDPK16*, and *ClCDPK17*) were down-regulated, while gene *ClCRK2* was continuously up-regulated at all treatment times. Compared to *ClCDPK3* and *ClCDPK18*

down-regulated by drought stress, many more genes were obviously up-regulated, such as *ClCDPK1*, *ClCDPK2*, *ClCDPK8*, *ClCDPK9*, *ClCDPK12*, and *ClCDPK14*. In response to NaCl stress, the majority of *ClCDPK* and *ClCRK* genes were up-regulated, with few exceptions (Figure 6).

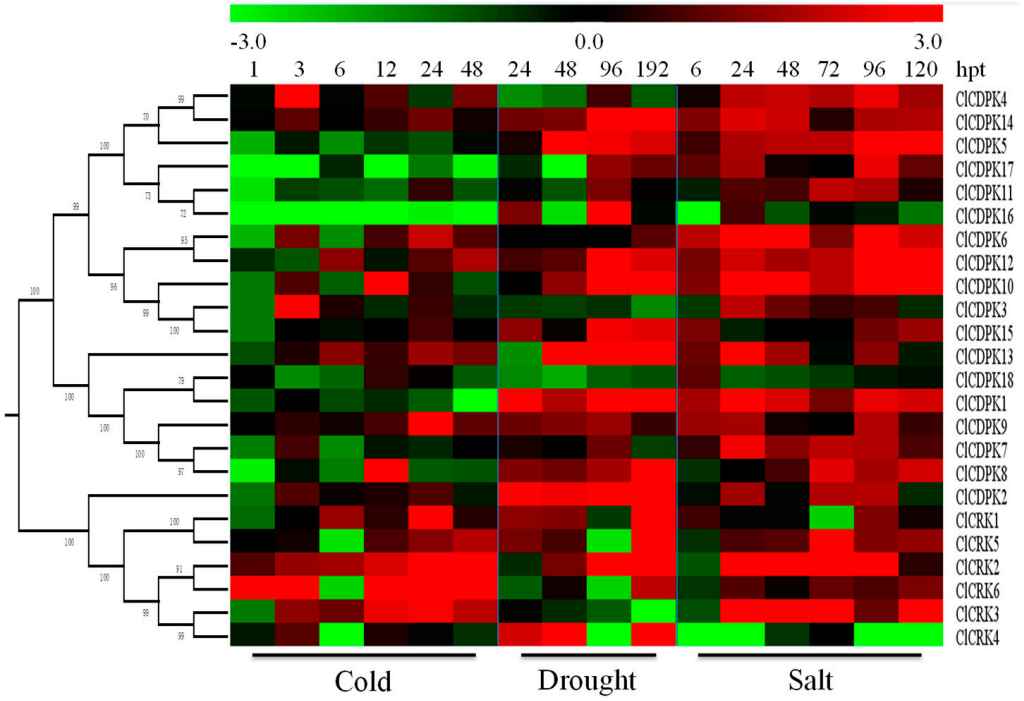


Figure 6. Expression patterns of *ClCDPK* and *ClCRK* genes under abiotic stresses. The abiotic stresses are displayed at the low end. The relative transcript level was $\log_2$ transformed and visualized as a heat map via Mev4.8.1, using red to indicate increased expression level and green to indicate decreased expression level.

2.9. Expression Patterns of ClCDPK and ClCRK Genes under Hormone Treatments

Previous studies have indicated that *CDPKs* and *CRKs* are involved in the signaling pathways of various plant hormones [19,33]. Here, expression profiles of *ClCDPKs* and *ClCRKs* were investigated in response to four plant hormones ABA, SA, MeJA, and ETH.

Increasing evidence has shown that *CDPK* and *CRK* genes could participate in ABA-mediated signal transduction in plants [17,34,35]. In the present study, the majority of *ClCDPKs* were found to be induced by ABA treatment, with transcript abundance retaining at higher levels from 6 to 24 hpt (Figure 7a). Interestingly, all *ClCRKs* except for *ClCRK1*, were sharply down-regulated at 6 hpt, which is similar to observations in melon [11]. Following SA treatment, expression levels of most *ClCDPKs* and *ClCRKs* remained either unchanged or slightly changed at 0.5 and 1 hpt (Figure 7b). However, SA application significantly induced or reduced their transcript levels at 6 or 12 hpt. Notably, *ClCRKs* were also found to have decreased at 6 hpt, which is similar to the tendency mentioned above. In contrast to the responses to ABA and SA treatments, almost all *ClCDPKs* and *ClCRKs* showed continuous over-expression in response to ETH and MeJA stimuli, except for genes *ClCDPK16* and *ClCRK3* (Figure 8). Moreover, transcript abundances of most *ClCDPKs* and *ClCRKs* had either sharply increased or decreased at 0.5 hpt, implying their rapid response to ETH and MeJA stimuli. In summary, these expression analyses indicated that *ClCDPKs* and *ClCRKs* could be involved in the regulatory pathways of plant hormones, and thus participate in the plant defense against environmental stresses.

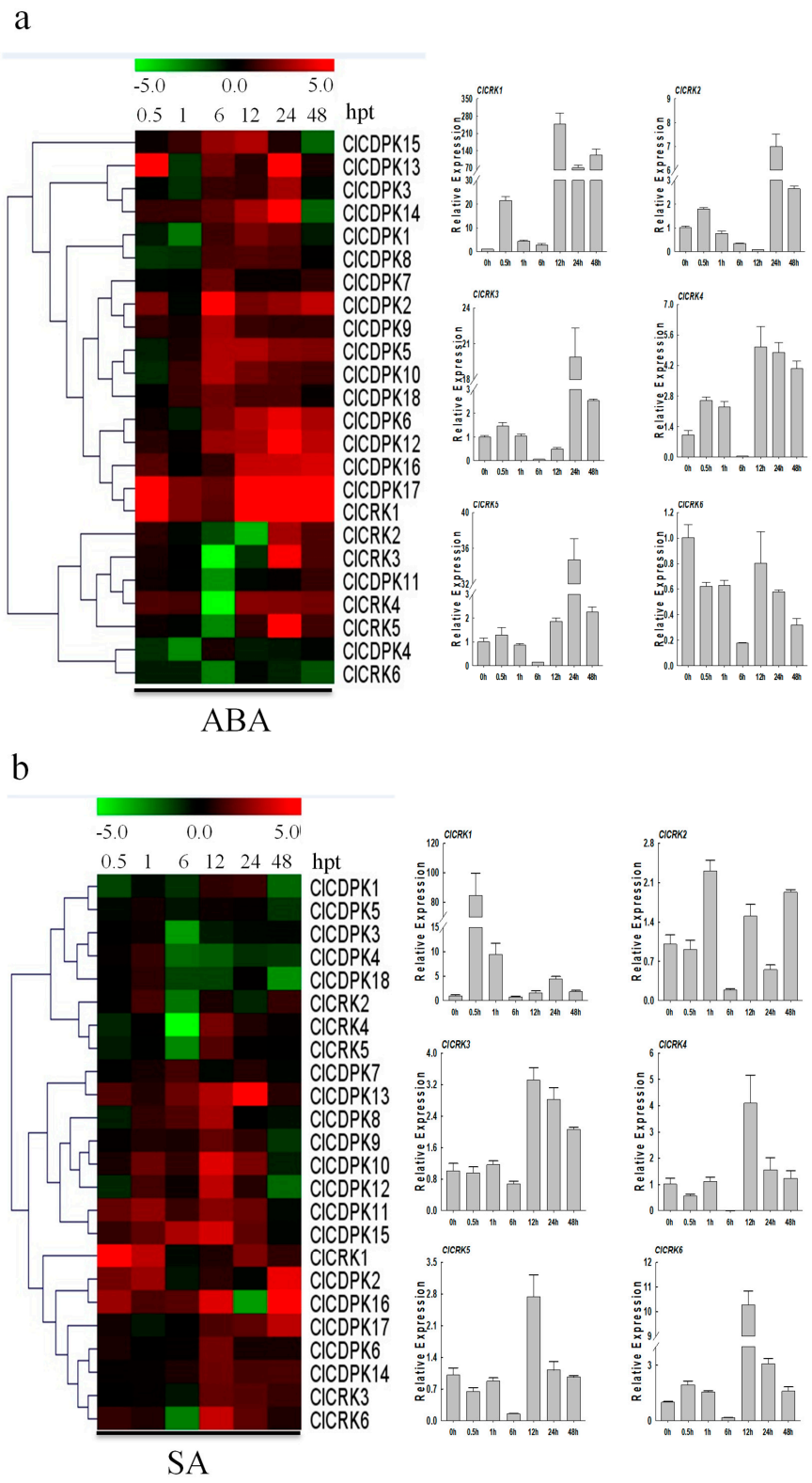


Figure 7. Expression patterns of *ClCDPK* and *ClCRK* genes under ABA and SA hormone treatments. (**a**) Expression levels of *ClCDPKs* and *ClCRKs* under ABA stress visualized as a heat map (Left). Detailed expression patterns of *ClCRKs* under ABA stress (Right). (**b**) Expressions of *ClCDPKs* and *ClCRKs* under SA stress visualized as a heat map (Left). Detailed expression patterns of *ClCRKs* under SA stress (Right). The relative transcript level was $\log_2$ transformed and visualized as a heat map via Mev4.8.1, using red to indicate increased expression level and green to indicate decreased expression level.

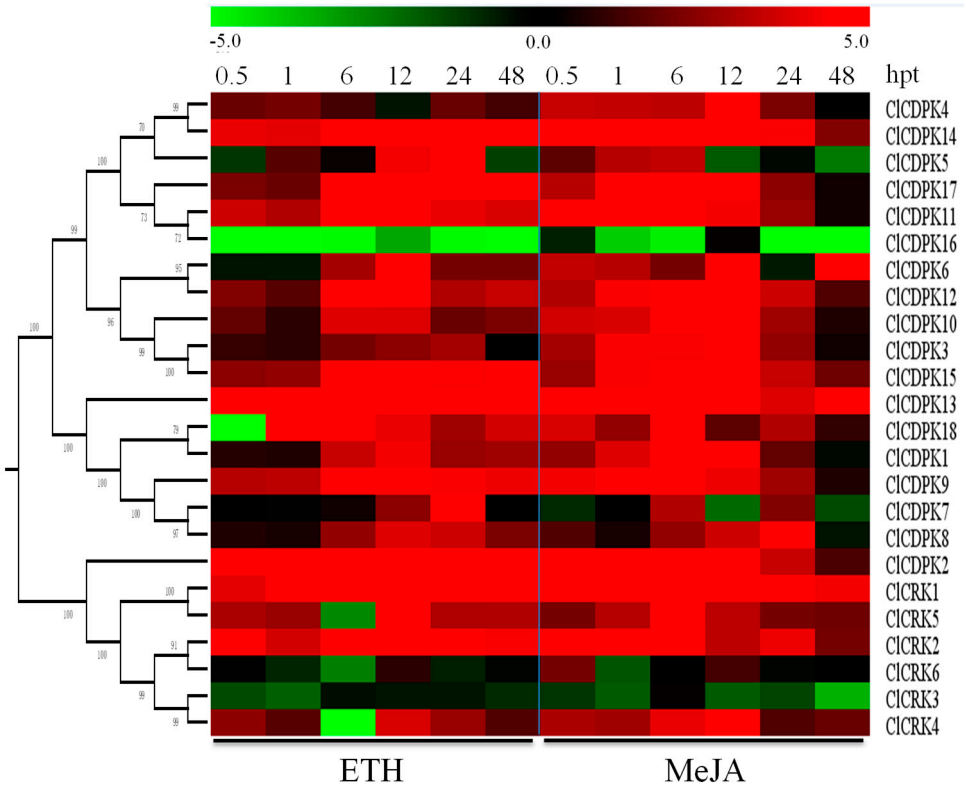


Figure 8. Expression patterns of *ClCDPK* and *ClCRK* genes under ETH and MeJA hormone stresses. The abiotic stresses are indicated at the low end. The relative transcript level was $\log_2$ transformed and visualized as a heat map by Mev4.8.1, using red to indicate increased expression level and green to indicate decreased expression level.

3. Discussion

3.1. Characteristic Features of CDPK and CRK Genes in Cucurbitaceae

Functioning as both Ca^{2+} sensors and effectors, the *CDPK* gene family has been identified throughout the plant kingdom, as well as in several protozoa, but are absent in animals [4,28]. For instance, 34 *CDPKs* and eight *CRKs* were identified genome-wide in *Arabidopsis* [13], 29 *CDPKs* and six *CRKs* in tomato [10], 31 *CDPKs* and five *CRKs* in pepper [15], 30 *CDPKs* and nine *CRKs* in poplar [36], and 31 *CDPKs* and five *CRKs* in rice [12,14]. Moreover, approximately 19, 41, and 40 *CDKPs* have been detected in the genomes of grape [19], cotton [30], and maize [37], respectively. In Cucurbitaceae, 18 *CDPKs* and seven *CRKs* have been found in the melon genome [11], while 19 *CDPKs* were identified in cucumber by Xu et al. [9]. In this study, a total of 128 *CDPK* and 56 *CRK* genes were identified in six Cucurbitaceae species, including *C. lanatus*, *C. sativus*, *C. moschata*, *C. maxima*, *C. pepo*, and *L. siceraria* (Table 1). The numbers of *CDPKs* and *CRKs* are much higher in three *Cucurbita* species than in four Benincaseae tribe species (*C. lanatus*, *C. melon*, *C. sativus*, and *L. siceraria*), which may be due to a WGD event during the origin of this genus [25,26]. Moreover, the numbers of *CRKs* in three *Cucurbita* genomes are much higher than in other species, such as tomato, pepper, and rice, although these species contain similar copies of *CDPKs*.

Tandem, segmental, and whole genome duplication events are confirmed to play important roles in the expansion of gene families. Approximately 12, 13, and seven segmental duplications were reported to exist in poplar, cotton, and rice genomes [14,30,36], while many more events were found in three *Cucurbita* species that mainly occurred between two sub-genomes (Table S5). However, only a few segmental duplication events (one or two) were detected in genomes of watermelon, cucumber, bottle gourd, and melon [11], which may cause the low copy numbers of *CDPK* and *CRK*

genes. In the present study, the majority of CDPKs contained four EF-hand motifs (Table S2), which is consistent with observations found in other species [7,11,19]. However, two CDPKs (CpCDPK19 and CpCDPK27) in *C. pepo* with detectable transcriptional levels had been confirmed to contain nine and eight EF-hands, respectively. Dot plot analysis showed that self-duplications of the STKs_CAMK protein kinase or EF-hand domains resulted in the gene structure variations (Figure 1), which may affect their functional specificity.

3.2. Conserved Evolution of CDPK and CRK Genes in Cucurbitaceae

Generally, *CDPK* and *CRK* genes are randomly distributed in genomes [10,11,15,36], which has also been validated in Cucurbitaceae species (Figure S3). Using watermelon chromosomes as reference, an integrated map was obtained and contained 16 *CDPK* and nine *CRK* loci, harboring almost all *CDPK* and *CRK* genes identified in Cucurbitaceae (Figure 2 and Table S4). Of these, nine loci (four *CDPK* and five *CRK*) exhibited Presence/Absence polymorphisms, while the remaining 16 loci (12 *CDPK* and four *CRK*) were shared by all seven species, implying that the flanking regions of most *CDPK* and *CRK* genes were conserved in these species during the evolutionary process.

Increasing evidence indicates that topological structures of *CDPK* and *CRK* gene families are conserved, with four *CDPK* and one *CRK* groups in phylogenetic trees [10,11,28,36]. In the present study, group CDPK IV was found to be close to CRK I rather than the other three CDPK groups in distance trees (Figure 3 and Figures S4, S5 and S7), confirming that CDPK IV and CRK I may originate from a common ancestor [10,11,30]. Moreover, the five groups can be further divided into 25 loci, according to the integrated map. The evolutionary scenario of seven modern Cucurbitaceae species revealed that watermelon diverged from bottle gourd around 10.4–14.6 Mya and from *Cucumis* 17.3–24.3 Mya; however, the progenitor B of *Cucurbita* diverged from Benincaseae around 25.5–27.0 Mya, and progenitor A diverged from the common ancestor of progenitor B and Benincaseae around 29.9–31.6 Mya [26,27]. In agreement with this evolutionary scenario, orthologs from watermelon and bottle gourd usually gathered together in phylogenetic trees, and genes from sub-genome B were preferentially clustered with those from Benincaseae tribe genomes (Figure 3 and Figure S7).

Both exon-intron structures and intron numbers can reflect the evolution, expansion, and functional relationships within a gene family, which were caused by three main types of mechanisms, including exon/intron gain/loss, exonization/pseudoexonization, and insertion/deletion [11,19,38]. Exon-intron organization analyses revealed that each group contained one or two major intron phase patterns (Figure S8). For instance, the majority of homologs in group CDPK I contained six introns with a distinct intron phase pattern of 111000, while most members in CDPK II had seven introns sharing a similar intron phase 1110020. As a peripheral sister clade of CDPK I and II, CDPK III contained two major intron phases and one of them was identical to that of group CDPK I. Moreover, the major intron pattern (02201010000) of eight members in CDPK IV is similar to that (0220110000) of most *CRK* homologs in CRK I (Figure S8). Combined with the topological structures of CDPK IV and CRK I, we infer that group CDPK IV is the ancient lineage of CDPK gene family, which may have diverged from the last common ancestor with CRK I before the divergence of monocots and dicots [28].

3.3. Functional Comparison of CDPK and CRK Genes

CDPKs and *CRKs* have been confirmed to play crucial roles in the signal pathways in response to various environmental stresses [4,5,28]. The systemic expression profiles of *CDPK* and *CRK* genes in *Cucumis* species under different stimuli have been reported recently [9,11]. Consequently, dynamic expression levels of *CDPKs* and *CRKs* under different stimuli were investigated in detail in watermelon. Compared to cold stimuli, far more *ClCDPKs* and *ClCRKs* were induced under drought and NaCl treatments (Figure 6), which differed from the expression trends of most homologs that were up-regulated by cold stress in cucumber and melon [9,11]. In watermelon, expression levels of *ClCDPK14* were up-regulated by all three stresses (cold, drought, and salt); its ortholog *CmCDPK1* in melon, could be also induced by cold and salinity treatments [11]. Completely different expression

trends were also observed among species. For example, both genes *ClCDPK6* and *ClCDPK12*, with a close relationship in group CDPK II, could be induced by salt stress (Figure S4 and Figure 6), while expression levels of their orthologs in melon (*CmCDPK9* and *CmCDPK11*) and cucumber (*CsCDPK14* and *CsCDPK5*) exhibited contrary responses to salinity stimuli [9,11]. Similarly, transcriptional levels of *ClCRK2* and its paralog *ClCRK3*, as well as their orthologs *CmCRK3* and *CmCRK6* in melon, were increased under low temperature; however, these two groups of genes showed different expression trends under salt stress between watermelon and melon. Taken together, we inferred that the functional fates of some orthologs may be diversified during species evolution. The phytohormone ABA has been reported to be widely involved in the response of plants to biotic and abiotic stresses [17,34,35]. For instance, AtCPK4 and AtCPK11 in *Arabidopsis* can positively mediate the CDPK/calcium-mediated ABA signaling pathways via phosphorylation of two ABA-responsive transcription factors, ABF1 and ABF4, and loss-of-function mutations decrease the tolerance of seedlings to salt stress [17]. As their closest homolog in the phylogenetic tree (Figure S7), transcript abundance of *ClCDPK14* was also up-regulated under ABA and salt stimuli (Figures 6 and 7). The biosynthesis of ABA could be induced by drought, leading to stomatal closure [17,39]. In *Arabidopsis*, the AtCPK4 and AtCPK11 are partially involved in ABA-induced stomatal closure, and the double mutant lost more water from leaves compared to single mutants [17]; gene *ClCDPK14*, as their closed homolog in watermelon, was also activated by drought (Figure 6). Additionally, gene *AtCPK6* has been proven to be involve in the response to drought and salt stress as a positive regulator [40]. Intriguingly, *ClCDPK5* in watermelon, sharing the highest sequence similarity with *AtCPK6*, was also up-regulated by drought (or salt) and ABA stresses, indicating that they might function in similar pathways. In this study, the transcription levels of most *ClCDPKs* were increased by exogenous ABA (Figure 7), similar to the expression trend in grape but different to that in *Cucumis* species [9,11,19].

In watermelon, transcription levels of four genes (*ClCDPK1*, *ClCDPK5*, *ClCDPK16*, and *ClCDPK17*) were down-regulated by cold treatment, while only one gene (*ClCRK2*) was continuously up-regulated at all treatment times (Figure 6). Moreover, complex expression patterns were also observed under cold stimuli, such as *ClCRK5* and *ClCRK6*. Three up-regulated genes (*ClCDPK1*, *ClCDPK2*, and *ClCDPK5*) and one down-regulated gene (*ClCDPK18*) under continuous drought treatment in this study have also been detected and regarded as different expression genes in our previous study [41]. Following salt treatment, gene *ClCDPK6* was significantly up-regulated at all treatment times, which is similar to its ortholog *CsCDPK14* in cucumber but in contrast with its ortholog *CmCDPK9* in melon [9,11]. Additionally, two pairs of segmental duplications were detected in watermelon: *ClCDPK7*/*ClCDPK8* and *ClCDPK6*/*ClCDPK16* (Table S5). Interestingly, *ClCDPK7* and *ClCDPK8* had similar expression patterns under most treatments, while expression tendency of *ClCDPK6* was usually opposite to *ClCDPK16* (Figures 6–8), inferring that genes *ClCDPK6* and *ClCDPK16* may have undergone sub-functionalization after duplication.

Generally, *CDPK* genes are ubiquitously expressed in plant organs, with some showing organ- or tissue-specific expression [19,21]. In the present study, most identified *ClCDPKs* and *ClCRKs* could be detected in at least one tissue, with at least five genes showing extremely high expression levels in specific organs (Figure 5). For instance, *ClCDPK17* in locus 24 showed a high expression level in the male flower, similar to its ortholog *CmCDPK6* in melon, which preferentially accumulated in male flowers [11]. In *Arabidopsis*, both genes *AtCPK17* and *AtCPK20* are reported to be preferentially expressed in mature pollen to regulate the growth of pollen tubes [21,42]. Interestingly, their phylogenetically-close homologs *ClCDPK6* and *ClCDPK16* in watermelon, as well as that in melon (*CmCDPK9* and *CmCDPK5*) and cucumber (*CsCDPK14* and *CsCDPK9*) (Figure S7), were also detected with high transcriptional abundances in male organs [9,11], indicating their conserved and important roles in the development of male flowers.

4. Materials and Methods

4.1. Identification and Biochemical Characterization of CDPKs and CRKs

For genome-wide identification of *CDPK* and *CRK* genes in Cucurbitaceae species, protein sequences of 18 CmCDPKs, seven CmCRKs, and 19 CsCDPKs were obtained according to recently published studies [9,11]. Then, using BLASTp program with an E-value setting of 1.0×10^{-5}, these protein sequences were used as queries to search against the predicted protein files of watermelon (*C. lanatus*, v1), cucumber (*C. sativus*, v1), *Cucurbita* genus (*C. moschata,* v1; *C. maxima,* v1.1; *C. pepo*, v4.1), and bottle gourd (*L. siceraria*, v1), which were downloaded from the Cucurbit Genomics Database (http://cucurbitgenomics.org/). Additionally, Hidden Markov Model (HMM) profiles of the core protein kinase domain (PF00069) and EF-hand_7 domain (PF13499) were downloaded from the Pfam database (http://pfam.xfam.org/), and were also subjected to searches for *CDPKs* and *CRKs* with software HMMER 3.0 (default parameters). The reliability of candidates was verified through searching against the NCBI nr database, and all non-redundant putative genes were tested for the presence of core domains using ScanProsite (http://prosite.expasy.org/scanprosite/). Finally, all candidate genes were further characterized with the following online tools: ProtParam (http://web.expasy.org/protparam/), ScanProsite (http://prosite.expasy.org/scanprosite/), SMART (http://smart.embl-heidelberg.de/), ExPASy (http://web.expasy.org/myristoylator/), and CSS-plam 4.0 (http://csspalm.biocuckoo.org/). Dot plot analysis was performed using the software Geneious (http://www.geneious.com).

4.2. Chromosomal Location of CDPK and CRK Genes

The genomic distributions of *CDPKs* and *CRKs* on chromosomes were drawn using the software TBtools (http://cj-chen.github.io/tbtools/). To construct an integrated map, all identified CDPKs and CRKs, as well as those reported in recent studies [9,11], were mapped onto watermelon chromosomes based on the syntenic relationships of their flanking genes using BLASTp method, with a stricter E-value setting to 1.0×10^{-10}. Then, in-house Perl scripts were used to parse the resulting files and to visualize the chromosomal locations of *CDPK* and *CRK* loci.

4.3. Phylogenetic, Gene Structure, and Syntenic Analyses of CDPKs and CRKs

Full-length CDPK and CRK protein sequences were aligned using the software Muscle [43], and then were used to construct phylogenetic trees via MEGA 6.0 using the neighbor-joining method with 1000 bootstrap replicates [11]. To perform gene structure analyses, genomic and cDNA sequences of *CDPKs* and *CRKs* were obtained from their corresponding genomes. Then, the exon-intron organization was displayed via the online tool GSDS 2.0 (http://gsds.cbi.pku.edu.cn/). The destination tabular (-m 8) files of the BLASTp program (with an E-value setting of 1.0×10^{-10}) and GFF profiles served as input documents for MCScanX to analyze the synteny relationships [44], which were then visualized using software CIRCOS (http://circos.ca/).

4.4. Plant Material and Treatments

The watermelon inbred line "Y34" provided by the Cucurbits Germplasm Resource Research Group at the Northwest A&F University in China was used in this study. For tissue-specific analysis, germinated seeds of watermelon "Y34" were directly sown in the experimental base, and six organs (roots, stems, leaves, tendrils, and both male and female flowers) were independently sampled during the fruit maturation period (approximately 60–70 days after sowing). For stress experiments, seedlings were cultured in plastic pots (8 cm × 7 cm × 7 cm) filled with commercial peat-based compost (Shaanxi Yufeng Seed Industry Co., Ltd., Yangling, China). All plants were grown under springtime natural light with temperatures of 28–35 °C/16–20 °C (day/night) in a greenhouse, which were uniformly watered and nourished weekly with half-strength Hoagland's solution before treatments. Four weeks after sowing, seedlings were used for the following treatments.

For the salinity treatment, seedlings irrigated with 300 mM NaCl solution (80 mL per plant) were sampled at 6, 24, 48, 72, 96, and 120 h post-treatment (hpt), while plants irrigated with distilled water were used as control. Compared to seedlings planted in a growth chamber at 27 ± 1 °C and 80% humidity under a light intensity of 300 mmol·m^{-2}·s^{-1} PPFD, leaves of plants kept at 4 °C were sampled at 1, 3, 6, 12, 24, and 48 hpt for cold treatment. To simulate a national drought treatment [45], all seedlings were uniformly well-watered to 70 ± 5% field capacity based on their weight. Then, leaves of drought-treated (unwatered) and control plants were sampled at 24, 48, 96, and 192 hpt. Leaves sprayed with 100 μM abscisic acid (ABA) [46], 1 mM salicylic acid (SA) [47], 100 μM methyljasmonate (MeJA), and 10 mM ethephon (ETH) [48] were collected at 0.5, 1, 6, 12, 24, and 48 hpt for hormone treatments, while control seedlings were only sprayed with equal volumes of the corresponding solution without hormones. In this study, leaves of four plants were pooled at each time point for each treatment with three biological replicates, which were immediately frozen in liquid nitrogen and stored at −80 °C until further analysis.

4.5. RNA Isolation and qRT-PCR

The total RNA of samples was extracted using the RNASimple Total RNA Kit (TIANGEN, China) following the manufacturer's instructions. Then, approximately 1 μg of total RNA was used to synthesize the first strand of cDNA using the FastKing RT Kit with gDNase (TIANGEN, China). Gene-specific primers for *ClCDPKs* and *ClCRKs* are listed in Table S1. Amplification was conducted in a 20 μL reaction volume, containing 10.0 μLSYBR Green Premix, 0.8 μL of each primer (10 μM), and 1.0 μL cDNA template (80 ng/μL), which was diluted with ddH_2O to 20 μL. The PCR conditions consisted of pre-denaturing at 95 °C for 5 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. The watermelon *β-actin* gene (*Cla007792*) was used as the internal control gene [49]. Each treatment was repeated thrice, and all data were calculated for relative expressions following the $2^{-\Delta\Delta Ct}$ method, as described by Livak and Schmittgen [50]. The relative expressions were then log2 transformed and visualized in a heat map using Mev 4.8.1 (http://www.mybiosoftware.com/). All data were analyzed via IBM SPSS Statistics 21 and values were presented as the means ± SD of three biological and three technical replicates (Table S7). The significance of expression between treatments and controls was evaluated by one-way ANOVA and Duncan's multiple range tests.

5. Conclusions

In the present study, a total of 128 *CDPK* and 56 *CRK* genes were identified in six Cucurbitaceae species. Using the watermelon genome as reference, an integrated map containing 25 loci (16 *CDPK* and nine *CRK* loci) was obtained, 16 of which (12 *CDPK* and four *CRK*) were shared by all seven Cucurbitaceae species. Combined with exon-intron organizations, topological analyses indicated an ancient origination of groups CDPK IV and CRK, which will contribute to elucidating the evolutionary history of these two gene families in Cucurbitaceae. Moreover, expression patterns of *ClCDPKs* and *ClCRKs* under different abiotic stresses were also performed in this study. Importantly, comparative analyses indicated a subset of valuable orthologous genes for future functional characterizations. For instance, both *ClCDPK6* and its ortholog *CsCDPK14* in cucumber could be induced by salinity, while *ClCDPK6* and *ClCDPK16*, as well as their orthologs in *Cucumis*, maintained high expression levels in male flowers, which may play important roles in plant organ development and response to environmental stresses.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/10/2527/s1.

Author Contributions: C.W. and X.Z. designed the study. C.W., R.Z., X.Y., and C.Z. contributed to the experiments and data analysis. J.M. provide the seed for the experiment. Y.Z. and J.Y. provided guidance throughout the study. C.W. wrote, and H.L. revised the manuscript. All authors reviewed and approved the final manuscript.

Funding: This work was supported by National Key R&D Program of China (2018YFD0100704), Scientific Startup Foundation for Doctors of Northwest A&F University (Z109021604), National Natural Science Foundation of China Grant No. 31701939, and the Modern Agro-industry Technology Research System of China (No. CARS-26-18).

Acknowledgments: We wish to thank all the reviewers and editors for their careful reading and helpful comments on this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CDPK	Calcium-dependent protein kinase
CRK	CDPK-related kinases
hpt	hours post-treatment
ABA	abscisic acid
SA	salicylic acid
MeJA	methyljasmonate
ETH	ethephon

References

1. Hepler, P.K. Calcium: A central regulator of plant growth and development. *Plant Cell* **2005**, *17*, 2142–2155. [CrossRef]
2. Boudsocq, M.; Sheen, J. CDPKs in immune and stress signaling. *Trends Plant Sci.* **2013**, *18*, 30–40. [CrossRef] [PubMed]
3. Ranty, B.; Aldon, D.; Galaud, J.P. Plant calmodulins and calmodulin-related proteins: Multifaceted relays to decode calcium signals. *Plant Signal. Behav.* **2006**, *1*, 96–104. [CrossRef] [PubMed]
4. Hamel, L.P.; Sheen, J.; Seguin, A. Ancient signals: Comparative genomics of green plant CDPKs. *Trends Plant Sci.* **2014**, *19*, 79–89. [CrossRef] [PubMed]
5. Ludwig, A.A.; Romeis, T.; Jones, J.D. CDPK-mediated signalling pathways: Specificity and cross-talk. *J. Exp. Bot.* **2004**, *55*, 181–188. [CrossRef]
6. Boudsocq, M.; Willmann, M.R.; McCormack, M.; Lee, H.; Shan, L.; He, P.; Bush, J.; Cheng, S.H.; Sheen, J. Differential innate immune signalling via Ca(2+) sensor protein kinases. *Nature* **2010**, *464*, 418–422. [CrossRef] [PubMed]
7. Cheng, S.H.; Willmann, M.R.; Chen, H.C.; Sheen, J. Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family. *Plant Physiol.* **2002**, *129*, 469–485. [CrossRef] [PubMed]
8. Liese, A.; Romeis, T. Biochemical regulation of in vivo function of plant calcium-dependent protein kinases (CDPK). *Biochim. Biophys. Acta* **2013**, *1833*, 1582–1589. [CrossRef] [PubMed]
9. Xu, X.; Liu, M.; Lu, L.; He, M.; Qu, W.; Xu, Q.; Qi, X.; Chen, X. Genome-wide analysis and expression of the calcium-dependent protein kinase gene family in cucumber. *Mol. Genet. Genom. MGG* **2015**, *290*, 1403–1414. [CrossRef] [PubMed]
10. Wang, J.P.; Xu, Y.P.; Munyampundu, J.P.; Liu, T.Y.; Cai, X.Z. Calcium-dependent protein kinase (CDPK) and CDPK-related kinase (CRK) gene families in tomato: Genome-wide identification and functional analyses in disease resistance. *Mol. Genet. Genom. MGG* **2016**, *291*, 661–676. [CrossRef]
11. Zhang, H.; Wei, C.; Yang, X.; Chen, H.; Yang, Y.; Mo, Y.; Li, H.; Zhang, Y.; Ma, J.; Yang, J.; et al. Genome-wide identification and expression analysis of calciumdependent protein kinase and its related kinase gene families in melon (Cucumis melo L.). *PLoS ONE* **2017**, *12*, e0176352.
12. Ray, S.; Agarwal, P.; Arora, R.; Kapoor, S.; Tyagi, A.K. Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (*Oryza sativa* L. ssp. indica). *Mol. Genet. Genom. MGG* **2007**, *278*, 493–505. [CrossRef] [PubMed]
13. Hrabak, E.M.; Chan, C.W.; Gribskov, M.; Harper, J.F.; Choi, J.H.; Halford, N.; Kudla, J.; Luan, S.; Nimmo, H.G.; Sussman, M.R.; et al. The Arabidopsis CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* **2003**, *132*, 666–680. [CrossRef] [PubMed]

14. Asano, T.; Tanaka, N.; Yang, G.; Hayashi, N.; Komatsu, S. Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: Comprehensive analysis of the CDPKs gene family in rice. *Plant Cell Physiol.* **2005**, *46*, 356–366. [CrossRef]
15. Cai, H.; Cheng, J.; Yan, Y.; Xiao, Z.; Li, J.; Mou, S.; Qiu, A.; Lai, Y.; Guan, D.; He, S. Genome-wide identification and expression analysis of calcium-dependent protein kinase and its closely related kinase genes in Capsicum annuum. *Front. Plant Sci.* **2015**, *6*, 737. [CrossRef]
16. Zhou, L.; Lan, W.; Jiang, Y.; Fang, W.; Luan, S. A calcium-dependent protein kinase interacts with and activates a calcium channel to regulate pollen tube growth. *Mol. Plant* **2014**, *7*, 369–376. [CrossRef]
17. Zhu, S.Y.; Yu, X.C.; Wang, X.J.; Zhao, R.; Li, Y.; Fan, R.C.; Shang, Y.; Du, S.Y.; Wang, X.F.; Wu, F.Q.; et al. Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. *Plant Cell* **2007**, *19*, 3019–3036. [CrossRef] [PubMed]
18. Ma, S.Y.; Wu, W.H. AtCPK23 functions in Arabidopsis responses to drought and salt stresses. *Plant Mol. Biol.* **2007**, *65*, 511–518. [CrossRef] [PubMed]
19. Zhang, K.; Han, Y.T.; Zhao, F.L.; Hu, Y.; Gao, Y.R.; Ma, Y.F.; Zheng, Y.; Wang, Y.J.; Wen, Y.Q. Genome-wide Identification and Expression Analysis of the CDPK Gene Family in Grape, Vitis spp. *BMC Plant Biol.* **2015**, *15*, 164. [CrossRef]
20. Chen, F.; Fasoli, M.; Tornielli, G.B.; Dal Santo, S.; Pezzotti, M.; Zhang, L.; Cai, B.; Cheng, Z.M. The evolutionary history and diverse physiological roles of the grapevine calcium-dependent protein kinase gene family. *PLoS ONE* **2013**, *8*, e80818. [CrossRef] [PubMed]
21. Myers, C.; Romanowsky, S.M.; Barron, Y.D.; Garg, S.; Azuse, C.L.; Curran, A.; Davis, R.M.; Hatton, J.; Harmon, A.C.; Harper, J.F. Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. *Plant J. Cell Mol. Biol.* **2009**, *59*, 528–539. [CrossRef]
22. Huang, S.; Li, R.; Zhang, Z.; Li, L.; Gu, X.; Fan, W.; Lucas, W.J.; Wang, X.; Xie, B.; Ni, P.; et al. The genome of the cucumber, Cucumis sativus L. *Nat. Genet.* **2009**, *41*, 1275–1281. [CrossRef]
23. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Mir, G.; Gonzalez, V.M.; Henaff, E.; Camara, F.; Cozzuto, L.; Lowy, E.; et al. The genome of melon (Cucumis melo L.). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11872–11877. [CrossRef]
24. Guo, S.; Zhang, J.; Sun, H.; Salse, J.; Lucas, W.J.; Zhang, H.; Zheng, Y.; Mao, L.; Ren, Y.; Wang, Z.; et al. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat. Genet.* **2013**, *45*, 51–58. [CrossRef]
25. Montero-Pau, J.; Blanca, J.; Bombarely, A.; Ziarsolo, P.; Esteras, C.; Marti-Gomez, C.; Ferriol, M.; Gomez, P.; Jamilena, M.; Mueller, L.; et al. De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the Cucurbita genus. *Plant Biotechnol. J.* **2018**, *16*, 1161–1171. [CrossRef]
26. Sun, H.; Wu, S.; Zhang, G.; Jiao, C.; Guo, S.; Ren, Y.; Zhang, J.; Zhang, H.; Gong, G.; Jia, Z.; et al. Karyotype Stability and Unbiased Fractionation in the Paleo-Allotetraploid Cucurbita Genomes. *Mol. Plant* **2017**, *10*, 1293–1306. [CrossRef]
27. Wu, S.; Shamimuzzaman, M.; Sun, H.; Salse, J.; Sui, X.; Wilder, A.; Wu, Z.; Levi, A.; Xu, Y.; Ling, K.S.; Fei, Z. The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. *Plant J. Cell Mol. Biol.* **2017**, *92*, 963–975. [CrossRef]
28. Valmonte, G.R.; Arthur, K.; Higgins, C.M.; MacDiarmid, R.M. Calcium-dependent protein kinases in plants: evolution, expression and function. *Plant Cell Physiol.* **2014**, *55*, 551–569. [CrossRef]
29. Zheng, Y.; Wu, S.; Bai, Y.; Sun, H.; Jiao, C.; Guo, S.; Zhao, K.; Blanca, J.; Zhang, Z.; Huang, S.; et al. Cucurbit Genomics Database (CuGenDB): A central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* **2019**, *47*, D1128–D1136. [CrossRef]
30. Liu, W.; Li, W.; He, Q.; Daud, M.K.; Chen, J.; Zhu, S. Genome-wide survey and expression analysis of calcium-dependent protein kinase in Gossypium raimondii. *PLoS ONE* **2014**, *9*, e98189. [CrossRef]
31. Vision, T.J.; Brown, D.G.; Tanksley, S.D. The origins of genomic duplications in Arabidopsis. *Science* **2000**, *290*, 2114–2117. [CrossRef]
32. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol.* **2004**, *4*, 10. [CrossRef]
33. Coca, M.; San Segundo, B. AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. *Plant J. Cell Mol. Biol.* **2010**, *63*, 526–540. [CrossRef]

34. Zou, J.J.; Wei, F.J.; Wang, C.; Wu, J.J.; Ratnasekera, D.; Liu, W.X.; Wu, W.H. Arabidopsis calcium-dependent protein kinase CPK10 functions in abscisic acid- and Ca2+-mediated stomatal regulation in response to drought stress. *Plant Physiol.* **2010**, *154*, 1232–1243. [CrossRef]
35. Li, Z.; Takahashi, Y.; Scavo, A.; Brandt, B.; Nguyen, D.; Rieu, P.; Schroeder, J.I. Abscisic acid-induced degradation of Arabidopsis guanine nucleotide exchange factor requires calcium-dependent protein kinases. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4522–E4531. [CrossRef]
36. Zuo, R.; Hu, R.; Chai, G.; Xu, M.; Qi, G.; Kong, Y.; Zhou, G. Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (Populus trichocarpa). *Mol. Biol. Rep.* **2013**, *40*, 2645–2662. [CrossRef]
37. Kong, X.; Lv, W.; Jiang, S.; Zhang, D.; Cai, G.; Pan, J.; Li, D. Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genom.* **2013**, *14*, 433. [CrossRef]
38. Xu, G.; Guo, C.; Shan, H.; Kong, H. Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1187–1192. [CrossRef]
39. Roelfsema, M.R.G.; Hedrich, R. In the light of stomatal opening: New insights into 'the Watergate'. *New Phytol.* **2005**, *167*, 665–691. [CrossRef]
40. Xu, J.; Tian, Y.-S.; Peng, R.-H.; Xiong, A.-S.; Zhu, B.; Jin, X.-F.; Gao, F.; Fu, X.-Y.; Hou, X.-L.; Yao, Q.-H. AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in Arabidopsis. *Planta* **2010**, *231*, 1251–1260. [CrossRef]
41. Yang, Y.; Mo, Y.; Yang, X.; Zhang, H.; Wang, Y.; Li, H.; Wei, C.; Zhang, X. Transcriptome Profiling of Watermelon Root in Response to Short-Term Osmotic Stress. *PLoS ONE* **2016**, *11*, e0166314. [CrossRef] [PubMed]
42. Gutermuth, T.; Lassig, R.; Portes, M.-T.; Maierhofer, T.; Romeis, T.; Borst, J.-W.; Hedrich, R.; Feijó, J.A.; Konrad, K.R. Pollen tube growth regulation by free anions depends on the interaction between the anion channel SLAH3 and calcium-dependent protein kinases CPK2 and CPK20. *Plant Cell* **2013**, *25*, 4525–4543. [CrossRef] [PubMed]
43. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
44. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]
45. Mo, Y.; Yang, R.; Liu, L.; Gu, X.; Yang, X.; Wang, Y.; Zhang, X.; Li, H. Growth, photosynthesis and adaptive responses of wild and domesticated watermelon genotypes to drought stress and subsequent re-watering. *Plant Growth Regul.* **2016**, *79*, 229–241. [CrossRef]
46. Song, Q.; Li, D.; Dai, Y.; Liu, S.; Huang, L.; Hong, Y.; Zhang, H.; Song, F. Characterization, expression patterns and functional analysis of the MAPK and MAPKK genes in watermelon (Citrullus lanatus). *BMC Plant Biol.* **2015**, *15*, 298. [CrossRef]
47. Jing-Hua, Y.; Yuan, G.; Yan-Man, L.; Xiao-Hua, Q.; Zhang, M.-F. Salicylic acid-induced enhancement of cold tolerance through activation of antioxidative capacity in watermelon. *Sci. Hortic.* **2008**, *118*, 200–205. [CrossRef]
48. Yang, X.; Li, H.; Yang, Y.; Wang, Y.; Mo, Y.; Zhang, R.; Zhang, Y.; Ma, J.; Wei, C.; Zhang, X. Identification and expression analyses of WRKY genes reveal their involvement in growth and abiotic stress response in watermelon (Citrullus lanatus). *PLoS ONE* **2018**, *13*, e0191308. [CrossRef]
49. Kong, Q.; Yuan, J.; Gao, L.; Zhao, S.; Jiang, W.; Huang, Y.; Bie, Z. Identification of suitable reference genes for gene expression normalization in qRT-PCR analysis in watermelon. *PLoS ONE* **2014**, *9*, e90612. [CrossRef]
50. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method. *Methods* **2001**, *25*, 402–408. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of Multiple Organellar RNA Editing Factor Family in Poplar Reveals Evolution and Roles in Drought Stress

Dongli Wang [1,†], Sen Meng [1,2,†], Wanlong Su [1], Yu Bao [1], Yingying Lu [1], Weilun Yin [1], Chao Liu [1,*] and Xinli Xia [1,*]

1 Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China; wangdongli1997@163.com (D.W.); mengsen021124@126.com (S.M.); bjfususan@163.com (W.S.); byaq0556@gmail.com (Y.B.); luyingying2018@bjfu.edu.cn (Y.L.); yinwl@bjfu.edu.cn (W.Y.)

2 State Key Laboratory of Tree Genetics and Breeding, Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China

* Correspondence: liuchao1306@bjfu.edu.cn (C.L.); xiaxl@bjfu.edu.cn (X.X.); Tel.: +86-10-6233-6400 (C.L. & X.X.)

† These authors contributed equally to this work.

Received: 22 January 2019; Accepted: 18 March 2019; Published: 21 March 2019

Abstract: Poplar (Populus) is one of the most important woody plants worldwide. Drought, a primary abiotic stress, seriously affects poplar growth and development. Multiple organellar RNA editing factor (*MORF*) genes—pivotal factors in the RNA editosome in *Arabidopsis thaliana*—are indispensable for the regulation of various physiological processes, including organelle C-to-U RNA editing and plasmid development, as well as in the response to stresses. Although the poplar genome sequence has been released, little is known about *MORF* genes in poplar, especially those involved in the response to drought stress at the genome-wide level. In this study, we identified nine *MORF* genes in the Populus genome. Based on the structural features of MORF proteins and the topology of the phylogenetic tree, the *P. trichocarpa* (Ptr) MORF family members were classified into six groups (Groups I–VI). A microsynteny analysis indicated that two (22.2%) *PtrMORF* genes were tandemly duplicated and seven genes (77.8%) were segmentally duplicated. Based on the d_N/d_S ratios, purifying selection likely played a major role in the evolution of this family and contributed to functional divergence among *PtrMORF* genes. Moreover, analysis of qRT-PCR data revealed that *PtrMORFs* exhibited tissue- and treatment-specific expression patterns. *PtrMORF* genes in all group were involved in the stress response. These results provide a solid foundation for further analyses of the functions and molecular evolution of *MORF* genes in poplar, and, in particular, for improving the drought resistance of poplar by genetics manipulation.

Keywords: *Populus trichocarpa*; multiple organellar RNA editing factor; drought stress; RNA editing; genome

1. Introduction

Poplar (Populus) has enormous economic and ecological benefits. It has a relatively small genome (485 ± 10 Mb; $2n = 38$) and is the model forest species for genomic and genetic studies of woody plants owing to the relative ease of experimental manipulation and range of available genetic tools [1,2]. It is characterized by its height, thickness, and rapid growth, but biotic and abiotic stresses have limited its growth [3,4]. Among these stresses, drought is commonly pervasive with the characteristics of repressing stomata, photosynthesis, respiration, altering gene expression, and reducing biomass [5,6].

Drought is destructive and economically damaging for poplar, with important research and economic values [1,2,7,8]. Accordingly, investigating the stress resistance mechanisms of poplar has research and practical implications.

RNA editing primarily occurs in mitochondria and plastids of land plants and plays an important role in transcript maturation by insertions/deletions and conversion editing [9–11]. Conversion editing includes C-to-U, U-to-C, and A-to-I editing. U-to-C and A-to-I editing are likely less than C-to-U, which account for the vast majority of RNA editing events [12,13]. The cis-regulatory region, between nucleotide positions −40 and −5 of upstream of the edited site, interacts with trans-regulatory elements involved in identifying the site and catalyzing the conversion, from cytidine to uridine [14]. The organelle RNA recognition motif-containing protein (ORRM), organelle zinc finger (OZ), protoporphyrinogen IX oxidase (PPO), pentatricopeptide repeat (PPR), and RNA editing factor interacting protein (RIP)/multiple organellar RNA editing factor (MORF) families are key trans-acting elements for RNA editing [11,15,16]. The *ORRM* genes are required for RNA editing; the family contains six genes named *ORRM1* to *ORRM6*. Plastid and mitochondrial editing sites are impaired in almost all *orrm*-mutants [16–20]. OZ1 and PPO1 have also recently been implicated in RNA editing and are all located in plastids [21,22]. PPR proteins include many family members and have been studied extensively [23]. Several PPR functions are in a broad range of events, including developmental and physiological processes and response to various biotic and abiotic stresses [24]. For instance, a PPR96 deficiency in *A. thaliana* is associated with insensitivity to ABA and oxidative stress [23]. Mutations in SLO2 of the E+ subclass of the P-L-S subfamily of PPRs retard leaf emergence, restrict root growth, and result in late flowering, and these parameters are enhanced in the absence of sucrose. Additionally, four RNA editing defects and reduced editing at three sites in *slo2* mutants have been identified [25]. The conversion of nucleotides can not succeed without protein–protein interactions between certain PPRs and MORFs. *$(PLS)_3$PPR* and *LPA66*—two PPR genes—are associated with increased RNA-binding activity based on the presence of *MORF9* binding in *A. thaliana* [26]. MORF proteins interact with specific PPR proteins based on pull down in vivo and yeast two-hybrid assays [15].

MORFs are a small protein family in land plants (nine members in *A. thaliana* and seven members in maize) [27,28]. The name of DAG-like (DAL) gene family in maize was previously adopted based on the first identified member (DAG) of the gene family in *Antirrhinum majus* [27,28]. In *A. thaliana*, MORF proteins were also named RNA editing factor interacting proteins (RIP). There are nine members, defined as MORF1–9, while there are ten RIPs. RIPs or MORFs have been used interchangeably, except for RIP1, which corresponds to MORF8, and RIP8, which has been referred to MORF1. Only one gene, *RIP10* (At1g53260), was specially not defined as belonging to the MORF family [9]. In *A. thaliana*, all MORF proteins have no annotated domains but shared a similarly conserved domain. MORF2 and MORF9 are targeted to plastids, and MORF8 is located in chloroplasts and mitochondria, and the others are located in mitochondria. Some *MORF* genes work together and influence each other during some RNA editing events, i.e., they exhibit homo- and heteromeric interactions. For example, both mitochondrial MORF1 and plastid MORF2 proteins can interact with the dual-targeted MORF8 protein. MORF2 and MORF9 are both required for editing at several sites. The presence or absence of MORF8 influences edited sites targeted by MORF2 and/or MORF9 [27,29].

Several members of the RNA editosome interacted with *MORF* family genes towards their important roles in plant growth, development, and RNA editing efficiency [30,31]. Plant development would be negatively affected without MORF proteins. In rice, owing to impaired chloroplast development, the *wsp1* mutant has the variegated phenotype and reduced chlorophyll content. Further, photosynthetic efficiency, CO_2 gas conductance, and transpiration rate of *wsp1* plants are lower than those of the wild type [32]. In *A. thaliana*, *morf2* and *morf9* mutants show a lack of chlorophyll in leaves, and the T-DNA insertional *rip1* (RIP1 also named as MORF8) mutant demonstrates dwarfism [27,33].

In poplar, functional studies of MORF proteins focused on biotic and abiotic stresses are sparse. In this study, we predicted nine putative *MORF* genes in the *P. trichocarpa* genome. A comprehensive

analyses of the poplar MORF family, including phylogenetic, gene structure, chromosomal distribution, and synteny analyses, were performed. The expression profiles of *PtrMORF* genes under drought were determined using public microarray and quantitative RT-PCR data. Our results provide insight into the fascinating properties and biological functions of *MORF* genes in response to drought stress in poplar.

2. Results

2.1. Identification and Sequence Analysis of the PtrMORF Gene Family

We searched the poplar genome with known *A. thaliana* MORF proteins as queries. Initially, nine putative *PtrMORF* genes were obtained—*PtrMORF1.1*, *PtrMORF1.2*, *PtrMORF1.3*, *PtrMORF2.1*, *PtrMORF2.2*, *PtrMORF3*, *PtrMORF8.1*, *PtrMORF8.2*, and *PtrMORF9* based on a phylogenetic analysis using the amino acid sequences of all MORFs of *P. trichocarpa* as well as those of *A. thaliana* (Figure 1A). Poplar MORF proteins were predicted using TargetP and Wolf PSORT to enter mitochondria, chloroplasts, or nuclei, like their homologs in *A. thaliana* (Figure 1A). No known motif in poplar MORF proteins was found in the PFAM and INTERPRO databases, but the MORF box was identified, as in previous studies [27,28,33]. Novel putative motifs were explored using the MEME server. By selecting a motif length of between 15 and 50 aa, we identified four conserved motifs located in the MORF domain (Figure 1B–D).

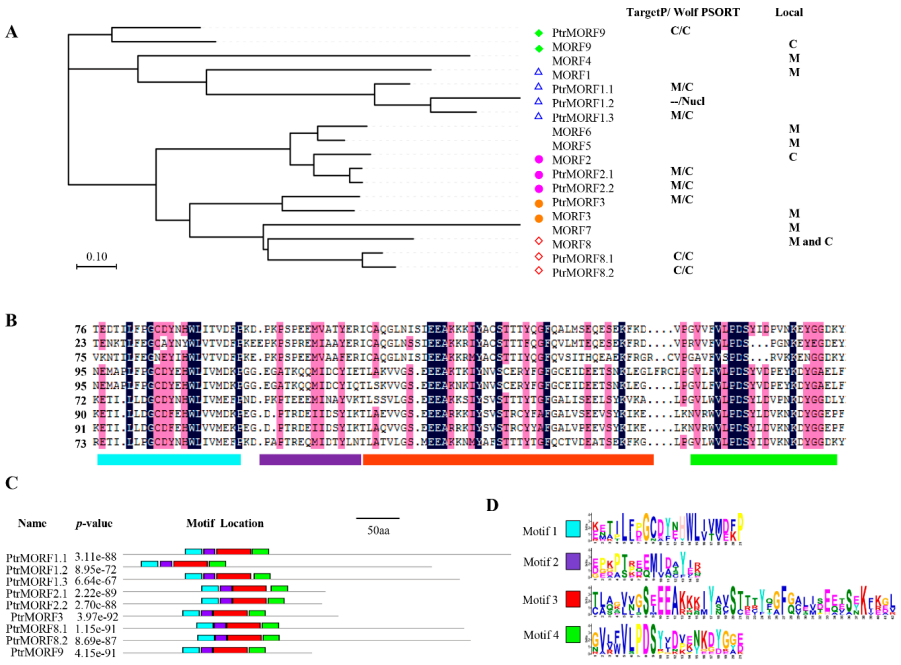


Figure 1. Poplar multiple organellar RNA editing factor (MORF) proteins and their conserved motifs. (**A**) Multiple sequences alignment of *A. thaliana* and *P. trichocarpa* MORF proteins was carried out using MUSCLE, and the Neighbor-Joining (NJ) tree was built using MEGA v7.0. And the chloroplast, mitochondrial or nuclei transit peptides of poplar MORF proteins were predicted using TargetP (http://www.cbs.dtu.dk/services/TargetP/) and Wolf PSORT (http://wolfpsort.org/). M, mitochondria; C, chloroplast; Nucl, nuclei. Green diamond, MORF9 and PtrMORF9; Blue triangle, MORF1, PtrMORF1.1, PtrMORF1.2 and PtrMORF1.3; Red circle, MORF2, PtrMORF2.1 and PtrMORF2.2; Orange circle, MORF3 and PtrMORF3; Orange diamond, MORF8, PtrMORF8.1 and PtrMORF8.2. Local meant the gene where was located in (**B**). Alignment of conserved MORF domains in poplar MORF proteins was conducted using DNAMAN (https://www.lynnon.com/). Lake blue, Motif1; purple, Motif2; red, Motif3; green, Motif4. (**C**,**D**) Putative motifs were explored using the MEME server with the parameters of between 15 and 50 aa in length and sharing of each motif among all PtrMORF proteins.

We further analyzed the sequence structures of the nine *PtrMORF* genes. An alignment of the genomic sequences to predicted CDS sequences of *PtrMORF* genes, showed that *PtrMORF* genes had a conserved gene structure. The *PtrMORF* genes had three introns with intron phases 0, 1, and 2. In all *PtrMORF* genes, motif 1 was encoded by exon 1, motif 2 was encoded by exon1 and exon 2, and motif 3 was encoded by exons 3 and 4, but motif 4 was only located in exon 4 of *PtrMORF1.1* (Figure 2A,B). The nine *PtrMORF* genes ranged from 1515 to 4441 bp and contained four or five exons (Figure 2A). All identified poplar *MORF* genes encoded proteins ranging from 229 to 470 amino acids, and their sequences contained zero to two transmembrane domains (TMDs). The molecular weight (MW) of the nine putative proteins ranged from 26.0 to 51.7 kDa. The GRAVY values of putative MORFs were negative and ranged from −1.382 to −0.602 (Table S1).

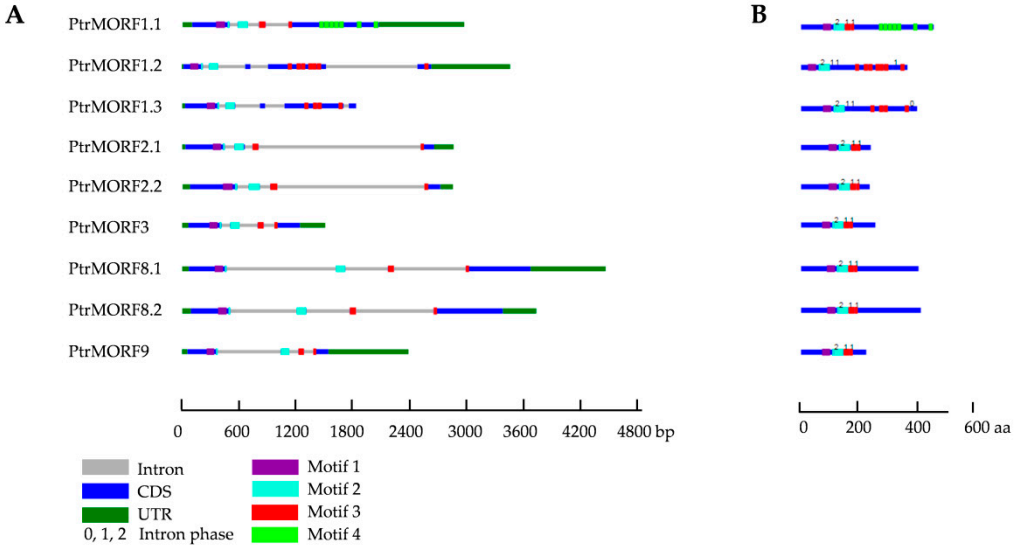


Figure 2. Gene structures of poplar *MORF* genes. (**A**,**B**) The gene structures of *PtrMORF* genes were built using GSDraw (http://wheat.pw.usda.gov/piece/GSDraw.php) by submitting both genomic sequences obtained from PopGenIE (http://popgenie.org/) and coding sequences (CDS) of *PtrMORF* genes.

2.2. Phylogenetic Comparison of the MORF from Different Species

To investigate their molecular evolution and functions in poplar, a phylogenetic analysis of MORF proteins was performed. We used the HMMER 3.0 package to build a Hidden Markov Model (HMM) file (morf.hmm) with 19 MORF domain sequences of MORF proteins from *A. majus*, *P. trichocarpa*, and *A. thaliana* (Table S2). To mine additional MORF domain-encoding genes in other plants, we used the morf.hmm algorithm to query the genomes of six species, representing major evolutionary lineages, including *Arabidopsis lyrata*, *Brachypodium distachyon*, *Glycine max*, *Oryza sativa Japonica*, *Prunus persica*, and *Vitis vinifera*. The numbers of MORF genes in eight species were comparable, ranging from six (in *P. persica*) to 11 (in *A. lyrata* and *G. max*).

In total, 69 MORF genes were identified in eight plant genomes to build an unrooted tree using MEGA7.0 by employing the neighbor-joining (NJ) method. As shown in Figure 3, sequences were classified into six groups (I, II, III, IV, V, and VI). Each group included *MORF* genes from diverse plant taxa. Among these, classes III and VI were larger than the others, containing 30 members and accounting for 42.5% of all predicted *MORF* genes. *PtrMORF* genes were found in all classes other than group III.

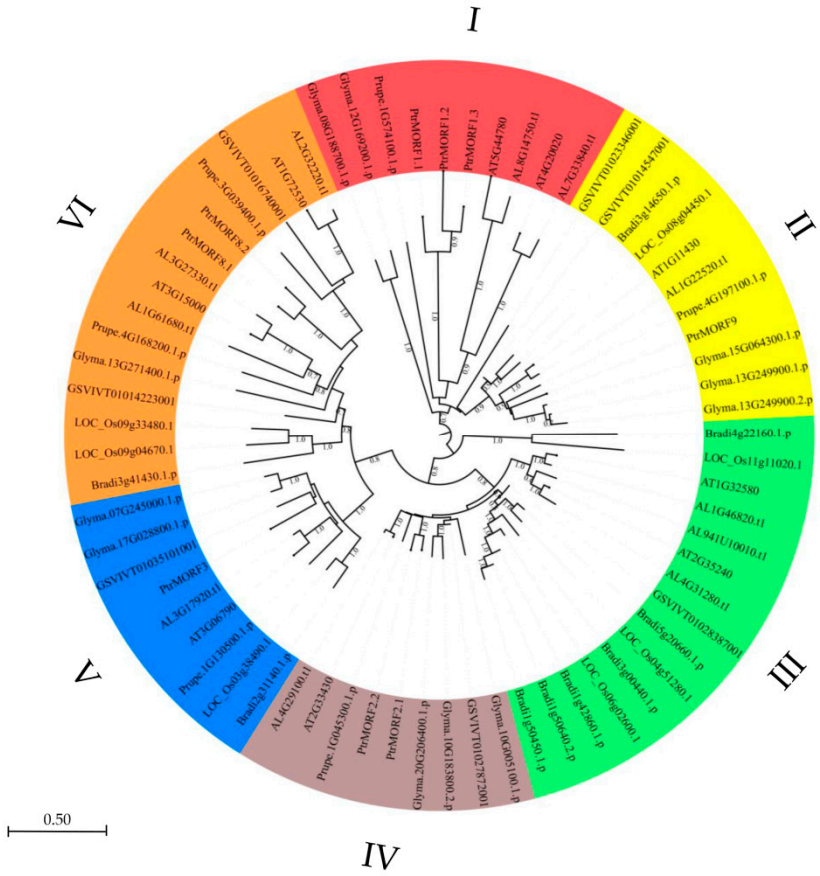


Figure 3. Phylogenetic relationships of the *MORF* gene family members from poplar (*PtrMORFs*) *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine max*, *Oryza sativa Japonica*, *Prunus persica*, and *Vitis vinifera*. The phylogenetic tree was constructed with MEGA 7.0 (http://www.megasofware.net/mega.html/) program by the neighbor-joining method.

2.3. Chromosomal Distribution, Synteny, and Evolution of PtrMORF Genes

Figure 4 showed that the genes were distributed on six poplar chromosomes, including chromosomes 1, 3, 4, 8, 10, and 11. Half of the chromosomes had two *PtrMORF* members and the other half had a single *MORF*. Tandem duplication was defined as different members of the gene family occurring within the same or neighboring intergenic region [34]. One tandem duplication event involving two *MORF* genes (*PtrMORF1.2*/*PtrMORF1.3*) was identified by BLASTP and MCScanX methods. In addition to the tandem duplication, seven *PtrMORF* genes were assigned to three segmental duplication events (*PtrMORF1.1/PtrMORF1.2/PtrMORF1.3*, *PtrMORF2.1/PtrMORF2.2*, and *PtrMORF8.1/PtrMORF8.2*) in Populus linkage groups 1, 3, 4, 8, 10, and 11. Notably, *PtrMORF1.2* and *PtrMORF1.3* occurred in both tandem and segmental duplications. A cross-matching event was also found in three doubling blocks of *PtrMORF* genes; the chromosomal fragment in which a gene was located was identical to more than one nonself chromosomal segment. *PtrMORF1.1*, *PtrMORF1.2*, and *PtrMORF1.3* synteny blocks corresponded to the third, fourth, and fourth chromosomes, respectively (Figure 5A and Table S3). Figure 5B showed five of the nine *MORF* genes were involved in three segmental duplication events (*MORF1/MORF4*, *MORF5/MORF6*, and *MORF8*/AT1G53260) in *A. thaliana*. It was worth noting that the gene, AT1G53260, was not included in the scope of *MORF* genes because of partial MORF box in our research [27]. The *MORF* gene of *A. thaliana* was also shown to be highly segmental duplicated, which was similar to that in poplar.

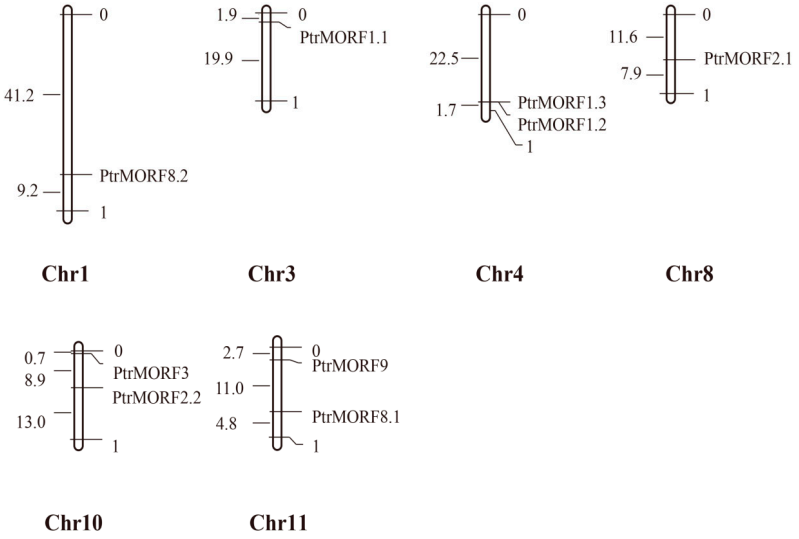


Figure 4. Chromosomal locations of Populus *MORF* family members. The chromosomal locations of the *MORF* genes were mapped with MapDraw program. The number to the left of each chromosome represented the size of the chromosome in Mbp.

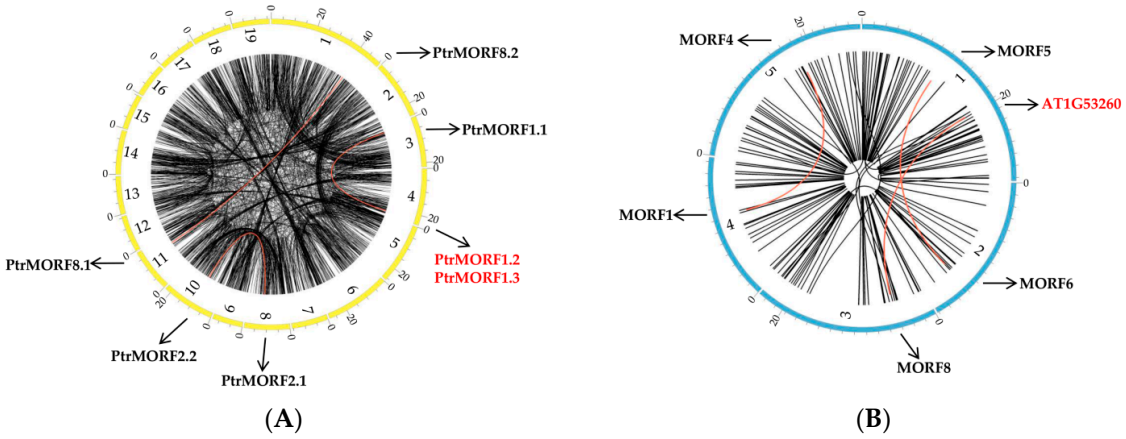


Figure 5. Interchromosomal relationships of *MORF* genes. (**A**) Interchromosomal relationships of *MORF* genes in poplar. (**B**) Interchromosomal relationships of *MORF* genes in *A. thaliana*. Black lines indicated all synteny blocks in the Populus or *A. thaliana* genome and the red indicated synteny blocks where *MORF* duplicated gene pairs were. The chromosome number was indicated at the bottom of each chromosome and each cell on the outside of chromosomes represented 5 Mbp in *Populus* and 1.5 Mbp in *A. thaliana*.

We further inferred the phylogenetic relationships of *PtrMORFs*, based on the diverse roles of DYW-PPR protein binding to MORFs on RNA editing sites in *A. thaliana* [9]. We constructed two syntenic maps of poplar with *A. thaliana* and rice. A total of four *PtrMORF* genes showed syntenic relationships with *MORFs* in *A. thaliana* (Figure 6A,C and Table S3). However, no synteny was in rice (Figure 6B). We performed phylogenetic analyses of these MORF proteins in Populus and *A. thaliana*. The ratio of nonsynonymous/synonymous substitution rates of *PtrMORFs* and *MORFs* in *A. thaliana* was determined to evaluate the selection pressure on amino acid substitutions ($\omega = d_N/d_S$) and the role of Darwinian positive selection in driving gene divergence after duplication [35–37]. Generally, $\omega > 1$ indicates positive selection, $\omega < 1$ provides evidence for negative or purifying selection, and $\omega = 1$ supports neutral evolution.

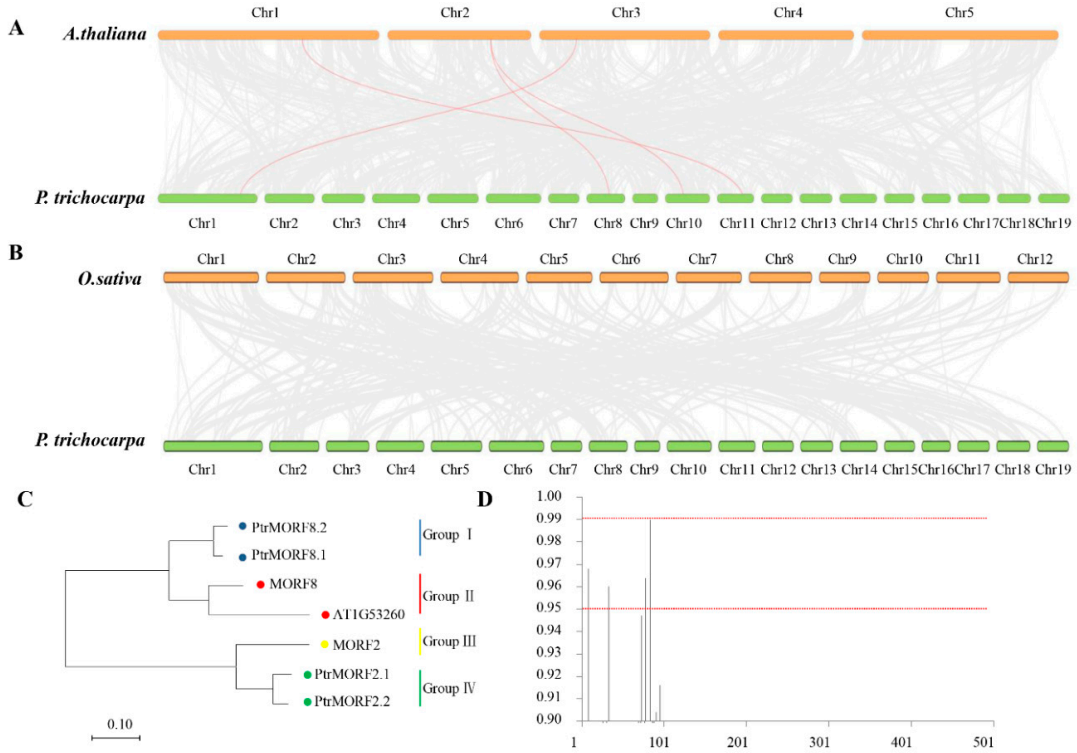


Figure 6. Synteny analysis of MORF genes between poplar and two representative plant species. (**A**,**B**) Gray lines in the background indicate the collinear blocks within poplar and other plant genomes, while the red lines highlight the syntenic MORF gene pairs. (**C**) The phylogenetic relationship of 7 members of four pairs. (**D**) The Bayes Empirical Bayes (BEB) probabilities for sites in the positively selected class ($\omega > 1$). The x-axis denotes position in the amino acid alignment.

Using the maximum likelihood method and codon substitution models implemented in PAML, the selection pressure in the four groups of *MORF* genes was evaluated by likelihood ratio tests (LRTs). The estimated ω value for all MORF genes was 0.055 using a one-ratio model (M0). We then detected the positive selection acting on particular groups using a branch model in which each clade had its own ω value. The LRT statistic suggested that the ω values for groups II, III, and IV were significantly different from that of group I, and the ω estimates for all groups other than group I still suggested purifying selection (Table 1).

Thus, we estimated the evolutionary forces acting on individual codon sites using site-specific likelihood models of codon substitution because positive selection was unlikely to affect all sites over prolonged time periods. Three pairs of models—M1 (neutral) and M2 (selection), M0 (one ratio) and M3 (discrete), and M7 (beta) and M8 (beta & ω)—formed three LRTs. As shown in Table 2, model M1 was not significantly worse than M2, although it suggested that 6.6% of sites were nearly neutral with $\omega = 1$. Model M3 with K = 8 suggested that 0.7% of sites were under positive selection, and model M3 was significantly better than the one rate model. Model M8, in which an additional ω ratio was estimated from the data, was not significantly better than M7, indicating that no sites were under positive selection.

To better detect positive selection, the branch-site model was also applied to evaluate selection on all amino acids of MORF proteins in specific groups (Table 3). LRT showed that model A fitted the data significantly better than the site-specific model M1 ($p < 0.01$) in group II, implying positive selection on 29.7% of sites in group II. At the posterior probability (p) > 95%, four sites were likely to be under positive selection in group II. Referring to first sequence in group II, *MORF8*, these positively selected sites were 8T, 33R, 78D, and 83V (Figure 6D). However, no positive selection was found in the branch including *PtrMORFs*.

Table 1. Parameter estimates and likelihood ratio tests for the branch model.

Model	P [a]	LnL [b]	Estimates of Parameters	ΔLRT [c]	df	*p*	Positively Selected Sites
M0 (one ratio model)	1	−1366.204	$\omega = 0.055$	-	-	-	None
Branch-specific model (Model 2: two ratios)							
Estimate ω for I	2	−1366.136	$\omega_{I1} = 0.067$, $\omega_0 = 0.051$	Model 2 vs. M0: 0.136	1	0.712	-
Estimate ω for II	2	−1357.726	$\omega_{II1} = 0.177$, $\omega_0 = 0.028$	Model 2 vs. M0: 16.956	1	0.000	-
Estimate ω for III	2	−1361.616	$\omega_{III1} = 0.018$, $\omega_0 = 0.079$	Model 2 vs. M0: 9.176	1	0.002	-
Estimate ω for IV	2	−1363.519	$\omega_{IV1} = 0.010$, $\omega_0 = 0.083$	Model 2 vs. M0: 5.370	1	0.002	-

a: The number of free parameters for the ω ratios; b: lnL value; c: likelihood ratio test.

Table 2. Parameter estimates and likelihood ratio tests for the site model.

Model	P [a]	LnL [b]	Estimates of Parameters	ΔLRT [c]	df	*p*	Positively Selected Sites
M0 (one ratio model)	1	−1366.204	ω = 0.0546	-	-	-	None
Site-specific models							
M1 (K = 2)	1	−1363.770	*p*0 = 0.926, (*p*1 = 0.074)	-	-	-	Not allowed
M2 (K = 3)	4	−1363.770	*p*0 = 0.926, *p*1 = 0.008, (*p*2 = 0.066), ω2 = 1.000	M2 vs. M1: 0.000	2	1.000	None
M3 (K = 8)	5	−1346.619	*p*0 = 0.485, *p*1 = 0.001, *p*2 = 0.192, *p*3 = 0.174, *p*4 = 0.029, *p*5 = 0.040, *p*6 = 0.074, *p*7 = 0.007; ω0 = 0.006, ω1 = 0.062, ω2 = 0.062, ω3 = 0.062, ω4 = 0.266, ω5 = 0.266, ω6 = 0.266, ω7 = 0.266	M3 vs. M0: 39.170	14	0.000	None
M7	2	−1347.154	*p* = 0.441, q = 6.096	-	-	-	Not allowed
M8	4	−1347.154	*p*0 = 0.999, *p* = 0.441, q = 6.096, (*p*1 = 0.000), ω = 1.000	M8 vs. M7: 0.000	2	1.000	None

a: The number of free parameters for the ω ratios; b: lnL value; c: likelihood ratio test.

Table 3. Parameter estimates and likelihood ratio tests for the branch-site model.

Model	P [a]	LnL [b]	Estimates of Parameters	ΔLRT [c]	df	*p*	Positively Selected Sites
M1: Neutral	1	−1363.770	$p0 = 0.926$, ($p1 = 0.074$)	-	-	-	-
Branch-site models							
Model A (I)	3	−1361.803	$p0 = 0.929$, $p1 = 0.000$, ($p2a + p2b = 0.071$), $\omega2 = 1.706$	Model A vs. M1: 3.934	2	1.40×10^{-1}	-
Model A (II)	3	−1346.578	$p0 = 0.684$, $p1 = 0.018$, ($p2a + p2b = 0.297$), $\omega2 = 1.053$	Model A vs. M1: 34.384	2	3.40×10^{-8}	Site for foreground lineage: 4 (at $p > 0.95$)
Model A (III)	3	−1363.521	$p0 = 0.927$, $p1 = 0.057$, ($p2a + p2b = 0.015$), $\omega2 = 17.935$	Model A vs. M1: 0.500	2	7.78×10^{-1}	-
Model A (IV)	3	−1363.522	$p0 = 0.925$, $p1 = 0.064$, ($p2a + p2b = 0.012$), $\omega2 = 2.630$	Model A vs. M1: 0.496	2	7.81×10^{-1}	-

a: The number of free parameters for the ω ratios; b: lnL value; c: likelihood ratio test.

2.4. Expression of PtrMORF Genes and Six Genes from Chloroplasts and Mitochondria under Drought Stress

Although the role of MORF proteins in plastid development and RNA editing in *A. thaliana* and rice have been studied, little is known about how *MORFs* respond to abiotic stimuli, such as drought stress, particularly in poplar [27,32].

Quantitative real-time reverse transcription-PCR (qRT–PCR) for long-term water deficiency stress in black poplar (Populus × *euramericana cv. 'Neva'*) was further performed to evaluate the differential expression of each *PtrMORF* gene. Nine genes exhibited significantly different expression under limited water stress except *PtrMORF2.1*, and these could be preliminarily considered as drought-responsive genes. Among them, the two *PtrMORF* genes *PtrMORF1.3* and *PtrMORF8.1* were highly expressed followed by low expression. The expression levels of *PtrMORF1.1* and *PtrMORF1.2*, *PtrMORF2.2*, *PtrMORF3*, and *PtrMORF8.2* under drought conditions were significantly lower than those in control conditions. The mRNA accumulation of *MORF9* fluctuated obviously, exhibiting decreased expression after 3-day and 9-day drought, but increased approximately 1.5-fold after 6-day and 12-day drought. However, the pattern of the response to drought was not consistent. For example, the expression level of *PtrMORF1.1* had downregulated after 3-day drought stress. At the same time, *PtrMORF1.2* and *PtrMORF8.2* were significantly downregulated until 12 days of limited water stress. Furthermore, *PtrMORF* genes that belonged to the same group in the phylogenetic tree had different expression patterns under stress treatment. For instance, *PtrMORF1.1* and *PtrMORF1.3* were downregulated and upregulated in 3-day drought treatments, while *PtrMORF1.2*, also orthologous to group I members and *MORF1* in *A. thaliana*, was not affected by limiting water until 12 days post-treatment (Figure 7).

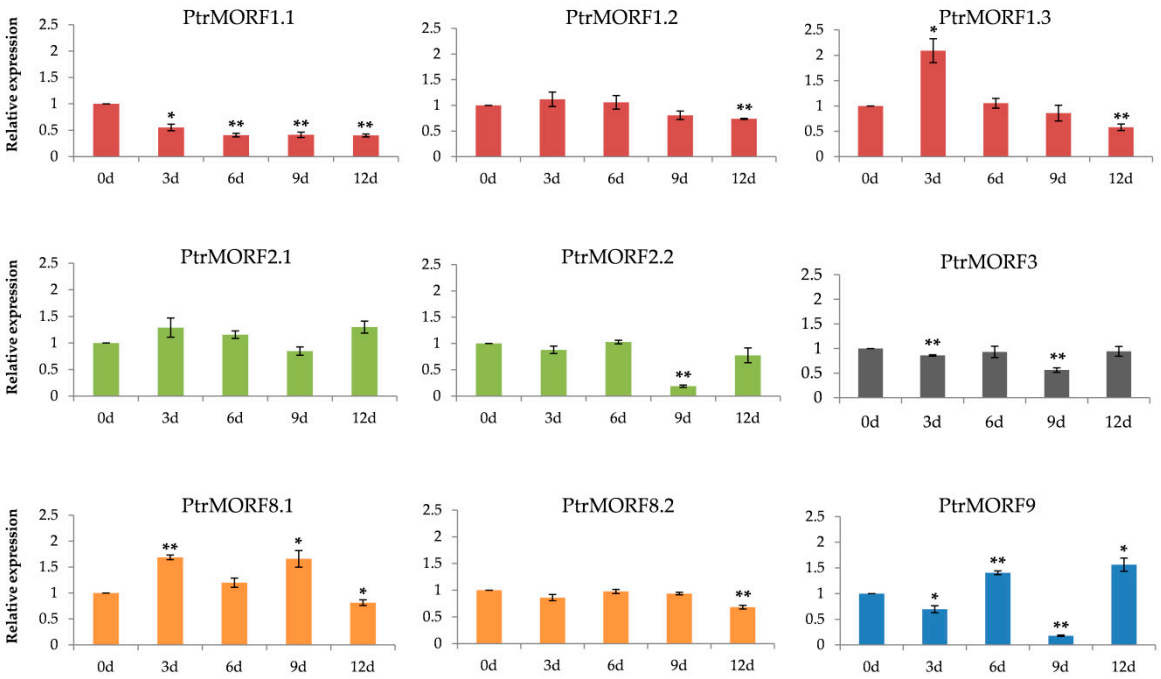


Figure 7. Quantitative RT-PCR analysis of *PtrMORF* genes expression in response to drought stress. 0d, 3d, 6d, 9d, and 12d, drought for 0, 3, 6, 9, and 12 days in a greenhouse environment, respectively. Data were normalized to UBQ gene. The sample without drought treatment was defined as 1 in the figure. The data were presented as the mean ± SE of three separate measurements. Asterisks denote significant differences: * $p < 0.05$; ** $p < 0.01$.

In order to further explore the possible relationship between RNA editing and plant stress, we selected a total of six genes from chloroplasts and mitochondria to evaluate their expression, based on previous studies [27,38,39]. The *RPS14* gene represented important chloroplast proteins, and the fluctuant editing efficiency and gene expression of *RPS14* were associated with cytokinins

against stress [39,40]. Under salt stress, *PSBF* and *NDHB* likely showed an elevated editing percentage, linked to PSII repair and increase in *NDHB* gene translation [39]. In rice, *atp6*, *atp9*, and *ccmC* were related to mitochondrial electron transport chain and had been selected to test RNA editing during stress exposure [38]. In our study, these reported genes in poplar were obtained from National Center for Biotechnology information and showed different expression owing to drought. Six genes, except *PSBF*, were significantly upregulated or downregulated after drought stress (Figure 8). The three mitochondrial genes—*atp6*, *atp9*, and *ccmC*—were significantly downregulated after 3-day drought. The level of *RPS14* was highly expressed on 3-day drought followed by a low expression on 6-day drought, and the expression of poplar *NDHB* was significantly increased under 12-day drought.

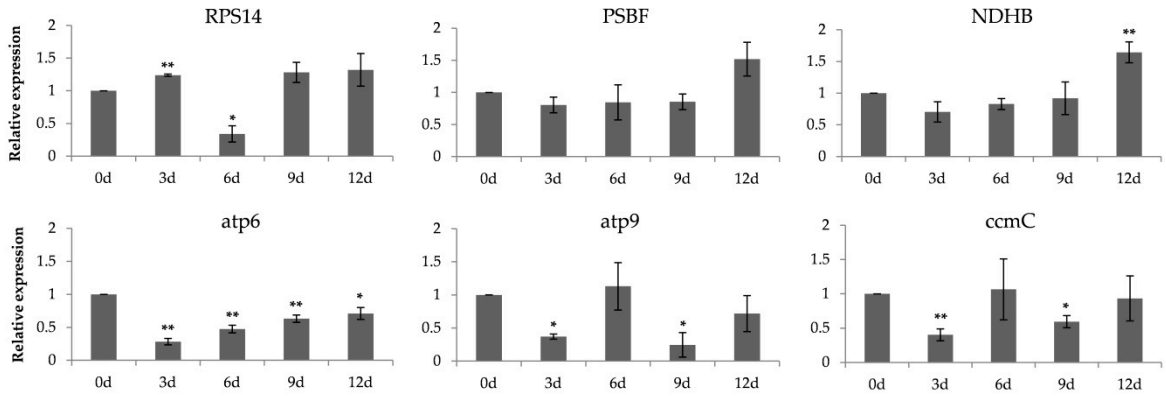


Figure 8. Expressions of three plastid (RPS14, PSBF, and NDHB) and three mitochondrial genes (atp6, atp9, and ccmC) in poplar on the 0, 3rd, 6th, 9th, and twelfth day after drought tolerance. Data were normalized to UBQ gene. The sample without drought treatment was defined as 1 in the figure. The data were presented as the mean ± SE of three separate measurements. Asterisks denote significant differences: * $p < 0.05$; ** $p < 0.01$.

2.5. Expression Profiles of Populus MORF Gene

We then examined the tissue-specific expression of nine Populus *MORF* genes by qRT-PCR: we evaluated various tissue types, including buds, freshly expanded leaves, expanding young leaves, mature leaves, old leaves, cortex, xylem, and roots. The expression levels of *PtrMORFs* in these tissues were comparable to those in buds. Some *PtrMORF* genes exhibited clear tissue-specific expression (Figure 9). Five *PtrMORF* genes—*PtrMORF1.1*, *PtrMORF8.1*, *PtrMORF2.2*, *PtrMORF3*, and *PtrMORF9*—were higher or weakly expressed in over four tissues significantly. Among them, the latter three genes were highly expressed level in almost all leaves types, and *PtrMORF8.1* and *PtrMORF1.1* had higher expression levels in freshly expanded leaves and old leaves, respectively. All but one gene—*PtrMORF2.1*—had higher expression in xylem. *PtrMORF1.2*, *PtrMORF1.3*, and *PtrMORF8.2* had no significant difference in their expression in leaves and merely higher expressed in xylem compared to buds. Additionally, The *PtrMORF* genes with closest evolutionary relationship had different expression patterns. *PtrMORF1.1*, *PtrMORF1.2*, and *PtrMORF1.3* had the closest homology relationship with *MORF1* of *A. thaliana.* Among them, *PtrMORF1.1* had high expression levels in old leaves, cortex, xylem, and roots, while *PtrMORF1.2* and *PtrMORF1.3* only in xylem. With respect to *PtrMORF2.1* and *PtrMORF2.2*, which were orthologous to *MORF2* of *A. thaliana*, the latter had high expression level in four leaves types and the former exhibited high expression only in old leaves. A similar inconsistency in expression was observed between *PtrMORF8.1* and *PtrMORF8.2*.

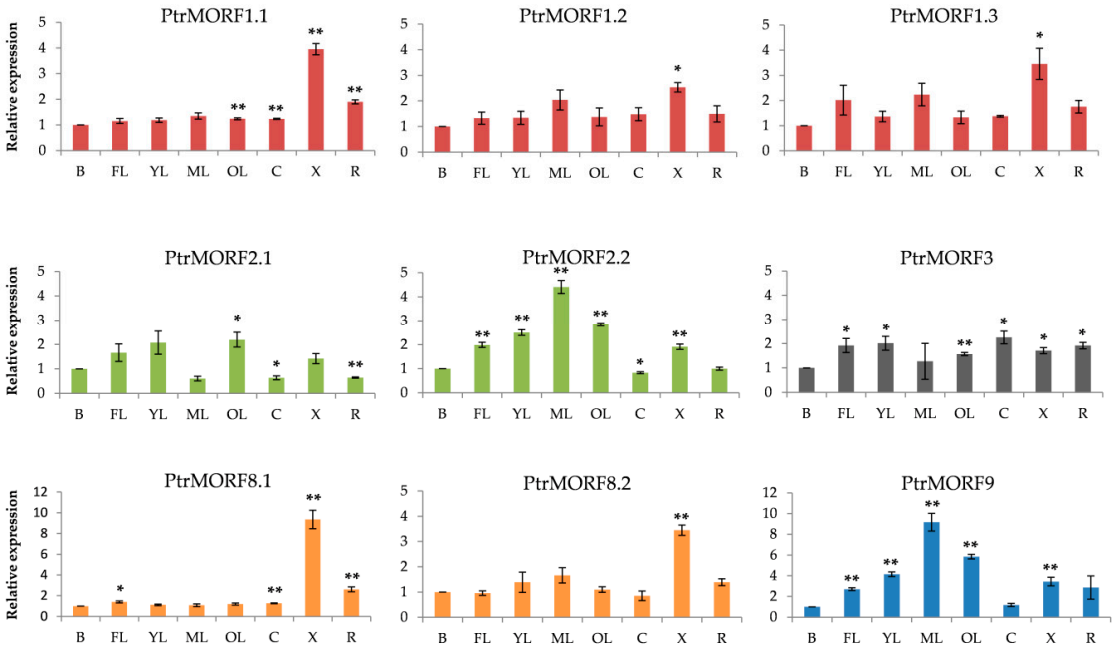


Figure 9. Expression analysis of 9 *MORF* genes in eight representative samples by qRT-PCR. Eight various tissues including buds (B), freshly expanded leaves (FL), expanding young leaves (YL), mature leaves (ML), old leaves (OL), cortex (C), xylem (X), and roots (R). Data were normalized to UBQ gene. The buds were defined as 1 in the figure. The data were presented as the mean ± SE of three separate measurements. Asterisks denote significant differences: * $p < 0.05$; ** $p < 0.01$.

3. Discussion

RNA editing plays an irreplaceable role in plant growth and development and C-to-U RNA editing events occur frequently in vascular plants. Trans-acting factors (RNA editosome) are required to recognize nucleotides to be edited, including OZ1, PPR, ORRM, and others. MORF proteins that interact with these factors have been found in *A. thaliana* and *O. sativa* [22,26,32]. *MORF* genes, which are important subunits of the RNA editosome, play a vital role in the regulation of RNA editing [11]. It is worth noting that the functions of most MORF proteins in response to stress, especially drought, remain unclear in woody plant. In this study, we identified the whole *MORF* gene family in poplar and examined the expression patterns of these genes in different plant tissues and in response to drought. The structure of the MORF family, evolutionary events, transcriptional changes responded to drought, and tissue-specific expression pattern are discussed below.

Members of the MORF family have been identified and characterized in many taxa. For example, 10 *MORF* genes have been identified in *A. thaliana*, but At1g53260 (RIP10) may exhibit a partial lack of partial functionality owing to incomplete *MORF* box. We focus on nine *MORF* genes in *A. thaliana* in this study [9,27]. In maize, seven putative the DAG-like (*DAL*) genes have also have been identified. It is worth noting that DAL, MORF, and RIP genes are the same [28]. We screened the *P. trichocarpa* genome for putative *MORF* genes using a tailor-made HMM file derived from multiple MORF domain alignments. We identified nine *PtrMORF* genes consistent with the findings of previous report [28]. Compared with *MORF* genes in *A. thaliana* and in maize (*Zea mays*), the family has not been extensively studied in poplar, indicating that *P. trichocarpa* MORF proteins might perform functions similar to those preformed by their homologs in herbaceous plants.

In our study, to better understand the evolution of the *MORF* gene family in poplar, the structure, conserved motifs, phylogenetic relationships, and collinearity of *PtrMORF* genes were characterized. Four conserved motifs were located in the MORF domain, suggesting that the MORF domain is conserved among *A. thaliana* and *Populus* proteins. Most of the *PtrMORF* genes exhibited similar numbers of exons. A phylogenetic analysis revealed that *PtrMORF* genes and putative *MORF* genes

from other species could be classified into six subgroups. The distribution of genes among the subclasses indicated that the expansion of the MORF family occurred before the divergence of the species. Most of *PtrMORF* genes were grouped with *MORF* genes from *A. thaliana* and *P. persica* genomes, indicating a close relationship among *MORF* genes from these species. Gene duplication is a major mechanism underlying the evolution of novel protein functions. We detected 2 (22.2%) *PtrMORF* genes that were tandemly duplicated and seven genes (77.8%) that were segmentally duplicated, implying low tandem and high segmental duplication rates in *PtrMORF* genes. Three homologous pairs of chromosomes included seven of the nine *PtrMORF* genes, with segmental duplication in the poplar genome. For example, homologous chromosomes 8 and 10 both contained one *MORF* gene each; similar findings were obtained for homologous chromosomes 1 and 11, as well as 3 and 4. Two *MORF* genes—Potri.010G007200 and Potri.011G032900—were detected on chromosomes 10 and 11, but their homologous chromosomes 8 and 1 lacked *PtrMORF* genes. Tandem duplication was detected (Table S3). Furthermore, the synteny block including *PtrMORF1.1*, *PtrMORF1.2*, and *PtrMORF1.3* was attributed to multiple copies of the chromosome. These results indicated that some *PtrMORF* genes were possibly generated by gene duplication and segmental duplication events likely served as driving force in *PtrMORF* evolution.

Additionally, synteny maps between two representative species and poplar were constructed to better understand the phylogenetic relationships. Four pairs were detected in *A. thaliana* and none was detected in rice indicating a weak homology relationship between poplar and rice. The ω values for duplicated gene pairs between *PtrMORF* genes and *MORFs* in *A. thaliana* were calculated to explore selective pressures (Tables 1–3). In general, the *MORF* genes in the two plants were under strong purifying selection using the branch model. The ω values for four groups were less than 1, except for group II, *MORF8* (At3G15000) and At1G53260 (RIP10), indicating that purifying selective pressure was strong. Positively selected sites might cause adaptive changes after gene duplication and during the evolution of *MORFs* in *A. thaliana*.

RNA editing could potentially contribute to plant resistance to abiotic tolerance on the basis of previous studies [38,41,42]. And some genes closely related to RNA editing, such as *PPR* genes in *A. thaliana* or rice, their mutants changed morphological characteristics due to environmental forces [32,43]. Additionally, *MORF* genes were interacted with *PPR* genes to establish complex editosomes in plant [26,27]. Therefore, we made an attempt to confirm whether the *PtrMORF* genes responded to stress. The eight *PtrMORF* genes distributed in all groups were upregulated or downregulated significantly under drought treatment. It was indicated that they may respond to drought stress. Additionally, the sensitivity of *PtrMORF* genes responding to drought stress was different. The expressions of *PtrMORF1.2*, *PtrMORF2.2*, and *PtrMORF8.2* changed significantly after nine or even 12 days of drought, while the other five genes had lower or higher expression only in three-day drought restriction. This might implied that the former three genes are less sensitive to drought than the latter five. However, not all the genes, which were responded to drought on day three, had consistent response as drought stress increases. This was similar to the response of rice some *PPR* genes to drought stress [44]. Additionally, five of the six genes from poplar chloroplasts and mitochondria showed obviously higher or lower expression compared with no drought treatment, and the edited site efficiency of these genes were affected when the *MORF* genes were mutated out in *A. thaliana* [27], suggesting that the *PtrMORF* gene family, as important component of RNA editing, might be involved in the response to drought stress. However, many questions were worth exploring: (1) How would RNA editing perform under stress? (2) What were the deeper mechanisms of *MORF* genes, as important editosome members of RNA editing?

The functional divergence of *MORF* genes was speculated depended on their tissue-specific expression. Firstly identified member in *A. majus*, the *DAG* gene is essential for chloroplast development in the leaves and etioplast formation of cotyledons [45]. In *A. thaliana*, the *DAG-like* gene is involved in early chloroplast differentiation [46]. Our study showed that six *PtrMORF* genes are expressed highly in the leaves of black poplar (Populus × *euramericana cv. 'Neva'*), suggesting that

*PtrMORF*s might play an important role in chloroplast differentiation and development. Some genes that were highly expressed highly in both the leaves and other tissues, such as the cortex, xylem, and roots, indicating the expanded function of *PtrMORF*s. The expression of three genes, merely higher expressed in xylem compared to buds. There are homologous genes in *A. thaliana*; *MORF1* was located in both mitochondria and *MORF8* in both mitochondria and chloroplasts and possessed an extended C-terminus with unknown function. In previous reports, glycine-rich regions were observed in the C-termini of MORF1 and MORF8 in *A. thaliana* [16]. The genes exhibited glycine-rich regions may play key roles in the biotic and abiotic stresses [47]. In addition, the firstly discovered glycine-rich protein—*GRP-1*—was highly expressed in buds and vascular tissue in the petunia [48]. Taken together, these reports suggested this hypothesis that glycine-rich regions observed in *MORFs* might support them in responding to adversity and be highly expressed in xylem in poplar. Xylem was important for water transportation in plants and *PtrMORF* genes expressed highly in xylem might be closely related to water regulation, making the *MORF* genes likely respond to drought [49]. However, we had to point out that the exact roles within *PtrMORFs* need further experiments. Additionally, there were still living parenchyma cells in the xylem and the activity of parenchyma cells provided a critical metabolic and energetic role in woody stem [49]. They played a major role in editing implying that the three *MORF* genes in poplar may play unlikely similar roles in annual plants and were probably required in other critical functions in perennial woody plants.

Therefore, this research preliminarily provided insight into the roles of *PtrMORF*s in stress response and investigations of functions for these genes required further experimental validation in future studies.

4. Materials and Methods

4.1. Plant Materials and Treatments

Black poplar (Populus × *euramericana cv. 'Neva'*) were grown in a greenhouse at 25 °C under a 16/8 h light/dark cycle. The 3.5-month-old plants were subjected to drought stress. Twelve plants were encountered water-limited treatment, ranging from 3 to 12 days of drought in which the soil RWC was reduced from 70%; the other three plants were provided abundant water, with a the soil RWC of greater than 30% [50]. Three biological replicates were performed.

Total RNA was isolated from poplar mature leaves (sixth to twelfth) after drought stress. Different organs and tissues of 3.5-month-old plants, including buds, freshly expanded leaves (second to third), expanding young leaves (fourth to fifth), mature leaves (sixth to twelfth), old leaves (leaves below the mature leaves), cortex, xylem (the stem without cortex), and roots, were collected at the same time and immediately immersed in liquid nitrogen.

4.2. Genome-Wide Identification and Sequence Analysis of MORF Genes in P. trichocarpa

The potential *MORF* genes in *P. trichocarpa* were queried using a local BLASTP search with an E-value threshold of $<10^{-10}$ and a bit score of >100 in poplar genome annotation data (Ptrichocarpa_210_v3.0.protein.fa.gz downloaded from https://phytozome.jgi.doe.gov/pz/portal.html) based on preexisting MORF/RIP genes, including At1g11430, At1g32580, At1g72530, At2g33430, At2g35240, At3g06790, At3g15000, At4g20020, and At5g44780 from *A. thaliana*, which was then confirmed in previous studies [27,28]. Multiple sequence alignments of PtrMORF proteins and the known MORF proteins were generated using DNAMAN. The alignment results were used to build protein Hidden Markov Models (HMMs) to mine the conserved domain; the file was named morf.hmm and was generated using the hmmbuild program in the HMMER 3.0 package (version 3.1b2). No known motifs in PtrMORF proteins and MORF proteins of other plants were detected by screening the PFAM (http://pfam.sanger.ac.uk/) and INTERPRO (http://www.ebi.ac.uk/interpro/) databases [51,52]. The HMM file was used as a probe to search genome files of representative species downloaded from Phytozome V11.0, including *A. lyrata*, *B. distachyon*, *G. max*, *O. sativa Japonica*,

P. persica, and *V. vinifera* [53]. Protein hits with an E-value of $<10^{-10}$ and sequence score of "best 1 domain" >100 were collected.

MEME (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) was used to investigate the putative conserved motifs among PtrMORF proteins with the following parameters: length between 15 and 50 aa, maximum number of motifs = 4, and one per sequence. To obtain the intact conserved MORF domain, different limits for the length of each motif were used between 100 and 120 aa [54]. In addition, TargetP, and Wolf PSORT were used to predict the putative organelle localization of PtrMORF proteins [55,56].

4.3. Phylogenetic Analysis

A multiple sequence alignment of 69 MORF proteins from *P. trichocarpa* and other species including *A. lyrata*, *A. thaliana*, *B. distachyon*, *G. max*, *O. sativa Japonica*, *P. persica*, and *V. vinifera* was generated using the MUSCLE method. A phylogenetic tree was constructed by using the NJ method implemented in MEGA V7.0 [57]. The parameters for tree construction were as follows. Phylogeny test and options: bootstrap (1000 replicates); gaps/missing data: pairwise deletion; model: Dayhoff model; pattern among lineages: same (homogeneous); and rates among sites: uniform rates. Finally, the phylogenetic tree was visualized using itol (http://itol.embl.de/) [58].

4.4. Chromosome Location and Gene Structure Analysis

Positional information and gene structures of *PtrMORF* genes on chromosomes of *P. trichocarpa* were obtained from the PopGenIE database [59]. The chromosomal locations were displayed with MapDraw program [60]. The numbers and organization of introns and exons and gene structures were drawn and displayed using the online PIECE2 server of GSDraw [61].

4.5. Colinearity of PtrMORF Gene

The chromosomal locations of *PtrMORF* genes were obtained from PopGenIE. Multiple Collinearity Scan Toolkit (MCScanX) was used to analyze gene duplication events with the default parameters and the colinearity information was showed by Circos [62,63]. Using Dual Systeny Plotter software (https://github.com/CJ-Chen/TBtools), the synteny relationships of orthologous *MORF* genes among poplar, *A. thaliana*, and rice were evaluated [64]. Nonsynonymous (d_N) and synonymous (d_S) substitutions, as well as d_N/d_S ratio (ω) for duplicated *MORF* genes, were calculated, using branch-specific, site-specific (one, neutral, selection, discrete, beta, and beta & $\omega > 1$), and branch-site (Model A) of codon substitution models, as implemented in PAML version 4 [65–67]. Likelihood ratio tests (LRTs) were used to compare model fits. The sites under positive selection were identified using Bayesian methods [68].

4.6. RT-PCR and qRT-PCR Analyses

Total RNA was extracted using a Plant RNA EASYspin Plus Kit (Aidlab, Beijing, China) and first-strand cDNA synthesis was performed using approximately 2 μg of RNA using the FastQuant RT Kit (with gDNase) (TIANGEN, Beijing, China) following the manufacturer's protocols. qRT-PCR was conducted on the ABI StepOnePlus Real-Time PCR System (ABI, Foster City, CA, USA) based on the SYBR Green II method. Twenty microliters of cDNA was diluted 1:10 with nuclease-free water. Each reaction contained 10 μL of SYBR Green qPCR Mix (Aidlab), 0.2 μL of ROX Reference Dye (Aidlab), 1 μL of cDNA (corresponding to 10 ng of total RNA), 7.8 μL of nuclease-free water, and 0.25 μM each primer. The thermal profile for qRT-PCR: 3 min at 94 °C, 40 cycles of 10 s at 94 °C, 20 s at 60 °C, and 72 °C for 30 s. Primers were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) [69]. A list of primers used was given (Table S4). Each experiment was performed in three biological replicates. The poplar housekeeping gene, UBQ, was used as an internal control. Relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method [70]. A Student's *t*-test was used to generate every *p*-value for

statistical analyses, and R project was used to identify significant variance (R version 3.5.1, one sample and two sample *t*-test: * $p < 0.05$; ** $p < 0.01$).

5. Conclusions

Genome-wide identification and evolutionary, gene structure, and expression analyses under drought stress of poplar *MORF* genes provide a deep insight into the gene family. Moreover, analyses of *MORF* genes based on qRT-PCR validation of poplar tissues provide functional information. Our results provide an important resource for advancing our understanding of the roles of *MORF* genes in poplar.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/6/1425/s1.

Author Contributions: D.W. and S.M. designed and conducted all of the experiments. X.X., C.L. and W.Y. conceived the project. W.S., Y.B. and Y.L. helped perform the experiments. D.W. contributed to analyze the experimental results. D.W., S.M., C.L. and X.X. wrote the paper.

Funding: This research was sponsored by grants from the National Key Program on Transgenic Research (2018ZX08021001) and the National Natural Science Foundation of China (31770649, 31570308, and 31600484).

Acknowledgments: We thank Jinbu Jia for bioinformatics analysis, experimental design, and useful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tuskan, G.A.; DiFazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* **2006**, *313*, 1596–1604. [CrossRef]
2. Tuskan, G.A.; Difazio, S.P.; Teichmann, T. Poplar Genomics is Getting Popular: The Impact of the Poplar Genome Project on Tree Research. *Plant Biol.* **2004**, *6*, 2–4. [CrossRef] [PubMed]
3. Bradshaw, H.D.; Ceulemans, R.; Davis, J.; Stettler, R. Emerging Model Systems in Plant Biology: Poplar (Populus) as A Model Forest Tree. *J. Plant Growth Regul.* **2000**, *19*, 306–313. [CrossRef]
4. Suzuki, N.; Rivero, R.M.; Shulaev, V.; Blumwald, E.; Mittler, R. Abiotic and biotic stress combinations. *New Phytol.* **2014**, *203*, 32–43. [CrossRef] [PubMed]
5. Zhao, Z.; Zhang, C.; Meng, S.; Li, M. Transcriptomic insight into nitrogen uptake and metabolism of Populus simonii in response to drought and low nitrogen stresses. *Tree Physiol.* **2018**, *38*, 1672–1684. [CrossRef]
6. Shinozaki, K.; Yamaguchi-Shinozaki, K. Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* **2006**, *58*, 221–227. [CrossRef] [PubMed]
7. Zhu, J.K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **2002**, *53*, 247–273. [CrossRef] [PubMed]
8. Zhu, J.-K. Abiotic Stress Signaling and Responses in Plants. *Cell* **2016**, *167*, 313–324. [CrossRef]
9. Bentolila, S.; Oh, J.; Hanson, M.R.; Bukowski, R. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* **2013**, *9*, e1003584. [CrossRef]
10. Oldenkott, B.; Yamaguchi, K.; Tsuji-Tsukinoki, S.; Knie, N.; Knoop, V. Chloroplast RNA editing going extreme: More than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte Selaginella uncinata. *RNA* **2014**, *20*, 1499–1506. [CrossRef]
11. Sun, T.; Bentolila, S.; Hanson, M.R. The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant. Sci.* **2016**, *21*, 962–973. [CrossRef] [PubMed]
12. Chateignerboutin, A.L.; Small, I. Plant RNA editing. *RNA Biol.* **2010**, *7*, 213–219. [CrossRef]
13. Freyer, R.; Kiefermeyer, M.C.; Kössel, H. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6285–6290. [CrossRef] [PubMed]
14. Takenaka, M.; Neuwirt, J.; Brennicke, A. Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* **2004**, *32*, 4137–4144. [CrossRef] [PubMed]
15. Yan, J.; Zhang, Q.; Yin, P. RNA editing machinery in plant organelles. *Sci. China Life Sci.* **2018**, 1–8. [CrossRef]

16. Bayer-Csaszar, E.; Haag, S.; Jorg, A.; Glass, F.; Hartel, B.; Obata, T.; Meyer, E.H.; Brennicke, A.; Takenaka, M. The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochim. Biophys. Acta Gene Regul. Mech.* **2017**, *1860*, 813–828. [CrossRef]
17. Shi, X.; Castandet, B.; Germain, A. ORRM5, an RNA recognition motif-containing protein, has a unique effect on mitochondrial RNA editing. *J. Exp. Bot.* **2017**, *68*, 2833–2847. [CrossRef]
18. Hackett, J.B.; Shi, X.; Kobylarz, A.T.; Lucas, M.K.; Wessendorf, R.L.; Hines, K.M.; Bentolila, S.; Hanson, M.R.; Lu, Y. An Organelle RNA Recognition Motif Protein is Required for Photosynthetic Subunit psbF Transcript Editing. *Plant. Physiol.* **2017**, *173*, 2278–2293. [CrossRef]
19. Shi, X.; Bentolila, S.; Hanson, M.R. Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. *Plant Signal. Behav.* **2016**, *11*, 294–309. [CrossRef]
20. Shi, X.; Hanson, M.R.; Bentolila, S. Two RNA recognition motif-containing proteins are plant mitochondrial editing factors. *Nucleic Acids Res.* **2015**, *43*, 3814–3825. [CrossRef]
21. Zhang, F.; Tang, W.; Hedtke, B.; Zhong, L.; Liu, L.; Peng, L.; Lu, C.; Grimm, B.; Lin, R. Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2023–2028. [CrossRef] [PubMed]
22. Sun, T.; Shi, X.; Friso, G.; Van Wijk, K.; Bentolila, S.; Hanson, M.R. A zinc finger motif-containing protein is essential for chloroplast RNA editing. *PLoS Genet.* **2015**, *11*, e1005028. [CrossRef] [PubMed]
23. Liu, J.M.; Zhao, J.Y.; Lu, P.P.; Chen, M.; Guo, C.H.; Xu, Z.S.; Ma, Y.Z. The E-Subgroup Pentatricopeptide Repeat Protein Family in Arabidopsis thaliana and Confirmation of the Responsiveness PPR96 to Abiotic Stresses. *Front. Plant Sci.* **2016**, *7*, 1825. [CrossRef] [PubMed]
24. Xing, H.; Fu, X.; Yang, C.; Tang, X.; Guo, L.; Li, C.; Xu, C.; Luo, K. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Sci. Rep.* **2018**, *8*, 2817. [CrossRef] [PubMed]
25. Zhu, Q.; Dugardeyn, J.; Zhang, C.; Takenaka, M.; Kuhn, K.; Craddock, C.; Smalle, J.; Karampelias, M.; Denecke, J.; Peters, J.; et al. SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant. J.* **2012**, *71*, 836–849. [CrossRef]
26. Yan, J.; Zhang, Q.; Guan, Z.; Wang, Q.; Li, L.; Ruan, F.; Lin, R.; Zou, T.; Yin, P. MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat. Plants* **2017**, *3*, 17037. [CrossRef]
27. Takenaka, M.; Zehrmann, A.; Verbitskiy, D.; Kugelmann, M.; Hartel, B.; Brennicke, A. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5104–5109. [CrossRef]
28. Luo, M.; Cai, M.; Zhang, J.; Li, Y.; Zhang, R.; Song, W.; Zhang, K.; Xiao, H.; Yue, B.; Zheng, Y.; et al. Functional divergencse and origin of the DAG-like gene family in plants. *Sci. Rep.* **2017**, *7*, 5688. [CrossRef]
29. Zehrmann, A.; Hartel, B.; Glass, F.; Bayer-Csaszar, E.; Obata, T.; Meyer, E.; Brennicke, A.; Takenaka, M. Selective homo- and heteromer interactions between the multiple organellar RNA editing factor (MORF) proteins in Arabidopsis thaliana. *J. Biol. Chem.* **2015**, *290*, 6445–6456. [CrossRef]
30. Du, L.; Zhang, J.; Qu, S.; Zhao, Y.; Su, B.; Lv, X.; Li, R.; Wan, Y.; Xiao, J. The Pentratricopeptide Repeat Protein Pigment-Defective Mutant2 is Involved in the Regulation of Chloroplast Development and Chloroplast Gene Expression in Arabidopsis. *Plant Cell Physiol.* **2017**, *58*, 747–759. [CrossRef]
31. Xiao, H.; Xu, Y.; Ni, C.; Zhang, Q.; Zhong, F.; Huang, J.; Liu, W.; Peng, L.; Zhu, Y.; Hu, J. A rice dual-localized pentatricopeptide repeat protein is involved in organellar RNA editing together with OsMORFs. *J. Exp. Bot.* **2018**, *69*, 2923–2936. [CrossRef] [PubMed]
32. Zhang, Z.; Cui, X.; Wang, Y.; Wu, J.; Gu, X.; Lu, T. The RNA Editing Factor WSP1 Is Essential for A Chloroplast Development in Rice. *Mol. Plant* **2017**, *10*, 86–98. [CrossRef] [PubMed]
33. Bentolila, S.; Heller, W.P.; Sun, T.; Babina, A.M.; Friso, G.; Wijk, K.J.V.; Hanson, M.R. RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8372–8373. [CrossRef]
34. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol.* **2004**, *4*, 10. [CrossRef] [PubMed]

35. Xu, Y.; Wang, Y.; Mattson, N.; Yang, L.; Jin, Q. Genome-wide analysis of the Solanum tuberosum (potato) trehalose-6-phosphate synthase (TPS) gene family: Evolution and differential expression during development and stress. *BMC Genom.* **2017**, *18*, 926. [CrossRef] [PubMed]
36. Yang, H.-L.; Liu, Y.-J.; Wang, C.-L.; Zeng, Q.-Y. Molecular evolution of trehalose-6-phosphate synthase (TPS) gene family in Populus, Arabidopsis and rice. *PLoS ONE* **2012**, *7*, e42438. [CrossRef] [PubMed]
37. Ma, B.; Yuan, Y.; Gao, M.; Xing, L.; Li, C.; Li, M.; Ma, F. Genome-wide Identification, Classification, Molecular Evolution and Expression Analysis of Malate Dehydrogenases in Apple. *Int. J. Mol. Sci.* **2018**, *19*, 3312. [CrossRef]
38. Xiong, J.; Tao, T.; Luo, Z.; Yan, S.; Liu, Y.; Yu, X.; Liu, G.; Xia, H.; Luo, L. RNA Editing Responses to Oxidative Stress between a Wild Abortive Type Male-Sterile Line and Its Maintainer Line. *Front. Plant Sci.* **2017**, *8*, 2023. [CrossRef]
39. Rodrigues, N.F.; Fonseca, G.C.; Kulcheski, F.R.; Margis, R. Salt stress affects mRNA editing in soybean chloroplasts. *Genet. Mol. Biol.* **2017**, *40*, 200–208. [CrossRef]
40. Yamburenko, M.V.; Zubo, Y.O.; Vanková, R.; Kusnetsov, V.V.; Kulaeva, O.N.; Börner, T. Abscisic acid represses the transcription of chloroplast genes. *J. Exp. Bot.* **2013**, *64*, 4491–4502. [CrossRef]
41. Kurihara-Yonemoto, S.; Handa, H. Low temperature affects the processing pattern and RNA editing status of the mitochondrial cox2 transcripts in wheat. *Curr. Genet.* **2001**, *40*, 203–208. [CrossRef]
42. Nakajima, Y.; Mulligan, R.M. Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr. Genet.* **2001**, *40*, 209–213. [CrossRef] [PubMed]
43. Tan, J.; Tan, Z.; Wu, F.; Sheng, P.; Heng, Y.; Wang, X.; Ren, Y.; Wang, J.; Guo, X.; Zhang, X.; et al. A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. *Mol. Plant* **2014**, *7*, 1329–1349. [CrossRef] [PubMed]
44. Chen, G.; Zou, Y.; Hu, J.; Ding, Y. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genom.* **2018**, *19*, 720. [CrossRef] [PubMed]
45. Chatterjee, M.; Sparvoli, S.; Edmunds, C.; Garosi, P.; Findlay, K.; Martin, C. DAG, a gene required for chloroplast differentiation and palisade development in Antirrhinum majus. *EMBO J.* **1996**, *15*, 4194–4207. [CrossRef] [PubMed]
46. Bisanz, C.; Begot, L.; Carol, P.; Perez, P.; Bligny, M.; Pesey, H.; Gallois, J.L.; Lerbs-Mache, S.; Mache, R. The Arabidopsis nuclear DAL gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Mol. Biol.* **2003**, *51*, 651–663. [CrossRef] [PubMed]
47. Ciuzan, O.; Hancock, J.; Pamfil, D.; Wilson, I.; Ladomery, M. The evolutionarily conserved multifunctional glycine-rich RNA-binding proteins play key roles in development and stress adaptation. *Physiol. Plant* **2015**, *153*, 1–11. [CrossRef] [PubMed]
48. Condit, C.M.; McLean, B.G.; Meagher, R.B. Characterization of the expression of the petunia glycine-rich protein-1 gene product. *Plant Physiol.* **1990**, *93*, 596–602. [CrossRef]
49. Spicer, R. Symplasmic networks in secondary vascular tissues: Parenchyma distribution and activity supporting long-distance transport. *J. Exp. Bot.* **2014**, *65*, 1829–1848. [CrossRef]
50. Wang, C.; Liu, S.; Dong, Y.; Zhao, Y.; Geng, A.; Xia, X.; Yin, W. PdEPF1 regulates water-use efficiency and drought tolerance by modulating stomatal density in poplar. *Plant Biotechnol. J.* **2016**, *14*, 849–860. [CrossRef]
51. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215. [CrossRef] [PubMed]
52. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]
53. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef] [PubMed]
54. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **2006**, *34*, W369. [CrossRef] [PubMed]

55. Emanuelsson, O.; Brunak, S.; Von, H.G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [CrossRef] [PubMed]
56. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adamscollier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, 585–587. [CrossRef]
57. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
58. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [CrossRef]
59. Sandberg, G. The Populus Genome Integrative Explorer (PopGenIE): A new resource for exploring the Populus genome. *New Phytol.* **2009**, *182*, 1013–1025. [CrossRef]
60. Liu, R.H.; Meng, J.L. MapDraw: A microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Yi Chuan* **2003**, *25*, 317–321.
61. Wang, Y.; Xu, L.; Thilmony, R.; You, F.M.; Gu, Y.Q.; Colemanderr, D. PIECE 2.0: An update for the plant gene structure comparison and evolution database. *Nucleic Acids Res.* **2017**, *45*, 1015–1020. [CrossRef] [PubMed]
62. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]
63. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef] [PubMed]
64. Liu, C.; Xie, T.; Chen, C.; Luan, A.; Long, J.; Li, C.; Ding, Y.; He, Y. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus). *BMC Genom.* **2017**, *18*, 503. [CrossRef]
65. Yang, Z.; Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [CrossRef] [PubMed]
66. Yang, Z.; Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **1998**, *46*, 409–418. [CrossRef] [PubMed]
67. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
68. Yang, Z.; Wong, W.S.; Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [CrossRef]
69. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]
70. Livak, K.J.; Schmittgen, T.D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Identification and Expression Profiling of Protein Phosphatases (*PP2C)* Gene Family in *Gossypium hirsutum* L.

Hamna Shazadee [1,†], Nadeem Khan [2,†], Jingjing Wang [1], Chencan Wang [1], Jianguo Zeng [1], Zhongyi Huang [1] and Xinyu Wang [1,*]

[1] College of Life Science, Nanjing Agricultural University, Nanjing 210095, China; 2018116157@njau.edu.cn (H.S.); 2017116115@njau.edu.cn (J.W.); 2018116102@njau.edu.cn (C.W.); 2016116113@njau.edu.cn (J.Z.); 2018116101@njau.edu.cn (Z.H.)

[2] State Key Laboratory of Crop Genetics and Germplasm Enhancement, Ministry of Science and Technology/College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China; 2016104235@njau.edu.cn

* Correspondence: xywang@njau.edu.cn

† These authors contributed equally to this work.

Received: 28 February 2019; Accepted: 18 March 2019; Published: 20 March 2019

Abstract: The protein phosphatase (*PP2C*) gene family, known to participate in cellular processes, is one of the momentous and conserved plant-specific gene families that regulate signal transduction in eukaryotic organisms. Recently, PP2Cs were identified in *Arabidopsis* and various other crop species, but analysis of PP2C in cotton is yet to be reported. In the current research, we found 87 (*Gossypium arboreum*), 147 (*Gossypium barbadense*), 181 (*Gossypium hirsutum*), and 99 (*Gossypium raimondii*) PP2C-encoding genes in total from the cotton genome. Herein, we provide a comprehensive analysis of the *PP2C* gene family in cotton, such as gene structure organization, gene duplications, expression profiling, chromosomal mapping, protein motif organization, and phylogenetic relationships of each species. Phylogenetic analysis further categorized *PP2C* genes into 12 subgroups based on conserved domain composition analysis. Moreover, we observed a strong signature of purifying selection among duplicated pairs (i.e., segmental and dispersed) of *Gossypium hirsutum*. We also observed the tissue-specific response of *GhPP2C* genes in organ and fiber development by comparing the RNA-sequence (RNA-seq) data reported on different organs. The qRT-PCR validation of 30 *GhPP2C* genes suggested their critical role in cotton by exposure to heat, cold, drought, and salt stress treatments. Hence, our findings provide an overview of the *PP2C* gene family in cotton based on various bioinformatic tools that demonstrated their critical role in organ and fiber development, and abiotic stress tolerance, thereby contributing to the genetic improvement of cotton for the resistant cultivar.

Keywords: protein phosphatase (PP2C); cotton; syntenic relationships; expression patterns; evolutionary analysis

1. Introduction

The protein kinases (PKs) and protein phosphatases (PPs) are known to regulate the protein function, and are the fundamental molecular mechanism, by reversing protein phosphorylation during cellular signaling. Thus, it is involved in many biological processes, such as signal transduction, development, and environmental stimuli [1]. The PKs phosphorylate largely serine (Ser), threonine (Thr), and tyrosine (Tyr), whereas PPs can reverse this functioning by eliminating the phosphate group [2]. Mainly, the PPs are subcategorized into three major groups based on their requirement

for substrate specificity, namely Ser/Thr phosphatases (STPs), protein Tyr phosphatases (PTPs), and dual-specificity phosphatases (DSPTPs) [3,4]. Moreover, based on crystalline structure, amino acid sequence and compassion to specific inhibitors (okadaic acid and cyclosporine-A), the PTPs are further classified into phosphoprotein metallophosphatase (PPM) and phosphoprotein phosphatases (PPP) [5]. The PPM family mainly involves Mn^{2+}- or Mg^{2+}-dependent protein phosphatase (PP2C) and pyruvate dehydrogenase phosphate, however, the PPP family includes different types of protein phosphatase, including PP1, PP4, PP5, PP6, PP7, PP2A, and PP2B [5]. In the PPP family, PP2A rigorously affects root hair growth during the elongation phase by denaturing the shape of the cells [6].

PP2C are evolutionarily conserved from Archaea to higher plants that pointedly modulate stress signaling pathways and reverse the stress-induced PK cascades to complex environmental stimuli [3]. From various literature, several key stress-responsive protein kinases genes have been extensively studied and proven to respond in diverse stress conditions including biotic and abiotic factors [7,8]. In *Arabidopsis*, several members of PP2c such as PLL4 and PLL5 (POL-like gene) are known to adjust leaf development, though with no obvious functions within the meristem [9].

In *Arabidopsis* and rice, 112 and 132, respectively, candidate *PP2C* genes have been characterized into various groups [10,11]. In particular, group A of *Arabidopsis* PP2Cs (e.g., ABI1 and ABI2), is accompanied by abscisic acid (ABA) signaling, group B is known to stimulate mitogen-activated protein kinase (MAPK) signaling, and group C is involved in the regulation of flower development [12]. Moreover, several members of PP2C have been implicated as a negative regulator in mediated stress signaling within ABA [13–15]. In higher plants, PP2Cs act as negative regulators of the ABA signaling pathway and reduce tolerance against oxidative stress [16]. Proteins encoded by these candidate genes play a critical role in various abiotic stress signaling such as salt, drought, and freezing [17–20]. Crop productivity encounters a number of abiotic stress signaling including salt, drought, heat, and osmotic stress in the environment. Thus, plants have developed complex molecular mechanisms to implement and survive during adverse growth conditions. Intriguingly, it can be contingent on quite a few factors, like adaptive alterations in their structure, physiology, and gene expression of some regulatory proteins [16]. In reaction to drought stress, the interaction of PP2C i.e., MPK3 and MPK6, assists plants in stomatal assimilation to thwart water loss [21], such as in the *Arabidopsis* [22]. Taken together, the above studies have established the diverse role of *PP2C* genes in plant development and environmental stresses. Hence, it is indispensable to probe into the identification and functional description of the *PP2C* gene family, which will cement the base for understanding its essential molecular mechanism in stress signaling.

Cotton, as an oil crop and an important source of natural textile fiber, plays a crucial role in agriculture and industry all around the world. However, its production is mainly constrained due to various abiotic and biotic stress conditions. The release of *Gossypium* whole-genome data in four different cotton species, such as *Gossypium arboreum* [2], *Gossypium barbadense* [23], *Gossypium hirsutum* [24], and *Gossypium raimondii* [25] and their publicly available database allows us to comprehensively characterize the *PP2C* gene family based on bioinformatic tools. We further compared the *PP2C* genes between *Gossypium* and *Arabidopsis*, to explore and identify both shared and specific subgroups. Following the gene structure organization analysis, conserved protein motifs, and cis-elements, we traced the duplication gene pairs and their evolutionary divergence that likely resulted in the widespread extension of the *PP2C* gene family. In order to shed light on some critical *PP2C* genes, their associated indigenous functional roles were further exposed to heat, cold, drought, and salt stress conditions. Additionally, the transcriptional profiling of the *PP2C* genes for various tissues and fiber development and qRT-PCR analysis of 30 genes were analyzed and compared. Therefore, to our knowledge, this is the first systematic report of the genome-wide expression dynamics of the *PP2C* genes in cotton species, and it is necessary to carry out a comprehensive study to understand the regulation of phosphatases in cotton during stress and development.

2. Results

2.1. Characterization of PP2C Gene Family in Cotton

The *Arabidopsis* 94 *PP2C* genes were obtained from TAIR (http://www.arabidopsis.org) with the help of Interpro scan domain (IPR001932) and then used as queries against *G. arboreum*, *G. barbadense*, *G. hirsutum*, and *G. raimondii* databases. This search resulted in 87, 147, 181, and 99 *PP2C* genes in *Gossypium* species i.e., *G. arboreum*, *G. barbadense*, *G. hirsutum*, and *G. Raimondii*, respectively. In the current study, we used the following nomenclature system for the *PP2C* genes to distinguish each PP2C from the homology of *Arabidopsis* and *Gossypium* species: GaPP2C (GaPP2C1–GaPP2C87) *G. arboreum*, GbPP2C (GbPP2C1–GbPP2C147) *G. barbadense*, GhPP2C (GhPP2C1–GhPP2C181) *G. hirsutum*, and GrPP2C (GrPP2C1–GrPP2C99) *G. raimondii*. We have studied different gene features of the *PP2C* genes including the chromosomal location, coding sequence length (CDS), protein length (aa), molecular weight (MW), isoelectric point (PI), grand average of hydropathicity (GRAVY), aliphatic index, and subcellular localization (Supplementary Tables S1–S4). In general, most of the PP2C proteins were in the range of 254–899 (GaPP2C1–GaPP2C66), 240–1630 (GbPP2C3–GbPP2C27), 263–794 (GhPP2C38–GhPP2C135), and 264–1101 (GrPP2C27–GrPP2C67) amino acids. The MWs of the proteins ranged from 27.99–100.38 (GaPP2C1–GaPP2C66), 26.45–181.58 (GbPP2C38–GbPP2C27), 28.65–100.47 (GhPP2C38–GhPP2C133), and 28.83–122.57 (GrPP2C27–GrPP2C1), the PIs ranged from 4.79–9.18 (GaPP2C28–GaPP2C7), 4.41–9.27 (GbPP2C96–GbPP2C135), 4.53–9.27 (GhPP2C57–GhPP2C177), and 4.64–9.27 (GrPP2C68–GrPP2C86), and the GRAVY ranged from −0.59–0.04 (GaPP2C66–GaPP2C64), −0.70–0.06 (GbPP2C27–GbPP2C102), −0.591–0.056 (GhPP2C132–GhPP2C127) and −0.587–0.071 (GrPP2C78–GrPP2C71). In our study, most of the *PP2C* genes showed hydrophilic properties with negative values, only a few of them were hydrophobic in nature with a positive value. The predicted subcellular localization results displayed that most of the PP2C proteins were confined in different organelles such as the chloroplast, the nuclear region, mitochondria, cytoplasm, and others (Supplementary Tables S1–S4).

2.2. Phylogenetic and Genomic Distribution and Organization Analysis of PP2C Gene Family

To study the phylogenetic relationship of the *PP2C* genes among cotton plants (*G. arboreum*, *G. barbadense*, *G. hirsutum*, and *G. raimondii*), a maximum likelihood (ML) tree was constructed with *Arabidopsis* using MEGA 7.0. The phylogenetic tree revealed that the *PP2C* genes can be further divided into 12 subgroups, as previously reported [26,27] (Supplementary Figures S1–S3). Moreover, 181 *GhPP2C* genes were clustered into 12 subgroups (A–L) with *Arabidopsis*. In the phylogenetic tree, subgroup D had the most members of the *PP2C* genes (38), followed by subgroup E (27), and the least number of *PP2C* genes was observed in subgroup L, having 4 *GhPP2C* genes (Figures 1 and 2a). Moreover, we also analyzed the conserved motifs and gene structure based on phylogenetic relationships to provide insight into the structural features of the PP2C members in *G. hirsutum*. For GhPP2C proteins, 10 conserved motifs were acquired with the help of MEME. The majority of the PP2C family contained motifs 7 and 2, except a few genes showed motifs 3 and 4. This indicates that all the identified *PP2C* genes have typical family features and the proteins classified into the same subgroup share similar amino acid sequences (Figure 2b). Alongside, their logos were obtained by online MEME server. A total of 10 types of consensus motifs were obtained in all of the GhPP2C proteins and their distribution patterns are presented in Supplementary Figure S4. To better understand the gene structure of *PP2C* genes in cotton, exon-intron organizations of these genes were also tested (Figure 2c) [26], and most of the subgroups contained 1-10 introns. These results specify that *PP2C* genes in the same subgroup show more or less similar exon-intron organization.

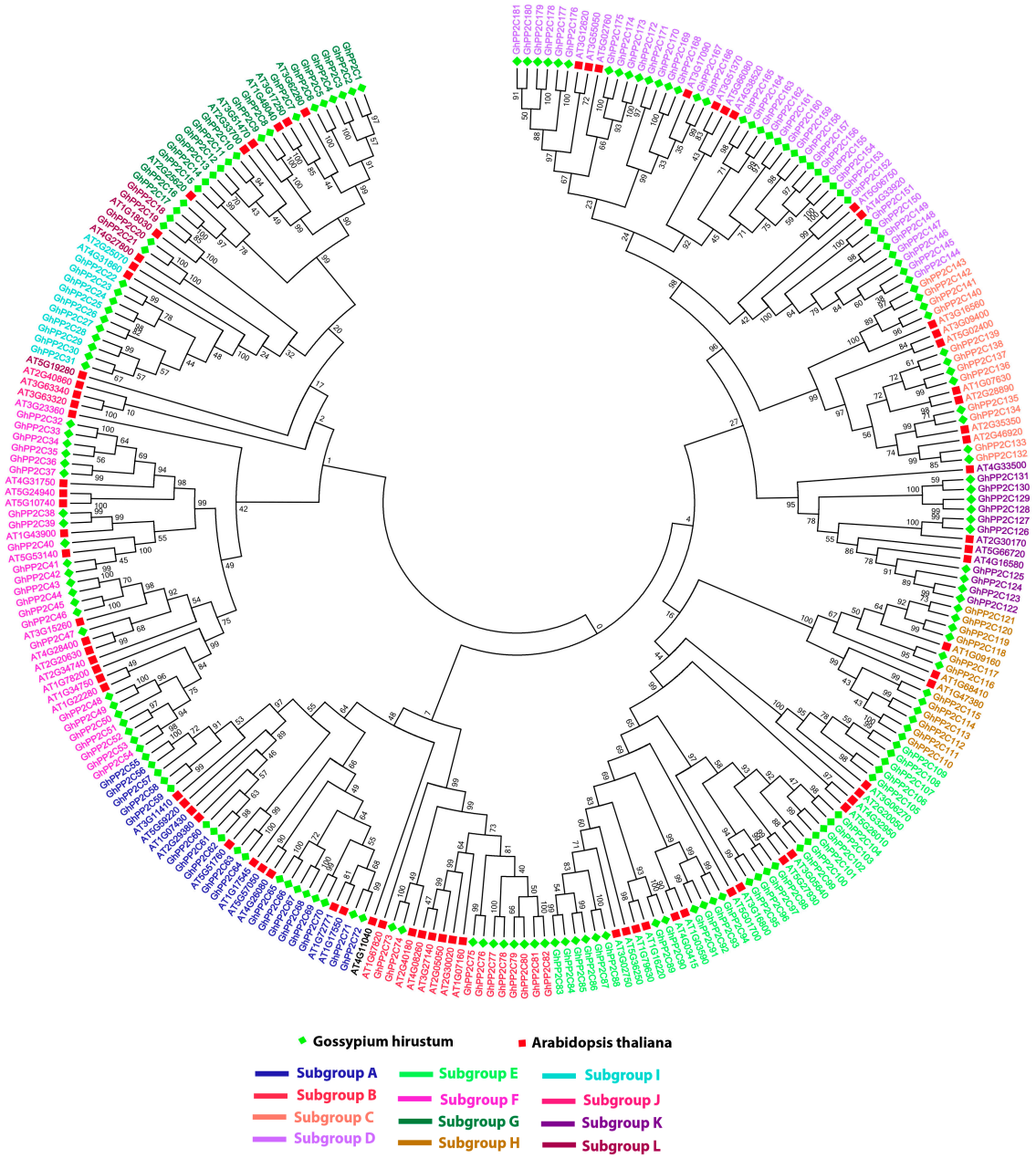


Figure 1. Phylogenetic relationship of *PP2C* genes between *G. hirsutum* and *A. thaliana*. The phylogenetic tree was constructed by MEGA 7 using the Maximum Likelihood Method (1000 bootstrap). Genes of protein phosphatase (PP2C) from different subgroups are marked with various color.

Figure 2. (**a**) (**b**) and (**c**). Phylogenetic relationships, gene structure, the exon-intron, and upstream/downstream region are represented, respectively. The phylogenetic tree was constructed by MEGA 7 using the Maximum Likelihood Method (1000 bootstrap). At the bottom of the figure, the relative position is proportionally displayed based on the kilobase scale.

2.3. Chromosomal Localization and Syntenic Relationships of PP2C Gene Family

The chromosomal localization of A and D genomes of *G. hirsutum* was analyzed using Tbtools software. A total of 89 *GhPP2C* genes were allocated in D genome (D01–D13) ranging from 2–14 genes per chromosome and only 2 genes were found on the scaffold. Moreover, every chromosome showed variation in a number of genes, such as D04 exhibited the highest number of *GhPP2C* genes (14), followed by 12 genes in D02, and the least number of genes (2) were recorded for D07 (Figure 3).

On the other hand, we also demonstrated the chromosomal localization for A genome (A01–A12). A total of 79 genes were found ranging from 2–15 per chromosomes and 9 of them were located on the scaffold. The highest number of genes was found on A05 (15), followed by A11 (8), and the least number of *PP2C* genes (2) were found on A13 (Figure 4). These findings suggested that GhPP2C were allocated unevenly to different chromosomal locations.

Moreover, a collinear correlation was also demonstrated between *G. hirsutum* and *Arabidopsis* (A and D genomes) (Figure 5a,b). To validate our results, we also performed a collinear relationship of *PP2C* genes using only *G. hirsutum* (Supplementary Figure S5). The results exhibited high conservation among PP2C members between the A and D genomes of cotton.

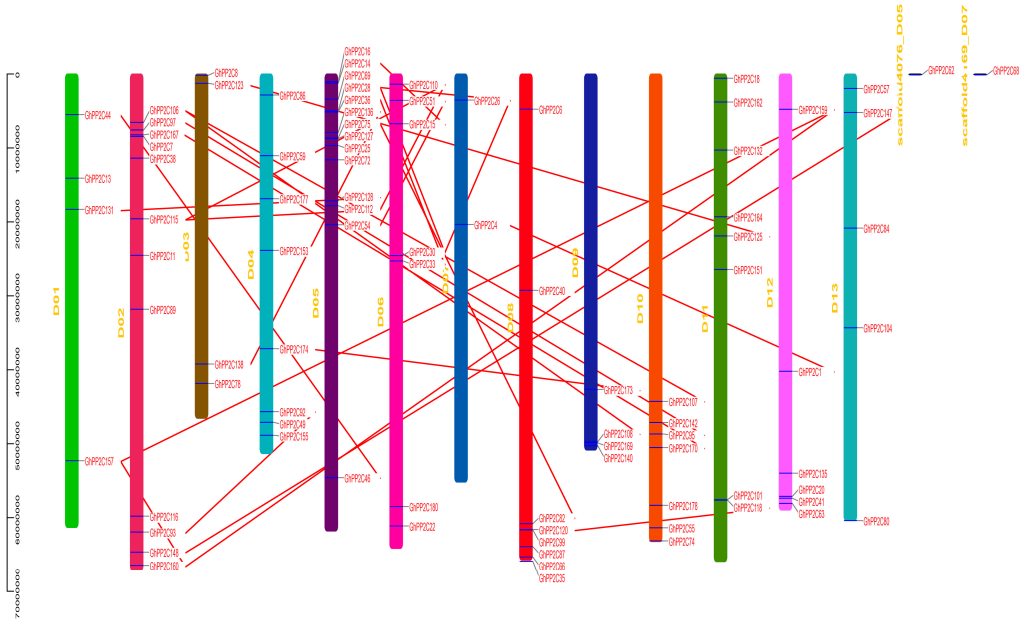


Figure 3. Chromosomal locations of the PP2C D genome of cotton were obtained from the Generic File Format (GFF) file and displayed using TBtools software. The red line indicates the collinear relationship for different chromosomes.

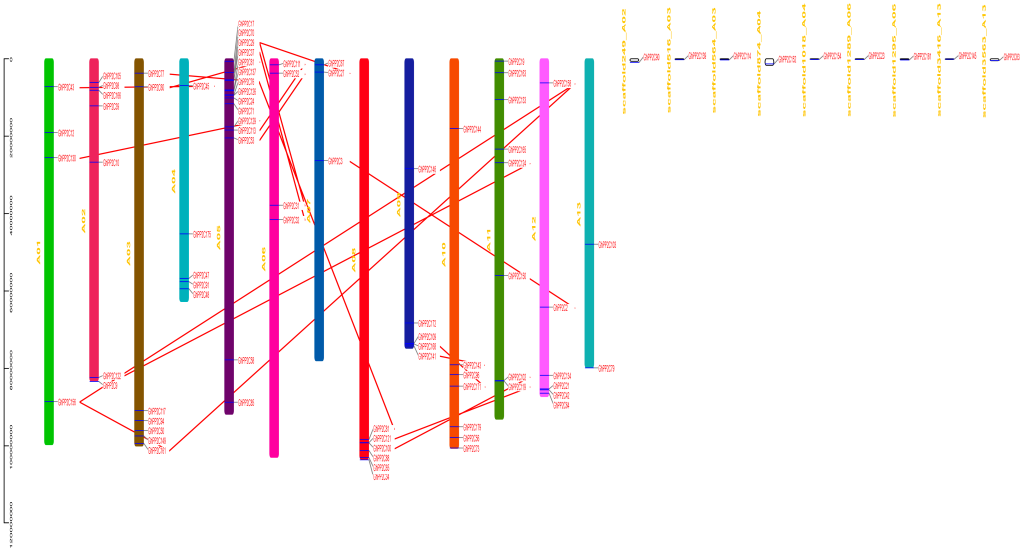


Figure 4. Chromosomal locations of PP2C A genome of cotton were obtained from the GFF file and displayed using TBtools software. The red line indicates the collinear relationship for different chromosomes.

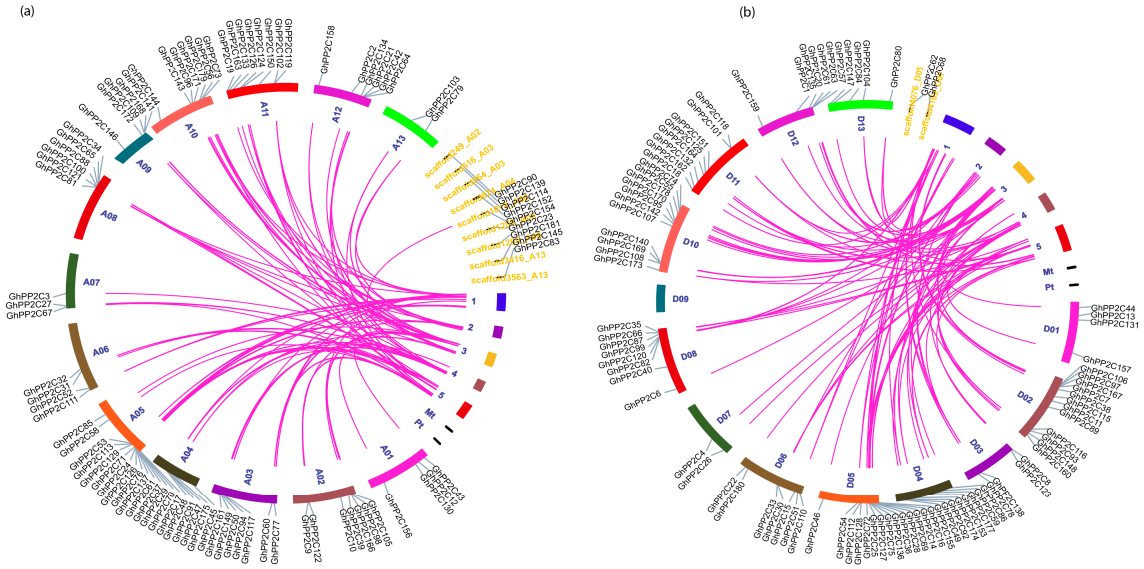


Figure 5. (**a**) and (**b**). Chromosomal locations of PP2C (A and D) genome of cotton were obtained from the GFF file and displayed using TB tools software. The pink line indicated the collinear relationship for different chromosomes.

2.4. Analysis of Putative Regulatory Cis-Element and Gene Duplication Analysis of PP2C Gene Family in Cotton

In the promoter region of *PP2C* genes, cis-acting elements play a critical role as stress-adaptive signaling in plants by interacting with their cognate transcription factor (TF). For instance, abscisic acid (ABA)-responsive elements (ABREs) are involved in salt, drought, and ABA signaling [28]. LTR is crucial to chilling response and regulation [29]. Likewise, TCA-elements and TGACG-motif are responsive to salicylic acid (SA) and MeJA treatments [30]. The exploration of the cis-acting element of *G. hirsutum PP2C* genes was executed by using the PlantCARE database and seven common cis-regulatory elements were briefly summarized (Figure 6 and Supplementary Table S5). The results unveiled that most of the genes contributed in various signaling pathways such as phytohormones and biotic and abiotic regulatory stress factors. On the other hand, *PP2C* genes are recognized to show a major part in both biotic and abiotic stress phenomena. Most of the genes were highly (34.88%) responsive to light (AE-BOX, BOX-4, LAMP-ELEMENTS, GAG-motif, GATA-motif), followed by (24.32%) hormones (ABRE, CGTCA, TGA, TCA, AuxRe, GARE-motif), (18.61%) and other regulatory cis-elements (HD-ZIP3, o2-site, AT-Rich elements, CAT-BOX, A-Box, EIRE), while the fewest genes (0.93% and 0.85%) were mainly counted in circadian and enhancer elements (5UTR Py-rich stretch, TA-Rich Region, and GC-motif), respectively. In plants, the circadian cis-regulatory element is known to control the circadian rhythms. Moreover, these results indicate that *PP2C* genes are vital in various biotic-abiotic/hormone signaling which might be hypothesized by their diversity in nature.

For gene duplication analysis, we used MCScanX to determine the types of gene duplications and the results suggested most of the genes were segmental (167) and few of them were dispersed (12); thus, indicating that segmental duplication plays a major contribution to the expansion of the *PP2C* gene family. Moreover, the *PP2C* genes might have experienced functional discrepancy due to gene duplications, and few of them might have lost their unique functions, developed novel functions, or preserved partition of innovative functions [31,32]. During the evolutionary processes, genes are often exposed to various selective pressures such as positive, neutral, and purifying selection. Additionally, for a better understanding of the selection pressure between the duplicated genes, we calculated the *Ka/Ks* ratios among selected genes from segmental and dispersed (Figure 7 and Table 1). It was shown that only two pairs have a positive selection (>1) *Ka/Ks*, while the rest were purifying in nature with <1.00, reducing divergence after duplication.

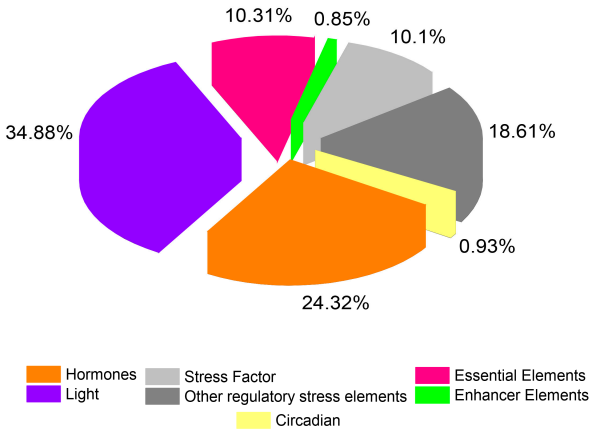


Figure 6. The ratio of various cis-elements in cotton.

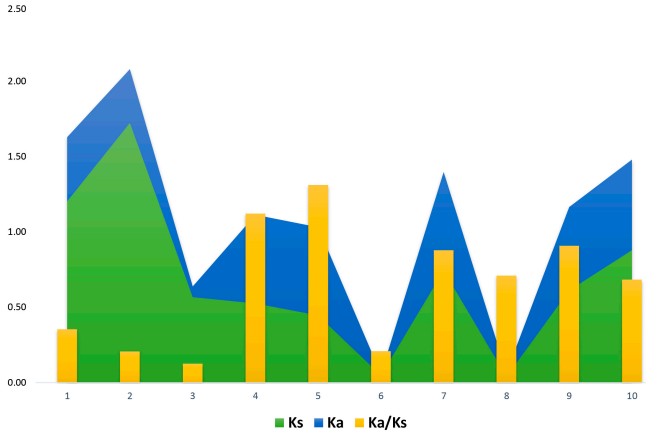


Figure 7. The correlation between *Ks* and *Ka* for duplicated genes (segmental and dispersed).

Table 1. The gene with outlier *Ka/Ks* values and types of duplications i.e., segmental and dispersed genes.

Gene 1	Gene 2	*Ks*	*Ka*	*Ka/Ks*	Selection Pressure	Gene Duplications
GhPP2C9	*GhPP2C3*	1.20	0.43	0.354114713	Purifying Selection	Segmental
GhPP2C10	*GhPP2C6*	1.72	0.36	0.207076566	Purifying Selection	Segmental
GhPP2C4	*GhPP2C1*	0.57	0.07	0.125220459	Purifying Selection	Segmental
GhPP2C7	*GhPP2C8*	0.53	0.59	1.121904762	Positive Selection	Segmental
GhPP2C11	*GhPP2C2*	0.45	0.59	1.311659193	Positive Selection	Segmental
GhPP2C89	*GhPP2C90*	0.05	0.01	0.208333333	Purifying Selection	Dispersed
GhPP2C114	*GhPP2C146*	0.74	0.65	0.879032258	Purifying Selection	Dispersed
GhPP2C83	*GhPP2C84*	0.04	0.03	0.710526316	Purifying Selection	Dispersed
GhPP2C139	*GhPP2C180*	0.61	0.56	0.908346972	Purifying Selection	Dispersed
GhPP2C62	*GhPP2C23*	0.88	0.60	0.683731513	Purifying Selection	Dispersed

2.5. Expression Profiling of PP2C Gene Family in Different Tissues of G. hirsutum

To gain insight into the tissue-specific expression patterns of the cotton, previously reported [24] transcriptome data was utilized for various tissues (root, stem, leaf, patel, and stamen) and fiber development (3, 6, 9, 12, and 15 days post anthesis). For instance, Figure 8a reveals the expression levels of *PP2C* genes drastically varied in various tissues and most numbers of them were highly expressed. However, the expression of a few genes (GhPP2C12, 16, 46, 45, 47, 50, 57, 74, 79, 95, 97, 122, 124, 125, 132, 141, 154, 159, and GhPP2C170) did not show any striking expression in any tissues.

Though, some of the *PP2C* genes were expressed in one or more tissues (GhPP2C15, 17, 18, 27, 40, 44, 46, 64, 68, 72, 78, 117, 127, 129, 138, 151, 168, and GhPP2C171). Intriguingly, the tissue-specific clustering (Figure 8b) showed that root (2), stamen (3), patel (3), and leaf (1) genes were commonly involved in the tissue developmental role in cotton. To gain further insights into the connection between these *GhPP2C* genes in tissue-specific responses, a correlation analysis was established based on the Pearson correlation coefficients (PCCs) ($p = 0.05$). Results showed a higher positive correlation among various specific tissues (Figure 8c and Supplementary Table S6).

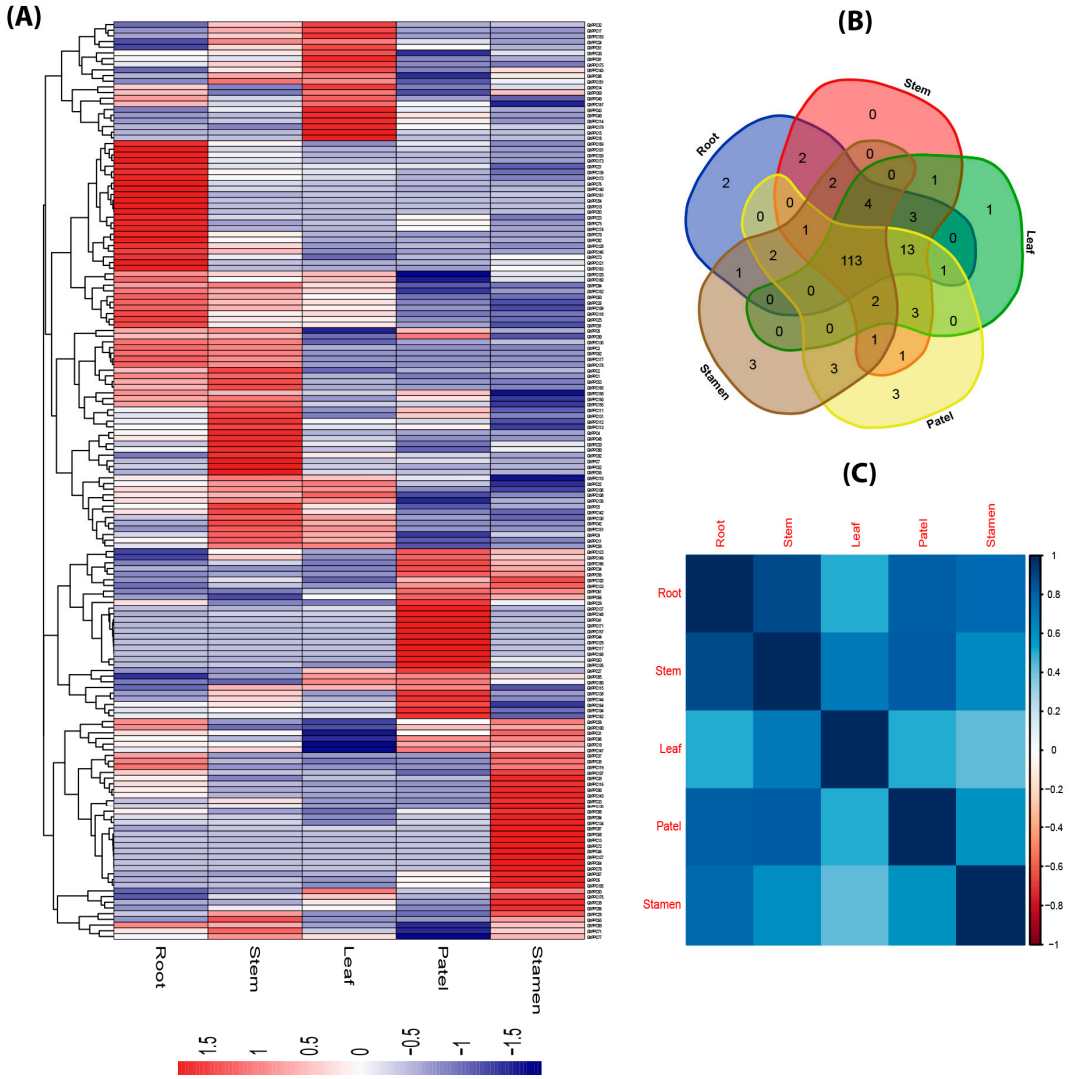


Figure 8. (**A**) Heat map of expression profiles (in $\log_2$-based fragments per kilobase of transcript per million fragments mapped (FPKM)) for PP2C in the five various tissues—root, stem, leaf, patel, and stamen. The expression levels are indicated by the color bar. (**B**) Venn diagram analysis of the tissue expression of PP2C. (**C**) Pearson's correlation coefficients (PCCs) of FPKM-based values of 181 GhPP2Cs in different tissues using Rstudio.

Furthermore, the expression patterns of fiber developmental stages (3, 6, 9, 12, and 15 days post anthesis) exhibited a dynamic expression level. Majority of the *PP2C* genes were highly expressed (Figure 9a), but some of them were not expressed, such as GhPP2C11, 21, 25, 26, 31, 40, 42, 45, 54, 76, 80, 86, 89, 91, 98, 100, 110, 117, 137, 148, 155, 171, and GhPP2C178, respectively. As shown in Figure 9b, the clustering analysis indicated the common developmental genes in 3 days post anthesis (DPA) (4), 6DPA (2), 9DPA (3), and 15DPA (1), respectively. The PCCs-based correlation analysis of the relative

gene expression of selected genes suggested a high positive correlation and low inverse correlation between selected genes. In addition, some genes exhibited an inverse correlation among various fiber developmental stages (Figure 9c and Supplementary Table S6). These findings suggested that *PP2C* genes share diverse and high expression patterns in tissues and fiber development, which implied *PP2C* genes are conserved.

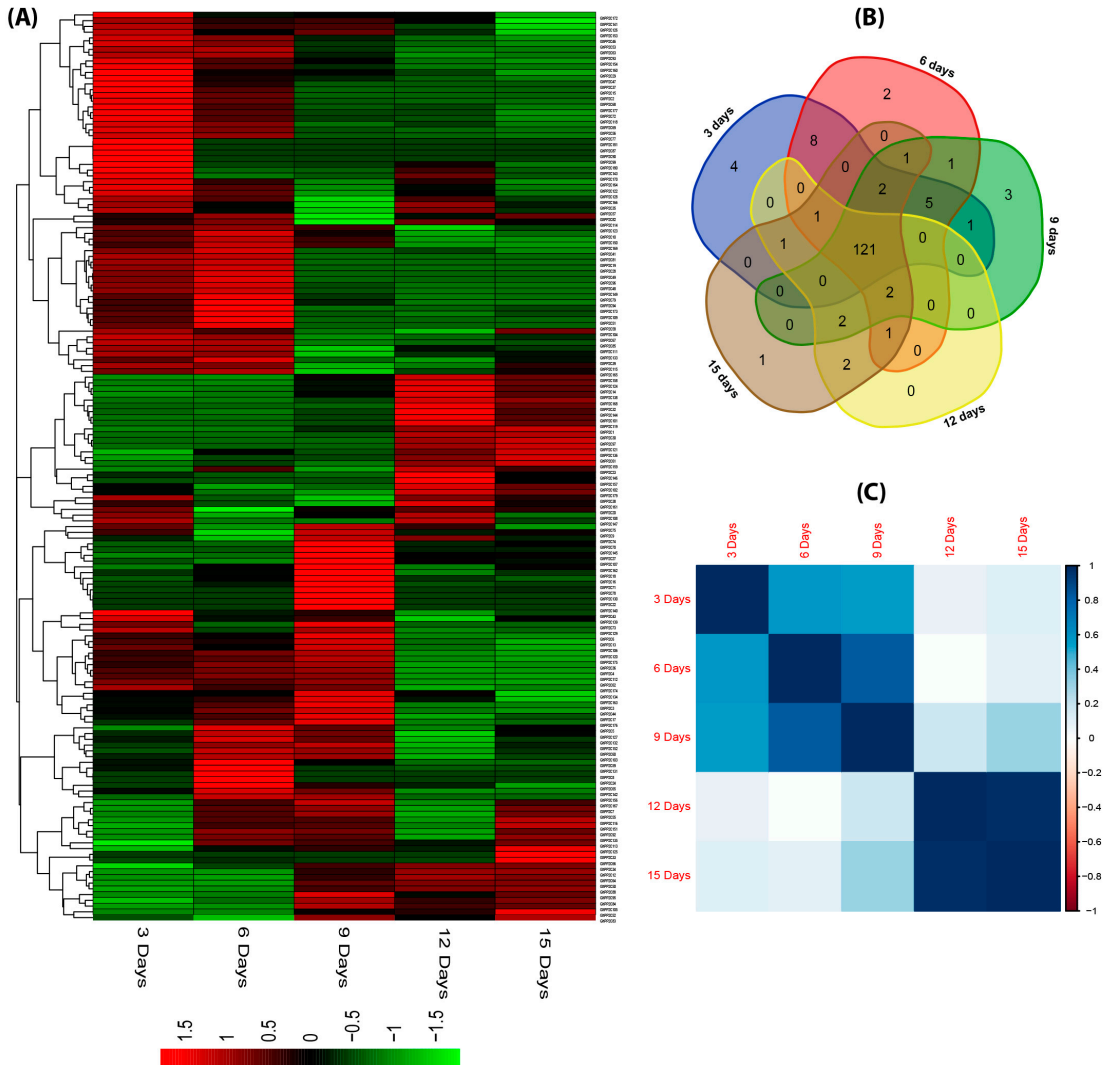


Figure 9. (**A**) Heat map of expression profiles (in $\log_2$-based FPKM) for PP2C in the five fiber developments: 3 days post anthesis (DPA), 6, 9, 12, and 15 DPA. The expression levels are indicated by the color bar. (**B**) Venn diagram analysis of the fiber development of PP2C. (**C**) Pearson's correlation coefficients (PCCs) of FPKM-based values of 181 GhPP2Cs in different fiber development using Rstudio.

2.6. qRT-PCR Analysis of the Candidate PP2C Gene Family in Response to Various Stresses

A prediction of the cis-regulatory elements indicated that *GhPP2C* genes may participate in responses to heat, cold, drought, and NaCl stress tolerance. Moreover, the expression profiling of PP2C has been studied in various species after exposure to abiotic–biotic and hormones stresses [3,26]. To verify this hypothesis, we subjected the cotton seedlings to four various abiotic stress treatments such as heat, cold, drought, and NaCl stress, and then carefully selected 30 genes for qRT-PCR. The results showed that some *GhPP2C* genes exhibited high transcript levels after exposure to multiple treatments, but a few of them were induced by one or more treatments. For instance, heat stress possesses the dominant portion of down-regulated genes (52%). However, cold stress showed 70% of the genes were up-regulated and 30% decreased in transcript level, followed by drought stress,

which exhibited about 53% and 47% of up- and down-regulated *GhPP2C* genes, respectively. On the other hand, exposure to salt stress resulted in 56% up-regulation and 44% down-regulation in genes (Figure 10). Among these 30 genes, we also calculated the Pearson correlation coefficient (PCC) based on the expression by making three categories (i.e., highly positive >0.5, mild positive <0.5 and >0, and negative correlation <0). The results emphasized that both cold and salt stress showed 12 each PCC values having a highly positive correlation, while negative correlation was recorded in heat and drought with 6 each PCC values (Figure 11 and Supplementary Table S6). Taken together, all the 30 genes were induced by different abiotic stresses, but the diversity in the expression profiling of *GhPP2C* genes may suggest that these genes may be critical to abiotic-stress responses.

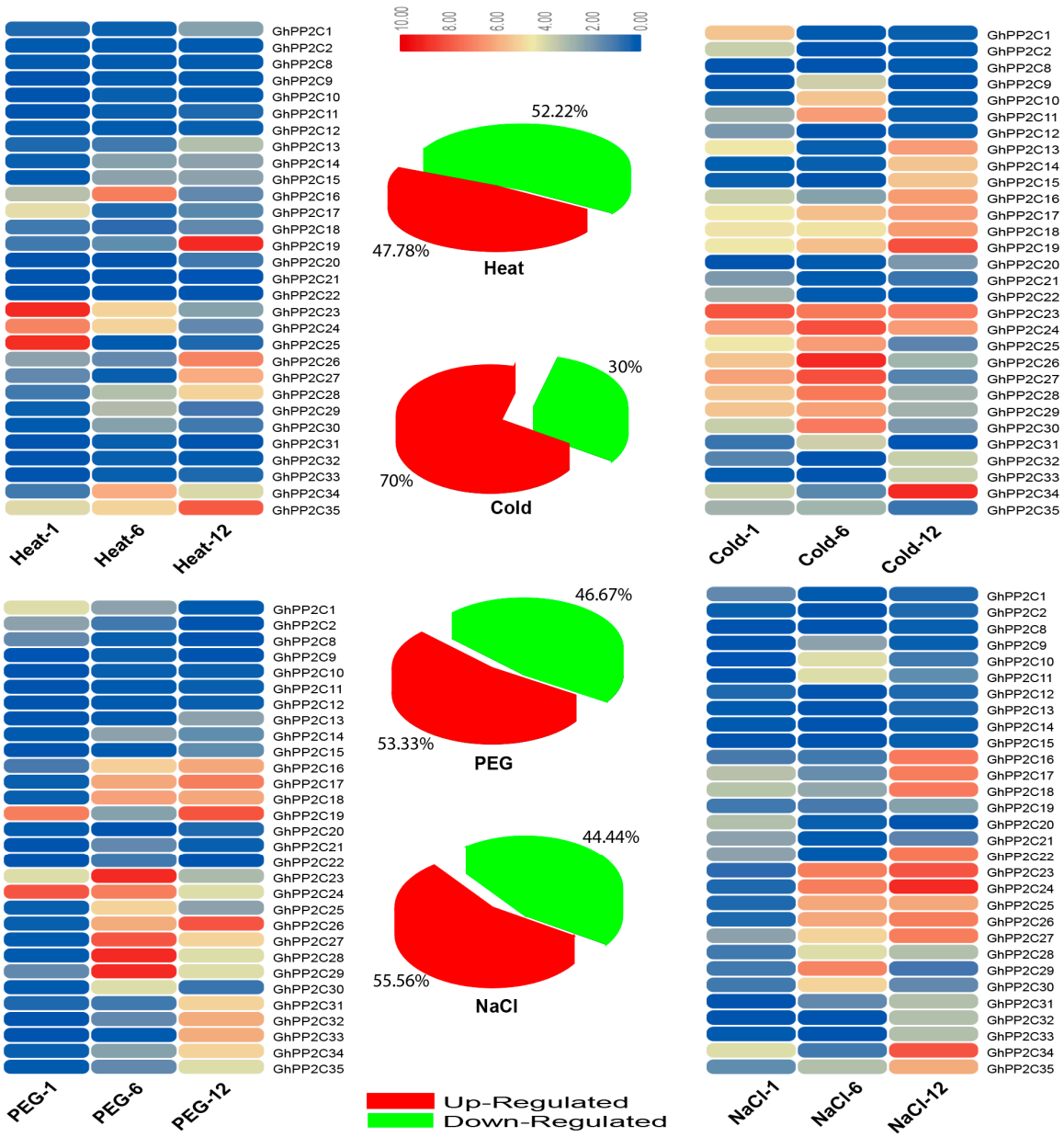


Figure 10. Relative expression analysis by qRT-PCR of *PP2C* genes under heat, cold, drought, and salt stressing in *G. hirsutum*.

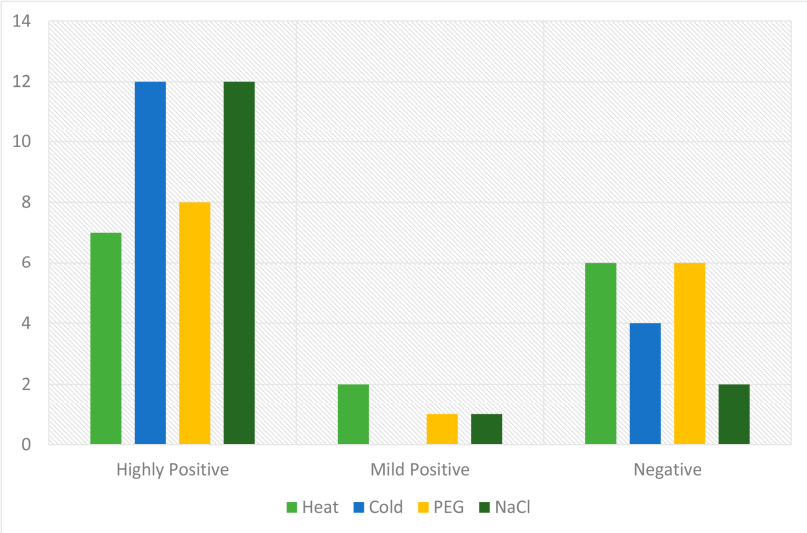


Figure 11. The Pearson's correlation coefficient (PCC) values under response to heat, cold, drought, and salt treatments.

3. Discussion

In this study, we systematically investigate the *PP2C* genes based on the genome-wide analysis. Although the *PP2C* gene family has been in many species, the knowledge of *PP2C* genes in cotton is yet to be elucidated and their systematic analysis has not been reported. Previously, the *PP2C* gene family has been reported in different plant species such as maize [12], rice [11], *Arabidopsis* [10], hot pepper [33], banana [34], and *Brachypodium distachyon* [3]. Thus, we observed high variations in the number of *PP2C* genes that might be present during whole-genome duplication (WGD) events. For exploring new biological functions, evolutionary implications and its expansions are mainly based on gene duplications [35]. Therefore, to study evolutionary analysis and polyploid formation, cotton is an excellent model and ideal crop by being typically allotetraploid [36]. The importance of evolutionary analysis is further reflected in the fact that most of the angiosperms have undergone either one or multiple polyploidization events [37–39]. Herein, we systematically found PP2C-encoding genes in four cotton species—87 (*Gossypium arboreum*), 147 (*Gossypium barbadense*), 181 (*Gossypium hirsutum*), and 99 (*Gossypium raimondii*). Cotton species represent the high number of genes, specifying that the *PP2C* gene family have undergone extensive expansion during the evolution of cotton. In the current study, a comprehensive genome-wide analysis was executed, such as gene identification, gene structure organization, phylogenetic characterization, syntenic relationships, chromosomal localization, and gene duplications. Moreover, transcriptional profiling of 30 *PP2C* genes was also performed, after exposure to heat, cold, drought, and salt stress was also carried out [12,27].

In this study, the subcellular predictions for most of the members of GaPP2C, GbPP2C, GhPP2C, and GrPP2C were mainly located in different organelles such as the chloroplast, nuclear region, mitochondria, cytoplasm, and others. In addition, the gene characteristics of GaPP2C, GbPP2C, GhPP2C, and GrPP2C drastically vary including the protein length (aa), molecular weight (MW), predicted isoelectric point (PI), and a grand average of hydropathicity (GRAVY), signifying that different PP2C proteins may play complex roles in variable microenvironments.

The estimate of evolutionary patterns, such as the rate of computing the selection pressure analysis (*Ka/Ks*), provides useful information about purifying, positive, and neutral selections of gene pairs during the rate of divergence [40]. In evolutionary events, if the value of the *Ka/Ks* ratio is <1.00, this represents purifying selection, a *Ka/Ks* ratio of 1.00 indicates neutral selection, and/or *Ka/Ks* >1.00 depicts positive selection [41,42]. Similarly, during evolutionary processes and expansion of a gene family, these indicators are used for the selection history. In this study, we calculated these values among segmental and dispersed pairs of *PP2C* genes with the help of the MEGA7.0 program. Thus,

we estimated the selection pressure analysis of duplicated genes (i.e., segmental and dispersed) pairs. The majority of the GhPP2C pairs showed *Ka/Ks* ratios of less than 1.00, indicating the purifying selection, and only two pairs showed values more than 1.00, speculating the positive selection [31].

Plant productivity is always uncertain due to a range of climatic challenges such as heat, cold, drought, and salt stress being the major factors in limiting crop production. Previous studies reported that *AtPP2CG1* regulates positively against salt tolerance in *Arabidopsis* and is induced by drought, salt, or exogenous ABA treatment [43]. In *Arabidopsis*, two members of *PP2C* genes responded inversely; for example, *AP2C1* expression was powerfully tempted by drought, wounding, and cold but *AP2C2* was slightly prompted by these treatments too [44]. These findings highlighted the significance of critical members of the *PP2C* gene family in the model plant *Arabidopsis*, but their specific functions may be significantly varied in cotton. For achieving gene expression patterns in diverse growth phases of GhPP2C, we systematically analyzed the previously reported transcriptome data in various tissues (root, stem, leaf, patel, and stamen) and fiber development (3, 6, 9, 12, and 15 DPA). These results showed that most of the PP2C were highly varied, speculating the high diversity in their functions. However, few genes have shown tissue-specific expression, indicating their common importance to plant development. In *GhPP2C* genes, we also explored the promoter regions for identification of common conserved cis-regulatory elements. As a consequence, these results exhibit the participation of *PP2C* genes in various hormone signaling, including both biotic and abiotic stress factors. In the present study, to provide validation to our hypothesis, we also tested 30 candidate genes under various stresses using qRT-PCR. The majority of *PP2C* genes showed high striking transcriptional changes by exposure to heat, cold, drought, and salt stress, suggesting that *PP2C* genes might be crucial to stress tolerance in cotton. Noticeably, recent studies in *Arabidopsis* and rice also uncovered the pivotal role of PP2C candidate genes [11,45], however, in cotton, it is largely obscured. Moreover, 30 candidate genes involved in expression profiling demonstrated their critical role in *Gossypium hirsutum*.

Therefore, taken together, the results of our study provided valuable insight, signifying that PP2C has provided a stepping stone to the molecular mechanism in cotton. Alongside, the differential expression profiling and gene duplications analysis might have experienced functional divergence, and their further study will help considerably in improving the crop yield and quality and cultivating new resistant varieties.

4. Conclusions

We analyzed the genome analysis, evolutionary rates, and molecular characterization of *PP2C* genes in the *Gossypium* genome. Herein, we identified 87 (*Gossypium arboreum*), 147 (*Gossypium barbadense*), 181 (*Gossypium hirsutum*), and 99 (*Gossypium raimondii*) *PP2C* genes by bioinformatics analysis in cotton species. Gene synteny analysis showed that GhPP2C are highly conserved, while the gene duplications analysis reflected that only segmental duplication plays a starring role in the *PP2C* gene extension in cotton. Also, the results of the phylogenetic analysis categorized the *PP2C* genes into 12 subgroups. We further explored the previously published RNA-sequence (RNA-seq) data to compare the tissue-specific response of *PP2C* genes and their critical role in organ development and fiber. Various *PP2C* genes responded promptly to abiotic stresses, including heat, cold, salt, and drought, suggesting their crucial role in abiotic stress tolerance. As a consequence, our findings will facilitate advanced research on the functional analysis of *PP2C* genes regarding their critical role in tissues, fiber development, and abiotic stress tolerance.

5. Materials and Methods

5.1. Data Resources for Sequence Retrieval

For identification of *PP2C* genes in *Gossypium* and other species, we utilized the Plaza 4.0 database (https://bioinformatics.psb.ugent.be/plaza/) with the help of InterPro PP2C domain "IPR001932". The cotton genome sequences were downloaded from (https://www.cottongen.org/), and *A. thaliana*

sequences were retrieved from TAIR (http://www.arabidopsis.org/). The domains of obtained GhPP2C proteins were further verified using the NCBI-Conserved Domain database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) search program and SMART databases (http://smart.embl-heidelberg.de/) [46]. Those proteins which lack PP2C domains were removed from further analysis. In addition, protein sequences that were found with obvious errors in their gene length or having less than 100 lengths were eliminated.

5.2. Multiple Sequence Alignment and Phylogenetic Analysis

The amino acid sequences of the GhPP2C proteins were used for further investigation, and multiple sequence alignment was performed by MUSCLE [47] using MEGA 7 software with the default options [48]. The phylogenetic trees were constructed using the maximum likelihood (ML) method. In order to determine the reliability of the resulting tree, bootstrap values of 1000 replications were performed with the Jones, Taylor, and Thornton amino acid substitution model (JTT model), while keeping the other parameters as a default.

5.3. Calculation of the Ka/Ks for Duplicated Genes

The *Ka/Ks* ratios were calculated for duplicated (segmental and dispersed) gene pairs using MEGA 7.0 [48].

5.4. Conserved Motifs, Exon-Intron Structure Analysis, and Physicochemical Parameters of PP2C Proteins

Conserved motif scanning of GhPP2C proteins was carried out through local MEME Suite Version 5.0.3. For this purpose, parameter settings were calibrated as follows: Maximum number of motifs 10, with a minimum width of 100 and a maximum of 150. The other parameters were set as default [49]. For the exon-intron structure, we used the Gene Structure Display Server (GSDS 2.0) (http://gsds.cbi.pku.edu.cn) [50]. The physicochemical properties of the proteins, including molecular weight (MW), isoelectronic points (pI), aliphatic index, and GRAVY values for each gene, were calculated using the ExPASY PROTPARAM tools (http://web.expasy.org/protparam/). The subcellular localization was predicted using the WOLF PSORT (https://wolfpsort.hgc.jp/) website.

5.5. Cis-Elements Predictions of GhPP2C

Every GhPP2C promoter sequence (selected as 2000 upstream bp) was imported in Generic File Format (GFF) file from the cotton genome. Then, the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [51] was utilized to identify the cis-regulatory elements for promoters of each gene.

5.6. Chromosomal Location and Synteny Correlation Analysis

The chromosomal location of *PP2C* genes was illustrated from top to bottom concerning their position in the genome annotation by using TBtools software [52]. For synteny gene analysis, the relationships were verified between the homologs of A. *thaliana* and *Gossypium hirsutum*. Circos (using TBtools software) program was applied to exhibit the syntenic relationships among the chromosomes of *G. hirsutum* and *A. thaliana* [52].

5.7. Pearson Correlation Analyses (PCC)

Pearson correlation analysis was performed with the help of Excel 2013 in order to evaluate the PCC values that were used for qRT-PCR according to a previously reported study [53]. As well, the PCC of fragments per kilobase of transcript per million fragments mapped (FPKM) values was implemented using RStudio (R program) at 0.05 (p-value) significance level.

5.8. Plant Material and Treatments

In the present study, the germinated seeds of *G. hirsutum* cv. Junmian 1 were grown in plastic pots containing a mixture of soil and vermiculite (3:1). The pots were then placed in an artificial growth chamber for five weeks. The growth conditions were as follows: The temperature was set to 24/16 ºC, the photoperiod was 16/8 h, and the relative humidity was 65%–70%. The two-week-old seedling was exposed to specific treatment, as follows: For heat and cold treatments, seedlings were exposed to 38 ºC and 4 ºC, respectively. For NaCl and drought stress treatments, the seedlings were cultured in a nutrient solution medium with 250 μM and 6000 PEG (*w*/*v*). All treatments were carried out in continuous time intervals of 1, 6, and 12 h, respectively. Additionally, for every specific treatment, five randomly chosen whole seedlings were pooled to form a biological replicate. After that leaf samples were quick-frozen in liquid nitrogen and stored at −80 °C for further use.

5.9. RNA Isolation and Transcriptional Profiling of GhPP2C under Various Stresses

Total RNA was isolated from the treated frozen leaves with Trizol (Invitrogen) following the manufacturer's instructions. RNA was reverse-transcribed into cDNA using the Primer Script RT reagent kit (TAKARA, Dalian China) according to their instructions. Specific primers were designed using Becan Designer 7.9 and are presented in Supplementary Table S7. In order to check the specificity of the primers, we used the BLAST tool against the *Gossypium hirsutum* genome for confirmation. RT-PCR was performed according to the guidelines of previous studies [54]. Relative fold expression was calculated with the comparative Ct-method. The expression patterns of all *GhPP2C* genes were analyzed based on a previous study [55,56]. The cotton histone3 (AF024716) gene was used as the reference gene for qRT- PCR. In brief, the real-time PCR amplification reactions were performed on an ABI 7500 Real-Time PCR System (Applied Biosystems, California, CA, USA) using SYBR Green (Vazyme, Nanjing, China) with three replicates. The amplification parameters were denaturation at 95 °C for 10 min, 40 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 15 s, and extension at 72 °C for 15 s.

High-throughput RNA-sequencing data [24] was utilized for various vegetative (root, stem, and leaf), floral (stamen and patel), and fiber tissues (3, 6, 9, 12, and 15 days post anthesis), respectively. Furthermore, gene expression levels were quantified by FPKM (fragments per kilobase of transcript per million fragments mapped) values, and heat maps were generated using an online omicshares tool (http://www.omicshare.com/) and TBtools software [52].

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/6/1395/s1.

Author Contributions: Conceptualization and methodology, writing—original draft preparation H.S.; review and editing, N.K.; software, J.W.; validation, C.W., Z.W., and Z.H.; supervision, X.W.

Funding: This study was financially supported by National Key Research and Development Program for Crop Breeding (2017YFD0102000) and Natural Science Foundation in Jiangsu Province (BK20160712).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lessard, P.; Kreis, M.; Thomas, M. Protein phosphatases and protein kinases in higher plants. *Comptes Rendus de l'Academie des Sciences. Serie III, Sciences de la Vie* **1997**, *320*, 675–688. [PubMed]
2. Li, F.; Fan, G.; Wang, K.; Sun, F.; Yuan, Y.; Song, G.; Li, Q.; Ma, Z.; Lu, C.; Zou, C.; et al. Genome sequence of the cultivated cotton Gossypium arboreum. *Nat. Genet.* **2014**, *46*, 567–572. [CrossRef] [PubMed]
3. Cao, J.; Min, J.; Peng, L.; Chu, Z. Genome-wide identification and evolutionary analyses of the PP2C gene family with their expression profiling in response to multiple stresses in Brachypodium distachyon. *BMC Genom.* **2016**, *17*, 175. [CrossRef] [PubMed]

4. Kerk, D.; Templeton, G.; Moorhead, G.B.G. Evolutionary Radiation Pattern of Novel Protein Phosphatases Revealed by Analysis of Protein Data from the Completely Sequenced Genomes of Humans, Green Algae, and Higher Plants. *Plant Physiol.* **2008**, *146*, 351–367. [CrossRef]
5. Cohen, P. The structure and regulation of protein phosphatases. *Annu. Rev. Biochem.* **2003**, *58*, 453–508. [CrossRef] [PubMed]
6. Zhou, H.W.; Delong, A. Disparate roles for the regulatory A subunit isoforms in Arabidopsis protein phosphatase 2A. *Plant Cell* **2004**, *16*, 709–722. [CrossRef]
7. Boudsocq, M.; Barbier-Brygoo, H.; Lauriere, C. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. *J. Boil. Chem.* **2004**, *279*, 41758–41766. [CrossRef]
8. Ma, S.Y.; Wu, W.H. AtCPK23 functions in Arabidopsis responses to drought and salt stresses. *Plant Mol. Biol.* **2007**, *65*, 511–518. [CrossRef]
9. Song, S.K.; Clark, S.E. POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in Arabidopsis. *Dev. Boil.* **2005**, *285*, 272–284. [CrossRef]
10. Kerk, D.; Bulgrien, J.; Smith, D.W.; Barsam, B.; Veretnik, S.; Gribskov, M. The complement of protein phosphatase catalytic subunits encoded in the genome of Arabidopsis. *Plant Physiol.* **2002**, *129*, 908–925. [CrossRef]
11. Singh, A.; Giri, J.; Kapoor, S.; Tyagi, A.K.; Pandey, G.K. Protein phosphatase complement in rice: Genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. *BMC Genom.* **2010**, *11*, 435. [CrossRef] [PubMed]
12. Wei, K.; Pan, S. Maize protein phosphatase gene family: Identification and molecular characterization. *BMC Genom.* **2014**, *15*, 773. [CrossRef] [PubMed]
13. Merlot, S.; Gosti, F.; Guerrier, D.; Vavasseur, A.; Giraudat, J. The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* **2001**, *25*, 295–303. [CrossRef] [PubMed]
14. Gonzalez-Garcia, M.P.; Rodriguez, D.; Nicolas, C.; Rodriguez, P.L.; Nicolas, G.; Lorenzo, O. Negative regulation of abscisic acid signaling by the Fagus sylvatica FsPP2C1 plays a role in seed dormancy regulation and promotion of seed germination. *Plant Physiol.* **2003**, *133*, 135–144. [CrossRef]
15. Saez, A.; Apostolova, N.; Gonzalez-Guzman, M.; Gonzalez-Garcia, M.P.; Nicolas, C.; Lorenzo, O.; Rodriguez, P.L. Gain-of-function and loss-of-function phenotypes of the protein phosphatase 2C HAB1 reveal its role as a negative regulator of abscisic acid signalling. *Plant J.* **2004**, *37*, 354–369. [CrossRef] [PubMed]
16. Haider, M.S.; Zhang, C.; Kurjogi, M.M.; Pervaiz, T.; Zheng, T.; Zhang, C.; Lide, C.; Shangguan, L.; Fang, J. Insights into grapevine defense response against drought as revealed by biochemical, physiological and RNA-Seq analysis. *Sci. Rep.* **2017**, *7*, 13134. [CrossRef] [PubMed]
17. Sheen, J. Mutational analysis of protein phosphatase 2C involved in abscisic acid signal transduction in higher plants. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 975–980. [CrossRef] [PubMed]
18. Soderman, E.; Mattsson, J.; Engstrom, P. The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. *Plant J.* **1996**, *10*, 375–381. [CrossRef] [PubMed]
19. Strizhov, N.; Abraham, E.; Okresz, L.; Blickling, S.; Zilberstein, A.; Schell, J.; Koncz, C.; Szabados, L. Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis. *Plant J.* **1997**, *12*, 557–569. [CrossRef]
20. Murata, Y.; Pei, Z.M.; Mori, I.C.; Schroeder, J. Abscisic acid activation of plasma membrane Ca^{2+} channels in guard cells requires cytosolic NAD(P)H and is differentially disrupted upstream and downstream of reactive oxygen species production in abi1-1 and abi2-1 protein phosphatase 2C mutants. *Plant Cell* **2001**, *13*, 2513–2523. [CrossRef]
21. Haider, M.S.; Kurjogi, M.M.; Khalilurrehman, M.; Fiaz, M.; Pervaiz, T.; Jiu, S.; Haifeng, J.; Chen, W.; Fang, J. Grapevine immune signaling network in response to drought stress as revealed by transcriptomic analysis. *Plant Physiol. Biochem.* **2017**, *121*, 187–195. [CrossRef] [PubMed]
22. Tahtiharju, S.; Palva, T. Antisense inhibition of protein phosphatase 2C accelerates cold acclimation in Arabidopsis thaliana. *Plant J.* **2010**, *26*, 461–470. [CrossRef]

23. Yuan, D.; Tang, Z.; Wang, M.; Gao, W.; Tu, L.; Jin, X.; Chen, L.; He, Y.; Zhang, L.; Zhu, L.; et al. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* **2015**, *5*, 17662. [CrossRef] [PubMed]
24. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [CrossRef] [PubMed]
25. Wang, K.; Wang, Z.; Li, F.; Ye, W.; Wang, J.; Song, G.; Yue, Z.; Cong, L.; Shang, H.; Zhu, S.; et al. The draft genome of a diploid cotton Gossypium raimondii. *Nat. Genet.* **2012**, *44*, 1098–1103. [CrossRef] [PubMed]
26. Yang, Q.; Liu, K.; Niu, X.; Wang, Q.; Wan, Y.; Yang, F.; Li, G.; Wang, Y.; Wang, R. Genome-wide Identification of PP2C Genes and Their Expression Profiling in Response to Drought and Cold Stresses in Medicago truncatula. *Sci. Rep.* **2018**, *8*, 12841. [CrossRef]
27. Xue, T.; Wang, D.; Zhang, S.; Ehlting, J.; Ni, F.; Jakab, S.; Zheng, C.; Zhong, Y. Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis. *BMC Genom.* **2008**, *9*, 550. [CrossRef]
28. Wei, L.; Xiao, C.; Zhaolu, M.; Xiahe, H.; Qi, X.; Heng, W.; Hailing, J.; Dabing, Z.; Wanqi, L. Transcriptional regulation of Arabidopsis MIR168a and argonaute1 homeostasis in abscisic acid and abiotic stress responses. *Plant Physiol.* **2012**, *158*, 1279–1292.
29. Maestrini, P.; Cavallini, A.; Rizzo, M.; Giordani, T.; Bernardi, R.; Durante, M.; Natali, L. Isolation and expression analysis of low temperature-induced genes in white poplar (*Populus alba*). *J. Plant Physiol.* **2009**, *166*, 1544–1556. [CrossRef]
30. Wen, F.; Zhu, H.; Li, P.; Jiang, M.; Mao, W.; Ong, C.; Chu, Z. Genome-wide evolutionary characterization and expression analyses of WRKY family genes in Brachypodium distachyon. *DNA Res.* **2014**, *21*, 327–339. [CrossRef]
31. Prince, V.E.; Pickett, F.B. Splitting pairs: The diverging fates of duplicated genes. *Nat. Rev. Genet.* **2002**, *3*, 827–837. [CrossRef] [PubMed]
32. Vandepoele, K.; Simillion, C.; Van de Peer, Y. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* **2003**, *15*, 2192–2202. [CrossRef] [PubMed]
33. Kim, S.; Park, M.; Yeom, S.I.; Kim, Y.M.; Lee, J.M.; Lee, H.A.; Seo, E.; Choi, J.; Cheong, K.; Kim, K.T.; et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat. Genet.* **2014**, *46*, 270–278. [CrossRef] [PubMed]
34. Hu, W.; Yan, Y.; Shi, H.; Liu, J.; Miao, H.; Tie, W.; Ding, Z.; Ding, X.; Wu, C.; Liu, Y.; et al. The core regulatory network of the abscisic acid pathway in banana: Genome-wide identification and expression analyses during development, ripening, and abiotic stress. *BMC Plant Boil.* **2017**, *17*, 145. [CrossRef] [PubMed]
35. Hurles, M. Gene Duplication: The Genomic Trade in Spare Parts. *PLoS Boil.* **2004**, *2*, e206. [CrossRef] [PubMed]
36. Paterson, A.H.; Wendel, J.F.; Gundlach, H.; Guo, H.; Jenkins, J.; Jin, D.; Llewellyn, D.; Showmaker, K.C.; Shu, S.; Udall, J.; et al. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* **2012**, *492*, 423–427. [CrossRef] [PubMed]
37. Edger, P.P.; Pires, J.C. Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **2009**, *17*, 699–717. [CrossRef]
38. Jiao, Y.; Shi, C.; Edil, B.H.; de Wilde, R.F.; Klimstra, D.S.; Maitra, A.; Schulick, R.D.; Tang, L.H.; Wolfgang, C.L.; Choti, M.A.; et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **2011**, *331*, 1199–1203. [CrossRef] [PubMed]
39. Ohno, S.; Wolf, U.; Atkin, N.B. Evolution from fish to mammals by gene duplication. *Hereditas* **1968**, *59*, 169–187. [CrossRef]
40. Juretic, N.; Hoen, D.R.; Huynh, M.L.; Harrison, P.M.; Bureau, T.E. The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.* **2005**, *15*, 1292–1297. [CrossRef]
41. Li, J.; Zhang, Z.; Vang, S.; Yu, J.; Wong, G.K.-S.; Wang, J. Correlation Between Ka/Ks and Ks is Related to Substitution Model and Evolutionary Lineage. *J. Mol. Evol.* **2009**, *68*, 414–423. [CrossRef] [PubMed]
42. Lynch, M.; Conery, J.S. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **2000**, *290*, 1151–1155. [CrossRef] [PubMed]
43. Liu, X.; Zhu, Y.; Zhai, H.; Cai, H.; Ji, W.; Luo, X.; Li, J.; Bai, X. AtPP2CG1, a protein phosphatase 2C, positively regulates salt tolerance of Arabidopsis in abscisic acid-dependent manner. *Biochem. Biophys. Res. Commun.* **2012**, *422*, 710–715. [CrossRef] [PubMed]

44. Chonnanit, C. Characterization and functional analysis of a novel PP2C phosphatase AP2C2 from Arabidopsis. Ph.D. Thesis, University of Vienna, Viennam, Austria, 2008.
45. Williams, R.W.; Wilson, J.M.; Meyerowitz, E.M. A possible role for kinase-associated protein phosphatase in the Arabidopsis CLAVATA1 signaling pathway. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10467–10472. [CrossRef] [PubMed]
46. Letunic, I.; Doerks, T.; Bork, P. SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **2012**, *40*, D302–D305. [CrossRef] [PubMed]
47. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
48. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Boil. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
49. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucl. Acids Res.* **2009**, *37*, W202–W208. [CrossRef]
50. Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
51. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucl. Acids Res.* **2002**, *30*, 325–327. [CrossRef]
52. Chen, C.; Xia, R.; Chen, H.; He, Y. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *bioRxiv* **2018**, 289660. [CrossRef]
53. Khan, N.; Hu, C.-M.; Amjad Khan, W.; Hou, X. Genome-Wide Identification, Classification, and Expression Divergence of Glutathione-Transferase Family in Brassica rapa under Multiple Hormone Treatments. *BioMed Res. Int.* **2018**, *2018*, 19. [CrossRef] [PubMed]
54. Khan, N.; Hu, C.-M.; Amjad Khan, W.; Naseri, E.; Ke, H.; Huijie, D.; Hou, X. Evolution and Expression Divergence of E2 Gene Family under Multiple Abiotic and Phytohormones Stresses in Brassica rapa. *BioMed. Res. Int.* **2018**, *2018*, 5206758. [CrossRef] [PubMed]
55. Khan, N.; Hu, C.-M.; Khan, W.A.; Wang, W.; Ke, H.; Huijie, D.; Zhishuo, Z.; Hou, X. Genome-wide Identification, Classification, and Expression Pattern of Homeobox Gene Family in Brassica rapa under Various Stresses. *Sci. Rep.* **2018**, *8*, 16265. [CrossRef] [PubMed]
56. Xu, J.; Xu, X.; Tian, L.; Wang, G.; Zhang, X.; Wang, X.; Guo, W. Discovery and identification of candidate genes from the chitinase gene family for Verticillium dahliae resistance in cotton. *Sci. Rep.* **2016**, *6*, 29022. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of *LIM* Family Genes in Foxtail Millet (*Setaria italica* L.) and Characterization of the Role of *SiWLIM2b* in Drought Tolerance

Rui Yang [1,2,†], Ming Chen [2,†], Jian-Chang Sun [3,†], Yue Yu [1,2], Dong-Hong Min [1], Jun Chen [2], Zhao-Shi Xu [2], Yong-Bin Zhou [2], You-Zhi Ma [2,*] and Xiao-Hong Zhang [1,*]

1 State Key Laboratory of Crop Stress Biology for Arid Areas, College of Life Sciences, College of Agronomy, Northwest A&F University, Yangling 712100, China; Yangr1213@126.com (R.Y.); yuyue5911@163.com (Y.Y.); mdh2493@126.com (D.-H.M.)

2 Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS)/National Key Facility for Crop Gene Resources and Genetic Improvement, Key Laboratory of Biology and Genetic Improvement of Triticeae Crops, Ministry of Agriculture, Beijing 100081, China; chenming02@caas.cn (M.C.); chenjun@caas.cn (J.C.); xuzhaoshi@caas.cn (Z.-S.X.); zhouyongbin@caas.cn (Y.-B.Z.)

3 Institute of Crop Research, Ningxia Academy of Agriculture and Forestry Sciences, Yongning 750105, China; nxsjch@163.com

* Correspondence: mayouzhi@caas.cn (Y.-Z.M.); zhxh2493@126.com (X.-H.Z.); Tel.: +86-010-8210-8750 (Y.-Z.M.)

† These authors contributed equally to this work.

Received: 23 January 2019; Accepted: 11 March 2019; Published: 15 March 2019

Abstract: LIM proteins have been found to play important roles in many life activities, including the regulation of gene expression, construction of the cytoskeleton, signal transduction and metabolic regulation. Because of their important roles in many aspects of plant development, *LIM* genes have been studied in many plant species. However, the *LIM* gene family has not yet been characterized in foxtail millet. In this study, we analyzed the whole genome of foxtail millet and identified 10 *LIM* genes. All *LIM* gene promoters contain MYB and MYC *cis*-acting elements that are related to drought stress. Based on the presence of multiple abiotic stress-related *cis*-elements in the promoter of *SiWLIM2b*, we chose this gene for further study. We analyzed *SiWLIM2b* expression under abiotic stress and hormone treatments using qRT-PCR. We found that *SiWLIM2b* was induced by various abiotic stresses and hormones. Under drought conditions, transgenic rice of *SiWLIM2b*-overexpression had a higher survival rate, higher relative water content and less cell damage than wild type (WT) rice. These results indicate that overexpression of the foxtail millet *SiWLIM2b* gene enhances drought tolerance in transgenic rice, and the *SiWLIM2b* gene can potentially be used for molecular breeding of crops with increased resistance to abiotic stress.

Keywords: genome-wide analysis; drought tolerance; foxtail millet; *LIM* genes; transgenic rice

1. Introduction

LIM domain proteins are widely found in eukaryotes [1] and have been found to play important roles in many aspects of development, including the regulation of gene expression, cytoskeleton assembly, signal transduction and the regulation of metabolism [2–6]. In animals, LIM proteins have also been found to play a role in cell adhesion and cell movement, participating in tumorigenesis and metastasis [7]. The LIM domain was first discovered in the three homeodomain proteins LIN11 (lineage 11), ISl-1 (insulin I) and MEC-3 (Mechanotransduction 3) [8–11], and was named based on the first letters of these protein names [11]. The LIM domain consists of a double zinc finger sequence,

which contains two conserved sequences (C-X2-C-X16-23-H-X2-C-X2-C-X16-23-Z-X2-C) 50–60 amino acids in length that are separated by two hydrophobic amino acid residues [12].

The first LIM domain protein in plants was identified in sunflower pollen tubes [13], and to date many LIM domain proteins have been identified in plant species including tobacco [14,15], *Arabidopsis thaliana* [16–18], poplar [18], cotton [19–21], and tomato [22]. Plants have two different LIM domain protein subfamilies: CRP and DA1&DAR. Members of the CRP subfamily contain two conserved LIM domains similar to those of the cysteine-rich protein (CRP) family members in animals [18]. The two LIM domains are separated by a long inter-LIM domain. Unlike animal CRP LIM domain proteins, plant LIM domain proteins contain a short C-terminus and lack a glycine-rich region.

CRP-like LIM domain proteins are divided into two groups based on their expression patterns: WLIMs (WLIM1 and WLIM2), which are widely expressed in various tissues, and PLIMs (PLIM1 and PLIM2), which are abundantly expressed in pollen tubes [1]. Arnaud et al. [18] re-grouped the plant CRP-like LIM proteins into four subgroups based on amino acid similarity: αLIM1, βLIM1, γLIM2 and δLIM2. PLIM1 and WLIM1 are clustered in the αLIM1 subfamily, WLIM2 belongs to the γLIM subgroup and PLIM2 belongs to the δLIM subgroup. Arnaud et al. also discovered a new group of CRP LIM-like proteins that formed a single cluster in a phylogenetic tree and named this group the βLIM1 subfamily [18]. Members of the second plant LIM subfamily, the DA1&DAR subfamily, contain a conserved LIM domain, one or more ubiquitin interaction motif (UIM) domains and a conserved C-terminal region [23]. Zhao et al. classified DA1&DAR-like LIM proteins into two categories, Class I and Class II, according to the amino acid sequence. So far, the DA1&DAR LIM proteins have only been found in land plants [23].

Plant LIM proteins are mainly involved in actin cytoskeleton remodeling and the phenylpropane secondary metabolic pathway. In the phenylpropanol pathway L-phenylalanine is used as a substrate for the synthesis of many different aromatic phenolic compounds, such as coumarin, flavonoids, stilbene, hydroxy cinnamate and lignin [24–26]. These substances greatly enhance the tolerance of gymnosperms and angiosperms to mechanical damage such as injury or environmental stress such as drought [24]. The biological functions of LIM domain proteins are closely related to their subcellular localization [27]. LIM domain proteins have three different localization patterns: cytoplasmic localization, nuclear localization, and co-localization in the cytoplasm and nucleus [27]. Nuclear-localized LIM proteins play a major role in tissue-specific gene regulation and cell differentiation [28]. For example, in tobacco, a nuclear-localized LIM1 protein was found to activate the expression of a β-glucuronidase reporter gene under the control of the horseradish peroxidase *prxC2* promoter [29], and suppression of *LIM1* resulted in down-regulated expression of Phe ammonia lyase (*PAL*)-box genes [5]. Based on these findings, Kawaoka and Ebinuma proposed that tobacco LIM1 is a transcription factor involved in lignin synthesis. Cytoplasmic-localized LIM proteins play a major role in regulating the cytoskeleton [28]. For example, in *Arabidopsis*, six LIM domain proteins are localized at the actin cytoskeleton and have been found to bind to actin, with binding activity regulated by pH and Ca^{2+} [17]. Nuclear-cytoplasmic co-localized LIM proteins can shuttle between the cytoplasm and nucleus to regulate actin stability and the regulation of gene expression, respectively [2]. The tobacco NtWLIM2 protein is localized in the nucleus and nucleolus, where it activates the transcription of histone genes, and is also localized at the actin cytoskeleton where it can directly bind to actin fibers, leading to linkage of the actin cytoskeleton and the aggregation of fibers into actin bundles [14]. In upland cotton (*Gossypium hirsutum*), a WLIM1a protein that is localized in the cytosol and nucleus can shuttle into the nucleus when induced by the reactive oxygen species (ROS) hydrogen peroxide (H_2O_2). *WLIM1a* is mainly expressed during the elongation and secondary wall synthesis stages. When cotton fibers rapidly elongate, WLIM1a acts as an actin binding protein, mainly participating in active intracellular transport. When WLIM1a induced by ROS, fiber elongation is blocked, and WLIM1a enters the nucleus where it may function as a transcription factor to activate the expression of genes in the phenylpropanol biosynthesis pathway, thereby regulating secondary wall synthesis [21].

In recent years, studies have shown that *LIM* genes up-regulated expression respond to abiotic stresses, including drought, high salinity and hormones, suggesting that *LIM* genes may play a role in plant resistance to abiotic stress [22,23,30]. The functions of *LIM* family genes in cytoskeleton formation, signal transduction and the phenylpropane secondary metabolic pathway [15,21] also suggest that this family is important for abiotic stress because these processes are also important for stress responses. For example, the cytoskeleton participates in various life activities in cells and also plays a key regulatory role in the resistance of plants to various stress responses [31,32]. Plant antitoxins synthesized via the phenylpropane metabolic pathway are important for resistance to stress as well as for growth and development [33,34]. Because of their potential importance in the response to abiotic stress, *LIM* genes have been the focus of studies in many plants.

Drought stress is one of the main factors limiting global agricultural production [35]. Foxtail millet (*Setaria italica* L.) is an important food and feed crop in arid regionsof the world [36] and has strong resistance to various abiotic stresses, especially drought and low nutrition. Therefore, foxtail millet is used as a model monocotyledon crop for abiotic stress resistance research [35]. However, the mechanism of abiotic stress resistance in foxtail millet is still largely unclear [37]. Although the *LIM* gene family potentially plays an important role in abiotic stress response, this family has not yet been characterized in foxtail millet. Here we took advantage of the foxtail millet genome sequence [38,39] to identify all the members of the *LIM* gene family. Based on the complexity of LIM protein function, we identified all *LIM* genes in foxtail millet. We identified 10 *SiLIM* genes and classified these genes into two main groups. Quantitative real-time PCR (qRT-PCR) analysis showed that a *LIM*-like gene, *SiWLIM2b*, was induced by various stresses, and overexpression of *SiWLIM2b* increased the drought resistance of transgenic rice. These results provide evidence for the functions of a valuable novel *LIM*-like gene, *SiWLIM2b*, which can potentially be used for molecular breeding of resistance to abiotic stresses in graminaceous crops.

2. Results

2.1. Phylogenetic Analysis and Chromosomal Distribution of 10 LIM-Like Genes

By searching the foxtail millet genome using the LIM domain (PF00412) as a keyword, we identifieda total of 10 putative *LIM* genes. The amino acid lengths of the LIMs ranged from 199 to 1441 residues, the predicted protein molecular weights ranged from 22.02 kDa to 154.76 kDa, and the isoelectric points ranged from 4.64 to 8.98 (Supplementary Table S1). To explore the phylogenetic relationships between the *LIM* genes, a phylogenetic tree was constructed using the LIM domain proteins from maize, rice, *Arabidopsis* and foxtail millet (Figure 1A). Phylogenetic analysis showed that foxtail millet LIMs could be divided into the DA1&DAR (four proteins) and the CPR-like (six proteins) subfamilies (Figure 1A). The DA1&DAR-like LIM proteins were further divided into Class I and Class II [23], and the CPR-like subfamily was further divided into the WLIM1, WLIM2, βLIM1 and PLIM2 subgroups (Figure 1A) [22]. We found that the *SiLIM* genes were more closely related to genes from maize and rice than to genes from *Arabidopsis* (Figure 1A).

To visualize the genome distribution of the *LIM* genes in foxtail millet, the MapGene2Chrom tool [40] was used to map the chromosomal locations. The *SiLIM*s were mapped to chromosomes 1, 3, 4, 7 and 9 (Figure 1B), with most genes located on chromosome 9 (*Seita.9G164800*, *Seita.9G201000*, *Seita.9G459200* and *Seita.9G458000*) (Figure 1B). Thus, the *LIM* genes are unevenly distributed in the foxtail millet genome. The precise locations of the 10 *LIM* genes are listed in Supplementary Table S1.

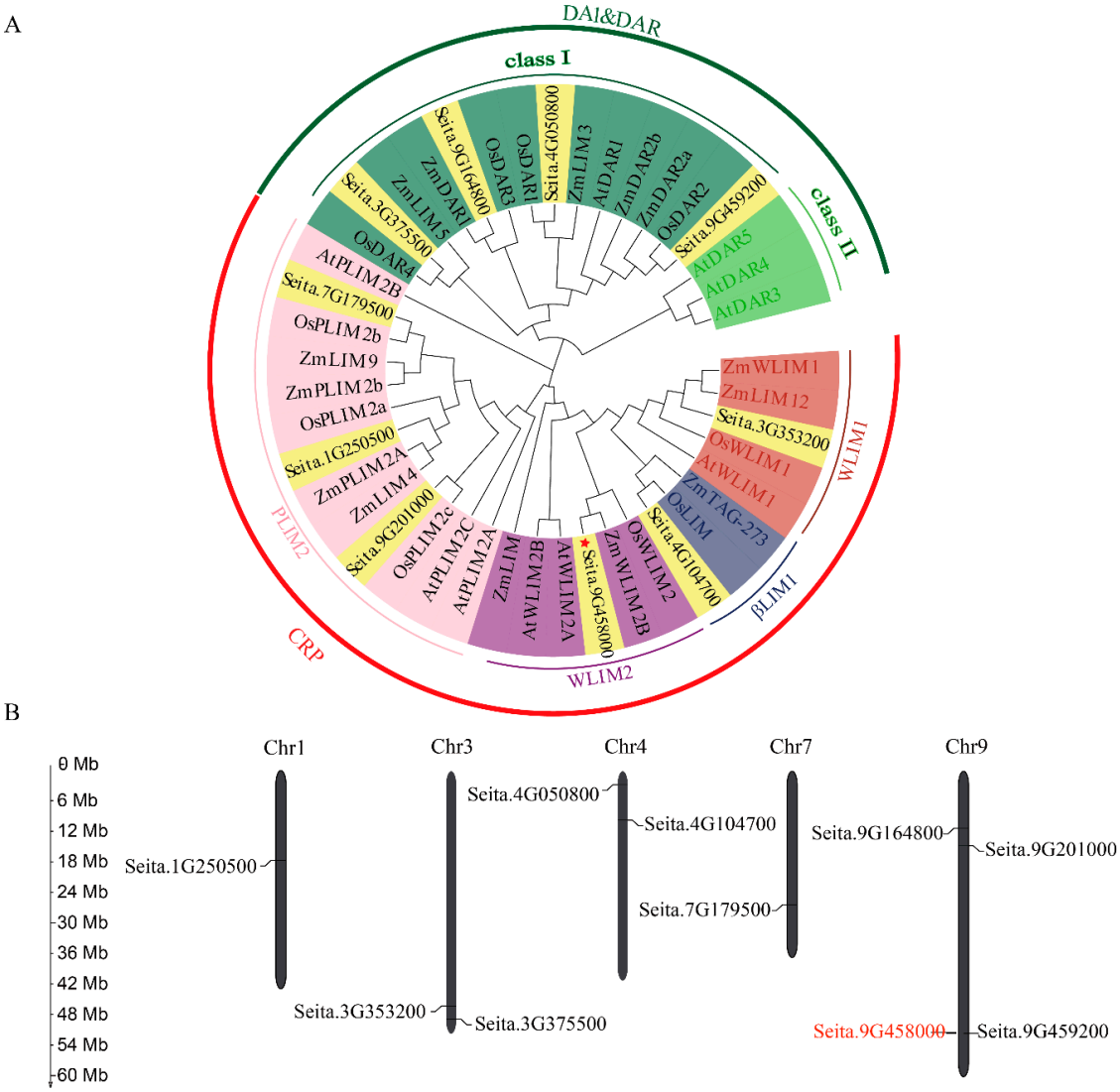


Figure 1. Phylogenetic analysis of foxtail millet and other plant LIM proteins and the chromosomal distribution of the 10 *SiLIM* genes. (**A**) Phylogenetic relationships of LIM family proteins in foxtail millet, *Arabidopsis*, maize and rice. The phylogenetic tree was constructed using the maximum likelihood method in Molecular Evolutionary Genetics Analysis 7.0 software based on the comparison of LIM amino acid sequences. Dark green, light green, pink, purple, dark blue and red represent the DA1&DAR class I, DA1&DAR class II, PLIM2, WLIM2, βLIM1 and WLIM1 subgroups, respectively. Yellow represents the foxtail millet *LIM* genes, and the red five-pointed star indicates the *SiWLIM2b* (*Seita.9G458000*) gene. The species acronym is shown before each LIM protein name: Seita, *Setaria italica*; At, *Arabidopsis thaliana;* Os, *Oryza sativa*; Zm, *Zea mays*. (**B**) Distribution of *SiLIM* genes in the foxtail millet genome. The *SiLIM* genes are located on chromosomes 1, 3, 4, 7 and 9. The *SiWLIM2b* gene is located on chromosome 9 and is marked in red. Chromosomal distances are given in Mb.

2.2. Functional Domains and Gene Structure Analysis of the 10 LIM-Like Genes

To further analyze the functional domains of *SiLIM* genes, we used the SMART website to obtain the conserved domains in the foxtail millet LIM proteins. All LIM proteins in the CRP subfamily contain two LIM domains, and members of the DA1&DAR subfamily contain a conserved LIM domain and two to three conserved UIM domains (Figure 2A).

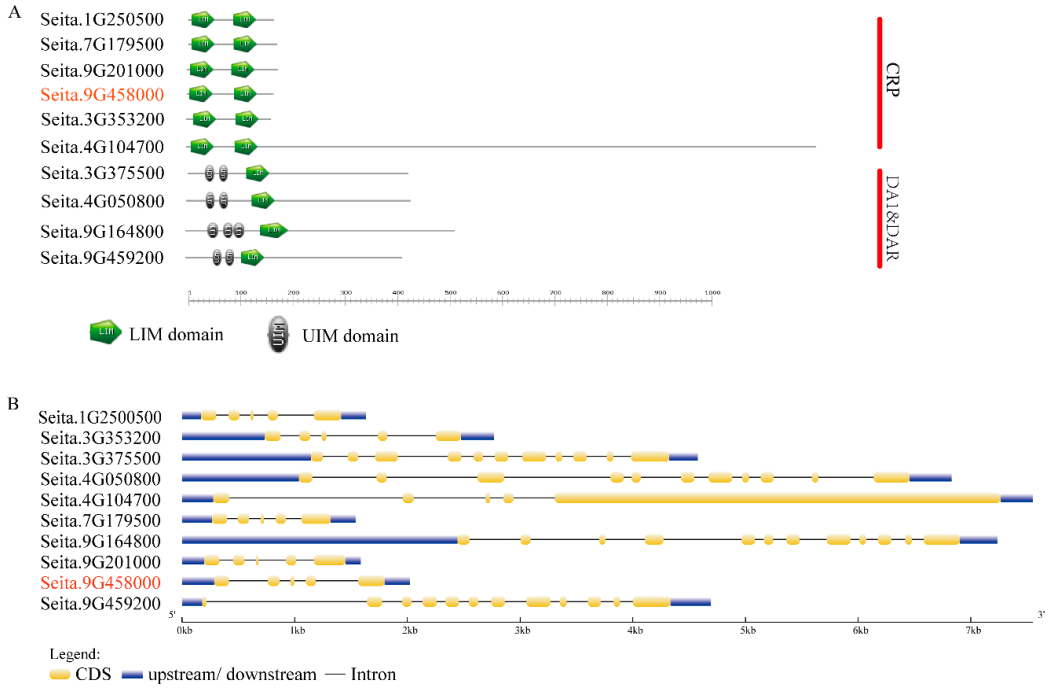


Figure 2. Functional domain and gene structure analysis of foxtail millet *LIM* genes. (**A**) Function domains in foxtail millet LIM proteins. The green pentagon represents the LIM domain, and the gray-black ellipse represents the UIM domain. Members of the PLIM and WLIM subgroups of foxtail millet each contain two LIM domains. Members of the DA1&DAR subgroup contain only one LIM domain and two to three UIM domains. (**B**) Intron–exon structures of *SiLIM* genes. The intron-exon structures were obtained by comparing the coding and genomic sequences using the GSDS website. Yellow and blue boxes represent coding regions and untranslated upstream/downstream regions, respectively. Lines indicate introns. Red represents the *SiWLIM2b* gene.

Exon-intron structural divergence within families plays a pivotal role in the evolution of multi-gene families [41]. We used the Gene Structure Display Server (GSDS) to construct a sketch of the exon-intron structure of each *SiLIM* gene. The gene structures are consistent with the classification of the proteins based on the phylogenetic tree. We found that the genes in each subfamily generally had similar exon-intron structures. The six *LIM* genes belonging to the CPR-like subfamily contain four introns, and two of the *LIM* genes in the DA1&DAR subfamily contain 10 introns. The other two of the *LIM* genes in the DA1&DAR subfamily genes contain 11 introns (Figure 2B). The number of exons in the 10 *LIM* genes ranged from 5 to 12 (Figure 2B), and we found examples of exon loss or acquisition during the evolution of the *LIM* genes in foxtail millet. The gain or loss of exons indicates that the function of these genes may change.

2.3. Promoter Analysis and Tissue-Specific Expression of the LIM Genes in Foxtail Millet

Cis-regulatory elements play an important role in regulating tissue-specific gene expression [42], and also participate in drought response, pathogen defense and cell wall metabolism [43]. Analysis of the foxtail millet *LIM* gene promoters revealed that there are many *cis*-acting elements associated with the response to hormones and abiotic stresses (Figure 3A) and that all of *SiLIM* gene promoters contain MYB and MYC elements, which are related to drought and ABA response (Figure 3A) [44–46]. Most *SiLIM* gene promoters have a light-responsive element (Sp1, TCCC-motif, or G-box) and the abscisic acid response element (ABRE, MBS) (Figure 3A). In addition, some foxtail millet *LIM* gene promoters contain hormone response elements, including gibberell in response elements (P-box, GARE-motif), the salicylic acid response element (TCA-element), methyl jasmonate response elements (CGTCA-motif, TGACG-motif) and the low-temperature responsive element (LTR) (Figure 3A).

There are differences of *cis*-elements in *LIM* genes from different subgroups. Among the *LIM* genes of CRP-like subgroup of foxtail millet, 83.3% of these genes contain auxin response elements, 83.3% of these genes contain GA response elements and 66.6% of these genes contain SA response elements. Among the *LIM* genes of DA1&DAR subgroup of foxtail millet, 50% of these genes contain auxin response elements, 75% of these genes contain GA response elements and 25% of these genes contain SA response elements, which were significantly lower than the CRP-like subgroup. In the CRP subgroup of *LIM* genes, 33.3% of these genes contain LTR elements, 66.6% of these genes contain JA response elements, while in the DA1-DAR subgroup of *LIM* genes, 75% of these genes contain LTR elements and 75% of these genes contain JA response elements, which is significantly higher than that in the CRP subgroup. More details about the *cis*-elements in *SiLIM* family gene promoters are shown in Supplementary Table S2.

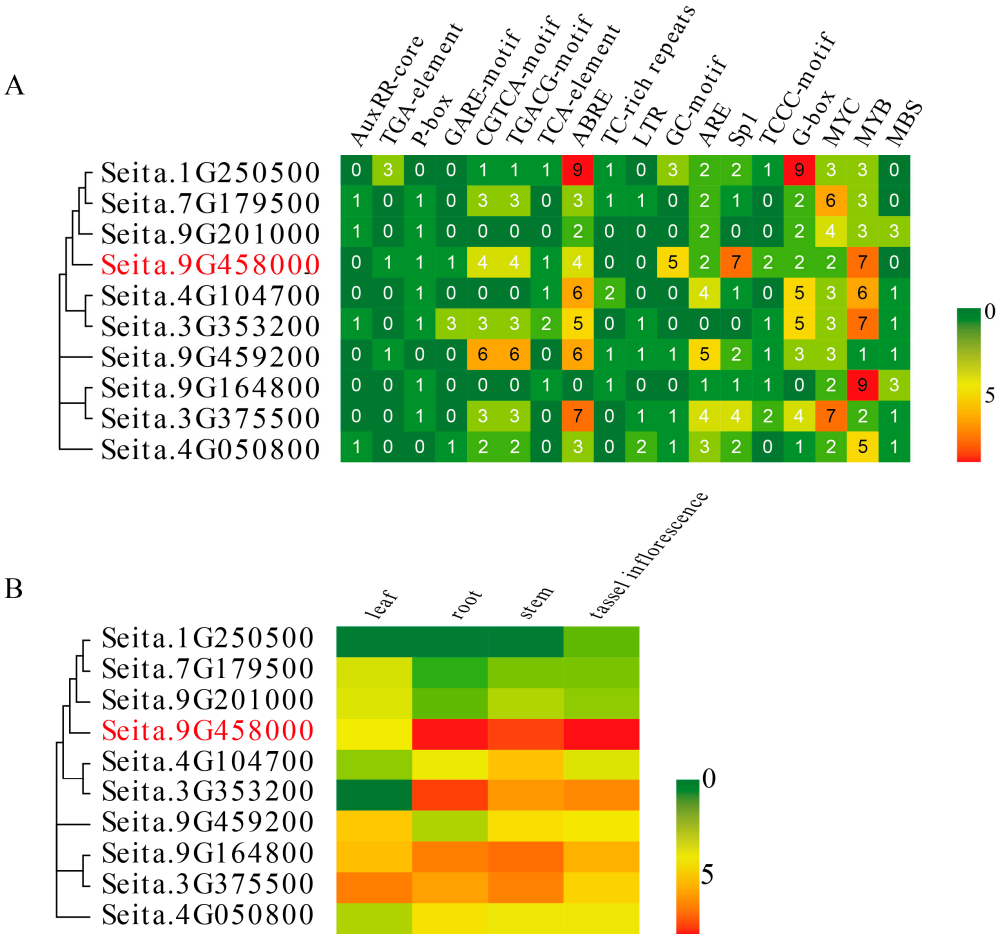


Figure 3. Promoter analysis of foxtail millet *LIM* genes and Tissue-specific expression profiles. (**A**) Promoter analysis of foxtail millet *LIM* genes. The *cis*-acting elements in the promoter regions were obtained from the PlantCARE database. Hormone response elements: AuxRR-core, TGA-element, P-box, GARE-motif, CGTCA-motif, TGACG-motif, TCA-element, ABRE; abiotic stress elements: TC-rich repeats, LTR, GC-motif, ARE, Sp1, TCCC-motif, G-box, MYC, MYB, MBS. Different colors represent the number of *cis*-acting elements within the promoter, with red indicating a higher number and green indicating a lower number. (**B**) Tissue-specific expression data were derived from published Illumina RNA-seq data. Data for roots, stems, leaves and tassel inflorescences were downloaded from the European Nucleotide Archive database [47,48]. The fragments per kilobase of exon model per million mapped reads (FPKM) values were $\log_2$ transformed, and a heat map was drawn using Evolview (http://www.evolgenius.info/evolview, accessed date: 7 September 2018). Different colors represent the level of expression, where red indicates a high level of gene expression, and green represents a low level of gene expression. The *SiWLIM2b* gene is marked in red.

We analyzed the expression profiles of the *LIM* genes in foxtail millet roots, stems, leaves and tassel inflorescences using the RNA-seq data in the European Nucleotide Archive database [47,48] (Figure 3B). This analysis showed that *Seita.1G250500* was specifically expressed in tassel inflorescences and that *Seita.3G353200* was mainly expressed in the root and not expressed in the leaf. The remaining genes were expressed in all four tissues. *Seita.3G375500*, *Seita.7G179500*, *Seita.9G201000* and *Seita.9G459200* were most highly expressed in leaves. The *Seita.9G458000* (*SiWLIM2b*) gene had the highest and lowest expression levels in roots and leaves, respectively. *Seita.4G104700* and *Seita.9G164800* were most highly expressed in stems and *Seita.4G050800* was most highly expressed in roots (Figure 3B). This expression profiling would facilitate combinatorial usage of *SiLIM*s in plant cell processes of different plant tissues, whereas *SiLIMs* expressed in all four tissues may participate in a broad set of plant cell processes. The *seita.9G458000* was more highly expressed in roots and tassels than other nine *LIM* genes, which indicated that *seita.9G458000* plays a more important role in roots and tassel inflorescences.

2.4. Analysis of SiWLIM2b Expression under Various Treatments

By comparing and analyzing the *cis*-elements in the promoter region of *LIM* genes in foxtail millet, we found that the promoter region of *seita.9G458000* contained the most *cis*-elements related to abiotic stress (27), while the promoter region of *seita.9G201000* contained the least *cis*-elements related to abiotic stress (14). Therefore, we selected the WLIM2 subfamily gene *SiWLIM2b* (Figure 1A) for further study. In order to explore the possible role of *SiWLIM2b* in foxtail millet, we used qRT-PCR to analyze the expression of the *SiWLIM2b* gene in response to different abiotic stresses and exogenous hormones. We found that *SiWLIM2b* was up-regulated in response tosalicylic acid (SA), methyl jasmonate (MeJA), gibberellic acid (GA), abscisic acid (ABA), high Ca^{2+} (120 ppm), and high salt (150 mM NaCl) (Figure 4A,B). *SiWLIM2b* was significantly down-regulated after 2 h of simulated drought (15% PEG) or high pH (8.0) (Figure 4A,B). Under nitrogen deficiency (ND), *SiWLIM2b* was slightly up-regulated (Figure 4B).

2.5. Subcellular Location of the SiWLIM2b Protein

In order to identify the subcellular localization of the SiWLIM2b protein, the full-length coding sequence (CDS) of the *SiWLIM2b* gene was cloned into an expression vector containing the GFP tag (16318hGFP) with expression of the SiWLIM2b-GFP fusion protein driven by the CaMV35S promoter. The plasmid SiWLIM2b-GFP was co-transformed into tobacco protoplasts with the nucleus marker AT2G03340-mCherry [49,50]. The SiWLIM2b-GFP fusion protein was mainly distributed in both the cytoplasm and nucleus, and GFP expressed from an empty 35S::GFP expression vector, which served as a control, was mainly distributed in the plasma membrane, cytoplasm and nucleus (Figure 4C). We speculated that the SiWLIM2b protein may act as a transcription factor and may shuttle between the cytoplasm and nucleus.

2.6. Overexpression of SiWLIM2b Enhances Drought Resistancein Transgenic Rice

Because the *SiWLIM2b* gene is regulated by drought stress, we wanted to further analyze the functions of *SiWLIM2b* in drought resistance. Phenotypic analysis was performed using homozygous T3 seeds of transgenic rice lines expressing *SiWLIM2b*. We analyzed the tolerance of *SiWLIM2b* transgenic rice under drought stress in a greenhouse and found that three independent transgenic rice lines grew better than wild type (WT) under drought stress (Figure 5A). Three transgenic lines were analyzed. At least 25 seedlings from each line were measured. qRT-PCR showed that the *SiWLIM2b* gene was more highly expressed in the three transgenic rice lines than in WT ($p < 0.01$) (Figure 5B). After drought treatment, WT plants withered and died earlier than the transgenic rice plants, and the survival rate of the transgenic rice lines was higher than that of WT ($p < 0.01$) (Figure 5C). Relative water content analysis showed there was no significant difference between the transgenic lines and WT under normal conditions, whereas under drought conditions, the three transgenic rice lines had significantly

higher relative water contents than WT ($p < 0.01$) (Figure 5D). We measured the malondialdehyde (MDA) content of transgenic rice and WT, and found that there was no significant difference in the MDA content under normal conditions, whereas under drought treatment, the MDA content of the three transgenic rice lines was significantly lower than that of WT ($p < 0.01$) (Figure 5E). The lower MDA content indicates that the transgenic lines had a lower degree of membrane damage and higher drought resistance than the WT.

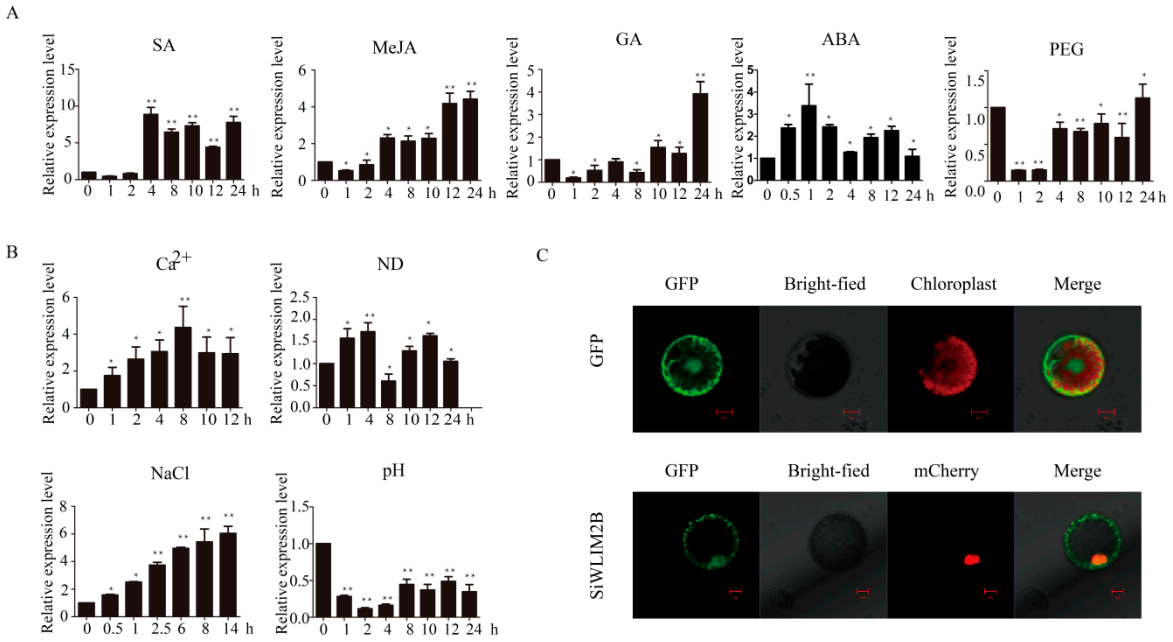


Figure 4. Molecular characteristics of the *SiWLIM2b* gene and Subcellular localization of the SiWLIM2b protein. (**A**) *SiWLIM2b* expression in seedlings treated with ABA, SA, GA, MeJA or PEG. (**B**) *SiWLIM2b* expression in seedlings in the presence of Ca^{2+}, nitrogen-deficient conditions, NaCl or high pH. (**C**) Subcellular localization analysis of the SiWLIM2b protein in tobacco protoplasts. The SiWLIM2b-GFP fusion protein is mainly distributedin the cytoplasm and nucleus. Vertical bars in (**B**,**C**) indicate ±SE of three replicates. ** indicates significant differences in comparison with the control lines at $p < 0.01$. * indicates significant differences in comparison with the control lines at $p < 0.05$. (C) Scale bars = 10 μm.

LIM genes have been shown to regulate the phenylpropanoid secondary metabolic pathway [5,6,21,29]. Therefore, we analyzed the expression of three genes encoding key rate-limiting enzymes in the phenylpropane pathway: *PAL* (phenylalanine ammonia-lyase), *4CL2* (4-coumarate coenzyme A (CoA) ligase 2) and *C3H* (coumarate 3-hydroxylase) [51]. The qRT-PCR data showed that these phenylpropane synthesis-related genes were more highly expressed in the transgenic rice lines than in WT under both normal growth conditions and under drought treatment (Figure 5F). In addition, the expression levels of the three genes were significantly higher under drought than under normal conditions.

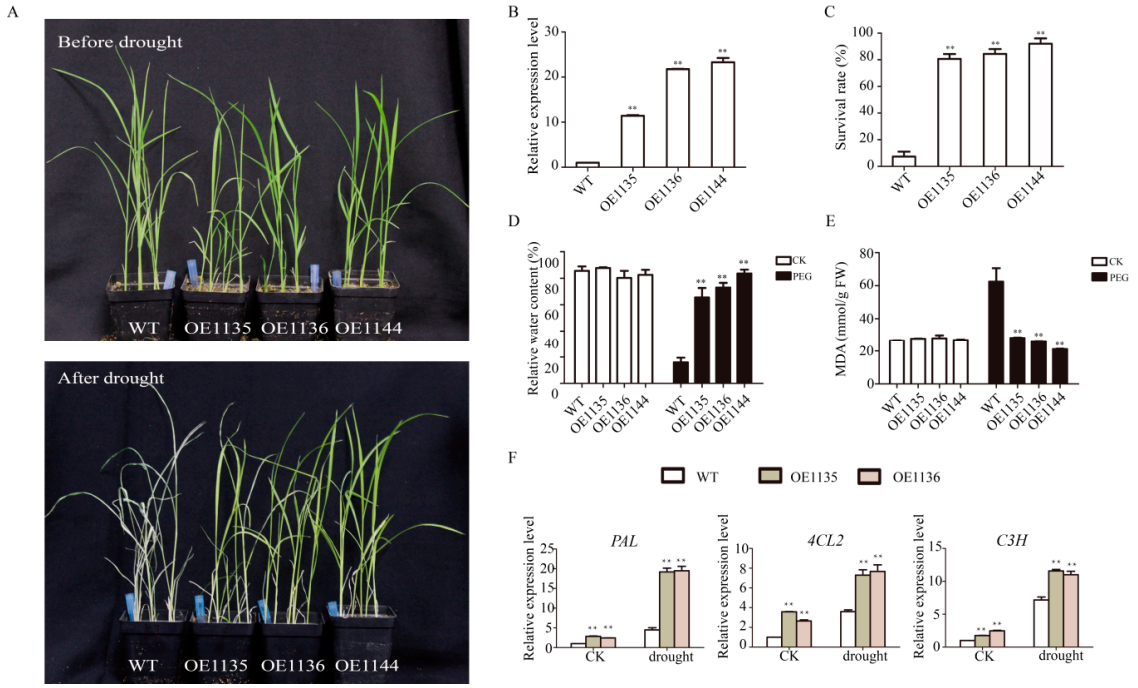


Figure 5. Phenotypic analysis of transgenic and control rice seedlings under drought stress. (**A**) Growth of transgenic and control rice seedlings under normal conditions and after drought treatment for 12 days. (**B**) The expression level of the *SiWLIM2b* gene in transgenic rice lines was determined by qRT-PCR. (**C**) The survival rates of control rice and transgenic rice lines after drought treatment. (**D**) The relative water contents of the control and transgenic rice lines under normal growth conditions and after drought treatment. (**E**) The MDA contents of the control and transgenic rice lines under normal growth conditions and after drought treatment. (**F**) The expression of genes related to the phenylpropane secondary metabolic pathway normal growth (CK) and after drought treatment. Vertical bars in (**B**–**F**) indicate ±SE of three replicates. ** indicates significant differences in comparison with the control lines at $p < 0.01$. * indicates significant differences in comparison with the control lines at $p < 0.05$.

3. Discussion

Plants encounter various biotic and abiotic stresses in the stage of growth and development, as one of the members of plant protein family, LIM proteins are heavily studied. Previous research showed that plant LIM proteins are mainly involved in actin cytoskeleton remodeling and the transcriptional regulation of gene expression. Han et al. found that cotton WLIM1a played a role in developing cotton fibers [21]. Recent studies have showed that *LIM* gene expression in plants was induced by various biotic and abiotic stresses. Although the LIM proteins family has been studied extensively in many crops, knowledge about foxtail millet stress tolerance was limited. Here, a genome-wide analysis of the *LIM* family was performed in foxtail millet and 10 *LIM* genes were identified. Among the 10 *LIM* genes, *SiWLIM2b* was screened to study its function in stress responses and this study will provide some information towards further study of the *LIM* gene in other species.

An accurate evolution history is the first step to understanding the evolution relationship of genes, and may provide some useful information to study the pathway and function they are involved in or regulated by [52]. To understand the relationship of foxtail millet LIM proteins with other LIM proteins in different species, a phylogenetic tree of the LIM domain proteins from rice, maize, *Arabidopsis* and foxtail millet was constructed. The result showed that foxtail millet LIM proteins shared genes more closely related with rice and maize LIM proteins than with *Arabidopsis* LIM proteins, which indicated that the LIM gene family differentiated after the divergence of monocotyledons and dicotyledons. These results are consistent with the present understanding of plant evolutionary history [53]. Zhang, G. Y. et al. revealed five major duplications between chromosomes 2 and 9, 4 and 1, 7 and 1, 6 and

2, and 5 and 3 [38]. In this study, chromosomal distributionanalysis of the *LIM* gene family of foxtail millet showed that the *LIM* genesof foxtail millet was only distributed on chromosomes 1, 3, 4, 7 and 9. This may indicate that the formation of *LIM* genes family of foxtail millet occurred after the chromosome duplication events.

The function of a gene is related to its structure, for example, introns can increase transcription levels by affecting the transcription rate, nuclear output and transcript stability [54]. Both the functional domain analysis and exon-intron structure analysis showed that *LIM* genes were conserved in the same subgroup, which indicated that *LIM* genes in the same subgroup have some functional similarities. Both of the functional domains and exon-intron structures of DA1&DAR subfamily genes is more complex than that of CRP subfamily genes, which may suggest that substantial differentiations might have occurred during the evolution of this gene family. The DA1&DAR protein has some roles that are different to CRP-like protein. For example, the DA1 family also plays a role in regulating the degradation of abnormal proteins [55,56] and in the cell cycle [57]. DA1 was also found to regulate the size of seeds and organs in *Arabidopsis* [58,59].

Cis-regulatory elements in the promoter regions of genes play significant roles in plant stress responses [60]. Hence, we analyzed the *cis*-acting elements in promoters of the *SiLIM* genes and found elements related to GA, auxin, MeJA, and SA response. These elements play an important role in many physiological processes, such as plant development, senescence and maturation, secondary metabolism and response to various environmental stresses [37,61,62]. For example, the P-box and GARE-motifsare associated with the GA signal transduction pathway, and the CGTCA-motif and TGACG-motif are associated with the response to MeJA [37,61,62]. In addition, we found the ABRE in the promoters of *SiLIM* genes, and confirmed that the bZIP transcription factors AREB1, AREB2, and ABF3 interact with the ABRE motifs in the DREB2A promoter to co-regulate plant responses to osmotic stress [63]. All promoters contained more than one element related to drought and ABA response, such as MYC and MYB elements [44]. Based on this evidence, we speculate that the *SiLIM* genes are probably involved in abiotic stress responses.

The *LIM* genes have been found to be induced by various stresses and hormones in other plants. The expression levels of *PbLIMs* were significantly induced by SA, ABA and MeJA in pear [30]. The *LIM* genes of tomato were induced by ABA, cold, drought, NaCl and heat treatment [22]. The *LIM* genes of Brassica rapa were induced by cold, ABA and pH (pH5, pH7 and pH9) treatments [64]. Similar results were found in this study. QPCR analysis showed that the *SiWLIM2b* gene was induced by various stresses (PEG, Ca^{2+}, ND, NaCl, pH) and exogenous hormones (ABA, SA, MeJA, GA). These studies may illustrate that the functions of *LIM* genes in different species appear to be similar. We chose the *SiWLIM2b* gene for further study and found that *SiWLIM2b* was induced by drought. In order to further study the role of *SiWLIM2b* in drought resistance, we studied the phenotype of transgenic rice heterologously by expressing the *SiWLIM2b* gene under drought conditions. We found that the *SiWLIM2b* gene can improve drought resistance when expressed in transgenic rice (Figure 5). MDA can inhibit the activity of cell protective enzymes and reduce the antioxidant content [65,66]. When the enzymes and membrane systems of plant tissues are destroyed, the MDA content is greatly increased. Therefore, MDA reflects the antioxidant capacity of plant tissues, and can also serve as a measure of plant resistance to external stresses [67]. Determination of the MDA content of transgenic lines and WT under drought stress showed that under drought stress transgenic lines had much less membrane damage and higher drought resistance than the kitaake (WT) [65,67]. These results indicate that the *SiWLIM2b* gene is involved in drought response and could be used as a new candidate gene in molecular breeding for crop stress resistance.

A recent study reported that the function of LIM proteins is closely related to its subcellular localization and LIM proteins with nuclear localized or co-localized in the cytoplasm and nucleus act in regulation of gene expression [6,14,21]. In this study, we found that the SiWLIM2b protein mainly accumulated in both the cytoplasm and nucleus, suggesting that it may function in the regulation of gene expression. In our study, the expression of three genes encoding key rate-limiting enzymes in the

phenylpropane pathway was up-regulated in the transgenic rice lines. This suggests that *SiWLIM2b* gene may be involved in the phenylpropane pathway. It was confirmed that the synthesis substances of the phenylpropanol pathway greatly enhance the tolerance to drought in plants [24]. Wei et al. confirmed that the *TaMyb1D* gene in transgenic tobacco regulated the expression of related genes in the phenylpropanol metabolism pathway, and over-expression of *TaMyb1D* enhanced drought resistance of transgenic tobacco [68]. Similar results were found in our study, three phenylpropane synthesis-related genes were up-regulated expression in transgenic rice seedlings after drought treatment. This suggests that over-expression of the *SiWLIM2b* gene may enhance drought resistance of transgenic rice by participating in the secondary metabolic pathway of phenylpropane. However, we found that *SiWLIM2b* did not have transcriptional activation activity (Supplemental Figure S1). We speculate that *SiWLIM2b* may also regulate the expression of genes by interacting with other transcription factors [28]. Further experiments are required to identify the relationship between the *SiWLIM2b* gene and the phenylpropanoid pathway and drought resistance.

4. Materials and Methods

4.1. Plant Materials

Foxtail millet seeds were grown in nutrient soil (1:1 mix of nutrient soil and vermiculite) at 28 °C for three weeks, and then subjected tovarious abiotic stresses, namely drought (15% PEG), high salinity (150 mM NaCl), high pH (8.0) and high Ca^{2+} (120 ppm), or treatment with exogenous hormones, namely ABA (20 μM), SA (20 μM), GA (20 μM) and MeJA (20 μM). Seedlings were sampled at different times after being transferred to a solution containing each substance. In order to subject foxtail millet seedlings to low nitrogen conditions (ND treatment), foxtail millet seeds were germinated at 28°C for 2 days, then transferred into Hoagland nutrient solution and grown to the four-leaf stage. Seedlings were then transferred into nitrogen-free nutrient solution and sampled after 0, 1, 4, 8, 10, 12 and 24 h. All samples were frozen in liquid nitrogen and stored at −80 °C.

4.2. Identification and Phylogenetic Analysis of LIMFamily Genes

The Hidden Markov Model (HMM) of the LIM domain (PF00412) was obtained from Pfam 31.0 (https://en.wikipedia.org/wiki/Pfam, accessed date: 11 August 2018). The putative *LIM* genes were identified in the foxtail millet genomic database (Phytozome v12.1, https://phytozome.jgi.doe.gov/pz/portal.html#, accessed date: 11 August 2018) using 'LIM' and 'PF00412' as key wordswith a threshold E values $\leq$ 1.0. Each putative *SiLIM* gene sequence was checked using SMART (http://smart.embl-heidelberg.de/, accessed date: 11 August 2018) to confirm the presence of the LIM domain. The protein sequences of the *Arabidopsis LIM* genes were acquired from TAIR (https://www.arabidopsis.org/, accessed date: 11 August 2018) and those of the rice and maize *LIM* genes were acquired from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html#, accessed date: 11 August 2018). In order to investigate the evolutionary relationships among the LIM domain proteins in different plant species, aphylogenetic tree was constructed [69]. All *Arabidopsis*, rice, maize and foxtail millet LIM protein sequences were imported using MEGA7.0 [70], and multiple sequence alignments were performed using ClustalW with the default multiple alignment parameters. The phylogenetic tree was constructed using the maximum likelihood method. Information of LIM indifferent species (*Arabidopsis*, rice, maize) is listed in Supplemental Table S3). The percentage of similarity compared with the sequences of LIM proteins in foxtail millet and the LIM proteins of maize, rice and arabidopsis thaliana on the same evolutionary branch is listed in Supplemental Table S4.

4.3. Chromosome Locations and Gene Structures of LIM Family Genes and Identification of Functional Domains and Cis-Acting Elements

Foxtail millet chromosome lengths were obtained from the file Setaria_italica.JGIv2.0.38.gff3, and the position of each *LIM* gene on the foxtail millet chromosomes was obtained from Phytozome

(https://phytozome.jgi.doe.gov/pz/portal.html#, accessed date: 11 August 2018). All *SiLIM* genes were mapped onto the nine foxtail millet chromosomes according their physical position (bp) using MapGene2Chromosome v2 (http://mg2c.iask.in/mg2c_v2.0/, accessed date: 12 August 2018). The exon-intron structures of the *LIM* genes were determined by comparing the CDS with the genomic sequences using the GSDS (http://gsds.cbi.pku.edu.cn/, accessed date: 12 August 2018) [71].

The SMART online tool (http://smart.embl-heidelberg.de/, accessed date: 11 August 2018) was used to analyze the functional domains in the SiLIM proteins, and functional domain sketches were drawn using the ExPAsy-PROSITE website (https://prosite.expasy.org/, accessed date: 13 August 2018). The molecular weights and the oretical isoelectric points of the SiWLIMs were calculated using the ExPASy online tool (https://web.expasy.org/protparam/, accessed date: 13 August 2018).

To obtain the promoter sequences we downloaded approximately 2.0 Kb of sequence upstream of each *SiLIM* coding sequence from the Phytozome database. *Cis*-acting elements were analyzed using the plant *cis*-acting element database PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed date: 15 August 2018) [72]. Heat map showing the number of *cis*-acting elements in the promoters were made using EVOLVIEW (http://www.evolgenius.info/evolview/#mytrees//, accessed date: 7 September 2018) [73].

4.4. Tissue-Specific Expression Profiling Using RNA-Seq Data

Illumina RNA-seq data for four tissues namely roots, stems, leaves and tassel inflorescences were downloaded from the European Nucleotide Archive database [SRX128223 (root); SRX12825 (stem); SRX128224 (leaf); SRX128226 (spica)] [47,48]. Four tissues were obtained from young seedlings of Zhang-gu, which grew in nutrient soil for 40 days in a greenhouse [38]. RNA-seq data were mapped onto the gene sequences of foxtail millet using CLC Genomics Workbench V4.7.1 tooland normalized by FPKM (reads per kilobase per million). A heat map showing tissue-specific expression profiles ($\log_2$RPKM values) was made using EVOLVIEW. The expression profiling data are shown in Supplementary Table S5.

4.5. Total RNA Extraction and qRT-PCR Analysis

Total RNA was extracted from foxtail millet using the Plant Total RNA Kit (Zhuangmeng, Beijing, China) according to the manufacturer's manual. The extracted RNA was detected by electrophoresis on a 1% agarose gel, and the first strand of cDNA was synthesized using the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech, Beijing, China). The relative expression level of the *SiWLIM2b* gene under various stresses was analyzed by qRT-PCR. The Primer3 website (http://bioinfo.ut.ee/primer3-0.4.0/, accessed date: 4 November 2017) was used to design primers for qRT-PCR [74,75], and the $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression level ofthe *SiWLIM2b* gene. The actin gene (Genbank number: AF288226) of foxtail millet was used as an internal control. QRT-PCR was performed for three biological replicates, with three technical replicates per biological replicate. All primer sequences used are shown in Supplementary Table S6.

4.6. Subcellular Localization of SiWLIM2b

We constructed a vector expressing a GFP-tagged fusion protein for subcellular localization analysis of the SiWLIM2b protein. A cDNA containing the coding sequence of *SiWLIM2b* was cloned into the 16318h GFP vector and expressed under the control of the CaMV35s promoter. The recombinant plasmids were confirmed by sequencing. The full-length cDNA coding sequences of AT2G03340 which were located in the nucleus [49] were cloned into the mCherry ORF (WRKY25-RFP) under the control of the CaMV 35S promoter [50]. The plasmid SiWLIM2b-GFP and AT2G03340-mcherry were co-transformed into tobacco protoplasts using the PEG-mediated method and observed under a confocal microscope 16h after transformation [76]. The 35S::GFP vector was transformed as the control.

4.7. Phenotypic Analysis of Transgenic Rice in a Greenhouse

The *SiWLIM2b* complete open reading frame (ORF) was amplified by PCR usingthe primers F1 (BamH I) and R1 (Sac I). The amplified PCR fragment was cloned into vector pMWB014 (driven by the ubiquitin promoter) digested with BamH I and Sac I. The pMWB014-*SiWLIM2b* was confirmed by sequencing and transformed into the kitaake rice (WT) cultivar by Agrobacterium-mediated transformation. The transformed callus was cultured on medium with 1.0 mg/L glufosinate. Six transgenic rice lines were obtained by using primers F2 (designed according to the CDS of *SiWLIM2b* gene) and R2 (designed according to NOS terminator sequence) for PCR verification. The resulting transgenic rice was cultured to the T3 generation. We selected three transgenic lines in the T3 generation (OE1135, OE1136, and OE1144) for subsequent functional analysis. To determine whether transgenic rice seedlings could resist drought stress, we conducted a pot experiment in the greenhouse. Rice seeds were soaked with 2.5% sodium hypochlorite (NaClO) for 30 min, then rinsed with tap water 5–6 times. Sterilized seeds were germinated in a 28-degree incubator, and the water was changed every 12 h. After germination, the seeds wereplanted in nutrient soil and grown to the three-leaf stage under a 16-h/8-h photoperiod at 30 °C. Rice seedlings were then subjected to drought treatment for 12 days and survival was scored by dividing the number of plants alive after drought treatment for 12 days by the total number of plants surveyed, using one green leaf as the standard for survival. Relative water content [77,78] and MDA content [66] were determined in transgenic and control rice after 5 days of drought treatment. All experiments were set up with three independent biological replicates.

4.8. Differential Expression of Phenylpropane Secondary Metabolic Pathway Genes in Transgenic and Control Rice Plants

Expression of phenylpropane secondary metabolic pathway-related genes in control and transgenic rice was analyzed by qRT-PCR. Rice were grown to the four-leaf stage in Hoagland nutrient solution, then treated with 6% PEG for three days. RNA was extracted and reverse transcribed into cDNA, which was used as template for qRT-PCR. Primers used in qRT-PCR are listed in Supplemental Table S4.

5. Conclusions

A total of 10 *LIM* protein-coding genes were identified in foxtail millet via a genome-wide analysis. Analysis of evolutionary relationships indicated that *SiLIM* genes could be divided into two large subgroups. The chromosome positions, gene structures, functional domains and promoters of the *SiLIM* genes were also analyzed. The *SiWLIM2b* gene was found to respond to various abiotic stresses and hormone treatments, and overexpression of the *SiWLIM2b* gene in rice improved drought tolerance.

Supplementary Materials: Supplementary Materials can be found at http://www.mdpi.com/1422-0067/20/6/1303/s1.

Author Contributions: M.C. coordinated the project, conceived and designed experiments, and edited the manuscript; J.-C.S. provided experimental data; J.C. and Y.-B.Z. conducted the bioinformatics work and analyzed the data; Z.-S.X., D.-H.M. and X.-H.Z. contributed valuable discussion; Y.Y. provided experimental assistance to R.Y.; R.Y. drafted the manuscript and prepared the figures and tables; Y.-Z.M. coordinated the project. All authors read and approved the final manuscript.

Funding: This work was supported by the National Key Project for Research on Transgenic Biology [2016ZX08002-002] and the Agricultural Science and Technology Innovation Program (ASTIP, Transgenic Technology and Application of Crops, and Development and Application of Molecular Markers in Crops).

Acknowledgments: We thank Xianmin Diao, Institute of Crop Science, Chinese Academy of Agricultural Sciences, for providing foxtail millet seeds.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

PAL	Phe ammonia lyase
ROS	Reactive oxygen species
MDA	malondialdehyde
4CL2	4-coumarate coenzyme A (CoA) ligase 2
C3H	Coumarate-3-hydroxylase
SA	salicylic acid
MeJA	methyl jasmonate
GA	gibberellic acid
ABRE	ABA-responsive element
qRT-PCR	quantitative real-time PCR
GFP	green fluorescent protein
ND	nitrogen deficiency
ABA	abscisic acid

References

1. Eliasson, A.; Gass, N.; Mundel, C.; Baltz, R.; Krauter, R.; Evrard, J.L.; Steinmetz, A. Molecular and expression analysis of a lim protein gene family from flowering plants. *Mol. Gen. Genet.* **2000**, *264*, 257–267. [CrossRef] [PubMed]
2. Kadrmas, J.L.; Beckerle, M.C. The lim domain: From the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 920–931. [CrossRef] [PubMed]
3. Srivastava, V.; Verma, P.K. The plant lim proteins: Unlocking the hidden attractions. *Planta* **2017**, *246*, 365–375. [CrossRef] [PubMed]
4. Wang, H.J.; Wan, A.R.; Jauh, G.Y. An actin-binding protein, lllim1, mediates calcium and hydrogen regulation of actin dynamics in pollen tubes. *Plant Physiol.* **2008**, *147*, 1619–1636. [CrossRef] [PubMed]
5. Kawaoka, A.; Ebinuma, H. Transcriptional control of lignin biosynthesis by tobacco lim protein. *Phytochemistry* **2001**, *57*, 1149–1157. [CrossRef]
6. Kawaoka, A.; Kaothien, P.; Yoshida, K.; Endo, S.; Yamada, K.; Ebinuma, H. Functional analysis of tobacco lim protein ntlim1 involved in lignin biosynthesis. *Plant J. Cell Mol. Biol.* **2000**, *22*, 289–301. [CrossRef]
7. Labalette, C.; Nouet, Y.; Levillayer, F.; Colnot, S.; Chen, J.; Claude, V.; Huerre, M.; Perret, C.; Buendia, M.A.; Wei, Y. Deficiency of the lim-only protein fhl2 reduces intestinal tumorigenesis in apc mutant mice. *PLoS ONE* **2010**, *5*, e10371. [CrossRef] [PubMed]
8. Way, J.C.; Chalfie, M. Mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in c. Elegans. *Cell* **1988**, *54*, 5–16. [CrossRef]
9. Freyd, G.; Kim, S.K.; Horvitz, H.R. Novel cysteine-rich motif and homeodomain in the product of the caenorhabditis elegans cell lineage gene lin-11. *Nature* **1990**, *344*, 876–879. [CrossRef] [PubMed]
10. Karlsson, O.; Thor, S.; Norberg, T.; Ohlsson, H.; Edlund, T. Insulin gene enhancer binding protein isl-1 is a member of a novel class of proteins containing both a homeo- and a cys-his domain. *Nature* **1990**, *344*, 879–882. [CrossRef]
11. Dawid, I.B.; Toyama, R.; Taira, M. Lim domain proteins. *C. R. l'Acad. Sci. Ser. III Sci.* **1995**, *318*, 295–306.
12. Perez-Alvarado, G.C.; Miles, C.; Michelsen, J.W.; Louis, H.A.; Winge, D.R.; Beckerle, M.C.; Summers, M.F. Structure of the carboxy-terminal lim domain from the cysteine rich protein crp. *Nat. Struct. Biol.* **1994**, *1*, 388–398. [CrossRef]
13. Baltz, R.; Evrard, J.L.; Bourdon, V.; Steinmetz, A. The pollen-specific lim protein plim-1 from sunflower binds nucleic acids in vitro. *Sex. Plant Reprod.* **1996**, *9*, 264–268. [CrossRef]
14. Moes, D.; Gatti, S.; Hoffmann, C.; Dieterle, M.; Moreau, F.; Neumann, K.; Schumacher, M.; Diederich, M.; Grill, E.; Shen, W.H.; et al. A lim domain protein from tobacco involved in actin-bundling and histone gene transcription. *Mol. Plant* **2013**, *6*, 483–502. [CrossRef] [PubMed]
15. Thomas, C.; Hoffmann, C.; Dieterle, M.; Van Troys, M.; Ampe, C.; Steinmetza, A. Tobacco wlim1 is a novel f-actin binding protein involved in actin cytoskeleton remodeling. *Plant Cell* **2006**, *18*, 2194–2206. [CrossRef]
16. Papuga, J.; Thomas, C.; Dieterle, M.; Moreau, F.; Steinmetz, A. Arabidopsis lim domain proteins involved in actin bundling exhibit different modes of regulation. *FEBS J.* **2009**, *276*, 245.

17. Papuga, J.; Hoffmann, C.; Dieterle, M.; Moes, D.; Moreau, F.; Tholl, S.; Steinmetz, A.; Thomas, C. Arabidopsis lim proteins: A family of actin bundlers with distinct expression patterns and modes of regulation. *Plant Cell* **2010**, *22*, 3034–3052. [CrossRef] [PubMed]
18. Arnaud, D.; Dejardin, A.; Leple, J.C.; Lesage-Descauses, M.C.; Pilate, G. Genome-wide analysis of lim gene family in populus trichocarpa, arabidopsis thaliana, and oryza sativa. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* **2007**, *14*, 103–116. [CrossRef] [PubMed]
19. Li, Y.; Jiang, J.; Li, L.; Wang, X.L.; Wang, N.N.; Li, D.D.; Li, X.B. A cotton lim domain-containing protein (ghwlim5) is involved in bundling actin filaments. *Plant Physiol. Biochem.* **2013**, *66*, 34–40. [CrossRef]
20. Li, L.; Li, Y.; Wang, N.N.; Li, Y.; Lu, R.; Li, X.B. Cotton lim domain-containing protein ghplim1 is specifically expressed in anthers and participates in modulating f-actin. *Plant Biol.* **2015**, *17*, 528–534. [CrossRef] [PubMed]
21. Han, L.B.; Li, Y.B.; Wang, H.Y.; Wu, X.M.; Li, C.L.; Luo, M.; Wu, S.J.; Kong, Z.S.; Pei, Y.; Jiao, G.L.; et al. The dual functions of wlim1a in cell elongation and secondary wall formation in developing cotton fibers. *Plant Cell* **2013**, *25*, 4421–4438. [CrossRef] [PubMed]
22. Khatun, K.; Robin, A.H.K.; Park, J.I.; Ahmed, N.U.; Kim, C.K.; Lim, K.B.; Kim, M.B.; Lee, D.J.; Nou, I.S.; Chung, M.Y. Genome-wide identification, characterization and expression profiling of lim family genes in solanum lycopersicum L. *Plant Physiol. Biochem.* **2016**, *108*, 177–190. [CrossRef] [PubMed]
23. Zhao, M.; He, L.; Gu, Y.; Wang, Y.; Chen, Q.; He, C. Genome-wide analyses of a plant-specific lim-domain gene family implicate its evolutionary role in plant diversification. *Genome Biol. Evol.* **2014**, *6*, 1000–1012. [CrossRef]
24. Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant.* **2010**, *3*, 2–20. [CrossRef] [PubMed]
25. Fraser, C.M.; Chapple, C. The phenylpropanoid pathway in arabidopsis. *Arabidopsis Book* **2011**, *9*, e0152. [CrossRef] [PubMed]
26. Zhang, X.; Liu, C.J. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol. Plant* **2015**, *8*, 17–27. [CrossRef]
27. Sala, S.; Ampe, C. An emerging link between lim domain proteins and nuclear receptors. *Cell. Mol. Life Sci.* **2018**, *75*, 1959–1971. [CrossRef] [PubMed]
28. Zheng, Q.; Zhao, Y. The diverse biofunctions of lim domain proteins: Determined by subcellular localization and protein-protein interaction. *Biol. Cell* **2007**, *99*, 489–502. [CrossRef]
29. Kaothien, P.; Kawaoka, A.; Ebinuma, H.; Yoshida, K.; Shinmyo, A. Ntlim1, a pal-box binding factor, controls promoter activity of the horseradish wound-inducible peroxidase gene. *Plant Mol. Biol.* **2002**, *49*, 591–599. [CrossRef]
30. Cheng, X.; Li, G.; Muhammad, A.; Zhang, J.; Jiang, T.; Jin, Q.; Zhao, H.; Cai, Y.; Lin, Y. Molecular identification, phylogenomic characterization and expression patterns analysis of the lim (lin-11, isl1 and mec-3 domains) gene family in pear (pyrus bretschneideri) reveal its potential role in lignin metabolism. *Gene* **2018**. [CrossRef] [PubMed]
31. Wasteneys, G.O.; Yang, Z. The cytoskeleton becomes multidisciplinary. *Plant Physiol.* **2004**, *136*, 3853–3854. [CrossRef] [PubMed]
32. Wang, C.; Zhang, L.J.; Huang, R.D. Cytoskeleton and plant salt stress tolerance. *Plant Signal. Behav.* **2011**, *6*, 29–31. [CrossRef]
33. Dixon, R.A.; Paiva, N.L. Stress-induced phenylpropanoid metabolism. *Plant Cell* **1995**, *7*, 1085–1097. [CrossRef] [PubMed]
34. Yamaguchi, M.; Valliyodan, B.; Zhang, J.; Lenoble, M.E.; Yu, O.; Rogers, E.E.; Nguyen, H.T.; Sharp, R.E. Regulation of growth response to water stress in the soybean primary root. I. Proteomic analysis reveals region-specific regulation of phenylpropanoid metabolism and control of free iron in the elongation zone. *Plant Cell Environ.* **2010**, *33*, 223–243. [CrossRef] [PubMed]
35. Fang, Y.J.; Xiong, L.Z. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell. Mol. Life Sci.* **2015**, *72*, 673–689. [CrossRef] [PubMed]
36. Lata, C.; Gupta, S.; Prasad, M. Foxtail millet: A model crop for genetic and genomic studies in bioenergy grasses. *Crit. Rev. Biotechnol.* **2013**, *33*, 328–343. [CrossRef] [PubMed]
37. Yu, T.F.; Zhao, W.Y.; Fu, J.D.; Liu, Y.W.; Chen, M.; Zhou, Y.B.; Ma, Y.Z.; Xu, Z.S.; Xi, Y.J. Genome-wide analysis of cdpk family in foxtail millet and determination of sicdpk24 functions in drought stress. *Front. Plant Sci.* **2018**, *9*, 651. [CrossRef]

38. Zhang, G.; Liu, X.; Quan, Z.; Cheng, S.; Xu, X.; Pan, S.; Xie, M.; Zeng, P.; Yue, Z.; Wang, W.; et al. Genome sequence of foxtail millet (setaria italica) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* **2012**, *30*, 549–554. [CrossRef] [PubMed]
39. Bennetzen, J.L.; Schmutz, J.; Wang, H.; Percifield, R.; Hawkins, J.; Pontaroli, A.C.; Estep, M.; Feng, L.; Vaughn, J.N.; Grimwood, J.; et al. Reference genome sequence of the model plant setaria. *Nat. Biotechnol.* **2012**, *30*, 555–561. [CrossRef]
40. Jiangtao, C.; Yingzhen, K.; Qian, W.; Yuhe, S.; Daping, G.; Jing, L.; Guanshan, L. Mapgene2chrom, a tool to draw gene physical map based on perl and svg languages. *Yi Chuan (Hereditas)* **2015**, *37*, 91–97.
41. Zhang, Y.; Mao, L.; Wang, H.; Brocker, C.; Yin, X.; Vasiliou, V.; Fei, Z.; Wang, X. Genome-wide identification and analysis of grape aldehyde dehydrogenase (aldh) gene superfamily. *PLoS ONE* **2012**, *7*, e32153. [CrossRef]
42. Wittkopp, P.J.; Kalay, G. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **2011**, *13*, 59–69. [CrossRef]
43. Wong, D.C.J.; Lopez Gutierrez, R.; Gambetta, G.A.; Castellarin, S.D. Genome-wide analysis of cis-regulatory element structure and discovery of motif-driven gene co-expression networks in grapevine. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* **2017**, *24*, 311–326. [CrossRef]
44. Zhu, C.F.; Schraut, D.; Hartung, W.; Schaffner, A.R. Differential responses of maize mip genes to salt stress and aba. *J. Exp. Bot.* **2005**, *56*, 2971–2981. [CrossRef]
45. Faraji, S.; Rasouli, S.H.; Kazemitabar, S.K. Genome-wide exploration of c2h2 zinc finger family in durum wheat (triticum turgidum ssp durum): Insights into the roles in biological processes especially stress response. *Biometals* **2018**, *31*, 1019–1042. [CrossRef]
46. Onishi, M.; Tachi, H.; Kojima, T.; Shiraiwa, M.; Takahara, H. Molecular cloning and characterization of a novel salt-inducible gene encoding an acidic isoform of pr-5 protein in soybean (glycine max [L.] merr.). *Plant Physiol. Biochem.* **2006**, *44*, 574–580. [CrossRef]
47. Cochrane, G.; Alako, B.; Amid, C.; Bower, L.; Cerdeno-Tarraga, A.; Cleland, I.; Gibson, R.; Goodgame, N.; Jang, M.; Kay, S.; et al. Facing growth in the european nucleotide archive. *Nucleic Acids Res.* **2013**, *41*, D30–D35. [CrossRef]
48. Muthamilarasan, M.; Bonthala, V.S.; Mishra, A.K.; Khandelwal, R.; Khan, Y.; Roy, R.; Prasad, M. C2h2 type of zinc finger transcription factors in foxtail millet define response to abiotic stresses. *Funct. Integr. Genom.* **2014**, *14*, 531–543. [CrossRef]
49. Miller, M.J.; Barrett-Wilt, G.A.; Hua, Z.; Vierstra, R.D. Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in arabidopsis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16512–16517. [CrossRef]
50. Du, Y.T.; Zhao, M.J.; Wang, C.T.; Gao, Y.; Wang, Y.X.; Liu, Y.W.; Chen, M.; Chen, J.; Zhou, Y.B.; Xu, Z.S.; et al. Identification and characterization of gmmyb118 responses to drought and salt stress. *BMC Plant Biol.* **2018**, *18*, 320. [CrossRef]
51. Liu, J.Y.; Osbourn, A.; Ma, P.D. Myb transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* **2015**, *8*, 689–708. [CrossRef]
52. Shen, X.X.; Salichos, L.; Rokas, A. A genome-scale investigation of how sequence, function, and tree-based gene properties influence phylogenetic inference. *Genome Biol. Evol.* **2016**, *8*, 2565–2580. [CrossRef]
53. Kellogg, E.A. Evolutionary history of the grasses. *Plant Physiol.* **2001**, *125*, 1198–1205. [CrossRef]
54. Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **2017**, *91*, 145–155. [CrossRef]
55. Yan, N.; Doelling, J.H.; Falbel, T.G.; Durski, A.M.; Vierstra, R.D. The ubiquitin-specific protease family from arabidopsis. Atubp1 and 2 are required for the resistance to the amino acid analog canavanine. *Plant Physiol.* **2000**, *124*, 1828–1843. [CrossRef]
56. Raasi, S.; Wolf, D.H. Ubiquitin receptors and erad: A network of pathways to the proteasome. *Semin. Cell Dev. Biol.* **2007**, *18*, 780–791. [CrossRef]
57. King, R.W.; Deshaies, R.J.; Peters, J.M.; Kirschner, M.W. How proteolysis drives the cell cycle. *Science* **1996**, *274*, 1652–1659. [CrossRef]
58. Li, Y.; Zheng, L.; Corke, F.; Smith, C.; Bevan, M.W. Control of final seed and organ size by the da1 gene family in arabidopsis thaliana. *Genes Dev.* **2008**, *22*, 1331–1336. [CrossRef]

59. Xia, T.; Li, N.; Dumenil, J.; Li, J.; Kamenski, A.; Bevan, M.W.; Gao, F.; Li, Y. The ubiquitin receptor da1 interacts with the e3 ubiquitin ligase da2 to regulate seed and organ size in arabidopsis. *Plant Cell* **2013**, *25*, 3347–3359. [CrossRef]
60. Zhao, W.; Liu, Y.W.; Zhou, J.M.; Zhao, S.P.; Zhang, X.H.; Min, D.H. Genome-wide analysis of the lectin receptor-like kinase family in foxtail millet (setaria italica L.). *Plant Cell Tissue Organ Culture* **2016**, *127*, 335–346. [CrossRef]
61. Pieterse, C.M.J.; van Wees, S.C.M.; van Pelt, J.A.; Knoester, M.; Laan, R.; Gerrits, N.; Weisbeek, P.J.; van Loon, L.C. A novel signaling pathway controlling induced systemic resistance in arabidopsis. *Plant Cell* **1998**, *10*, 1571–1580. [CrossRef]
62. Wang, Y.; Liu, G.J.; Yan, X.F.; Wei, Z.G.; Xu, Z.R. Meja-inducible expression of the heterologous jaz2 promoter from arabidopsis in populus trichocarpa protoplasts. *J. Plant Dis. Prot.* **2011**, *118*, 69–74. [CrossRef]
63. Kim, J.S.; Mizoi, J.; Yoshida, T.; Fujita, Y.; Nakajima, J.; Ohori, T.; Todaka, D.; Nakashima, K.; Hirayama, T.; Shinozaki, K.; et al. An abre promoter sequence is involved in osmotic stress-responsive expression of the dreb2a gene, which encodes a transcription factor regulating drought-inducible genes in arabidopsis. *Plant Cell Physiol.* **2011**, *52*, 2136–2146. [CrossRef]
64. Park, J.I.; Ahmed, N.U.; Jung, H.J.; Arasan, S.K.; Chung, M.Y.; Cho, Y.G.; Watanabe, M.; Nou, I.S. Identification and characterization of lim gene family in brassica rapa. *BMC Genom.* **2014**, *15*, 641. [CrossRef]
65. Gawel, S.; Wardas, M.; Niedworok, E.; Wardas, P. [malondialdehyde (mda) as a lipid peroxidation marker]. *Wiadomosci Lekarskie* **2004**, *57*, 453–455.
66. Tsikas, D. Assessment of lipid peroxidation by measuring malondialdehyde (mda) and relatives in biological samples: Analytical and biological challenges. *Anal. Biochem.* **2017**, *524*, 13–30. [CrossRef]
67. Karatas, I.; Ozturk, L.; Demir, Y.; Unlukara, A.; Kurunc, A.; Duzdemir, O. Alterations in antioxidant enzyme activities and proline content in pea leaves under long-term drought stress. *Toxicol. Ind. Health* **2014**, *30*, 693–700. [CrossRef]
68. Wei, Q.; Zhang, F.; Sun, F.; Luo, Q.; Wang, R.; Hu, R.; Chen, M.; Chang, J.; Yang, G.; He, G. A wheat myb transcriptional repressor tamyb1d regulates phenylpropanoid metabolism and enhances tolerance to drought and oxidative stresses in transgenic tobacco plants. *Plant Sci. Int. J. Exp. Plant Biol.* **2017**, *265*, 112–123. [CrossRef]
69. Mahapatro, G.; Mishra, D.; Shaw, K.; Mishra, S.; Jena, T. Phylogenetic tree construction for DNA sequences using clustering methods. *Procedia Eng.* **2012**, *38*, 1362–1366. [CrossRef]
70. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. Mega: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [CrossRef]
71. Hu, B.; Jin, J.P.; Guo, A.Y.; Zhang, H.; Luo, J.C.; Gao, G. Gsds 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
72. Lescot, M.; Dehais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouze, P.; Rombauts, S. Plantcare, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
73. He, Z.L.; Zhang, H.K.; Gao, S.H.; Lercher, M.J.; Chen, W.H.; Hu, S.N. Evolview v2: An online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **2016**, *44*, W236–W241. [CrossRef]
74. Koressaar, T.; Remm, M. Enhancements and modifications of primer design program primer3. *Bioinformatics* **2007**, *23*, 1289–1291. [CrossRef]
75. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef]
76. Yoo, S.D.; Cho, Y.H.; Sheen, J. Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2007**, *2*, 1565–1572. [CrossRef]
77. Jungklang, J.; Saengnil, K.; Uthaibutra, J. Effects of water-deficit stress and paclobutrazol on growth, relative water content, electrolyte leakage, proline content and some antioxidant changes in curcuma alismatifolia gagnep. Cv. Chiang mai pink. *Saudi J. Biol. Sci.* **2017**, *24*, 1505–1512. [CrossRef]
78. Smart, R.E. Rapid estimates of relative water content. *Plant Physiol.* **1974**, *53*, 258–260. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Identification, Expression Profile, and Alternative Splicing Analysis of the Brassinosteroid-Signaling Kinase (BSK) Family Genes in *Arabidopsis*

Zhiyong Li [1,2,*], Jinyu Shen [3] and Jiansheng Liang [2,*]

1 Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China
2 Department of Biology, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China
3 Co-Innovation Center for Modern Production Technology of Grain Crop, Yangzhou University, Yangzhou 225000, China; shenjinyu20160710@outlook.com
* Correspondence: lizy6@sustc.edu.cn (Z.L.); liangjs@sustc.edu.cn (J.L.); Tel.: +86-0755-8801-5711 (Z.L.); +86-0755-8801-0303 (J.L.)

Received: 17 February 2019; Accepted: 2 March 2019; Published: 6 March 2019

Abstract: Brassinosteroids (BRs) are steroid hormones essential for different biological processes, ranging from growth to environmental adaptation in plants. The plant brassinosteroid-signaling kinase (BSK) proteins belong to a family of receptor-like cytoplasmic kinases, which have been reported to play an important role in BR signal transduction. However, the knowledge of BSK genes in plants is still quite limited. In the present study, a total of 143 BSK proteins were identified by a genome-wide search in 17 plant species. A phylogenetic analysis showed that the BSK gene originated in embryophytes, with no BSK found in green algae, and these BSK genes were divided into six groups by comparison with orthologs/paralogs. A further study using comparative analyses of gene structure, expression patterns and alternative splicing of BSK genes in *Arabidopsis* revealed that all BSK proteins shared similar protein structure with some exception and post-translation modifications including sumolyation and ubiquitination. An expression profile analysis showed that most *Arabidopsis* BSK genes were constitutively expressed in different tissues; of these, several BSK genes were significantly expressed in response to some hormones or abiotic stresses. Furthermore, reverse transcription-polymerase chain reaction (RT-PCR) assays showed that *BSK5*, *BSK7*, and *BSK9* underwent alternative splicing in specific stress induced and tissue-dependent patterns. Collectively, these results lay the foundation for further functional analyses of these genes in plants.

Keywords: brassinosteroid-signaling kinase; gene family; expression profile; alternative splicing; intron retention

1. Introduction

As sessile organisms, plants need to proceed in a coordinated manner to adapt to the constantly changing environment and react to stress conditions for growth and development. Phytohormones play a crucial role during these processes. Among them, brassinosteroids (BRs) are generally known as important plant hormones that play fundamental roles in various cellular, physiological, and developmental processes during plant life cycle [1,2].

To date, the BR signaling pathway has been well established and a number of the intracellular components of this pathway have been identified by genetic, genomic and proteomic studies. In the current model of the BR signaling pathway, the BR signal is perceived by BRASSINOSTEROID INSENSITIVE 1 (BRI1), a membrane-localized receptor kinase [3,4]. The direct binding of the BR

ligand with BRI1 induces its association BRI1 with its co-receptor BRI1-ASSOCIATED RECEPTOR KINASE 1 (BAK1), enhancing the signaling output through reciprocal BRI1 transphosphorylation, inactivation of a glycogen synthase kinase-3 BRASSINOSTEROID INSENSITIVE 2 (BIN2) and activation of phosphatase BRI1 SUPPRESSOR 1 (BSU1), as well as the downstream transcription factors BRASSINAZOLE-RESISTANT 1 (BZR1) and BRI1-EMS-SUPPRESSOR 1 (BES1) [5–8]. Moreover, the transcription factor gene family, in turn, controls the expression of numerous target genes, which control various cellular, physiological, and developmental processes [6,9].

In the initial steps of BR signaling transduction, three receptor-like cytoplasmic kinases named BRASSINOSTEROID-SIGNALING KINASEs (BSKs) from *Arabidopsis* have been identified as BR-responsive proteins, including *BSK1*, *BSK2*, and *BSK3* [10]. *BSK1* and *BSK3* interact with BRI1 in vivo and are phosphorylated by BRI1 in vitro [10]. The phosphorylated BSK proteins further activate downstream phosphatase BSU1 for BR signaling transduction [6,11]. In *Arabidopsis*, there are 12 BSK proteins with putative kinase catalytic domain at the N-terminus and tetratricopeptide repeats (TPRs) at the C-terminus. However, the available results indicate that not all members are involved in BR signaling only. *BSK4*, *BSK6*, *BSK7*, and *BSK8* were reported to play a partial overlapping role in plant growth as well as in BR signaling with *BSK3* [12]. In contrast, *BSK3* was found as the only BSK member involved in BR-mediated plant root growth in a recent study [13]. Unexpectedly, the YODA mitogen-activated protein kinase pathway is activated by SHORT SUSPENSOR (SSP/BSK12) during embryogenesis, which has not been shown to be regulated by BRs [14]. In addition, the loss-of-function mutant *bsk5* is sensitive to salt stress and abscisic acid (ABA) hormone [15]. Silencing OsBSK1-2 inhibits flagellin- and chitin-triggered immune responses in rice [16]. Moreover, the *Arabidopsis BSK1* directly interacts with the immune receptor FLAGELLIN SENSING2 (FLS2) and further phosphorylates MAPKKK5 for the activation of pattern-triggered immunity (PTI) [17,18]. However, the detailed characterization of BSK family proteins and their functional importance in plants remains unclear.

In the present study, we screened the available genomes and identified a total of 143 BSK proteins from 17 plant species. We further performed a detailed analysis of their classification, phylogeny, and alternative splicing. Finally, we verified the expression profiles of the selected BSK genes in *Arabidopsis* by investigating their transcriptional levels upon exposure to abiotic stresses and hormones. Moreover, a novel post-transcription regulation pattern was found in several BSK genes, and potential significant functions of BSK genes were proposed. Our results provide important information about the evolution of the BSK gene family in plants and provide a basis for further studies of the functions of BSK family proteins.

2. Results

2.1. Identification and Characterization of the Brassinosteroid-Signaling Kinase (BSK) Genes in Plants

In this study, a genome-wide analysis of the BSK gene family was performed on the basis of the completed genome sequences. Using the *Arabidopsis* Information Resource (TAIR), PlantGDB, Phytozome, and National Center for Biotechnology Information (NCBI) databases, we first retrieved the available BSK sequences from the currently sequenced genomes. A total of 17 plant genomes were analyzed to identify potential orthologous genes of BSK. These plants, representing the major clades of plants, included eight dicots (*Aquilegia coerulea*, *Arabidopsis thaliana*, *Brassica rapa*, *Glycine max*, *Gossypium raimondii*, *Medicago truncatula*, *Populus trichocarpa*, and *Solanum tuberosum*), five monocots (*Brachypodium distachyon*, *Oryza sativa*, *Setaria italica*, *Sorghum bicolor*, and *Zea mays*), and another four plant species (*Zostera marina*, *Ananas comosus*, *Physcomitrella patens*, and *Marchantia polymorpha*). The genome-wide analysis led to the identification of 143 BSK proteins in these plants. By investigating the extent of lineage-specific expansion of the BSK genes in plants, we found that the BSK gene originated in embryophytes (Figure 1A). Furthermore, almost all selected plants were found to have at least five BSK genes, with *B. rapa* having the highest number (21) of BSK genes (Figure 1B) and

M. polymorpha having only one BSK gene. This result indicated that the BSK genes were subjected to a large-scale expansion in higher plants.

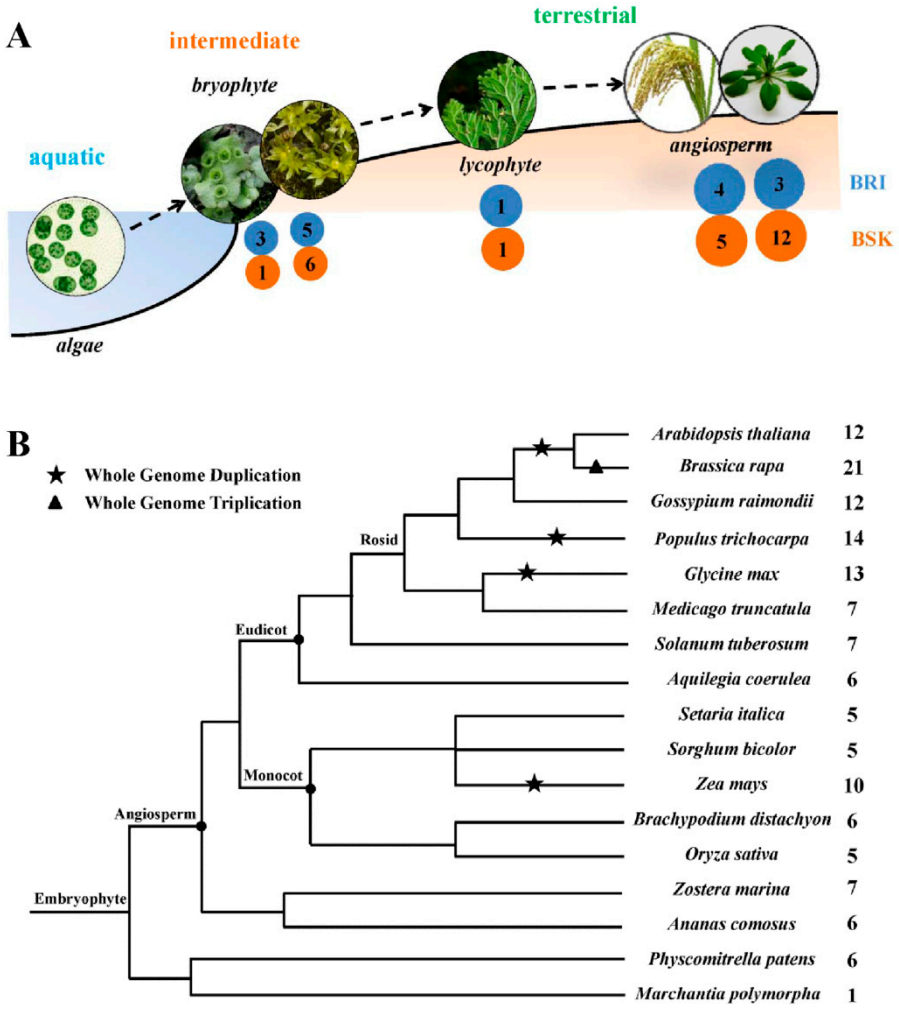


Figure 1. A comparative analysis of BSK genes in plants. (**A**) Evolution of core components of brassinosteroid (BR) signaling from aquatic plants to land plants indicated by dotted arrows. As representatives, component numbers of bryophyte, lycophyte and angiosperm were obtained from *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Arabidopsis thaliana*, and *Oryza sativa*, respectively. The numbers of BRI1-like and BSK proteins were indicated within blue and yellow circles, respectively. (**B**) The evolutionary relationships and the numerical details of the BSK family of each species.

2.2. Phylogenetic Analysis of BSK Family Genes

To explore phylogenetic relationships among the 143 BSK proteins identified in different plant species, a neighbor-joining phylogenetic tree was constructed. This analysis divided all BSK proteins into six major groups according to the bootstrap values and phylogenetic topology (Figure 2). Only one protein from maize, named Zm00008a014861_T01, was in group I, which contained protein with longer amino acid sequence (743 aa) and an additional MSP (Major Sperm Protein) domain along with a conserved kinase domain and tetratricopeptide repeats (TPRs) (Figure S1). Group II was composed of two putative BSK proteins from the monocot seagrass *Z. marina*. The other four groups consisted of the BSK proteins largely from both dicot and monocot species. In particular, the BSK proteins from bryophytes including *P. patens* and *M. polymorpha*, were integrated into the group III. *A. comosus* and *Z. marina* belong to the angiosperm species prior to the split of eudicots and monocots.

The phylogenetic analysis showed that the BSK proteins from *A. comosus* (Aco018845.1, Aco011823.1, Aco014133, Aco010223.1, and Aco000489.1) divided the BSK proteins from dicots and monocots in each group. Moreover, other five BSK proteins from *Z. marina* (Zosma313g00120, Zosma1g02160, Zosma37g01020, Zosma41g01020, and Zosma7g01140) further divided the BSK members from dicots into smaller groups. These results could be considered as evidence for lineage-specific expansion of the BSK genes after the divergence of dicots and monocots.

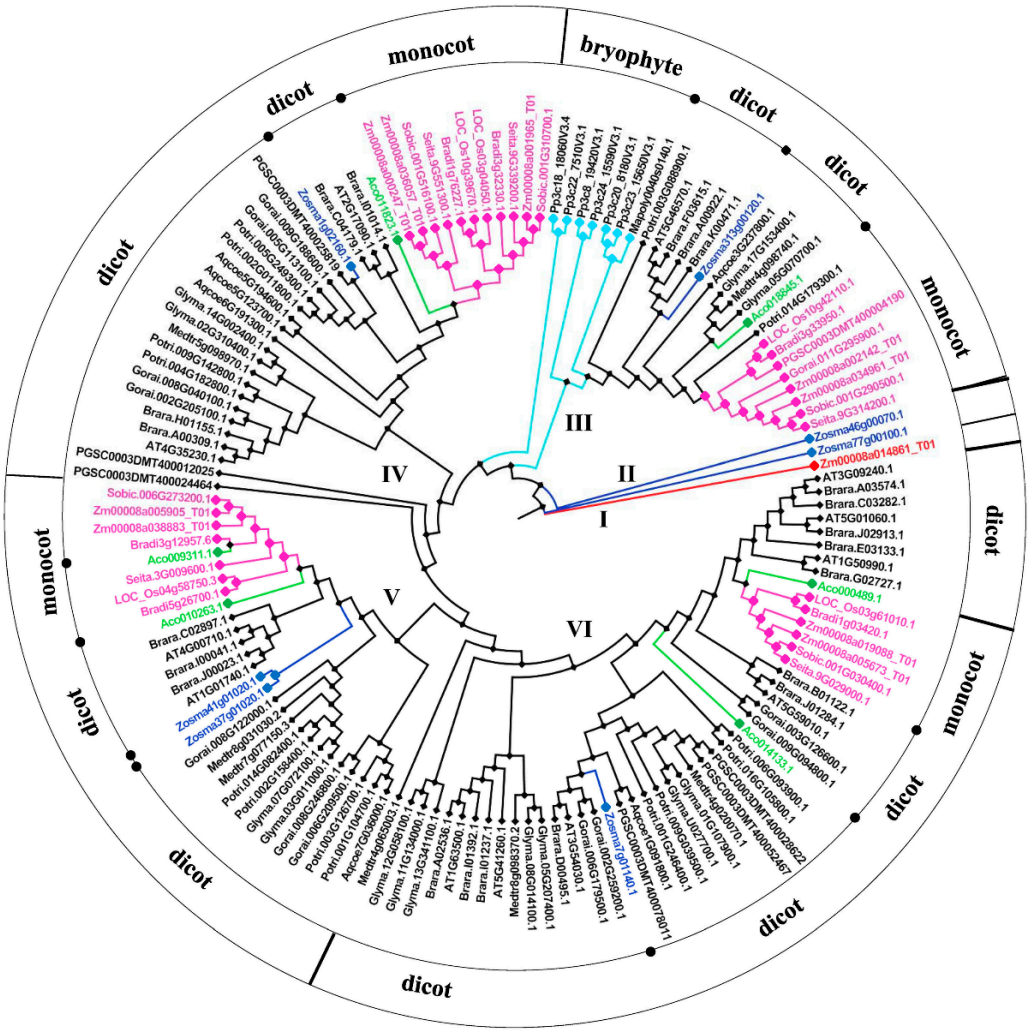


Figure 2. Neighbor-joining phylogenetic analysis of BSK genes. The gene tree was constructed using 143 BSK proteins and visualized using Figure Tree v1.4.2. Representative BSK proteins from different groups were marked with various colors. BSK proteins from dicot plants were in dark, and those from monocot were in pink. Blue and green ones were from *Zostera marina* and *Ananas comosus*, respectively.

2.3. Genomic Structure and Conserved Motif Analysis of the BSK Gene Family

The structure of the BSK genes was analyzed by comparing their full-length coding sequences (CDS) and the corresponding genomic DNA sequences using the software program GSDS 2.0 (http://gsds.cbi.pku.edu.cn/index.php). The exon/intron structure analysis showed that the structures of all BSK genes exhibited diversity both in their intron numbers and in their lengths (Figure 3). In particular, the intron numbers of the *Arabidopsis* BSK genes ranged between 6 and 9 (Figure 4A). Furthermore, we analyzed the different domain architectures, motif compositions, and gene structures of the BSK from *Arabidopsis*. The 12 BSK genes from *Arabidopsis* were located in all five chromosomes (Figure 4B). All BSK proteins had a putative kinase catalytic domain at the N-terminus and TPR

repeats at the C-terminus, whereas *BSK2* and *BSK12* contained only two TPR repeats, one less than that from the other BSK members (Figure 4C). In addition, a glycine-rich region and unknown motif were found at the N-terminus in *BSK1* and *BSK11*, respectively. Previous studies showed that *BSK1* was phosphorylated by BRI1 at S-230, and a mutation in the ATP-binding site K-104 led to enzyme inactivity [10,17]. These conserved sites were also consistent with other BSK family members, except for *BSK12*. Moreover, most of the BSK proteins, except for *BSK9*, had a myristoylation site at the N-terminus (Figure S2). Other kinds of protein post-translational modifications, including sumoylation and ubiquitination, were also analyzed. It was found that all BSK family proteins possessed at least one sumoylation site and multiple ubiquitination sites, some of which were conserved (Figure S2). For instance, most of the ubiquitination sites were located in the kinase catalytic or TPR domain.

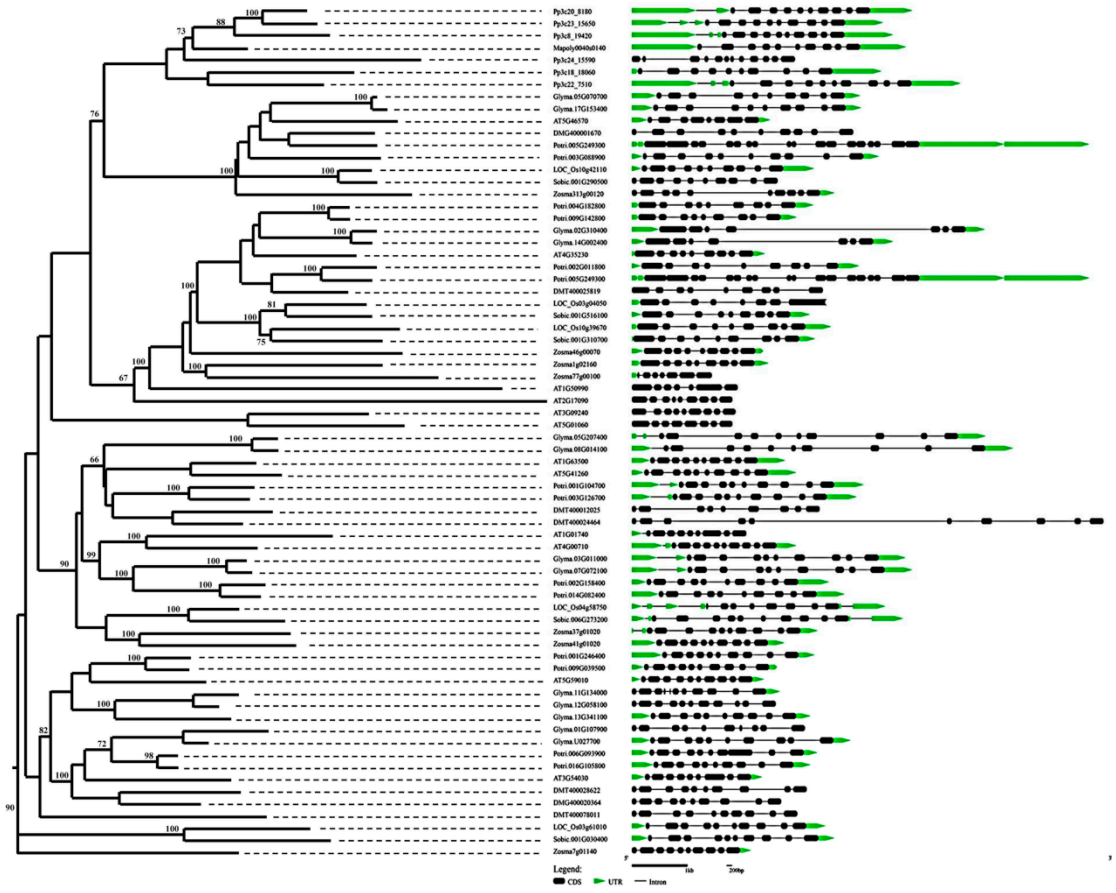


Figure 3. Phylogenetic relationship and exon-intron structure of BSK genes in representative plants. The BSKs were divided into five subgroups, which are indicated by different colors. The black boxes represent exons, solid lines represent introns, and green boxes represent untranslated regions (UTRs). The lengths of the BSK genes are indicated by horizontal lines (kb).

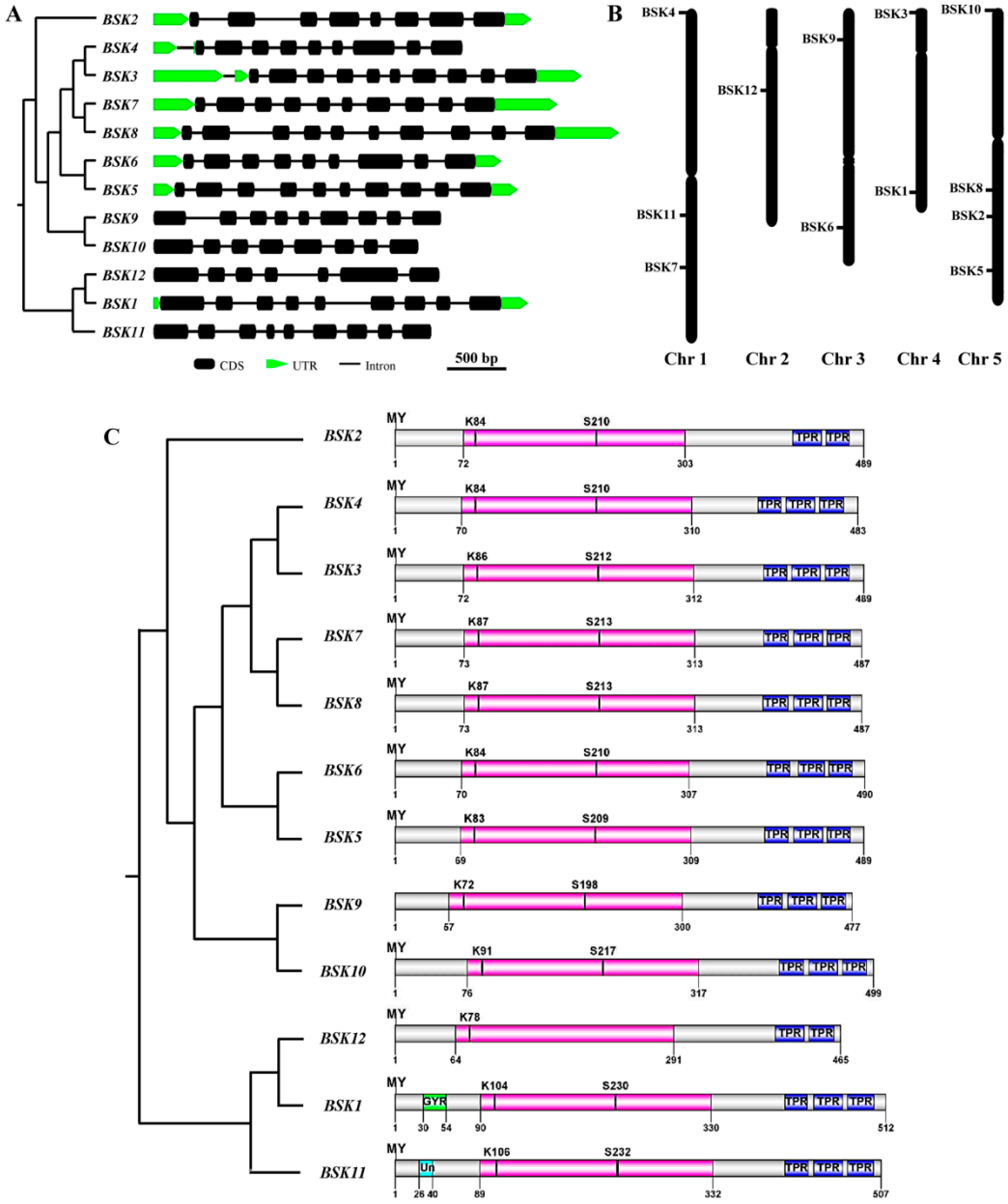


Figure 4. The features of the BSK genes in *Arabidopsis*. (**A**) Exon-intron structures of the BSK family genes. (**B**) Chromosomal location of BSK genes. (**C**) Schematic diagram of *Arabidopsis* BSK genes. The putative domains or motifs were identified using the Conserved Domain Database (CDD), Pfam and Simple Modular Architecture Research Tool (SMART) databases with the default parameters. The kinase domain was labeled in pink. MY, myrislation; TPR: tetratricopeptide repeat; GYR: glycine rich domain; Un: unknown domain.

2.4. Expression Profiles Analysis of BSK Family Genes from Arabidopsis

To compare the expression patterns of *Arabidopsis* BSK genes in different tissues or in responses to various abiotic stresses and hormones, the publicly available transcriptome sequencing and microarray datasets were analyzed. The results showed that most BSK members were constitutively expressed in different parts at different stages of plant development, while tissue-specific expression patterns were also observed (Figure 5A). For instance, *BSK1*, *BSK2*, *BSK3*, and *BSK5* were expressed at much lower levels in mature pollen, whereas *BSK11* and *BSK12* were more highly expressed. Among the 12 BSK genes, *BSK9* was specifically highly expressed in root, while *BSK5* was specifically expressed at very

low levels during seed development. A cell-specific expression pattern analysis for the BSK family genes showed that most BSK family members possessed different expression levels within three root zones (Figure 5B). *BSK11* and *BSK12* were undetectable, whereas *BSK9* and *BSK10* were found only in the differentiation zones. Moreover, the latest RNA-seq transcriptome data analysis showed that the expression of some BSK family genes were light-dependent (Figure 5C). *BSK1*, *BSK3*, *BSK5*, and *BSK8* were found to be highly expressed at night, while their expression was significantly depressed upon exposure to light. By contrast, *BSK6* was expressed in a rhythmic manner independent of light.

Under various abiotic and biotic stress treatments, significant alterations in the transcript amounts of some BSK family members were observed at certain time points (Figure 6A). For instance, *BSK5* and *BSK9* showed significant responses to multiple stimuli, while at most times, these transcripts were down-regulated. *BSK6* was upregulated after one hour of exposure to salt or oxidative stress, whereas *BSK2* was strongly repressed after one hour of heat stress. Under phytohormone treatments, the transcript levels of *BSK5*, *BSK9*, and *BSK12* showed rapid responses with either up- or down-regulation (Figure 6B). Surprisingly, only *BSK12* was rapidly induced by BL, while the other three members (*BSK1*, *BSK2*, and *BSK3*) reported to be involved in BR signaling were insensitive to BL for three hours: this, however, remains to be further elucidated.

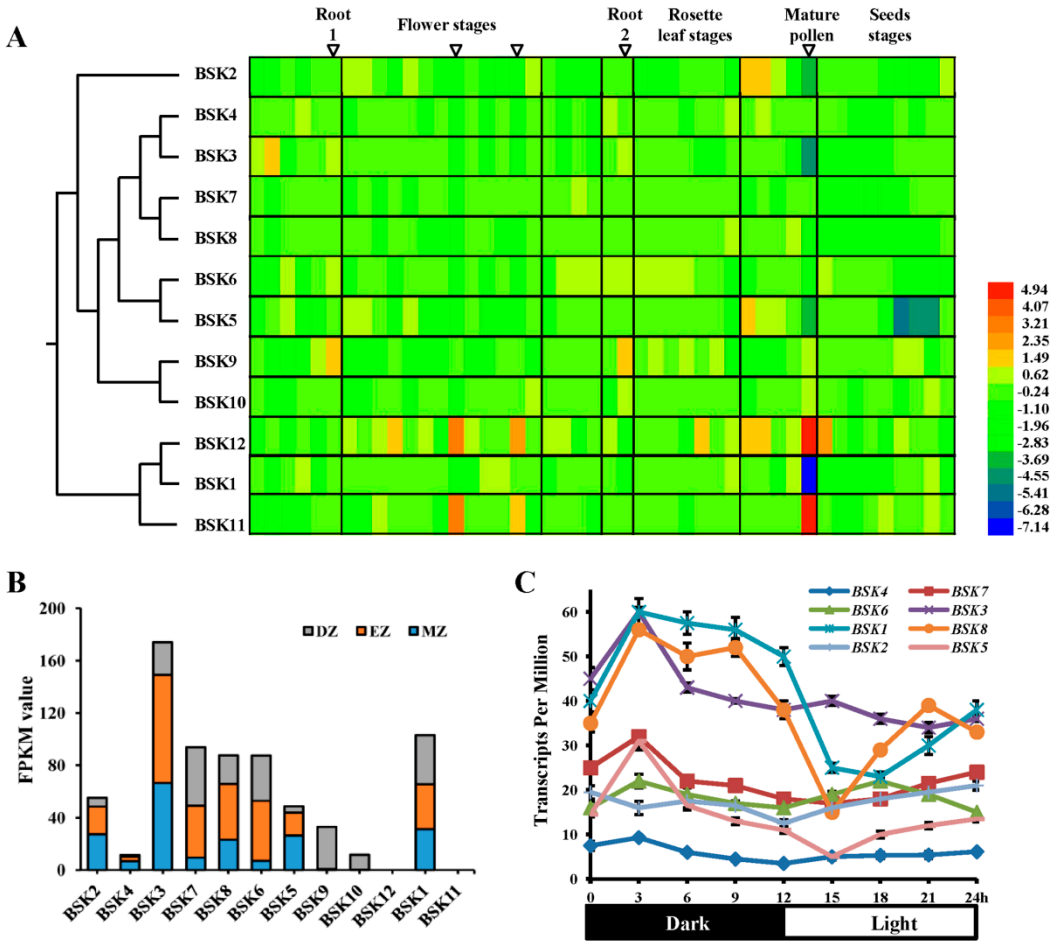


Figure 5. Development- and tissue-specific expression profiles of *Arabidopsis* BSK genes. (**A**) Heat map of the relative expression levels of all identified *Arabidopsis* BSK genes during development process. The colour bar represents the expression values. Samples from specific developing stages of *Arabidopsis* for gene expression analysis were highlighted with inverted triangles, such as roots from seedlings (1) and mature plants (2). (**B**) Expression profiles of BSK genes in specific root regions. DZ: division zone; EZ: elongation zone; MZ: mature zone. (**C**) Expression profiles of BSK genes across one light:dark cycle.

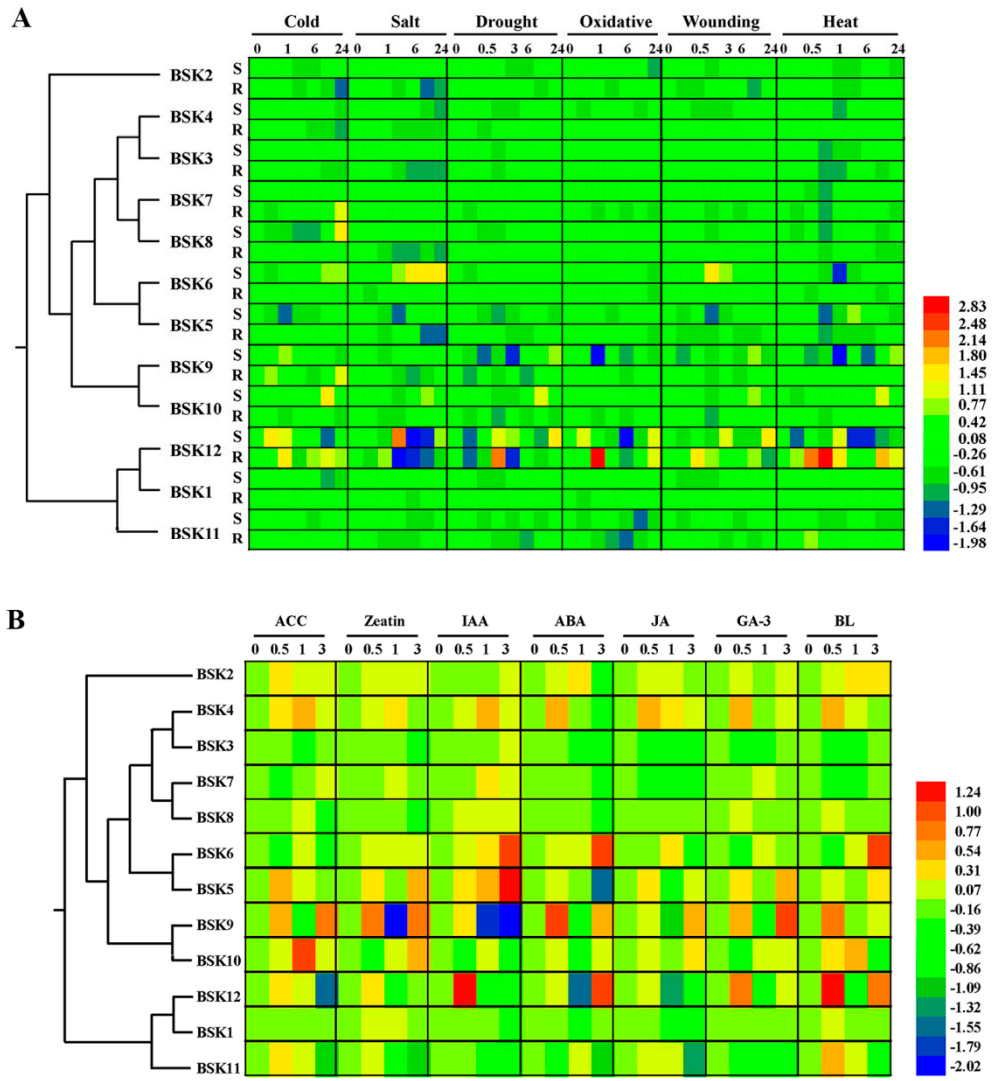


Figure 6. Expression patterns of all identified *Arabidopsis* BSK genes among hormone and abiotic stress. Public microarray data sets were obtained from BAR. Heat map representation the BSK genes under (**A**) six stress treatments, namely, cold, salt, drought, oxidative, wounding, and heat and (**B**) seven hormones, namely, ACC, zeatin, IAA, abscisic acid (ABA), methyl jasmonate (MeJA), gibberellic acid (GA-3), and brassinolide(BL). The colour bar represents the expression values.

2.5. Diverse Gene Expression Patterns of BSK Family from Arabidopsis Using Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

The expression profiles of the representative BSK family members that demonstrated a diverse pattern of induced gene expression were validated by RT-PCR. The results showed that the expressions of some BSK genes were, indeed, regulated by stress or hormone treatments (Figure 7A). For instance, *BSK1* and *BSK3* were shown to significantly respond to salt, cold, and heat stresses after 6 or 24 h. In response to heat stress, *BSK2* was first sharply inhibited (within 30 min of the stress treatment) and then highly accumulated within one hour, and then kept constant after incubation at the normal growth condition for an additional three hours. However, it was also found that stress responses in some BSK genes were inconstant to the trends observed with the microarray data. For example, *BSK2* was found to be significantly depressed by cold and salt stresses in previous studies, but was observed to be strongly induced in our study.

The tissue-specific expressional patterns of these BSK family genes were also investigated (Figure 7B). The results showed that promising expressional levels of *BSK2*, *BSK3*, *BSK5*, *BSK7*, and *BSK8* were found in leaves, flowers, fluorescents and siliques, which were consistent with the results from previous analysis. It also showed that all seven BSK genes were pronounced during reproductive growth. Our results further provided the information on the relative expression of these selected BSK genes. For instance, *BSK1* showed the strongest expression in flower and was dominantly expressed. Taken together, these results indicated the BSK genes are involved in multiple stress responses and developmental processes.

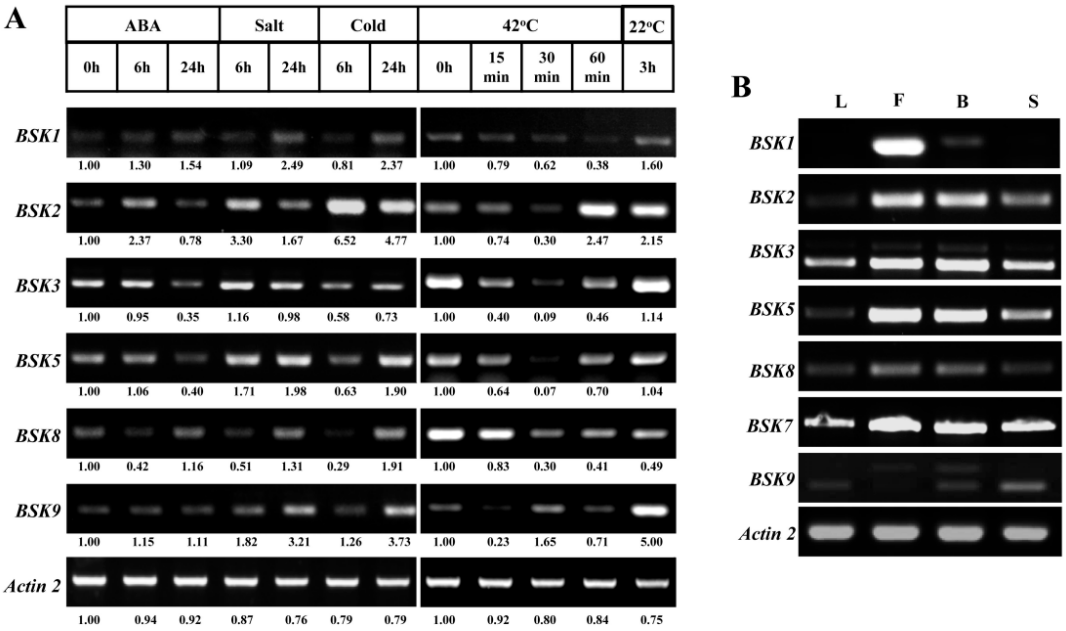


Figure 7. Reverse transcription-polymerase chain reaction (RT-PCR) validation of gene expression patterns of representative BSK genes under stresses and organs. Expression profiles of selected BSK genes under stresses (**A**) and in different tissues (**B**). For various stresses treatment, 7-day-old seedlings were subjected to 10 μM ABA, 150 mM NaCl and 4 °C on plates, respectively. Heat stress treatment was monitored with four-week-old seedlings under 42 °C for one hour and then recovered at 22 °C for three hours. The intensity corresponding to the individual PCR band was calculated by Image J with reactions from untreated samples as 1.00. L: leaves; F: flowers; B: buds; S: siliques.

2.6. Alternative Splicing (AS) Analysis for BSK Family Genes in Arabidopsis

The precursor mRNAs (pre-mRNAs) with introns can be spliced by alternative splicing (AS) to generate multiple mRNAs that are translated into different proteins. By analyzing the genome annotation obtained from the Phytozome v12.1 database (https://phytozome.jgi.doe.gov), we found that many BSK genes from different species possessed several transcripts resulting from AS (Figure S3). In particular, the BSK genes from *P. patens* underwent extensive AS events, resulting in an increase in the complexity of the BSK gene family transcriptome (Figure S3K). On the contrary, only one BSK gene was found to exhibit AS in *Arabidopsis* and rice.

As annotated in the Phytozome database, *BSK5* had two spliced transcripts in *Arabidopsis* (Figure S3A). *BSK5.1* encodes a complete BSK protein, while the spliced transcript *BSK5.2* encodes an N-terminal-truncated BSK with the first intron retained from *BSK5.1*. To investigate the expression patterns of the two transcripts, specific primers were designed for RT-PCR (Figure 8A). Our results showed that *BSK5.1* was the dominant transcript, which was expressed in different tissues and responded to various environmental stimuli (Figure 8B,C). To our surprise, a third isoform *BSK5.3*

was found in siliques (Figure 8B). A sequence analysis indicated two introns retained in *BSK5.3*, implying that *BSK5* spliced in a tissue-dependent manner.

In addition to *BSK5*, we also detected a novel splicing transcript of *BSK9* in flowers and buds (Figure 7B). Further analysis showed that the intron between exons 7 and 8 was retained in *BSK9* (Figure 9). We further screened the ASIP (Alternative Splicing in Plants) database (http://www.plantgdb.org/ASIP/) for intron retention (IR) events among the BSK family genes in *Arabidopsis*, and there was only one record for *BSK7*. However, no IR event was detected in *BSK7* from the leaves under normal growth conditions (Figure 9). Instead, this IR event within *BSK7* was strongly induced by cold stress, and this same phenomenon was also found in *BSK5* and *BSK9*. Notably, *BSK5.3* was also found to be specifically induced by heat stress. These results indicated a novel post-transcriptional regulation pattern in the BSK genes.

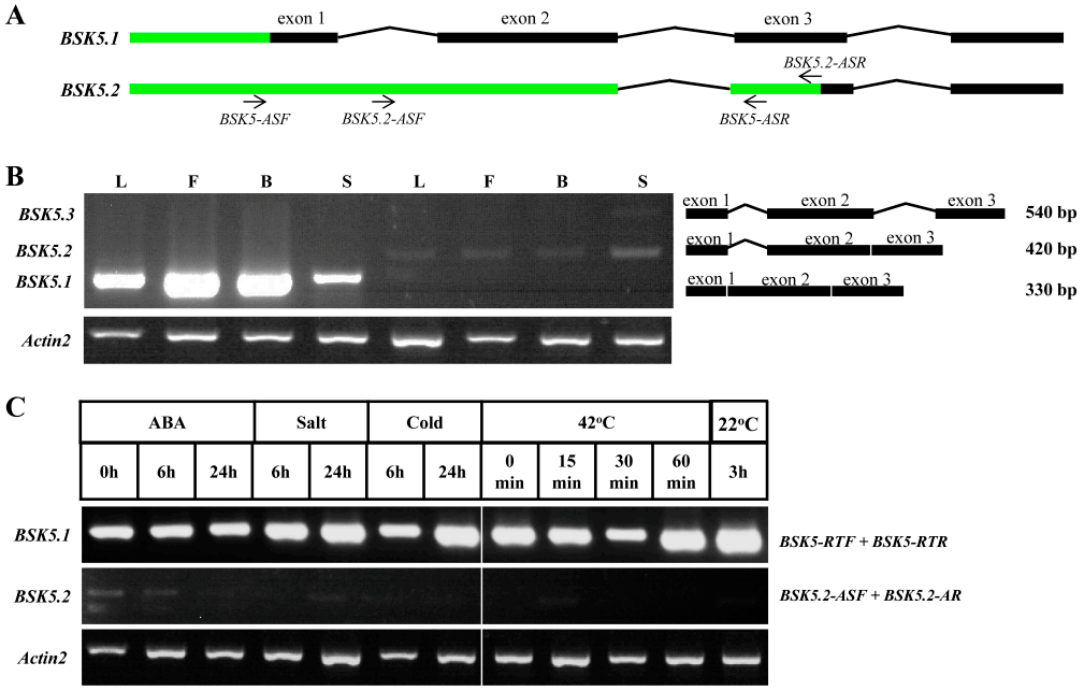


Figure 8. Alternative splicing analysis of *BSK5* in *Arabidopsis*. (**A**) The scheme shows the primers (black arrow) used in the PCR. The black boxes represent exons, solid lines represent introns, and green boxes represent untranslated regions (UTRs). BSK5-ASF, BSK5.3-ASF: forward primers; BSK5-ASR, BSK5.2-ASR: reverse primers. (**B**) The mRNA corresponding to the three variants of *BSK5* in different tissues was detected by RT-PCR. The numbers indicate the lengths (bp) of the spliced transcripts and schematic diagram outlining the organization of the transcript variants were on the right. L: leaves; F: flowers; B: buds; S: siliques. (**C**) RT-PCR analysis of the expression profiles of *BSK5* variants under stress treatment.

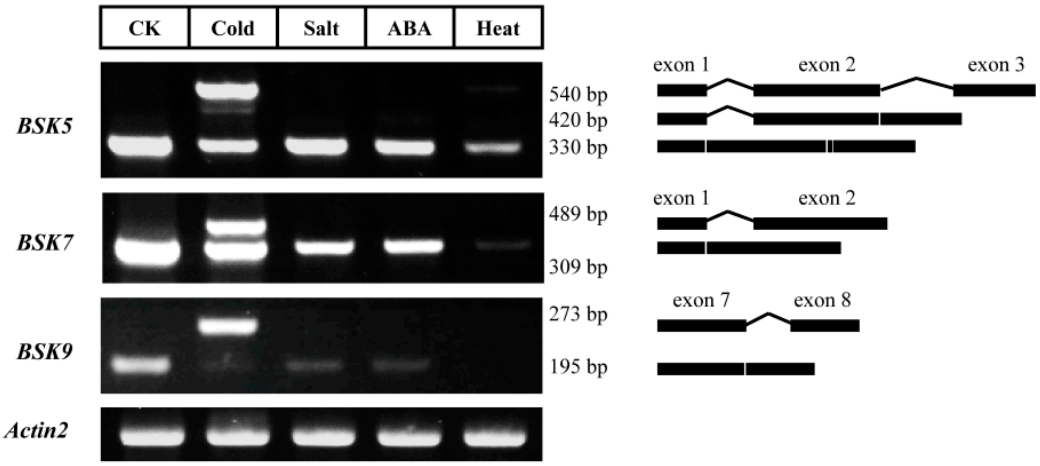


Figure 9. Intron retention analysis of BSK genes under stresses. Plants were stressed for 6 h under ABA/salt/cold and one hour under heat. Samples were then collected for mRNA isolation. Specific primers in exons were designed for RT-PCR in BSK genes. The detected PCR products were outlined on the right.

3. Discussion

Several studies have reported the roles of BSK proteins from *Arabidopsis* and rice in BR signaling and immunity, as well as in abiotic stress responses [10–19]. However, the knowledge of BSK proteins in other plant species is still quite limited. In the present study, we identified a total of 143 BSK proteins from 17 plant species. Our analysis showed the evolutionary origin of BSK genes and BR receptor BRI1 or receptor-like genes in embryophytes, indicating the origin of BR signaling in plants (Figure 1A). A previous study reveals that plant BRI1 is highly conserved across taxa [20]. Taken together, the origin and development of the BR signaling system seems to be highly relevant with the evolution from aquatic to terrestrial plants, which have been observed in the ABA signaling system [21]. Bryophytes have an intermediate position between aquatic and terrestrial plants, and the establishment of the BR signaling pathway might have a great impact on the explanation of the movement of plants from water to land as an adaptation to environmental conditions. We also found evidence that genome duplication likely contributed the most to the expansion of the BSK gene family in many plant lineages. Prior to the split of eudicots and monocots, there was no evidence of whole-genome duplication (WGD). However, several rounds of WGD or triple genome duplication (WTD) have been reported in the 12 angiosperms studied here [22,23]. This might result in the expansion of the BSK genes from angiosperms. Indeed, we found that approximately two-fold expansion in the BSK genes was identified in the plants with WGD, including *A. thaliana*, *G.*, *P. trichocarpa*, and *Z. mays*. Furthermore, in *B. rapa*, we observed over three-fold BSK expansion, which was likely due to WTD.

In *Arabidopsis*, the BSK gene family includes 12 members, which share many similarities with each other with some exceptions. For instance, additional motifs were found in *BSK1* and *BSK11* (Figure 4C). The glycine-rich (GXXXG) motif, which provides conformational flexibility in the unstructured region, has been identified in many proteins, but has not yet been characterized in the BSK gene family [24]. It is thus, possible that this motif in *BSK1* might confer flexibility in the interaction of BSK with *BRI1* or in homo-dimmer conformation. Post-translational protein modifications, such as phosphorylation and myristoylation, have been reported in the BSK gene family. In the current study, no myristoylation site, which is critical for membrane association of BSK was found in *BSK9* [17]. A prediction of the subcellular localization of all BSK proteins from *Arabidopsis* showed that *BSK9*, as well as *BSK10* and *BSK11*, were absent from the plasma membrane, indicating that myristoylation might be sufficient, but not necessary, to anchor a protein to the membrane (Table S1). Further validation of the subcellular localization is required to support the accurate role of the myristoylation site. Other kinds

of post-translational protein modification in this gene family were also investigated. The prediction results showed that conserved sumoylation and ubiquitination sites were found in all BSK family proteins (Figure S2). Overall, these post-translational protein modification sites were mainly located in the N- and C-terminals, where the residues were probably the targets of a SUMO or ubiquitin complex. Many proteins are both sumoylated and ubiquitinated, and often at the same lysine residues. In most BSK proteins, some conserved residues were found to serve as both ubiquitination and sumoylation sites. These two protein modifications can act either synergistically or antagonistically, indicating the existence of a fine-tuned post-translational regulatory mechanism of the BSK family, which remains to be elucidated.

In the present study, comprehensive microarray and RNA-seq analyses of the *Arabidopsis* BSK proteins at several developmental stages showed that some of these BSK proteins might extensively function in many tissues (Figure 5). Our RT-PCR data also showed that the selected BSK genes, including *BSK3*, *BSK7*, and *BSK8*, were expressed in leaves, buds, wide open flowers, and siliques; this result was consistent with the previous findings (Figure 7B). Some BSK proteins seemed to function in specific tissues, likely resulting in their sub-functionalization. For instance, *BSK12* was upregulated only at the early stages of see development, whereas the expression of *BSK5* was found to be reduced gradually; this was also verified by our RT-PCR data. BRs were reported to regulate seed size and shape in *Arabidopsis* by BZR1-mediated activation and repression of a number of known regulators of seed size [25]. AtBZR1-like genes are highly conserved in angiosperms, and both soybean GmBZL2 rice OsBZR1 play an important role in seed set/size [26–29]. Thus, it could be possible that *BSK5* and *BSK12* function upstream of BZR1 in seed development. BRs were also reported to play important roles in regulating root meristem maintenance and root elongation [30]. Here, significantly high expression of *BSK9* was observed in the meristematic zone of roots (Figure 6C). It could be exciting to examine the function of *BSK9* in root meristem development. Moreover, the expression of *BSK1* was the lowest in mature pollen among the BSK family genes (Figure 5A). On the contrary, *BSK11* and *BSK12*, which were in the same sub-group, showed much stronger expression. BRs have been reported to promote pollen germination and growth [31]. Some BSK downstream genes, including *Arabidopsis* BES1 and rice OsBZR1, have been found to be involved in anther and pollen development, indicating that these BSK members might play a role upstream of them in male gametophyte development [29,32]. In addition, BRs also regulate photomorphogenesis and link with light signals through BES1 [33,34]. Here, we found that several BSK family members showed light-dependent transcriptional regulation. For instance, the expressions of *BSK1*, *BSK3*, *BSK5*, and *BSK8* were significantly upregulated in the dark and repressed by light (Figure 5C). It will be interesting to investigate the function of these BSK members upstream of BZR1 in photomorphogenesis, broadening the functions of the BSK gene family.

A systematical microarray analysis of the BSK proteins under distinct abiotic stresses (cold, salt, drought, et al) and phytohormones (ABA, 1-aminocyclopropane-1-carboxylic acid (ACC), indole-3-acetic acid (IAA), BR, etc)) based on the *Arabidopsis* eFP Browser database indicated that most BSK proteins were involved in these abiotic stress and hormone responses (Figure 6). For instance, *BSK2*, *BSK3*, *BSK5*, *BSK6*, and *BSK12* were involved in the salt-stress response. Among these BSK genes, most were repressed by salt stress, whereas *BSK6* was constantly upregulated. In addition, *BSK5*, *BSK9,* and *BSK12* showed broad responses to multiple stresses. In contrast, only a small number of BSK genes were regulated by hormones. To our surprise, most BSK members, including *BSK1*, showed no response to BRs in 3 hours, whereas *BSK12* was sharply induced in 30 minutes, indicating an unexpected role of *BSK12* through BR perception (Figure 6B). Our RT-PCR results further confirmed that several selected BSK genes indeed showed a response to various abiotic stresses and hormones, although there was some inconsistency with the microarray data (Figure 7A). For instance, *BSK1* was upregulated under salt and cold stresses after 24 hours. Under the same situations, *BSK2* and *BSK9* were also highly induced. By contrast, all BSK genes were repressed by heat stress and ABA the inhibited expression of *BSK3*, *BSK5* and *BSK8*. Further experiments are required to elucidate the important roles of these BSK genes in various stress responses.

Alternative splicing, an important modulator of gene expression, can increase proteome diversity and regulate mRNA levels by generating multiple transcripts from the same gene [35,36]. In plants, this post-transcriptional mechanism is noticeably induced by environmental stimuli and plays an important role in the regulation of gene expression for biotic/abiotic stresse responses as well as for plant growth and development [37–42]. The RNA-seq analysis has revealed that in *Arabidopsis* more than 61% of intron-containing genes and even more undergo alternative splicing under normal growth and stress conditions, respectively [31]. Notably, IR is the most prevalent form of AS in higher plants. For instance, the AS forms of *Arabidopsis* contain an unusually high fraction of retained introns (above 30%), and more than half of the AS events belong to IR in rice [43–45]. IR is mostly accepted on account of mis-splicing and thought to be non-functional because they are likely to result in nonsense-mediated decay [46]. However, several studies have highlighted the functional importance of intron-retaining mRNAs in plants. IR in the *Arabidopsis* INDETERMINATE DOMAIN 14 (IDD14) transcription factor gene generates a competitive inhibitor that modulates starch accumulation in response to cold stress [47]. IR in the 5′ UTR of the Zinc-Induced Facilitator 2 gene (ZIF2) improves zinc tolerance in *Arabidopsis* [48]. *HAB1*, a group A protein type 2C phosphatase (PP2C), undergoes IR to produce a splice variant that plays opposing roles in fine-tuning ABA signaling [49]. Thus, IR should no longer be underestimated and exploring its roles in the development and/or stress response is of increasing importance. Here we found that the IR event in some *Arabidopsis* BSK genes was tissue-specific and responded to stress/hormone. The intron-containing form of *BSK9* preferred to express in development stages from buds to flowers, but was spliced completely in siliques. We also found that cold and heat stresses promoted the IR rate within the first two introns of *BSK5*, whereas ABA operated only in the first intron to generate and accumulate differently spliced transcripts. IR in *BSK7* and *BSK9* were also specifically induced by cold stress. Although the role of an individual transcript is largely unknown currently, these observations indicate the post-transcriptional regulation of these BSK genes in response to environmental stimuli. Future work will be performed to elucidate the mechanism and physiological significance of tissue-specific and stress-induced IR in these BSK genes.

Overall, our study represented diverse features of the BSK gene family in *Arabidopsis*, providing valuable insights into their important functions.

4. Materials and Methods

4.1. Identification of BSK Family Genes and Phylogenetic Analysis of BSKs in Plants

To identify ortholog(s) of AtBSK1 from *Arabidopsis* in the plant genome, the amino acid sequence of AtBSK1 was used as the seed sequence to perform a BLASTp search of the database whole genome sequences in the Phytozome database (https://phytozome.jgi.doe.gov). A similar method was applied on AtBRI1. Multiple protein sequence alignment for the deduced amino acid sequences of BSK proteins from different plant species was performed using the ClustalW software (http://www.clustal.org/clustal2/) with default settings. Then, the phylogenetic tree was constructed using MEGA software (Tokyo Metropolitan University, Tokyo, Japan) using the neighbor-joining (NJ) method and the bootstrap test carried out with 1000 replicates. The phylogenetic trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

4.2. Conserved Domain Recognition and Gene Structure Analysis

The CDS and the corresponding genomic sequences of the BSKs from the selected plant species were retrieved from the databases. Analysis of the exon-intron structures for the BSKs was carried out using the Gene Structure Display Server (GSDS v2.0, http://gsds.cbi.pku.edu.cn/index.php). Conserved domain for BSK proteins from *Arabidopsis* were identified using Conserved Domain Search Service (CD Search) from NCBI. The protein post-modification analysis was performed using SUMOgo (http://predictor.nchu.edu.tw/SUMOgo/) and UbiSite (http://csb.cse.yzu.edu.tw/UbiSite) with high

specificity level of 95%. Subcellular prediction for BSK family genes were performed using Cell eFP Browser (http://bar.utoronto.ca/cell_efp/cgi-bin/cell_efp.cgi).

4.3. Expression Profiles Analysis Using Microarray and RNA-seq Data

For expression profiling of the BSK genes in *Arabidopsis*, we utilized the *Arabidopsis* ATH1 microarray data (*Arabidopsis* eFP Browser 2.0), root RNA-seq data bases (https://sites.lsa.umich.edu/pgrp-roots/) and AtRTD2 [50]. Different tissues of *Arabidopsis* and the seedlings under various stresses were analyzed. The gene expression patterns of each tissue were analyzed, and the expression values were log2 transformed from The Bio-Analytic Resource for Plant Biology (BAR). Finally, heat maps of hierarchical clustering were visualized using the Heml software. For data retrieved from RNA-seq data bases, the transcript abundance was expressed as fragments per kilobase of exon model per million mapped reads (FPKM) or Transcript Per Million.

4.4. Plant Materials, Growth Conditions and Stress Treatments

Seeds of *Arabidopsis thaliana* ecotype Columbia (Col-0) were used in this research. The seeds were surface sterilized with bleach and washed extensively with sterilized water for three times, and then incubated in a growth chamber after cold treatment for two days. After growth for one week, the seedlings were exposed to various stress conditions in plates containing 150 mM NaCl and 10 μM ABA, respectively. For cold and heat treatment, 7-day-old plants were placed at 4 °C or 42 °C incubator with normal illumination. Samples from leaves, buds, flowers and siliques were collected from well-growing plants at the same time. All the plant materials were harvested by liquid nitrogen cooling before stored at −80 °C.

4.5. RNA Extraction and RT-PCR

Isolation of total RNA from treated samples was performed using an RNA extraction kit (Promega, Madison, WI, USA). The cDNA synthesis was performed with total RNA (2 μg) reverse transcribed using All-In-One RT MasterMix (Applied Biological Materials, Zhenjiang, China). RT-PCR analysis was conducted using 2 × T5 Super PCR Mix (Tsingke, Beijing, China) and Taq Master MixTaq mix (Vazyme Biotech, Nanjing, China). All primers used in this study are listed in Supplementary Table S2. Quantification for gel intensity was carried out using Image J software (https://imagej.nih.gov/ij/).

5. Conclusions

In this study, we identified a total of 143 BSK proteins from 17 plant species. The phylogenetic analysis showed that the expansion of the BSK genes originated from embryophytes. A further comparative study revealed that most of the BSK genes in *Arabidopsis* were constitutively expressed and responded to some hormones or abiotic stresses. We also found some interesting post-transcriptional regulation patterns in *BSK5*, *BSK7*, and *BSK9*. Our results will further provide clues for the functional analysis of the important functions of BSK family genes in plants.

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/20/5/1138/s1.

Author Contributions: Z.L. and J.L. conceived and designed the experiments; Z.L. and J.S. conducted the experiments analyzed the data; Z.L. wrote the manuscript and J.L. reviewed the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (31801277).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Clouse, S.D.; Sasse, J.M. BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **1998**, *49*, 427–451. [CrossRef] [PubMed]

2. Fàbregas, N.; Caño-Delgado, A.I. Turning on the microscope turret: A new view for the study of brassinosteroid signaling in plant development. *Physiol Plant.* **2014**, *151*, 172–183. [CrossRef] [PubMed]
3. Li, J.; Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **1997**, *90*, 929–938. [CrossRef]
4. Wang, Z.Y.; Seto, H.; Fujioka, S.; Yoshida, S.; Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **2001**, *410*, 380–383. [CrossRef] [PubMed]
5. Li, J.; Wen, J.; Lease, K.A.; Doke, J.T.; Tax, F.E.; Walker, J.C. BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **2002**, *110*, 213–222. [CrossRef]
6. Kim, T.W.; Guan, S.; Sun, Y.; Deng, Z.; Tang, W.; Shang, J.X.; Sun, Y.; Burlingame, A.L.; Wang, Z.Y. Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat. Cell Biol.* **2009**, *11*, 1254–1260. [CrossRef] [PubMed]
7. He, J.X.; Gendron, J.M.; Yang, Y.; Li, J.; Wang, Z.Y. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10185–10190. [CrossRef] [PubMed]
8. Yin, Y.; Wang, Z.Y.; Mora-Garcia, S.; Li, J.; Yoshida, S.; Asami, T.; Chory, J. BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **2002**, *109*, 181–191. [CrossRef]
9. Kim, T.W.; Wang, Z.Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant. Biol.* **2010**, *61*, 681–704. [CrossRef] [PubMed]
10. Tang, W.; Kim, T.W.; Oses-Prieto, J.A.; Sun, Y.; Deng, Z.; Zhu, S.; Wang, R.; Burlingame, A.L.; Wang, Z.Y. BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. *Science* **2008**, *321*, 557–560. [CrossRef] [PubMed]
11. Mora-García, S.; Vert, G.; Yin, Y.; Caño-Delgado, A.; Cheong, H.; Chory, J. Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in *Arabidopsis*. *Genes Dev.* **2004**, *18*, 448–460. [CrossRef] [PubMed]
12. Sreeramulu, S.; Mostizky, Y.; Sunitha, S.; Shani, E.; Nahum, H.; Salomon, D.; Hayun, L.B.; Gruetter, C.; Rauh, D.; Ori, N.; et al. BSKs are partially redundant positive regulators of brassinosteroid signaling in Arabidopsis. *Plant J.* **2013**, *74*, 905–919. [CrossRef] [PubMed]
13. Ren, H.; Willige, B.C.; Jaillais, Y.; Geng, S.; Park, M.Y.; Gray, W.M.; Chory, J. BRASSINOSTEROID-SIGNALING KINASE 3, a plasma membrane-associated scaffold protein involved in early brassinosteroid signaling. *PLoS Genet.* **2019**, *15*, e1007904. [CrossRef] [PubMed]
14. Bayer, M.; Nawy, T.; Giglione, C.; Galli, M.; Meinnel, T.; Lukowitz, W. Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science.* **2009**, *323*, 1485–1488. [CrossRef] [PubMed]
15. Li, Z.Y.; Xu, Z.S.; He, G.Y.; Yang, G.X.; Chen, M.; Li, L.C.; Ma, Y.Z. A mutation in Arabidopsis BSK5 encoding a brassinosteroid-signaling kinase protein affects responses to salinity and abscisic acid. *Biochem. Biophys. Res. Commun.* **2012**, *426*, 522–527. [CrossRef] [PubMed]
16. Wang, J.; Shi, H.; Zhou, L.; Peng, C.; Liu, D.; Zhou, X.; Wu, W.; Yin, J.; Qin, H.; Ma, W.; et al. OsBSK1-2, an Orthologous of AtBSK1, Is Involved in Rice Immunity. *Front Plant Sci.* **2017**, *8*, 908. [CrossRef] [PubMed]
17. Shi, H.; Shen, Q.; Qi, Y.; Yan, H.; Nie, H.; Chen, Y.; Zhao, T.; Katagiri, F.; Tang, D. BR-SIGNALING KINASE1 physically associates with FLAGELLIN SENSING2 and regulates plant innate immunity in Arabidopsis. *Plant Cell* **2013**, *25*, 1143–1157. [CrossRef] [PubMed]
18. Yan, H.J.; Zhao, Y.F.; Shi, H.; Li, J.; Wang, Y.C.; Tang, D.Z. BRASSINOSTEROID-SIGNALING KINASE1 Phosphorylates MAPKKK5 to Regulate Immunity in Arabidopsis. *Plant Physiol.* **2018**, *176*, 2991–3002. [CrossRef] [PubMed]
19. Yu, M.-H.; Zhao, Z.-Z.; He, J.-X. Brassinosteroid Signaling in Plant–Microbe Interactions. *Int. J. Mol. Sci.* **2018**, *19*, 4091. [CrossRef] [PubMed]
20. Navarro, C.; Moore, J.; Ott, A.; Baumert, E.; Mohan, A.; Gill, K.S.; Sandhu, D. Evolutionary, Comparative and Functional Analyses of the Brassinosteroid Receptor Gene, BRI1, in Wheat and Its Relation to Other Plant Genomes. *PLoS ONE* **2015**, *10*, e0127544. [CrossRef] [PubMed]
21. Umezawa, T.; Nakashima, K.; Miyakawa, T.; Kuromori, T.; Tanokura, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Molecular basis of the core regulatory network in ABA responses: Sensing, signaling and transport. *Plant Cell Physiol.* **2010**, *51*, 1821–1839. [CrossRef] [PubMed]

22. Panchy, N.; Lehti-Shiu, M.; Shiu, S.H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef] [PubMed]
23. Del Pozo, J.C.; Ramirez-Parra, E. Whole genome duplications in plants: An overview from Arabidopsis. *J. Exp. Bot.* **2015**, *66*, 6991–7003. [CrossRef] [PubMed]
24. Anand, S.; Sharma, C. Glycine-rich loop encompassing active site at interface of hexameric, M. tuberculosis Eis protein contributes to its structural stability and activity. *Int. J. Biol. Macromol.* **2018**, *109*, 124–135. [CrossRef] [PubMed]
25. Jiang, W.B.; Huang, H.Y.; Hu, Y.W.; Zhu, S.W.; Wang, Z.Y.; Lin, W.H. Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant Physiol.* **2013**, *162*, 1965–1977. [CrossRef] [PubMed]
26. Zhang, Y.; Zhang, Y.J.; Yang, B.J.; Yu, X.X.; Wang, D.; Zu, S.H.; Xue, H.W.; Lin, W.H. Functional characterization of GmBZL2 (AtBZR1 like gene) reveals the conserved BR signaling regulation in Glycine max. *Sci. Rep.* **2016**, *8*, 31134. [CrossRef] [PubMed]
27. Manoli, A.; Trevisan, S.; Quaggiotti, S.; Varotto, S. Identification and characterization of the BZR transcription factor family and its expression in response to abiotic stresses in *Zea mays* L. *Plant Growth Regul.* **2018**, *84*, 423. [CrossRef]
28. Li, Y.Y.; He, L.L.; Li, J.; Chen, J.G.; Liu, C.G. Genome-Wide Identification, Characterization, and Expression Profiling of the Legume BZR Transcription Factor Gene Family. *Front. Plant Sci.* **2018**, *9*, 1332. [CrossRef] [PubMed]
29. Zhu, X.; Liang, W.; Cui, X.; Chen, M.; Yin, C.; Luo, Z.; Zhu, J.; Lucas, W.J.; Wang, Z.; Zhang, D. Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. *Plant J.* **2015**, *82*, 570–581. [CrossRef] [PubMed]
30. González-García, M.P.; Vilarrasa-Blasi, J.; Zhiponova, M.; Divol, F.; Mora-García, S.; Russinova, E.; Caño-Delgado, A.I. Brassinosteroids control meristem size by promoting cell cycle progression in Arabidopsis roots. *Development* **2011**, *138*, 849–859. [CrossRef] [PubMed]
31. Vogler, F.; Schmalzl, C.; Englhart, M.; Bircheneder, M.; Sprunck, S. Brassinosteroids promote Arabidopsis pollen germination and growth. *Plant Reprod.* **2014**, *27*, 153–167. [CrossRef] [PubMed]
32. Ye, Q.; Zhu, W.; Li, L.; Zhang, S.; Yin, Y.; Ma, H.; Wang, X. Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. *Proc. Natl. Acad. Sci. USA* **2010**, *13*, 6100–6105. [CrossRef] [PubMed]
33. Wang, Z.Y.; Bai, M.Y.; Oh, E.; Zhu, J.Y. Brassinosteroid signaling network and regulation of photomorphogenesis. *Annu. Rev. Genet.* **2012**, *46*, 701–724. [CrossRef] [PubMed]
34. Wang, W.; Lu, X.; Li, L.; Lian, H.; Mao, Z.; Xu, P.; Guo, T.; Xu, F.; Du, S.; Cao, X.; Wang, S. Photoexcited CRYPTOCHROME1 Interacts with Dephosphorylated BES1 to Regulate Brassinosteroid Signaling and Photomorphogenesis in Arabidopsis. *Plant Cell* **2018**, *30*, 1989–2005. [CrossRef] [PubMed]
35. Mastrangelo, A.M.; Marone, D.; Laido, G.; De Leonardis, A.M.; De Vita, P. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* **2012**, *185*, 40–49. [CrossRef] [PubMed]
36. Marquez, Y.; Brown, J.W.; Simpson, C.; Barta, A.; Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* **2012**, *22*, 1184–1195. [CrossRef] [PubMed]
37. Staiger, D.; Brown, J.W. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* **2013**, *25*, 3640–3656. [CrossRef] [PubMed]
38. Ding, F.; Cui, P.; Wang, Z.Y.; Zhang, S.D.; Ali, S.; Xiong, L.M. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genomics* **2014**, *15*, 431. [CrossRef] [PubMed]
39. Feng, J.; Li, J.; Gao, Z.; Lu, Y.; Yu, J.; Zheng, Q.; Yan, S.; Zhang, W.; He, H.; Ma, L.; Zhu, Z. SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. *Mol. Plant* **2015**, *8*, 1038–1052. [CrossRef] [PubMed]
40. Thatcher, S.R.; Danilevskaya, O.N.; Meng, X.; Beatty, M.; Zastrow-Hayes, G.; Harris, C.; Van Allen, B.; Habben, J.; Li, B. Genome-Wide Analysis of Alternative Splicing during development and Drought Stress in Maize. *Plant Physiol.* **2016**, *170*, 586–599. [CrossRef] [PubMed]

41. Zhu, F.Y.; Chen, M.X.; Ye, N.H.; Shi, L.; Ma, K.L.; Yang, J.F.; Cao, Y.Y.; Zhang, Y.; Yoshida, T.; Fernie, A.R.; et al. Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in Arabidopsis seedlings. *Plant J.* **2017**, *91*, 518–533. [CrossRef] [PubMed]
42. Laloum, T.; Martín, G.; Duque, P. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci.* **2018**, *23*, 140–150. [CrossRef] [PubMed]
43. Ner-Gaon, H.; Halachmi, R.; Savaldi-Goldstein, S.; Rubin, E.; Ophir, R.; Fluhr, R. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J.* **2004**, *39*, 877–885. [CrossRef] [PubMed]
44. Min, X.J.; Powell, B.; Braessler, J.; Meinken, J.; Yu, F.; Sablok, G. Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. *BMC Genomics* **2015**, *16*, 721. [CrossRef] [PubMed]
45. Wang, B.B.; Brendel, V. Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7175–7180. [CrossRef] [PubMed]
46. Lejeune, F.; Maquat, L.E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **2005**, *17*, 309–315. [CrossRef] [PubMed]
47. Seo, P.J.; Kim, M.J.; Ryu, J.Y.; Jeong, E.Y.; Park, C.M. Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* **2011**, *2*, 303. [CrossRef] [PubMed]
48. Remy, E.; Cabrito, T.R.; Batista, R.A.; Hussein, M.A.; Teixeira, M.C.; Athanasiadis, A.; Sá-Correia, I.; Duque, P. Intron retention in the 5′UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in Arabidopsis. *PLoS Genet.* **2014**, *10*, e1004375. [CrossRef] [PubMed]
49. Wang, Z.; Ji, H.; Yuan, B.; Wang, S.; Su, C.; Yao, B.; Zhao, H.; Li, X. ABA signalling is fine-tuned by antagonistic HAB1 variants. *Nat. Commun.* **2015**, *6*, 8138. [CrossRef] [PubMed]
50. Calixto, C.P.G.; Guo, W.B.; Jams, A.B.; Tzioutziou, N.A.; Entizne, J.C.; Panter, P.E. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. *Plant Cell* **2018**, *30*, 1424–1444. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of Glycoside Hydrolase Family 1 β-glucosidase Genes in *Brassica rapa* and Their Potential Role in Pollen Development

Xiangshu Dong [1,*], Yuan Jiang [1] and Yoonkang Hur [2,*]

[1] School of Agriculture, Yunnan University, Kunming 650091, China; jiangyuan@mail.ynu.edu.cn
[2] Department of Biological Sciences, Chungnam National University, Daejeon 34141, Korea
* Correspondence: dongxiangshu@ynu.edu.cn (X.D.); ykhur@cnu.ac.kr (Y.H.);
Tel.: +86-871-6503-1539 (X.D.); +82-42-821-6279 (Y.H.)

Received: 28 February 2019; Accepted: 1 April 2019; Published: 3 April 2019

Abstract: Glycoside hydrolase family 1 (GH1) β-glucosidases (BGLUs) are encoded by a large number of genes, and are involved in many developmental processes and stress responses in plants. Due to their importance in plant growth and development, genome-wide analyses have been conducted in model plants (*Arabidopsis* and rice) and maize, but not in *Brassica* species, which are important vegetable crops. In this study, we systematically analyzed *B. rapa BGLU*s (*BrBGLU*s), and demonstrated the involvement of several genes in pollen development. Sixty-four *BrBGLU*s were identified in *Brassica* databases, which were anchored onto 10 chromosomes, with 10 tandem duplications. Phylogenetic analysis revealed that 64 genes were classified into 10 subgroups, and each subgroup had relatively conserved intron/exon structures. Clustering with *Arabidopsis* BGLUs (AtBGLUs) facilitated the identification of several important subgroups for flavonoid metabolism, the production of glucosinolates, the regulation of abscisic acid (ABA) levels, and other defense-related compounds. At least six BrBGLUs might be involved in pollen development. The expression of *BrBGLU10/AtBGLU20*, the analysis of co-expressed genes, and the examination of knocked down *Arabidopsis* plants strongly suggests that *BrBGLU10/AtBGLU20* has an indispensable function in pollen development. The results that are obtained from this study may provide valuable information for the further understanding of β-glucosidase function and *Brassica* breeding, for nutraceuticals-rich *Brassica* crops.

Keywords: β-glucosidase; *Brassica rapa*; BrBGLU10; pollen development; co-expression analysis

1. Introduction

Glycoside hydrolases (EC 3.2.1) are classified into a group of enzymes that hydrolyze the glycosidic bonds of carbohydrates [1]. At the end of March in 2019, 161 families have been identified and classified in the CAZy (Carbohydrate-Active enZYmes) database [2,3]. Among these families, the glycoside hydrolase (GH) family 1 is recognized for its β-glycosidase activity, which largely contributes to various developmental processes and stress responses in plants [4,5]. Genome-wide analysis of GH1 β-glycosidase genes (*BGLU*s) has been conducted in three plant species: *Arabidopsis*, with 48 genes grouped into 10 subfamilies [6]; rice, with 40 genes grouped into eight subfamilies [5]; and maize, with 26 genes grouped into four subfamilies [7,8]. Recently, a comparison between the *Arabidopsis* and rice *BGLU*s with respect to sequence identity and expression revealed that these exhibited substantial tissue specificity and differential responses to various stress treatments, although these have a high degree of similarity [9]. However, no systematic analysis of *BGLU*s in *Brassica rapa*, which is an important vegetable crop, has been performed to date.

In addition to classifications based on genomic DNA organization, *Arabidopsis* BGLUs (*AtBGLU*s) could be classified in relation to their known functions, which shows that genes within the same subfamily may function in similar processes. A large number of *AtBGLU*s are involved in flavonoid metabolism: *AtBGLU1-6* for flavonol accumulation [10,11], *AtBGLU7-11* for anthocyanin glucosyltransferase [11,12], and *AtBGLU12-17* for flavonoid utilization [10,13]. Seven genes (*AtBGLU26*, *AtBGLU34-39*) function as myrosinases for chemical defense against herbivores and pathogen attacks [14–16]. *AtBGLU18* and *AtBGLU33* regulate ABA responses by increasing ABA levels through the hydrolysis of glucose-conjugated ABA (ABA-GE) [17,18]. Scopolin, which is specifically produced in the roots, and which plays a role in a defense against pathogen attack and abiotic stresses [19,20], is controlled by *ArBGLU21-23* [21,22]. The gene products encoded by *AtBGLU45* and *AtBGLU46* hydrolyze monolignol glucosides, thereby regulating lignin biosynthesis [23]. *AtBGLU42* is involved in the induction of systemic resistance to bacterial disease, and the release of iron-mobilizing phenolic metabolites during iron deficiency [24]. However, no gene has been reported, with respect to pollen development.

During pollen development, the tapetum secretes various components, such as lipidic precursors and lipidics onto the pollen surface, leading to the formation of sculptured exine and exine cavities by hydrolyzation and other reactions [25]. In addition to lipid components, pollen wall development requires the regulation of polysaccharide metabolism [26], suggesting a possible involvement of the hydrolysis of glycosidic bonds of carbohydrates. Glycoside hydrolase has been reported involved in the cell wall polysaccharide degradation [27] and their coding genes were downregulated in the *OsTDR* (*Tapetum Degeneration Retardation*) mutant [28] and the sterile floral buds of *B. rapa* [29], indicating a possibility that β-glucosidase may play a role in pollen development.

In this study, we systematically identified *Brassica rapa* β-glycosidase genes (*BrBGLU*s) and analyzed their expression patterns and phylogenetic relationships. In addition, in silico analyses indicated that *BrBGLU10/AtBGLU20* have conserved functions during pollen development, and knocking down *AtBGLU20* using antisense oligos in *Arabidopsis* results in the production of aborted pollen grains. Furthermore, bioinformatics and molecular analyses provide valuable information on the function of *BrBGLUs* during pollen development.

2. Results

2.1. Identification and Chromosomal Distribution of BrBGLUs

After a HMM (Hidden Markov Model) search, 64 *BrBGLU* genes were identified and designated as *BrBGLU1* to *BrBGLU64*, according to their positions on the chromosomes (Figure 1). The locus ID, genome location, coding sequence (CDS) length, and the protein length of the *BrBGLUs* are listed in Table 1. The genomic DNA sequences of the *BrBGLU*s ranged from 390 bp to 9617 bp. While the average length was 1293 bp, the length of the CDS of the *BrBGLU*s ranged from 267 bp to 2058 bp. The *BrBGLU* genes were heterogeneously distributed among all 10 chromosomes of *B. rapa*. Chromosome 5 contained the largest number of *BrBGLU* genes, comprising 15 members (23.4%), whereas chromosome 2 contained only one gene member. We also detected tandem arrays of the *BrBGLU* genes among the 10 *B. rapa* chromosomes. The tandem array was defined as 'multiple *BrBGLU* genes located in neighboring or the same intergenic region' [30]. Ten *BrBGLU* gene clusters were found on chromosomes A01, A03, A05, A07, and A09. Chromosome 5 contained the maximum number of clusters, comprising 11 *BrBGLU*s.

Table 1. Characteristics of the GH1 (Glycoside hydrolase family 1) gene family in *Brassica rapa*.

Locus ID	Gene Name	CDS Length (bp)	Protein Length (aa)	Chromosome	Gene Start	Gene End	gDNA Length (bp)	No. of Exons	Best Hit to Arabidopsis (BLASTP)		
									ID	Gene Name	E-Value
BraA01g012490.3C	*BrBGLU1*	1290	430	Chr 01	6,516,500	6,519,450	2950	11	*AT4G21760*	*BGLU47*	0
BraA01g029610.3C	*BrBGLU2*	1452	484	Chr 01	19,673,693	19,677,620	3927	12	*AT1G61820*	*BGLU46*	0
BraA01g029670.3C	*BrBGLU3*	1551	517	Chr 01	19,772,218	19,775,132	2914	12	*AT1G61810*	*BGLU45*	0
BraA01g032340.3C	*BrBGLU4*	873	291	Chr 01	22,083,906	22,086,749	2843	8	*AT1G52400*	*BGLU18*	6.38×10^{-85}
BraA01g034680.3C	*BrBGLU5*	1545	515	Chr 01	23,747,770	23,750,431	2661	12	*AT3G18080*	*BGLU44*	0
BraA01g034690.3C	*BrBGLU6*	1464	488	Chr 01	23,754,677	23,757,145	2468	10	*AT3G18070*	*BGLU43*	0
BraA01g040820.3C	*BrBGLU7*	1374	458	Chr 01	27,455,063	27,458,182	3119	12	*AT3G09260*	*BGLU23*	0
BraA01g041990.3C	*BrBGLU8*	1926	642	Chr 01	28,048,456	28,052,048	3592	9	*AT3G06510*	*BGLU48*	0
BraA01g043570.3C	*BrBGLU9*	1566	522	Chr 01	28,909,991	28,913,218	3227	13	*AT3G03640*	*BGLU25*	0
BraA02g023150.3C	*BrBGLU10*	1653	551	Chr 02	13,570,012	13,572,993	2981	13	*AT1G75940*	*BGLU20*	0
BraA03g011770.3C	*BrBGLU11*	894	298	Chr 03	5,059,601	5,062,343	2742	8	*AT1G45191*	*BGLU1*	2.53×10^{-60}
BraA03g011780.3C	*BrBGLU12*	1023	341	Chr 03	5,063,808	5,066,446	2638	12	*AT1G60090*	*BGLU4*	9.15×10^{-82}
BraA03g024570.3C	*BrBGLU13*	846	282	Chr 03	12,073,161	12,075,780	2619	10	*AT4G22100*	*BGLU3*	5.01×10^{-68}
BraA03g033950.3C	*BrBGLU14*	1398	466	Chr 03	16,798,347	16,801,182	2835	11	*AT3G09260*	*BGLU23*	0
BraA03g041420.3C	*BrBGLU15*	729	243	Chr 03	20,778,980	20,781,062	2082	5	*AT4G22100*	*BGLU3*	2.54×10^{-61}
BraA03g041430.3C	*BrBGLU16*	669	223	Chr 03	20,781,085	20,782,278	1193	7	*AT1G60090*	*BGLU4*	3.4×10^{-100}
BraA03g049730.3C	*BrBGLU17*	1563	521	Chr 03	25,428,252	25,430,947	2695	12	*AT4G21760*	*BGLU47*	0
BraA04g000610.3C	*BrBGLU18*	1431	477	Chr 04	408,734	411,303	2569	11	*AT4G27830*	*BGLU10*	0
BraA04g002030.3C	*BrBGLU19*	1497	499	Chr 04	1,226,615	1,230,317	3702	12	*AT3G60140*	*BGLU30*	0
BraA04g002040.3C	*BrBGLU20*	2058	686	Chr 04	1,238,401	1,245,729	7328	18	*AT3G60120*	*BGLU27*	2.2×10^{-149}
BraA04g010020.3C	*BrBGLU21*	1341	447	Chr 04	7,880,007	7,883,680	3673	13	*AT5G36890*	*BGLU42*	0
BraA04g020960.3C	*BrBGLU22*	891	297	Chr 04	15,776,965	15,780,643	3678	8	*AT5G44640*	*BGLU13*	1.5×10^{-105}
BraA04g023640.3C	*BrBGLU23*	1638	546	Chr 04	17,341,351	17,344,539	3188	9	*AT2G32860*	*BGLU33*	0
BraA04g031090.3C	*BrBGLU24*	1233	411	Chr 04	21,218,491	21,221,708	3217	12	*AT3G60120*	*BGLU27*	0
BraA04g031100.3C	*BrBGLU25*	1380	460	Chr 04	21,229,282	21,232,323	3041	11	*AT5G24550*	*BGLU32*	0
BraA04g031130.3C	*BrBGLU26*	267	89	Chr 04	21,248,071	21,248,622	551	4	*AT2G44450*	*BGLU15*	5.66×10^{-85}
BraA04g031140.3C	*BrBGLU27*	525	175	Chr 04	21,249,497	21,251,177	1680	6	*AT2G44450*	*BGLU15*	1.4×10^{-110}
BraA05g004330.3C	*BrBGLU28*	1623	541	Chr 05	2,194,953	2,197,671	2718	11	*AT3G60120*	*BGLU27*	0
BraA05g004340.3C	*BrBGLU29*	1527	509	Chr 05	2,201,548	2,211,165	9617	12	*AT2G44450*	*BGLU15*	0
BraA05g004350.3C	*BrBGLU30*	1518	506	Chr 05	2,216,705	2,220,674	3969	12	*AT5G44640*	*BGLU13*	0
BraA05g004360.3C	*BrBGLU31*	1326	442	Chr 05	2,223,901	2,227,525	3624	9	*AT2G44460*	*BGLU28*	2.7×10^{-131}
BraA05g004370.3C	*BrBGLU32*	1155	385	Chr 05	2,245,448	2,248,087	2639	7	*AT3G60140*	*BGLU30*	2.1×10^{-140}
BraA05g004380.3C	*BrBGLU33*	1281	427	Chr 05	2,255,905	2,258,985	3080	11	*AT5G24540*	*BGLU31*	5.1×10^{-145}
BraA05g004390.3C	*BrBGLU34*	1545	515	Chr 05	2,261,685	2,270,997	9312	11	*AT2G44490*	*BGLU26*	0

Table 1. *Cont.*

Locus ID	Gene Name	CDS Length (bp)	Protein Length (aa)	Chromosome	Gene Start	Gene End	gDNA Length (bp)	No. of Exons	Best Hit to Arabidopsis (BLASTP)		
									ID	Gene Name	E-Value
BraA05g012860.3C	*BrBGLU35*	1536	512	Chr 05	7,011,962	7,015,388	3426	11	*AT2G32860*	*BGLU33*	2.3×10^{-162}
BraA05g012870.3C	*BrBGLU36*	957	319	Chr 05	7,023,185	7,028,485	5300	8	*AT2G32860*	*BGLU33*	4.3×10^{-102}
BraA05g015060.3C	*BrBGLU37*	1461	487	Chr 05	8,601,511	8,604,284	2773	13	*AT5G36890*	*BGLU42*	0
BraA05g017770.3C	*BrBGLU38*	1278	426	Chr 05	10,758,114	10,760,718	2604	11	*AT1G52400*	*BGLU18*	0
BraA05g033960.3C	*BrBGLU39*	1434	478	Chr 05	24,329,347	24,333,054	3707	12	*AT4G27830*	*BGLU10*	0
BraA05g037140.3C	*BrBGLU40*	1332	444	Chr 05	25,685,745	25,688,976	3231	11	*AT3G09260*	*BGLU23*	0
BraA05g037150.3C	*BrBGLU41*	1374	458	Chr 05	25,691,345	25,694,889	3544	12	*AT3G09260*	*BGLU23*	0
BraA05g038920.3C	*BrBGLU42*	1782	594	Chr 05	26,547,600	26,550,524	2924	11	*AT3G06510*	*BGLU48*	0
BraA06g002000.3C	*BrBGLU43*	1347	449	Chr 06	1,220,707	1,223,925	3218	12	*AT1G52400*	*BGLU18*	1.5×10^{-173}
BraA06g011040.3C	*BrBGLU44*	1080	360	Chr 06	5,995,931	6,000,341	4410	8	*AT3G21370*	*BGLU19*	0
BraA06g024630.3C	*BrBGLU45*	1599	533	Chr 06	17,098,530	17,100,946	2416	4	*AT5G44640*	*BGLU13*	0
BraA06g038720.3C	*BrBGLU46*	312	104	Chr 06	25,758,774	25,759,164	390	2	*AT4G22100*	*BGLU3*	8×10^{-52}
BraA07g008030.3C	*BrBGLU47*	1434	478	Chr 07	8,145,282	8,147,911	2629	11	*AT1G60090*	*BGLU4*	0
BraA07g008050.3C	*BrBGLU48*	1428	476	Chr 07	8,161,734	8,164,666	2932	12	*AT4G22100*	*BGLU3*	0
BraA07g011940.3C	*BrBGLU49*	765	255	Chr 07	11,620,825	11,623,618	2793	8	*AT3G62750*	*BGLU8*	3.75×10^{-29}
BraA07g024150.3C	*BrBGLU50*	1545	515	Chr 07	18,998,283	19,001,907	3624	12	*AT3G60130*	*BGLU16*	0
BraA08g002600.3C	*BrBGLU51*	1515	505	Chr 08	1,915,015	1,917,839	2824	13	*AT1G47600*	*BGLU34*	0
BraA08g008860.3C	*BrBGLU52*	408	136	Chr 08	7,848,512	7,850,189	1677	4	*AT3G09260*	*BGLU23*	1.03×10^{-75}
BraA08g014870.3C	*BrBGLU53*	930	310	Chr 08	12,301,355	12,303,970	2615	7	*AT4G22100*	*BGLU3*	3.9×10^{-127}
BraA08g025770.3C	*BrBGLU54*	1506	502	Chr 08	18,552,796	18,556,470	3674	11	*AT1G26560*	*BGLU40*	0
BraA09g018020.3C	*BrBGLU55*	1524	508	Chr 09	11,385,273	11,386,948	1675	2	*AT5G44640*	*BGLU13*	0
BraA09g038410.3C	*BrBGLU56*	1527	509	Chr 09	30,292,880	30,295,941	3061	11	*AT1G26560*	*BGLU40*	0
BraA09g049950.3C	*BrBGLU57*	1542	514	Chr 09	37,157,612	37,160,275	2663	10	*AT3G60120*	*BGLU27*	0
BraA09g049960.3C	*BrBGLU58*	1248	416	Chr 09	37,164,182	37,167,906	3724	10	*AT3G60130*	*BGLU16*	2.2×10^{-165}
BraA09g049970.3C	*BrBGLU59*	1047	349	Chr 09	37,169,340	37,173,150	3810	8	*AT3G60130*	*BGLU16*	3.4×10^{-165}
BraA09g049980.3C	*BrBGLU60*	1326	442	Chr 09	37,178,403	37,186,233	7830	11	*AT3G60140*	*BGLU30*	2.6×10^{-179}
BraA09g052040.3C	*BrBGLU61*	1461	487	Chr 09	38,112,498	38,115,245	2747	11	*AT4G27830*	*BGLU10*	0
BraA09g052050.3C	*BrBGLU62*	1176	392	Chr 09	38,116,332	38,118,944	2612	11	*AT4G27830*	*BGLU10*	9.8×10^{-140}
BraA10g001490.3C	*BrBGLU63*	1281	427	Chr 10	776,354	778,932	2578	11	*AT1G02850*	*BGLU11*	0
BraA10g012660.3C	*BrBGLU64*	1569	523	Chr 10	10,414,966	10,417,449	2483	11	*AT5G54570*	*BGLU41*	0

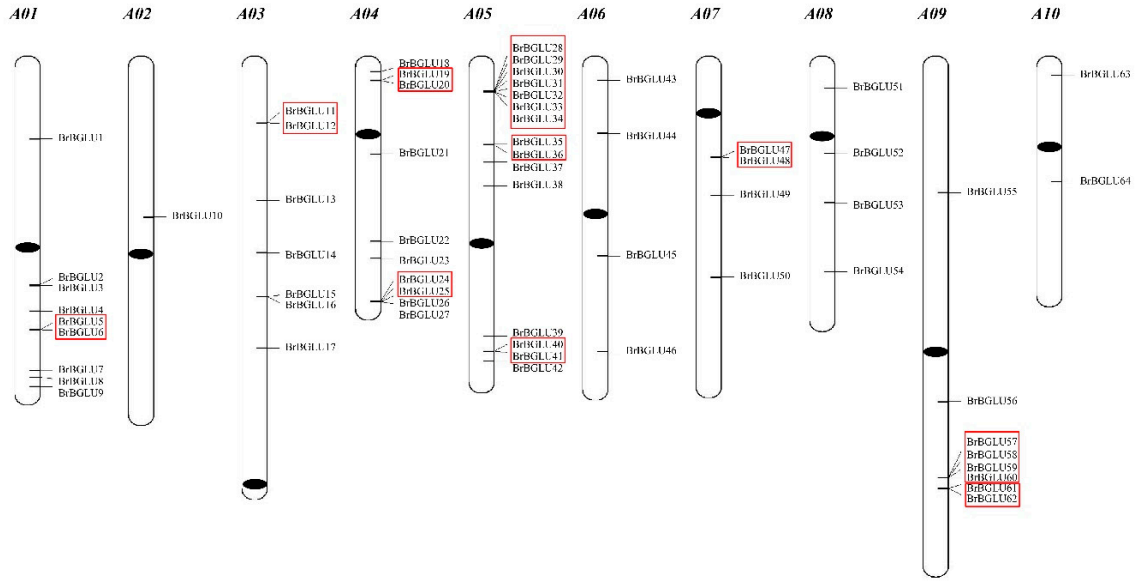


Figure 1. Chromosomal distribution of the 64 *BrBGLU* genes identified in this study. The chromosome number is indicated above each chromosome. Ten clusters of *BrBGLUs* are indicated in red boxes. Black ovals on each chromosome represent the centromeric regions.

2.2. Phylogenetic and Gene Structure Analysis of BrBGLUs

To understand the evolutionary relationship of the *BrBGLU* genes, phylogenetic analysis of the *BrBGLU* and *AtBGULU* genes was conducted. To obtain *AtBGLU*s, HMM searching was performed by using all of the putative protein sequences of the *Arabidopsis* genome (ARAPORT11, https://www.araport.org) as queries. A total of 48 *AtBGLU* genes were obtained, which agrees with the results of a previous study [6]. The 64 BrBGLUs and 48 AtBGLUs protein sequences were aligned using ClustaX2 [31]. An unrooted phylogenetic tree was constructed for the 64 BrBGLUs and 48 AtBGLUs, using the NJ method in MEGA6 with a Poisson model. All BGLU proteins were classified into 10 distinct subgroups, namely, BGLU-a to BGLU-j (Figure 2). The results of the phylogenetic analysis were relatively similar to the findings of a previous study using *Arabidopsis* [6], with a few exceptions. All *B. rapa* and *Arabidopsis* proteins are grouped into 10 subgroups, whereas *Arabidopsis* subgroups 8 and 9 were combined into a subgroup GH1-c in our analysis. In addition, AtBGLU48 (SENSITIVE TO FREEZING 2, SFR2), which belongs to a distinct lineage from 10 subgroups in a previous study [6,32], was incorporated into the GH1-j subgroup, together with BrBGLU8 and BrBGLU42 (Figure 2).

Phylogenetic analysis generated an interesting finding, that the clustering or groupings of genes were related to the chromosomal locus or function. Based on the functions of the *AtBGLUs*, flavonol accumulation (*AtBGLU1-6*) and anthocyanin glucosyltransferase (*AtBGLU7-11*)-related genes were highlighted by subgroup GH1-a [10–12]; flavonoid utilization-related *AtBGLUs* (*AtBGLU12-17*) were represented by the GH1-e subgroup [10,13]; myrosinase-encoding *AtBGLUs* (*AtBGLU34-39*) belonged to the GH1-d subgroup [14–16], and scopolin hydrolysis-related *AtBGLUs* (*AtBGLU21*-23) were grouped into GH1-i [21,22]. Most of the genes within the same clusters on a chromosome were grouped into the same subfamily, which is similar to the findings using *Arabidopsis*, i.e., *BrBGLU5/6*, *BrBGLU11/12*, *BrBGLU31/32/33*, *BrBGLU40/41*, *BrBGLU58/59*, and *BrBGLU61/62*. This clustering indicates that the *BGLU* genes may have evolved from an ancestral gene via gene duplication. However, *BrBGLU51* was grouped with six *AtBGLUs* (*AtBGLU34/35/36/37/38/39*) in the GH1-d subgroup, indicating the possible loss of some *BGLU* genes in *B. rapa*.

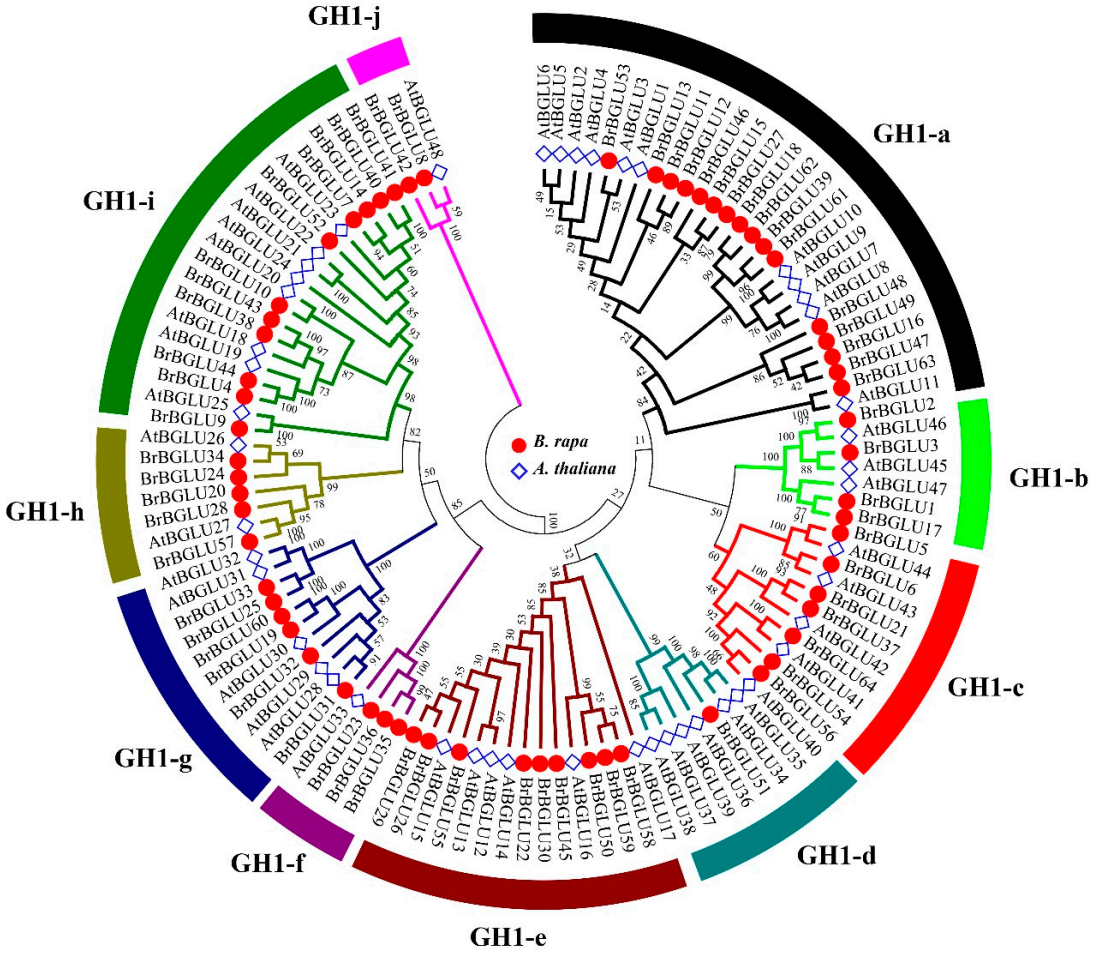


Figure 2. Phylogenetic reconstruction of GH1 genes of *Arabidopsis* and *Brassica rapa*. Multiple sequence alignment of GH1 proteins was performed using ClustalX2 with default parameters. The unrooted phylogenetic tree was constructed by MEGA 6 with the neighbor-joining (NJ) methods using the following parameters: bootstrap values (1,000 replicates) and Poisson model. The tree is divided into 11 phylogenetic subgroups, designated as GH1-a to GH1-k. Members of *Arabidopsis* and *B. rapa* are denoted by blue squares and red circles.

Gene structure was commonly diversified during the evolution of the large number of gene families. To expand our knowledge of BrBGLUs in relation to evolution and functional diversification, the gene structures of the *BrBGLU*s were analyzed on the basis of exon–intron organization, using GSDS 2.0 [33]. The *BrBGLU*s exhibited 12 distinct exon–intron organization patterns, and the most common organization was 11 exons separated by 10 introns, presenting 19 members (Table 1 and Figure 3). Most genes contained more than two introns, except for *BrBGLU46* and *BrBGLU55*, indicating the possible occurrence of alternative splicing during gene expression. The *AtBGLU*s exhibited 10 distinct exon–intron organization patterns, and the pattern with 13 exons was the most common [6]. This analysis is consistent with *Arabidopsis* and rice, where the intron size and number of the *BGLU*s genes are highly variable [5,6].

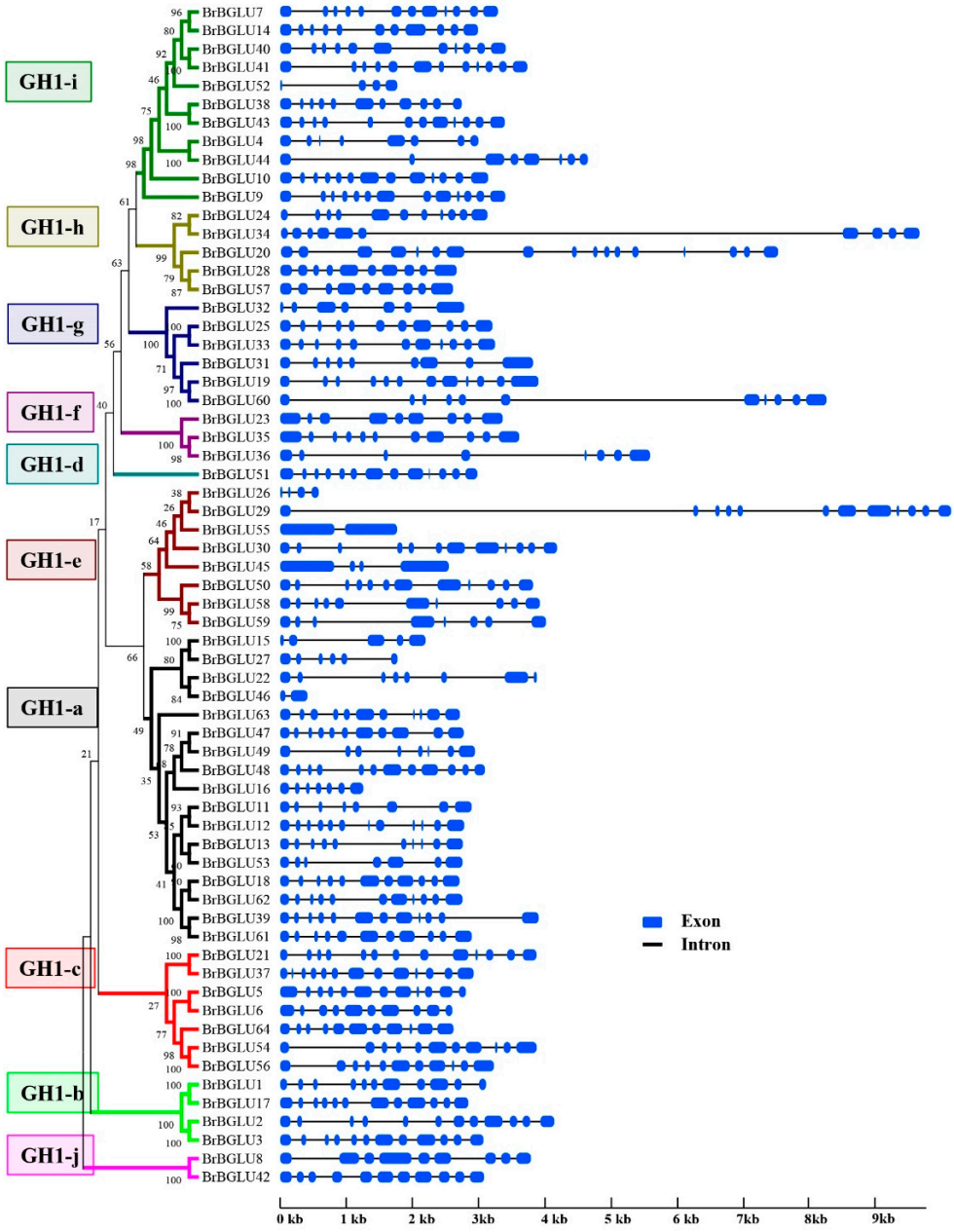


Figure 3. Exon–intron organization of *BrBGLUs* in different subgroups. Exons and introns are represented by blue boxes and black lines, respectively. The phylogenetic tree of each subfamily was constructed using MEGA6, as described in Figure 1.

2.3. Identification of BrBGLU Genes Involved in Pollen Development

Rice *TDR* (*Tapetum Degeneration Retardation*) mutant alters *BGLU1* expression with flower specificity [28], and *BGLU1* and *BGLU13* are found to be related to male organ development in *Calamus palustris* [34]. These previous reports lead to a hypothesis that *BrBGLUs* are involved in pollen development. To test this hypothesis, the previously published microarray data relating to male sterility in *B. rapa* [29] were re-annotated, using the improved *B. rapa* genome (version 3.0) [35] and analyzed based on pollen development (Table S1). A total of 36 *BrBGLUs*, represented by 88 probes (or 88 ESTs) showed significant hybridization values, of which 12 *BrBGLUs* showed over two-fold change in expression levels between fertile and sterile floral buds: six members were upregulated, and members were downregulated in fertile buds. Among these genes, four upregulated genes (*BrBGLU10/AtBGLU20*, *BrBGLU15/AtBGLU3*, *BrBGLU16/AtBGLU4*, and *BrBGLU64/AtBGLU41*) and

two downregulated genes (*BrBGLU2/AtBGLU46* and *BrBGLU19/AtBGLU30*) were described as good candidates that were associated with pollen development. The function of all four upregulated genes has not been known up to now, but at least three, *BrBGLU10*, *BrBGLU15*, and *BrBGLU64* appeared to be related to pollen wall development. In particular, we further analyzed *BrBGLU10/AtBGLU20*, as these showed hundred-fold changes between fertile and sterile buds.

2.4. Analysis of the Putative Functions of BrBGLU10/AtBGLU20 in Pollen Development

To gain more insights into the functions of the *BrBGLU*s during pollen development, *BrBGLU10*, which was highly and specifically expressed in fertile buds, was selected for further analysis. *AtBGLU20*, the *Arabidopsis* ortholog of *BrBGLU10*, was initially named as *ATA27*, which is one of the *A. thaliana* anther-specific expressed genes [36]. To confirm the expression patterns of *BrBGLU10* and *AtBGLU20*, RT-PCR was conducted (Figure 4A,B). The expression level of *BrBGLU10* was specifically detected at the F1–F3 stages, with highest levels at the F2 stage, representing the tetrad stage, and *AtBGLU20* was specifically expressed before floral stage 12. The RT-PCR results might imply its important role in pollen development.

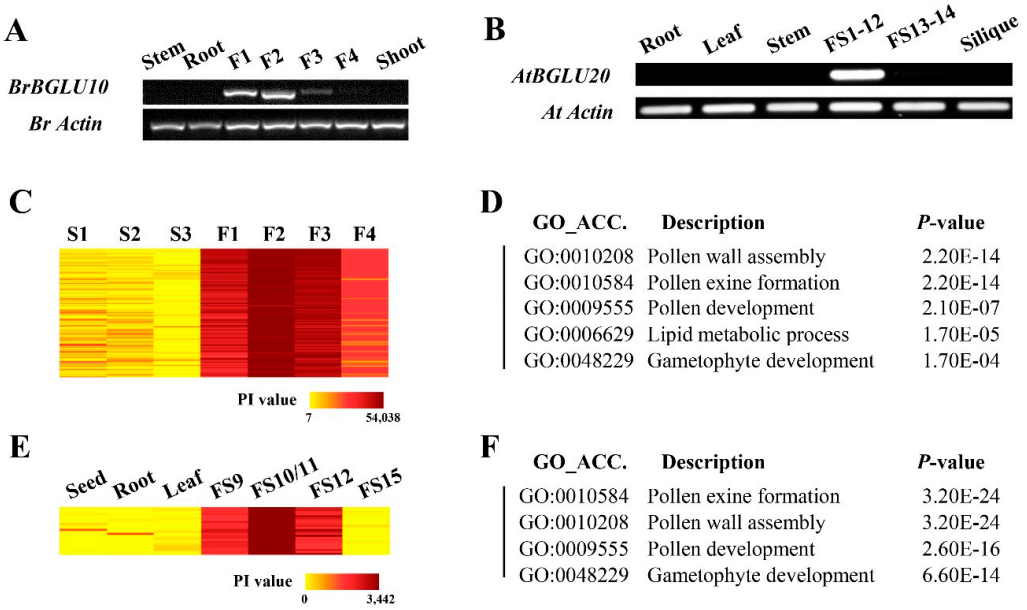


Figure 4. Analysis of expression of *BrBGLU10* and *AtBGLU20*, and Gene Ontology (GO) enrichment of co-expressed genes. **A**, Expression of BrBGLU10 in different tissues and floral bud stages in *B. rapa*. **B**, Expression of AtBGLU20 in different tissues and floral bud stages in *Arabidopsis*. **C**, Expression patterns of *BrBGLU10* and its co-expressed genes in sterile and fertile *B. rapa* floral buds, based on previously published microarray data [29]. **D**, GO enrichment analysis of genes co-expressed with *BrBGLU10*. **E**, Expression pattern of *AtBGLU20* and its co-expressed genes in various tissues of *Arabidopsis*, which was performed using the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). **F**, GO enrichment analysis of genes co-expressed with *AtBGLU20*. S1–S3 represent the floral buds from male-sterile *B. rapa*. S1, before the tetrad stage. S2, after the tetrad stage. S3, containing aberrant pollen grains. F1–F4 indicate fertile *B. rapa* floral buds before the tetrad stage (F1), at the tetrad stage (F2), after the tetrad stage, but before containing mature pollen (F3), and containing mature pollen (F4). For *Arabidopsis*, FS1–12, flower stage 1 to stage 12; FS13–14, flower stage 13 to stage 14. PI, probe intensity.

To demonstrate similar or conserved functions between *BrBLU10* and *AtBGLU20*, we isolated the co-expressed genes of *BrBGLU10*, using microarray data [29] and *AtBGLU20* from the *Arabidopsis* eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) [37]. With the Pearson's correlation coefficient (PCC) value above 0.90, 183 probes (107 genes) and 25 genes were determined to be

co-expressed with *BrBGLU10* and *AtBGLU20*, respectively (Figure 4C, E; Tables S2–S3). *BrBGLU10* and its co-expressers were upregulated at the fertile floral bud stage, and the highest expression level was detected at the F2 stage (Figure 4C), suggesting that *BrBGLU10* plays a role during pollen development, especially from the tetrad stage to that before the mature pollen stage. In *Arabidopsis*, flower and stamen development processes were divided into 14 stages and 12 stages, respectively [37–39]. *AtBGLU20* and its co-expressers were represented by a high probe intensity (PI) value at flower stages (FS) 9 to 12, indicating that *AtBGLU20* plays a role in *Arabidopsis* pollen development (Figure 4E). We also conducted Gene Ontology (GO) enrichment analysis to provide more information on the function of *BrBGLU10* and *AtBGLU20* (Figure 4D,F). The results showed that genes involved in pollen exine formation and pollen wall assembly were highly over-represented among genes co-expressed with *BrBGLU10* and *AtBGLU20.* Taken together, our analysis indicated that *BrBGLU10* and *AtBGLU20* may be required for pollen development in both *B. rapa* and *Arabidopsis*.

To validate *AtBGLU20* function in the pollen development, we generated knockdown mutants of *AtBGLU20* by introducing antisense constructs under the control of the *CaMV35S* promoter (Figure 5A). After screening, four independent knockdown lines were obtained with expression levels ranging from 55% to 85% (Figure 5B). However, the *AtBGLU20* downregulated plants showed normal vegetative growth based on morphology (Figure 5C), but produced defective pollen grains relative to the wild-type plants (Figure 5D). These results indicated that normal pollen development in *Arabidopsis* requires sufficient amounts of *AtBGLU20*. All data obtained from gene expression, co-expression analysis, and transgenesis led to the conclusion that *AtBGLU20* and *BrBGLU10* may have indispensable functions in pollen development in both *Arabidopsis* and *B. rapa*, respectively.

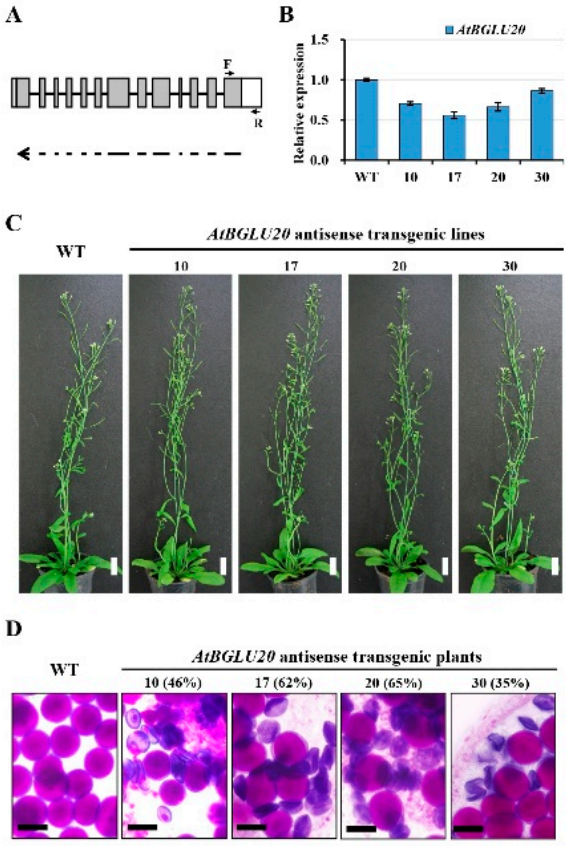


Figure 5. Analysis of WT and *AtBGLU20* antisense knockdown mutant *Arabidopsis* plants. **A**. Schematic representation of the *AtBGLU20* gene structure and DNA fragment regions for antisense constructs. The white box indicates the UTR region; gray boxes are exons; lines represent introns. The single arrow indicates the antisense orientation of the fragments in the constructs. F and R indicate the primer

positions used in qRT-PCR analysis. **B**, Analysis of the expression levels of *AtBGLU20*. Expression was normalized to that of *At*ACT7, and represented relative to the expression levels of the WT. Error bars represent the SD of three biological replicates. **C,** Morphologies of wild-type Arabidopsis plants and *AtBGLU20* knockdown transgenic plants, which showed no obvious differences in vegetative growth. Bar = 20 mm. **D**, Mature pollen grain of WT and *AtBGLU20* transgenic plants stained with modified Alexander solution (Peterson et al., 2010). The well-developed pollen grains were stained red. Bar = 20 μm. WT, wild-type. 10, 17, 20, and 30 indicate four independent transgenic lines. The number in the parentheses indicate the percentages of defective pollen grains.

3. Discussion

3.1. Identification and Analysis of BrBGLUs

GH1 family genes play an important role in regulating abiotic and biotic stress responses, as well as various developmental processes in plants [9,12,14,18,23,40]. Based on the results of an increase in the number of whole genome sequences from a large number of species, genome-wide analysis of various gene families has been extensively performed. However, genome-wide identification and characterization of the GH1 gene family has only been reported in a few plant species, and there is no information on *Brassica* species, which are important crops for production of functional foods, as well as health-promoting compounds. In this study, the isolation of *BrBGLU*s from *B. rapa* genome (Figure 1), the distribution of *BrBGLU* genes on chromosomes (Figure 1), phylogenetic analysis (Figure 2), and exon–intron structures (Figure 3) provides substantial information on the functions and roles of these genes.

Compared with the 49 *AtBGLUs* and 37 *OsBGLUs* in Arabidopsis and Rice, respectively [9], 64 *BrBGLUs* were isolated from the *B. rapa* genome, which is the largest number so far that has been reported in plants (Figure 1). The high number of *BGLU* family members in *B. rapa* could be related to the genome triplication event in this lineage [41]. To adapt different new functions that are suitable for changes in the environment, gene structure was commonly diversified during the evolution of multigene families [42]. For *BGLUs*, 13 exon–12 intron organization was considered as the ancestral gene structure, with the loss of certain introns leading to other gene structures [6]. The exons present in *BrBGLUs* varied from 2 to 13, and the most common organization was 11 exons (Figure 3). The introns in Arabidopsis vary from 0 to 13 [6]. This results suggested that little diversity exists in the gene structure of *BrBGLUs* when compared to *AtBGLUs*.

*BrBGLU*s may have originated from *Arabidopsis*, although duplication, gene loss, and functional diversification may have also occurred. This is supported by the fact that *BGLU*s from both species could be grouped into 10 subfamilies, with tandem arrays, as defined by Singh et al., 2013 [39], although some families were re-grouped or diverged into other subgroups. Figure 2 shows that *AtBGLU* subfamilies 8 and 9 [6] were incorporated into one *B. rapa* subfamily, GH1-c, and *BrBGLU51* is composed of GH1-d with six *AtBGLUs* (*AtBGLU34/35/36/37/38/39*), indicating the loss of some *BGLU*s in *B. rapa.* This phenomenon may result from the rapid evolution of genes similar to that previously observed between *Arabidopsis* and rice [5]. One more interesting finding was that *AtBGLU48* (*SFR2*) was incorporated into the GH1-j subgroup, with *BrBGLU8* and *BrBGLU42* (Figure 2). *AtBGLU42* is a β-glucosidase, but it is divergent from all other *AtBGLUs*, and more similar to several β-glycosidases from thermophilic archea and bacteria [32]. *SFR2* is involved in the lipid remodeling of the outer chloroplast membrane during freezing tolerance [43,44]. Because two *BrBGLU*s in the GH1-j subgroup had identities between 85% and 87% with *AtBGLU2*, *Brassica* genes may have a similar function of freezing tolerance as that in *AtSFR2*, although this requires further investigation.

On the basis of *Arabidopsis* study, most subfamilies of BGLUs in Figure 2 may be associated with specific functions: GH1-a for flavonoid and anthocyanin metabolism, GH1-e for flavonoid utilization, GH1-d for myrosinases, and GH1-i for scopolin hydrolysis. At least 12 genes are known to be involved in flavonoid metabolism in GH1-a: *AtBGLU1-6* for flavonol accumulation [10,11], *AtBGLU7-11* as anthocyanin glucosyltransferases [10–12], and *AtBGLU15* for flavonol bisglycoside

catabolism under abiotic stress [13]. *AtBGLU12-17* in the GH1-e subgroup code for flavonoid-utilizing BGLUs in legumes [10]. An examination of the functions of *BrBGLUs* that are clustered with *AtBGLUs* in subgroups GH1a and GH1-e may provide information and understanding into the regulation of flavonoid biosynthesis in *Brassica* species.

Several subfamilies may be related to abiotic and biotic stress resistance, such as GH1-b, GH1-c, GH1-d, GH1-f, and GH1-i. Myrosinases hydrolyze glucosinolates into active forms that are involved in plant defense against herbivory and pathogens, and in human health promotion [45–48]. *AtBGLU26* and *AtBGLU34-39* function as myrosinases [14–16]. Except for *AtBGLU26* (GH1–h), most genes belong to the GH1-d subgroup (Figure 2). Understanding myrosinase function in Brassicaceae, which is rich in glucosinolates, may provide an excellent strategy for breeding health-promoting *Brassica* crops [49,50]. ABA also functions in stress responses, including drought stress. *AtBGLU18* [17] and *AtBGLU33* [18] hydrolyze glucose-conjugated ABA, thereby increasing ABA levels and inducing ABA responses such as drought tolerance. However, these two proteins are separated into two subfamilies, implying the presence of more BGLUs for the regulation of ABA levels. Scopolin is one of the coumarins produced in roots [51], and it plays a role in the defense against pathogen attack and abiotic stresses [19,20]. Three β-glucosidases that hydrolyze scopolin and their encoding genes (*AtBGLU21, 22* and *23*) have been characterized [21,22]. The GH1-i subfamily includes these three genes and 11 *BrBGLU*s, which should be examined in relation to scopolin production. The GH1-b subfamily includes two monolignol glucoside hydrolases (AtBGLU45 and AtBGLU46) that control lignin content [23]. Because *OsBGLU14*, *16*, and *18* are involved in lignin biosynthesis with monolignol β-glucosidase activity and compensate for the *Arabidopsis bglu45* mutant [52], BrBGLUs in this subfamily may play similar roles. *AtBGLU42* in GH1-c is involved in the induction of systemic resistance to bacterial disease, and the release of iron-mobilizing phenolic metabolites under iron deficiency [24]. Several genes in this subfamily would thus be expected to contribute to eliciting defense responses. All of this information may contribute to future research directions in relation to *BrBGLUs*.

3.2. The Potential Functions of BrBGLUs During Pollen Development

Previous studies on rice and other plant species have indicated that β-glucosidases play roles in pollen development [34,36,53]. To identify the BrBGLUs responsible for pollen development, the previously published microarray data relating to male sterility in *B. rapa* were re-annotated and re-analyzed. Among the 36 *BrBGLU*s, 12 *BrBGLUs* showed over a two-fold change between fertile and sterile floral buds (Table S1). However, six genes (four upregulated and two downregulated genes) were more extensively studied in terms of their role in pollen development. We selected one *BrBGLU10* for investigation, the homolog of *AtBGLU20*, which showed hundreds-fold changes in its expression.

We examined the expression levels of *BrBGLU10/AtBGLU20* and analyzed the co-expressed genes in both *B. rapa* and *Arabidopsis* (Figure 4). An assessment of expression levels strongly suggests that *BrBGLU10/AtBGLU20* are involved in pollen development. The cellular contents from the degeneration of the tapetum supports pollen wall formation and subsequent pollen release [39]. Mutations in polysaccharide metabolism-related genes lead to defective pollen wall formation [26]. Glycoside hydrolase has been reported to be involved in the cell wall polysaccharide degradation [27]. The expression patterns of *BrBGLU10* and *AtBGLU20* suggest that they might play a role from the tetrad stage to mature pollen grains (Figure 4A, B), which corresponding to the tapetum degradation stage [29,39]. Co-expression analysis is a valuable approach for classifying and visualizing transcriptomic data to identify genes with similar cellular functions and regulatory pathways [54–56], although this is not always the case [57,58]. In plants, co-expression analysis under various experimental conditions has been used for predicting gene function [55,59]. Figure 4C,D shows that this gene possibly regulates pollen development. In particular, GO annotation of co-expressed genes reflects that pollen wall and exine formation are influenced by *BrBGLU10/AtBGLU20*, indicating that the hydrolysis of glucose moieties is necessary for proper pollen development.

Because BrBGLU10 had a high sequence identity with AtBGLU20 (87% at the nucleotide level and 84% at the amino acid sequence level), both genes may thus have similar functions. Therefore, knocking down *AtBGLU20* may provide direct evidence for its function in pollen development. Figure 5 shows that the suppression of *AtBGLU20* expression had no effect on plant growth and development, although this aborted pollen production. This result implies that BrBGLU10/AtBGLU20 are critical to pollen grain development.

4. Materials and Methods

4.1. Plant Materials and Growth Conditions

Seeds of *B. rapa* subsp. *pekinensis* (Chiifu) were germinated in Petri dishes in the dark at 23 °C for two days, then the germinated seeds were transferred to a 4 °C growth chamber with 16 h of light for 25 days to induce vernalization. After vernalization, the seedlings were transplanted into 15 cm × 15 cm × 18 cm pots containing potting soil and grown in a 23 °C growth chamber with 16 h of light. The floral buds were collected from 10 plants with three biological replicates, as previously described [29], and stored at −70 °C until use. Root and shoot tissues were collected from three-week-old seedlings without vernalization. Stem tissue was sampled from the plants one week after bolting.

A. thaliana (L.) Heynh var. Columbia (Col-0) plants were grown under 140 μmol/m^2/s light intensity at 23 ± 1 °C with a long day cycle with 16 h of light for plant transformation. Seeds were sown in 55 mm × 55 mm pots in potting soil, stratified for three days at 4 °C, and then transferred to the growth room. The plants were then kept under a transparent polythene lid for one week to increase humidity and support equal germination. The plate-cultured seeds were sterilized with 30% bleach and 0.1% Triton X-100 (Sigma, St. Louis, MO, USA), stratified for three days at 4 °C, and sown in Petri dishes with dimensions of 100 mm × 100 mm × 20 mm. The dishes contained half-strength MS media (Duchefa Biochemie, Netherlands) supplemented with 0.8% phytoagar and 1% sucrose.

4.2. Antisense Constructs and Plant Transformation

The full-length coding sequence of *AtBGLU20* was cloned from first-strand complementary DNA (cDNA), using the primers BGLU20F (Table S4). Then, the fragments were inserted into T&A cloning vectors (RBC T&A cloning kit, Real Biotech Corporation, Taiwan). After confirmation of the *AtBGLU20* sequence in the T&A vector by sequencing, the fragment was cloned into pCambina 3300-35S binary vectors and used in plant transformation. Col-0 were used for transformation with *Agrobacterium tumefaciens* GV3101 carrying the above binary plasmid using the floral dip method [60]. The transformants were selected on plates containing 25 mg/mL glufosinate in MS medium (Sigma, St. Louis, MO, USA), and also confirmed by genomic DNA PCR analysis.

4.3. Reverse Transcription PCR and qRT-PCR

Total RNA (1 μg) from each sample was used in reverse transcription. First-strand cDNA was synthesized with a PrimeScript™ RT reagent kit with a gDNA Eraser kit (TaKaRa, Japan). The concentration of the synthesized cDNA was determined, and the cDNA was diluted to 20 ng/μL for PCR analysis. Semi-RT-PCR was performed, which consisted of denaturation at 94 °C 5 min; followed by 25 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 60 s. The qRT-PCR conditions were pre-denaturation at 95 °C for 30 s; followed by 30 cycles of 95 °C for 5 s, 60 °C for 20 s, and 72 °C for 15 s. All primer sequences used in this study are listed in Table S4. The semi-RT-PCR products were separated on 1.5% agarose gels, and stained with ethidium bromide. The qRT-PCR results were analyzed using the $2^{-\Delta\Delta C_T}$ method, with three biological replicates.

4.4. Pollen Viability

For pollen viability and pollen developmental progression, flowers collected from Col-0 and *AtBGLU20* antisense transgenic plants were fixed in Carnoy's fixative (6:3:1 alcohol:chloroform:acetic acid) for 2 h. Then, the anthers were detected and stained with a solution of Malachite green, acid fuchsin, and Orange G for approximately 12 h, as previously described [61].

4.5. Identification of BrBGLUs and Phylogenetic Tree Construction

The protein sequence of 48 BGLU members were downloaded from TAIR (http://www.arabidopsis.org/tools/bulk/sequences/index.jsp) [6]. All putative protein sequences of *B. rapa* (version 3.0) were downloaded from BRAD (http://brassicadb.org/brad/index.php) [35] and used as queries to search against the Hidden Markov Model (HMM) profile (Version 3.1b2) with the Pfam HMM library (Pfam 32.0) [62]. A total of 64 protein sequences with PF00232 (E value below $1E^{-5}$) were obtained, and these sequences were considered as *BrBGLUs* candidates and used for further analysis. Multiple sequence alignment of full-length BGLU proteins and phylogenetic tree construction were conducted using ClustalX2 [31]. The phylogenetic tree was generated by the MEGA6 program, using the neighbor-joining method with the 'pairwise deletion' option and 'Poisson correction' model, with a bootstrap test of 1000 replications [63].

4.6. Chromosomal Location, Nomenclature, and Gene Duplication of BrBGLUs

The position of each *BrBGLU* on *B. rapa* chromosomes was identified from BRAD (http://brassicadb.org/brad/index.php). For nomenclature, the '*Br*' for *B. rapa* was added, followed by BGLU, and numbered according to its position from top to bottom on *B. rapa* chromosomes 1–10. MCScanX software was used to search potentially duplicated BrBGLUs [64]. All of the putative protein sequences of *B. rapa* (version 3.0) were compared with themselves by BLASTP, with a tabular format and an e-value of $< 10^{-5}$. Then, tandem, segmental, and dispersed duplications were identified using MCScanX, using default criteria.

4.7. Co-Expression and Gene Ontology Enrichment Analysis

AtBGLU20 was used as bait gene for genome-wide co-expression analysis to identify genes of similar function from Expression Angler [65]. *BrBGLU10* was represented by two EST probes *Brapa_ESTC004210* and *Brapa_ESTC007739*, which were used as bait for co-expression analysis. A cutoff threshold of 0.90 for the Pearson correlation coefficient was used. The expression pattern analysis was performed using the Arabidopsis eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) [37]. Clustering analysis for categorization was performed with the TIGR Multi-Experiment Viewer (http://www.tm4.org/mev.html). GO enrichment analysis was performed using agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) [66].

4.8. Microarray Analysis

To analyze the gene expression patterns of *BrBGLUs* in *B. rapa* during pollen development, the previously published microarray data relating to male sterility analysis were downloaded from NCBI's Gene Expression Omnibus (GSE47665) [29]. The microarray data were re-annotated using BLASTX by comparing with the newly improved *B. rapa* reference genome sequence (version 3.0) [35].

5. Conclusions

In conclusion, 64 *BrBGLUs* have been identified in *B. rapa* genome, which were classified into 10 subgroups with *Arabidopsis* counterparts, and the GH1-i subgroup included putative pollen development-related *BrBGLU10*. Base on its known function in *Arabidopsis*, BrBGLUs may participate in various defense responses against biotic and abiotic stresses, flavonoid metabolism, and pollen

development. This study has provided valuable information for a better understanding of BGLUs, and for their biotechnological application to crops.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/7/1663/s1.

Author Contributions: Designed the research scheme: X.D. and Y.H. Performed the experiments: X.D. and Y.J. Analyzed the data: X.D., Y.J., and Y.H. Wrote the manuscript: X.D. and Y.H. All authors read and approved the final manuscript.

Funding: This research was funded by the National Science Foundation of China, 31601771 and the Applied Basic Research Project of Yunnan, 2017FB056.

Acknowledgments: We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

GH1	Glycoside hydrolase family 1
BGLUs	β-glycosidase genes
BrBGLUs	*Brassica rapa* β-glycosidase genes
ABA	abscisic acid
OsTDR	Tapetum Degeneration Retardation
HMM	Hidden Markov Model
GO	Gene Ontology
CDS	coding sequence

References

1. Chandrasekar, B.; Colby, T.; Emran Khan Emon, A.; Jiang, J.; Hong, T.N.; Villamor, J.G.; Harzen, A.; Overkleeft, H.S.; van der Hoorn, R.A. Broad-range glycosidase activity profiling. *Mol. Cell. Proteomics* **2014**, *13*, 2787–2800. [CrossRef]
2. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The carbohydrate-active enzymes database (cazy): An expert resource for glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [CrossRef]
3. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (cazy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]
4. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **1991**, *280*, 309–316. [CrossRef] [PubMed]
5. Opassiri, R.; Pomthong, B.; Onkoksoong, T.; Akiyama, T.; Esen, A.; Ketudat Cairns, J.R. Analysis of rice glycosyl hydrolase family 1 and expression of os4bglu12 beta-glucosidase. *BMC Plant Biol.* **2006**, *6*, 33. [CrossRef] [PubMed]
6. Xu, Z.; Escamilla-Trevino, L.; Zeng, L.; Lalgondar, M.; Bevan, D.; Winkel, B.; Mohamed, A.; Cheng, C.L.; Shih, M.C.; Poulton, J.; et al. Functional genomic analysis of arabidopsis thaliana glycoside hydrolase family 1. *Plant Mol. Biol.* **2004**, *55*, 343–367. [CrossRef]
7. Gomez-Anduro, G.; Ceniceros-Ojeda, E.A.; Casados-Vazquez, L.E.; Bencivenni, C.; Sierra-Beltran, A.; Murillo-Amador, B.; Tiessen, A. Genome-wide analysis of the beta-glucosidase gene family in maize (zea mays l. Var b73). *Plant Mol. Biol.* **2011**, *77*, 159–183. [CrossRef] [PubMed]
8. Zhao, L.; Liu, T.; An, X.; Gu, R. Evolution and expression analysis of the β-glucosidase (glu) encoding gene subfamily in maize. *Genes Genom.* **2012**, *34*, 179–187. [CrossRef]
9. Cao, Y.Y.; Yang, J.F.; Liu, T.Y.; Su, Z.F.; Zhu, F.Y.; Chen, M.X.; Fan, T.; Ye, N.H.; Feng, Z.; Wang, L.J.; et al. A phylogenetically informed comparison of gh1 hydrolases between arabidopsis and rice response to stressors. *Front Plant Sci.* **2017**, *8*, 350. [CrossRef] [PubMed]
10. Roepke, J.; Bozzo, G.G. Arabidopsis thaliana β-glucosidase bglu15 attacks flavonol 3-*O*-β-glucoside-7-*O*-α-rhamnosides. *Phytochemistry* **2015**, *109*, 14–24. [CrossRef] [PubMed]

11. Ishihara, H.; Tohge, T.; Viehöver, P.; Fernie, A.R.; Weisshaar, B.; Stracke, R. Natural variation in flavonol accumulation in arabidopsis is determined by the flavonol glucosyltransferase bglu6. *J. Exp. Bot.* **2016**, *67*, 1505–1517. [CrossRef] [PubMed]
12. Miyahara, T.; Sakiyama, R.; Ozeki, Y.; Sasaki, N. Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in arabidopsis. *J. Plant Physiol.* **2013**, *170*, 619–624. [CrossRef] [PubMed]
13. Roepke, J.; Gordon, H.O.W.; Neil, K.J.A.; Gidda, S.; Mullen, R.T.; Freixas Coutin, J.A.; Bray-Stone, D.; Bozzo, G.G. An apoplastic β-glucosidase is essential for the degradation of flavonol 3-*O*-β-glucoside-7-*O*-α-rhamnosides in arabidopsis. *Plant Cell Physiol.* **2017**, *58*, 1030–1047. [CrossRef] [PubMed]
14. Barth, C.; Jander, G. Arabidopsis myrosinases tgg1 and tgg2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* **2006**, *46*, 549–562. [CrossRef] [PubMed]
15. Wittstock, U.; Burow, M. Glucosinolate breakdown in arabidopsis: Mechanism, regulation and biological significance. *Arabidopsis Book* **2010**, *8*, e0134. [CrossRef] [PubMed]
16. Zhou, C.; Tokuhisa, J.G.; Bevan, D.R.; Esen, A. Properties of β-thioglucoside hydrolases (tgg1 and tgg2) from leaves of arabidopsis thaliana. *Plant Sci.* **2012**, *191–192*, 82–92. [CrossRef] [PubMed]
17. Lee, K.H.; Hai, L.P.; Kim, H.Y.; Sang, M.C.; Fan, J.; Hartung, W.; Hwang, I.; Kwak, J.M.; Lee, I.J.; Hwang, I. Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell* **2006**, *126*, 1109–1120. [CrossRef] [PubMed]
18. Xu, Z.Y.; Lee, K.H.; Dong, T.; Jeong, J.C.; Jin, J.B.; Kanno, Y.; Kim, D.H.; Kim, S.Y.; Seo, M.; Bressan, R.A.; et al. A vacuolar beta-glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in arabidopsis. *Plant Cell* **2012**, *24*, 2184–2199. [CrossRef] [PubMed]
19. Costet, L.; Fritig, B.; Kauffmann, S. Scopoletin expression in elicitor-treated and tobacco mosaic virus-infected tobacco plants. *Physiol. Plant* **2002**, *115*, 228–235. [CrossRef] [PubMed]
20. Fabienne, B.; Patrice, D.R.; Serge, K. Molecular cloning and biological activity of alpha-, beta-, and gamma-megaspermin, three elicitins secreted by phytophthora megasperma h20. *Plant Physiol.* **2003**, *131*, 155–166.
21. Hino, F.; Okazaki, M.; Miura, Y. Effect of 2,4-dichlorophenoxyacetic acid on glucosylation of scopoletin to scopolin in tobacco tissue culture. *Plant Physiol.* **1982**, *69*, 810–813. [CrossRef] [PubMed]
22. Ahn, Y.O.; Shimizu, B.; Sakata, K.; Gantulga, D.; Zhou, Z.; Bevan, D.R.; Esen, A. Scopolin hydrolyzing beta-glucosidases in roots of arabidopsis. *Plant Cell Physiol.* **2010**, *51*, 132. [CrossRef] [PubMed]
23. Chapelle, A.; Morreel, K.; Vanholme, R.; Le-Bris, P.; Morin, H.; Lapierre, C.; Boerjan, W.; Jouanin, L.; Demont-Caulet, N. Impact of the absence of stem-specific beta-glucosidases on lignin and monolignols. *Plant Physiol.* **2012**, *160*, 1204–1217. [CrossRef] [PubMed]
24. Zamioudis, C.; Hanson, J.; Pieterse, C.M. Beta-glucosidase bglu42 is a myb72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in arabidopsis roots. *New Phytol.* **2014**, *204*, 368–379. [CrossRef] [PubMed]
25. Piffanelli, P.; Ross, J.H.E.; Murphy, D.J. Biogenesis and function of the lipidic structures of pollen grains. *Sex. Plant Reprod.* **1998**, *11*, 65–80. [CrossRef]
26. Jiang, J.; Zhang, Z.; Cao, J. Pollen wall development: The associated enzymes and metabolic pathways. *Plant Biol.* **2013**, *15*, 249–263. [CrossRef]
27. Minic, Z.; Jouanin, L. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* **2006**, *44*, 435–449. [CrossRef] [PubMed]
28. Zhang, D.S.; Liang, W.Q.; Yuan, Z.; Li, N.; Shi, J.; Wang, J.; Liu, Y.M.; Yu, W.J.; Zhang, D.B. Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. *Mol. Plant* **2008**, *1*, 599–610. [CrossRef] [PubMed]
29. Dong, X.; Feng, H.; Xu, M.; Lee, J.; Kim, Y.K.; Lim, Y.P.; Piao, Z.; Park, Y.D.; Ma, H.; Hur, Y. Comprehensive analysis of genic male sterility-related genes in brassica rapa using a newly developed br300k oligomeric chip. *PLoS ONE* **2013**, *8*, e72178. [CrossRef]
30. Singh, A.K.; Sharma, V.; Pal, A.K.; Acharya, V.; Ahuja, P.S. Genome-wide organization and expression profiling of the nac transcription factor family in potato (*Solanum tuberosum* L.). *DNA Res.* **2013**, *20*, 403–423. [CrossRef] [PubMed]

31. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal w and clustal x version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]
32. Thorlby, G.; Fourrier, N.; Warren, G. The sensitive to freezing2 gene, required for freezing tolerance in arabidopsis thaliana, encodes a beta-glucosidase. *Plant Cell* **2004**, *16*, 2192–2203. [CrossRef] [PubMed]
33. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. Gsds 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]
34. Ng, C.Y.; Wickneswari, R.; Choong, C.Y. Identification of floral genes for sex determination in *Calamus palustris* griff. By using suppression subtractive hybridization. *Genet. Mol. Res.* **2014**, *13*, 6037–6049. [CrossRef]
35. Zhang, L.; Cai, X.; Wu, J.; Liu, M.; Grob, S.; Cheng, F.; Liang, J.; Cai, C.; Liu, Z.; Liu, B.; et al. Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res.* **2018**, *5*, 50. [CrossRef]
36. Rubinelli, P.; Hu, Y.; Ma, H. Identification, sequence analysis and expression studies of novel anther-specific genes of *Arabidopsis thaliana*. *Plant Mol. Biol.* **1998**, *37*, 607–619. [CrossRef]
37. Winter, D.; Vinegar, B.; Nahal, H.; Ammar, R.; Wilson, G.V.; Provart, N.J. An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2007**, *2*, e718. [CrossRef] [PubMed]
38. Smyth, D.R.; Bowman, J.L.; Meyerowitz, E.M. Early flower development in arabidopsis. *Plant Cell* **1990**, *2*, 755–767. [CrossRef] [PubMed]
39. Ma, H. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu. Rev. Plant Biol.* **2005**, *56*, 393–434. [CrossRef] [PubMed]
40. Wang, P.; Liu, H.; Hua, H.; Wang, L.; Song, C.-P. A vacuole localized β-glucosidase contributes to drought tolerance in *arabidopsis*. *Chin. Sci. Bull.* **2011**, *56*, 3538–3546. [CrossRef]
41. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **2011**, *43*, 1035–1039. [CrossRef] [PubMed]
42. Nuruzzaman, M.; Manimekalai, R.; Sharoni, A.M.; Satoh, K.; Kondoh, H.; Ooka, H.; Kikuchi, S. Genome-wide analysis of nac transcription factor family in rice. *Gene* **2010**, *465*, 30–44. [CrossRef] [PubMed]
43. Fourrier, N.; Bedard, J.; Lopez-Juez, E.; Barbrook, A.; Bowyer, J.; Jarvis, P.; Warren, G.; Thorlby, G. A role for sensitive to freezing2 in protecting chloroplasts against freeze-induced damage in arabidopsis. *Plant J.* **2008**, *55*, 734–745. [CrossRef] [PubMed]
44. Moellering, E.R.; Bagyalakshmi, M.; Christoph, B. Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science* **2010**, *330*, 226–228. [CrossRef] [PubMed]
45. Andreasson, E.; Bolt Jorgensen, L.; Hoglund, A.S.; Rask, L.; Meijer, J. Different myrosinase and idioblast distribution in arabidopsis and brassica napus. *Plant Physiol.* **2001**, *127*, 1750–1763. [CrossRef] [PubMed]
46. Talalay, P.; Fahey, J.W. Phytochemicals from cruciferous plants protect against cancer by modulating carcinogen metabolism. *J. Nutr.* **2001**, *131*, 3027S–3033S. [CrossRef]
47. Jeffery, E.H.; Araya, M. Physiological effects of broccoli consumption. *Phytochem. Rev.* **2009**, *8*, 283–298. [CrossRef]
48. Borpatragohain, P.; Rose, T.J.; King, G.J. Fire and brimstone: Molecular interactions between sulfur and glucosinolate biosynthesis in model and crop brassicaceae. *Front. Plant Sci.* **2016**, *7*, 1735. [CrossRef] [PubMed]
49. Baskar, V.; Gururani, M.A.; Yu, J.W.; Park, S.W. Engineering glucosinolates in plants: Current knowledge and potential uses. *Appl. Biochem. Biotechnol.* **2012**, *168*, 1694–1717. [CrossRef] [PubMed]
50. Becker, T.M.; Jeffery, E.H.; Juvik, J.A. Proposed method for estimating health-promoting glucosinolates and hydrolysis products in broccoli (*Brassica oleracea* var. Italica) using relative transcript abundance. *J. Agric. Food Chem.* **2017**, *65*, 301–308. [CrossRef] [PubMed]
51. Kai, K.; Shimizu, B.; Mizutani, M.; Watanabe, K.; Sakata, K. Accumulation of coumarins in arabidopsis thaliana. *Phytochemistry* **2006**, *67*, 379–386. [CrossRef]
52. Baiya, S.; Mahong, B.; Lee, S.; Jeon, J.S.; Cairns, J.R.K. Demonstration of monolignol β-glucosidase activity of rice os4bglu14, os4bglu16 and os4bglu18 in arabidopsis thaliana bglu 45 mutant. *Plant Physiol. Biochem.* **2018**, *127*, 223. [CrossRef] [PubMed]

53. Zhu, J.; Chen, H.; Li, H.; Gao, J.F.; Jiang, H.; Wang, C.; Guan, Y.F.; Yang, Z.N. Defective in tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in arabidopsis. *Plant J.* **2008**, *55*, 266–277. [CrossRef]
54. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Correction: Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868. [CrossRef]
55. Usadel, B.; Obayashi, T.; Mutwil, M.; Giorgi, F.M.; Bassel, G.W.; Tanimoto, M.; Chow, A.; Steinhauser, D.; Persson, S.; Provart, N.J. Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant Cell Environ.* **2009**, *32*, 1633–1651. [CrossRef]
56. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868. [CrossRef] [PubMed]
57. Gillis, J.; Pavlidis, P. "Guilt by association" is the exception rather than the rule in gene networks. *PLoS Comput. Biol.* **2012**, *8*, e1002444. [CrossRef] [PubMed]
58. van Dam, S.; Vosa, U.; van der Graaf, A.; Franke, L.; de Magalhaes, J.P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinf.* **2018**, *19*, 575–592. [CrossRef]
59. Zheng, J.; He, C.; Qin, Y.; Lin, G.; Park, W.D.; Sun, M.; Li, J.; Lu, X.; Zhang, C.; Yeh, C.T.; et al. Co-expression analysis aids in the identification of genes in the cuticular wax pathway in maize. *Plant J.* **2019**, *97*, 530–542. [CrossRef] [PubMed]
60. Clough, S.J.; Bent, A.F. Floral dip: A simplified method for agrobacterium-mediated transformation of arabidopsis thaliana. *Plant J.* **1998**, *16*, 735–743. [CrossRef]
61. Peterson, R.; Slovin, J.P.; Chen, C. A simplified method for differential staining of aborted and non-aborted pollen grains. *Int. J. Plant Biol.* **2010**, *1*, e13. [CrossRef]
62. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The pfam protein families database in 2019. *Nucleic Acids Res.* **2018**, *47*, D427–D432. [CrossRef] [PubMed]
63. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. Mega6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef] [PubMed]
64. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. Mcscanx: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]
65. Toufighi, K.; Brady, S.M.; Austin, R.; Ly, E.; Provart, N.J. The botany array resource: E-northerns, expression angling, and promoter analyses. *Plant J.* **2005**, *43*, 153–163. [CrossRef]
66. Du, Z.; Zhou, X.; Ling, Y.; Zhang, Z.; Su, Z. Agrigo: A go analysis toolkit for the agricultural community. *Nucleic Acids Res.* **2010**, *38*, W64–W70. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Characterization, Evolution, and Expression Profiling of *VQ* Gene Family in Response to Phytohormone Treatments and Abiotic Stress in *Eucalyptus grandis*

Huifang Yan, Yujiao Wang, Bing Hu, Zhenfei Qiu, Bingshan Zeng * and Chunjie Fan *

Key Laboratory of State Forestry Administration on Tropical Forest Research, Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou 510520, China; yhuifangy@gmail.com (H.Y.); wangyujiao@caf.ac.cn (Y.W.); hubing@caf.ac.cn (B.H.); qiuzhenfei@caf.ac.cn (Z.Q.)

* Correspondence: b.s.zeng@vip.tom.com (B.Z.); fanchunjie@caf.ac.cn (C.F.); Tel.: +86-020-87032851 (B.Z.); +86-020-87032851 (C.F.)

Received: 24 February 2019; Accepted: 1 April 2019; Published: 10 April 2019

Abstract: *VQ* genes play important roles in plant development, growth, and stress responses. However, little information regarding the functions of *VQ* genes is available for *Eucalyptus grandis*. In our study, genome-wide characterization and identification of *VQ* genes were performed in *E. grandis*. Results showed that 27 *VQ* genes, which divided into seven sub-families (I–VII), were found, and all but two *VQ* genes showed no intron by gene structure and conserved motif analysis. To further identify the function of EgrVQ proteins, gene expression analyses were also developed under hormone treatments (brassinosteroids, methyl jasmonate, salicylic acid, and abscisic acid) and abiotic conditions (salt stress, cold 4 °C, and heat 42 °C). The results of a quantitative real-time PCR analysis indicated that the *EgrVQs* were variously expressed under different hormone treatments and abiotic stressors. Our study provides a comprehensive overview of *VQ* genes in *E. grandis*, which will be beneficial in the molecular breeding of *E. grandis* to promote its resistance to abiotic stressors; the results also provide a basis from which to conduct further investigation into the functions of *VQ* genes in *E. grandis*.

Keywords: *VQ* genes family; *Eucalyptus grandis*; expression pattern; plant hormones; abiotic stress

1. Introduction

The VQ protein family either positively or negatively regulates diverse developmental processes, including plant immunity, abiotic stresses, and growth and development [1]. The expression of genes, which are related to biological processes and abiotic stresses in plants, is altered in response to either internal or external signals [1,2]. The *VQ* genes are a type of plant-specific proteins with a dramatically conserved VQ motif, which possesses five conserved amino acids in its main protein sequences FxxVQxhTG [3,4]. Recently, *VQ* genes have been identified by genome-wide analysis in numerous plants. There are 34, 39, 51, 29, and 74 members in Arabidopsis [5], rice [6], poplar [7], moso bamboo [8], and tea [9], respectively. Previous studies have shown that *VQ* genes played important roles in resistance to abiotic and biotic stressors [10]. Meanwhile, accumulating researches showed that many VQ proteins interacted with WRKY transcription factors [1,11]. The WRKY transcription factors, which harbor a highly conserved WRKYGQK amino acid sequence that is followed by a zinc-finger motif at the N-terminal domain, are ubiquitous among higher plants [11,12]. The VQ proteins that contain 50–60 conserved amino acids interact with WRKY transcription factors by the residues V and Q [13,14]. For example, AtVQ15 plays a negative role in response to osmotic stress [15]. AtVQ23 and AtVQ16 transcript levels were strongly induced by *Botrytis cinerea*. A study has shown

that AtVQ22 negatively mediates AtWRKY28 and AtWRKY51 and it is involved with JA (jasmonic acid) defense [16]. In addition, the transcript levels of many *VQ* genes in rice differed when exposed to drought [17]. Moreover, *VQ* genes can regulate multiple biological processes including plant growth and development, senescence, and hormone signaling. The GmVQ1, GmVQ6, and GmVQ53 transcripts are highly expressed during seed development in soybean [18]. AtVQ14 and MINI3 interacted and regulated the abundance of mRNA-encoding proteins that participate in the process of seed development [16]. Specifically, *VQ* genes function by interacting with the group I or IIc WRKY transcription factors [4,19,20]. AtVQ23 and AtVQ16 activated AtWRKY33 by binding its C-terminal WRKY domain and further induced plant defenses [21]. Under salt stress, AtVQ9 expression was increased and it then interacted with WRKY8 for inhibiting the expression of its target genes to modulate salinity stress tolerance [22].

The *Eucalyptus* species are tropical/subtropical woody plants that belong to the Myrtaceae family of angiosperms. It is also the world's leading source of woody biomass [23]. Specifically, *Eucalyptus* not only provides fuel biomass and directly reduces atmospheric carbon dioxide levels [23,24], but it also performs a variety of indirect services through its essential oils that are used as a pesticidal agent and pest repellent [25]. However, *Eucalyptus* can be affected by a variety of biotic and abiotic stressors during growth and development, including heat, cold, salt stress, and disease. *VQ* genes played a vital role in resistance to abiotic and biotic stressors. Thus, it is necessary to understand the characteristics and functions of the *VQ* genes. However, there have been no reports on *VQ* genes in *E. grandis* until now. Fortunately, the availability of the complete *E. grandis* genome provides an opportunity to conduct a comprehensive analysis of *VQ* genes in *E. grandis* [26].

In this study, *VQ* genes in *E. grandis* were identified, and then a bioinformatics analysis was conducted, including phylogenetic relationships, conserved motifs, homologous pairs, gene structures, and promoter analysis. Furthermore, the expression of *EgrVQ* genes was investigated under biotic stress conditions, such as brassinosteroids (BRs), methyl jasmonate (MeJA), salicylic acid (SA), abscisic acid (ABA) treatments, and abiotic stresses, such as cold, heat, and NaCl treatments. These results will provide a solid foundation for elucidating the function of *EgrVQ* genes in response to biotic and abiotic stress and further molecular breeding in *E. grandis*.

2. Results

2.1. Identification of the VQ Gene Family in the E. grandis

In our study, 27 *VQ* genes were identified, which contained VQ domains using *AtVQ* genes performing BLASTP search in the *E. grandis* genome and VQ conserved domain (PF05678) identification. Table 1 lists detailed information regarding the *VQ* genes in *E. grandis*. These 27 *EgrVQ* genes were named from *EgrVQ1* to *EgrVQ27*. Their translated proteins ranged from 101 to 348 amino acids (AA), with an average of 217 AAs. The predicted molecular weight of the proteins varied from 11.3 to 37.6 kDa, and the pI values ranged from 5.38 to 10.66 (Table 1). By predicting their subcellular localization, it was also found that 12 *EgrVQ* genes were located in the chloroplasts, and 14 *EgrVQ* genes were located in other compartments.

2.2. Mapping EgrVQ Genes on Chromosomes, Gene Duplication, and Analysis of Paralogs and Orthologs

The chromosomal location illustrated that 27 *EgrVQ* genes were randomly and unequally distributed on nine chromosomes (Figure 1). Specifically, chromosome 6 contained eight *VQ* genes and chromosome 8 contained five genes, respectively. Meanwhile, the *EgrVQ13/EgrVQ14*, *EgrVQ19/EgrVQ20*, and *EgrVQ19/EgrVQ21*showed tandem duplication. Consistent with the gene duplication analysis, we also found that *EgrVQ13/EgrVQ14*, *EgrVQ19/EgrVQ20*, *EgrVQ19/EgrVQ21*, and *EgrVQ20/EgrVQ21* were paralogous pairs. Details from the analysis of paralogs and orthologs with *A. thaliana*, poplar, and rice are presented in Table 2.

Table 1. The summary of 27 identified *EgrVQ* genes.

Name	Gene ID	Chr	Location Coordinates (5′–3′)	ORF Length (bp)	Exons	Protein Length (aa)	Protein MW (Da)	Protein PI	Subcellular WoLF PSORT	Localization TargetP
EgrVQ1	Eucgr.A02182.1	1	37113647..37114778	687	1	228	22929.82	9.50	Nucleus	Chloroplast
EgrVQ2	Eucgr.A02242.1	1	37821284..37822147	468	1	155	16913.17	7.92	Mitochondrion	Chloroplast
EgrVQ3	Eucgr.A02450.1	1	39836460..39837326	567	1	188	21252.59	5.38	Nucleus	Any other location
EgrVQ4	Eucgr.A02544.1	1	40671118..40671423	306	1	101	11278.94	9.36	Cytoplasm	Chloroplast
EgrVQ5	Eucgr.B00035.1	6	35375963..35378557	969	1	322	34040.85	10.02	Nucleus	Chloroplast
EgrVQ6	Eucgr.C02621.1	3	51875030..51875674	645	1	215	22454.41	6.65	Mitochondrion	Chloroplast
EgrVQ7	Eucgr.C02629.1	3	51927510..51928462	831	2	276	28709.31	7.15	Nucleus	Chloroplast
EgrVQ8	Eucgr.D00026.1	4	216578..217627	62	1	207	22664.50	5.91	Nucleus	Any other location
EgrVQ9	Eucgr.D02200.1	4	35568137..35570178	972	1	323	35209.97	10.66	Nucleus	Any other location
EgrVQ10	Eucgr.F00302.1	6	5689343..5690714	966	1	321	34599.00	9.66	Nucleus	Chloroplast
EgrVQ11	Eucgr.F00371.1	6	5040555..5041763	1047	1	348	37574.61	9.99	Nucleus	Chloroplast
EgrVQ12	Eucgr.F01148.1	6	15726039..15727038	744	1	247	26541.68	9.01	Nucleus	Chloroplast
EgrVQ13	Eucgr.F01407.1	6	19292825..19293671	330	1	109	12303.87	7.79	Chloroplast	Any other location
EgrVQ14	Eucgr.F01408.1	6	19298931..19299260	330	1	109	12161.73	6.73	Cytoplasm	Any other location
EgrVQ15	Eucgr.F02723.1	6	39354381..39355067	687	1	228	24468.88	9.47	Nucleus	Any other location
EgrVQ16	Eucgr.F03907.1	6	49847862..49848569	669	1	222	23853.03	9.71	Nucleus	Any other location
EgrVQ17	Eucgr.G02312.1	7	43757342..43758655	675	1	224	24656.69	8.68	Nucleus	Chloroplast
EgrVQ18	Eucgr.G03218.1	7	52999148..52999886	510	1	169	17744.58	6.50	Nucleus	Any other location
EgrVQ19	Eucgr.H02317.1	8	29626932..29627764	522	1	173	18910.74	9.93	Chloroplast	Any other location
EgrVQ20	Eucgr.H02318.1	8	29632300..29632908	522	1	173	18883.63	9.93	Chloroplast	Any other location
EgrVQ21	Eucgr.H02319.1	8	29636679..29638276	546	2	181	19910.03	10.05	Chloroplast	Secretory pathway
EgrVQ22	Eucgr.H04261.1	8	57364319..57364817	363	1	120	13296.04	8.87	Nucleus	Chloroplast
EgrVQ23	Eucgr.H04807.1	8	67074212..67075837	846	1	281	29271.01	8.93	Nucleus	Any other location
EgrVQ24	Eucgr.I00016.1	9	223638..224213	543	1	180	19892.34	9.00	Nucleus	Any other location
EgrVQ25	Eucgr.I00226.1	9	4661485..4662947	726	1	241	25294.49	6.29	Nucleus	Chloroplast
EgrVQ26	Eucgr.K01342.1	11	16771278..16772474	747	1	248	25715.14	10.08	Nucleus	Any other location
EgrVQ27	Eucgr.K01686.1	11	19466093..19467503	840	1	279	28773.80	6.59	Nucleus	Any other location

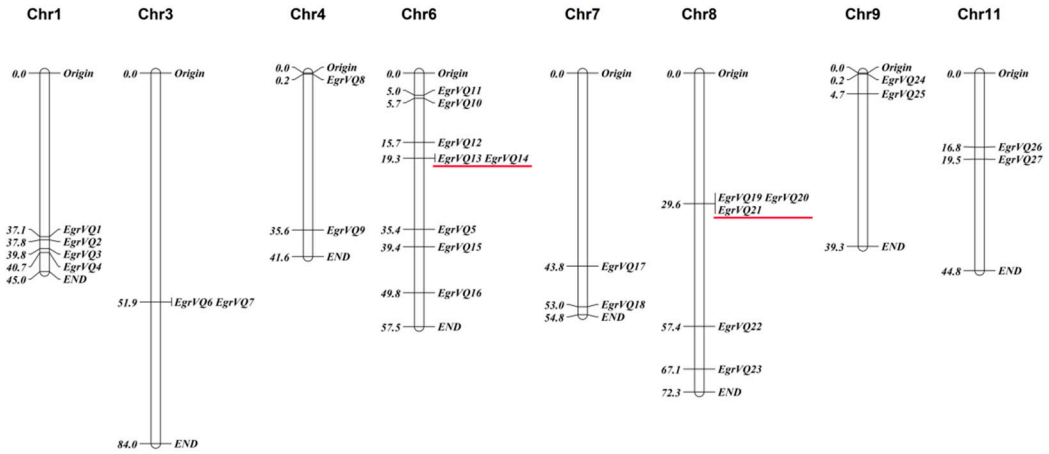


Figure 1. Chromosomal location of *E. grandis VQ* genes. Chromosome numbers were indicated above each chromosome. The size of a chromosome was indicated by its relative length. Gene positions and chromosome sizes were given in megabases (Mb) to the left of the figure. Tandem duplicated genes were underlined in red.

Table 2. The summary of paralogous (*Egr–Egr*) and orthologous (*Egr–At*, *Egr–Os* and *Egr-Pt*) gene pairs.

Egr–Egr	*Egr–At*	*Egr–Os*	*Egr-Pt*
EgrVQ6/EgrVQ7	*EgrVQ1/AtVQ24*	*EgrVQ12/OsVQ3*	*EgrVQ9/PtVQ3*
EgrVQ6/EgrVQ25	*EgrVQ5/AtVQ9*	*EgrVQ12/OsVQ21*	*EgrVQ9/PtVQ12*
EgrVQ19/EgrVQ20	*EgrVQ9/AtVQ6*	*EgrVQ15/OsVQ8*	*EgrVQ12/PtVQ7*
EgrVQ19/EgrVQ21	*EgrVQ11/AtVQ9*	*EgrVQ15/OsVQ15*	
EgrVQ20/EgrVQ21	*EgrVQ12/AtVQ4*	*EgrVQ24/OsVQ3*	
	EgrVQ12/AtVQ19		
	EgrVQ12/AtVQ33		
	EgrVQ13/AtVQ1		
	EgrVQ13/AtVQ10		
	EgrVQ14/AtVQ1		
	EgrVQ14/AtVQ10		
	EgrVQ15/AtVQ3		
	EgrVQ15/AtVQ21		
	EgrVQ16/AtVQ11		
	EgrVQ24AtVQ31		

2.3. Identification of Cis-Elements in the Promoter Regions of EgrVQ Genes

In this study, the identification of *cis*-regulatory elements was performed in the promoter regions of *EgrVQ* genes (Table 4). The results showed that many stress type *cis*-elements were widespread in the promoter region of *EgrVQ* genes. Table 3 shows the details information. ABA-stress, MeJA-stress, SA-stress, drought-stress, low temperature-stress, dehydration, and salt stress response elements were found in the *EgrVQs* promoter. Interestingly, all of the *EgrVQs* contained a CGTCA-motif and a TGACG-motif, which were involved in the response to MeJA. Moreover, all but *EgrVQ21* contained an ABRE element. In addition, the MBS, TCA-element, and LTR *cis*-elements, which were involved in the response to SA, drought, and low-temperature, respectively, were presented in most of the *EgrVQs* promoter regions. These results indicated that most of the *EgrVQ* genes might respond to plant biotic and abiotic stresses.

Table 3. The number of structural analysis of conserved motif FxxVQxxxG in *E. grandis* and other plants.

Gene Name	LTG	FTG	VTG	LTS	LTD	YTG	ITG	LTR	LTV	ATG	LTA	LSG
Arabidopsis	24	5	2	1	1	1	0	0	0	0	0	0
Poplar	39	11	2	0	0	0	0	0	0	0	0	0
Chinese Cabbage	43	8	3	1	0	1	0	0	1	0	0	0
Soybean	55	15	2	1	0	1	0	1	0	0	0	0
Grapevine	14	3	1	0	0	1	0	0	0	0	0	0
Maize	42	8	6	0	0	0	2	0	0	2	1	0
Rice	28	7	4	0	0	1	1	0	0	0	0	0
Eucalyptus grandis	20	4	2	0	0	0	0	0	0	0	0	1

2.4. Multiple Sequence Alignment and Phylogenetic of EgrVQ Proteins

To obtain the phylogenetic relationship of *E. grandis* VQ proteins, an unrooted tree was constructed, including 27 EgrVQ, 51 PtVQ, 34 AtVQ, and 40 OsVQ proteins (Figure 2). The results showed that EgrVQ proteins were divided into seven sub-families (I–VII). The multiple alignment analysis showed that the EgrVQ proteins presented four conserved motif variations: FxxVQxLTG(20/27), FxxVQxFTG(4/27), FxxVQxVTG(2/27), and FxxVQxLSG(1/27) (Figure 3). Among these motifs, LTG, FTG, and VTG were extensively presented in *A. thaliana* [20], rice [17], and poplar [7]. However, it was found that EgrVQ7 protein presented an LSG motif, which was not reported in previous studies. The details of the conserved motifs in *E. grandis* are presented in Table 3 and Figure 3.

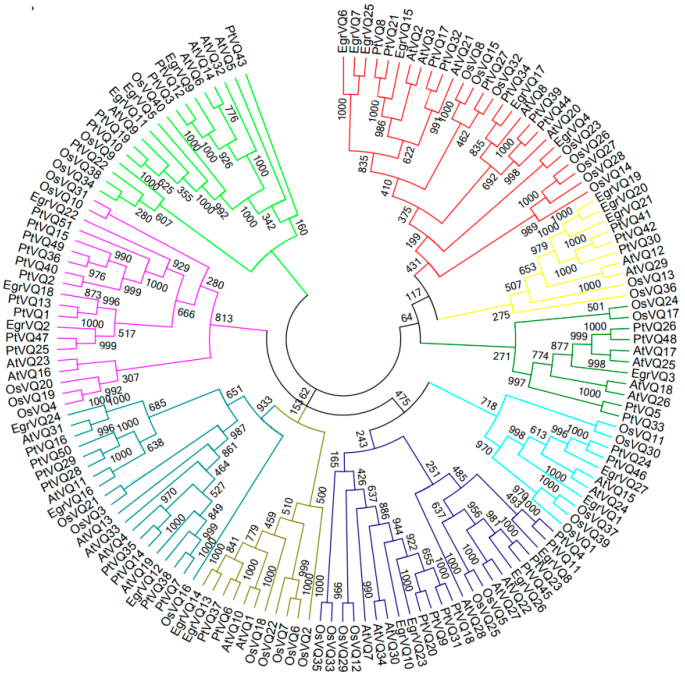


Figure 2. Phylogenetic tree of VQ proteins from *E. grandis*, Arabidopsis, rice, and poplar. ClustalW softwarealigned the complete amino acid sequences of 27 *E. grandis* (prefixed with 'Egr'), 34 Arabidopsis (prefixed with 'At'), 40 rice (prefixed with 'Os'), and 51 (prefixed with 'Pt') VQ proteins, and MEGA 7, with 1000 bootstrap replicates, constructed the neighbor-joining tree.

2.5. Analysis of Gene Structural and Conserved motifs of EgrVQ Genes

To further understand the structural features of the *EgrVQ* genes, exon/intron structural analysis was performed and results are shown in Figure 4 and Table 1. Interestingly, 25 *EgrVQ* genes had only one exon, while the *EgrVQ7* and *EgrVQ21* genes had two exons. Subsequently, the conserved motifs of the EgrVQ proteins were studied. All of the *EgrVQ* genes contained motif1, which was the DNA-binding domain of *EgrVQ* genes (Figure 5). Notably, some motifs appeared in one sub-family of

EgrVQ. For example, motif2, motif3, and motif17 only existed in sub-familyV-3, III, and V, respectively, which indicated that these motifs were the characteristic elements of these sub-families. Meanwhile, it was found that the same sub-family had similar motifs. For example, sub-familyV-3 genes contained motif4, motif9, motif1, motif5, motif7, and motif11, whereas sub-familyIII included motif4, motif12, motif7, motif8, motif1, motif6, and motif3. Figure 5 and Table 4 lists the details.

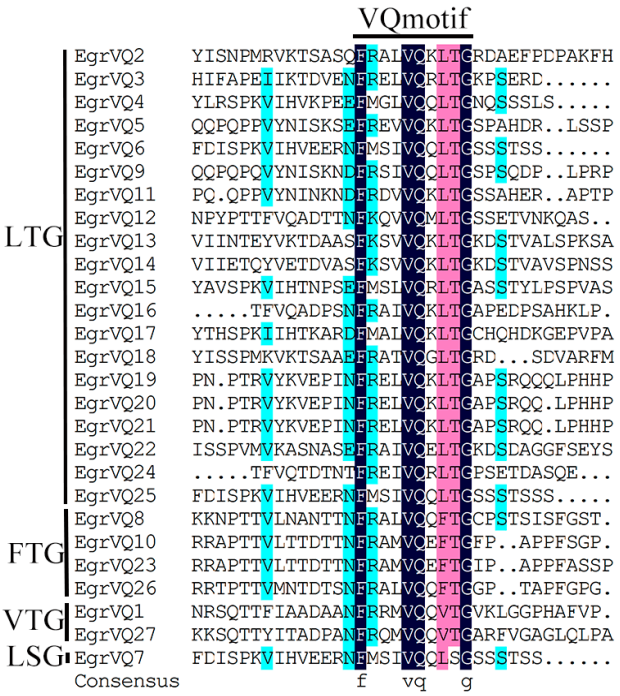


Figure 3. Multiple sequence alignment of VQ proteins in *E. grandis*. Sequences were aligned using DNAMAN software. The FxxxVQxxxG motif was highly conserved. The dark blue indicated the conserved section of EgrVQs protein family. The pink indicated the four conserved motif variations of LGT, FTG, VTG, and LSG.

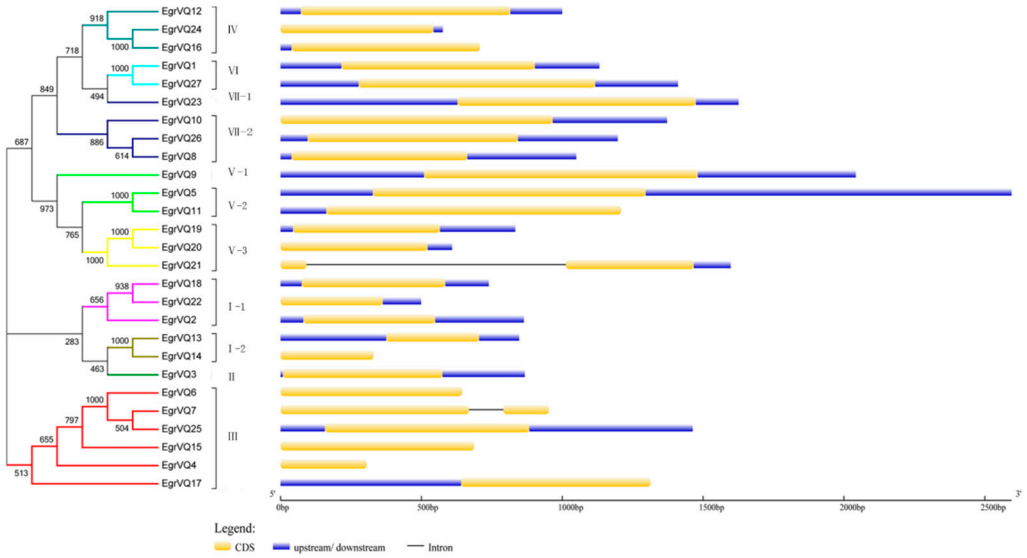


Figure 4. Gene structure of *VQ* genes in *E. grandis*. Exons were indicated by yellow rectangles. Upstream/downstream's sequences of *EgrVQs* were indicated by blue lines. Gray lines connecting two exons represented introns.

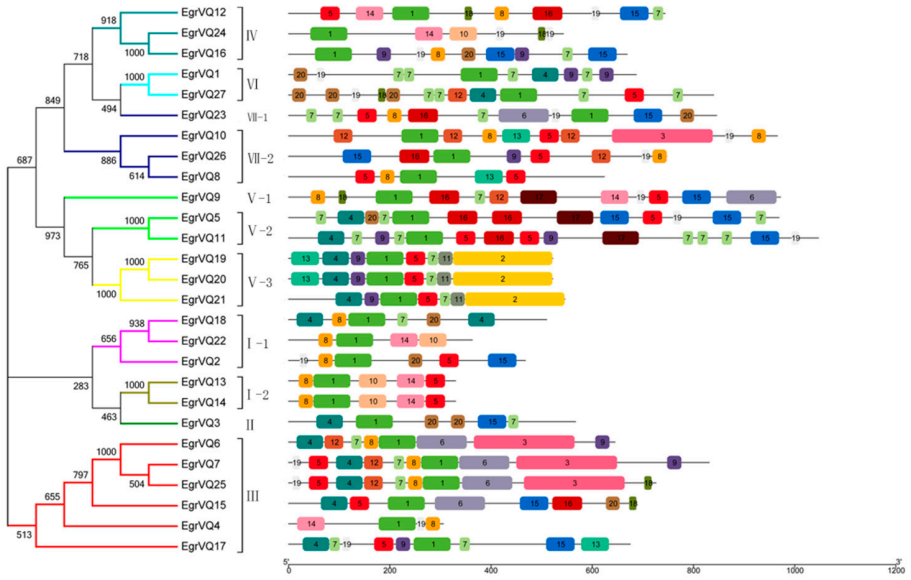


Figure 5. Phylogenetic relationships and conserved motifs of VQ proteins in *E. grandis*. The left panel showed the phylogenetic tree of EgrVQ, which was constructed by the neighbor-joining method based on the results of sequence alignment. Proteins were divided into seven subgroups (marked by different colors). The right panel showed the distribution of the 20 conserved motifs in the *EgrVQ* genes following analysis with Multiple Expectation Maximization for Motif Elicitation (MEME). Each specific motif was marked by a different colored box, and the motif numbers were included in the center of each box. The length of each box was proportional to the actual size of the motif.

Table 4. Summary of stresses inducible *cis*-elements in the promoter regions of *VQ* genes in *E. grandis*.

Gene Name	ABRE [1]	CGTCA-motif [2]	TGACG-motif [3]	TCA-element [4]	MBS [5]	LTR [6]	DRE [7]
EgrVQ1	1	1	1		1	1	
EgrVQ2	1	1	1	1	1		
EgrVQ3	1	1	1		1		
EgrVQ4	1	1	1		1	1	
EgrVQ5	1	1	1			1	
EgrVQ6	1	1	1		1		
EgrVQ7	1	1	1			1	
EgrVQ8	1	1	1	1			
EgrVQ9	1	1	1	1		1	
EgrVQ10	1	1	1	1	1	1	
EgrVQ11	1	1	1	1			
EgrVQ12	1	1	1				
EgrVQ13	1	1	1			1	
EgrVQ14	1	1	1			1	
EgrVQ15	1	1	1		1	1	
EgrVQ16	1	1	1	1		1	
EgrVQ17	1	1	1			1	
EgrVQ18	1	1	1		1	1	
EgrVQ19	2	1	1		1	1	
EgrVQ20	1	1	1				1
EgrVQ21		1	1	1			1
EgrVQ22	1	1	1		1		
EgrVQ23	1	1	1	1		1	
EgrVQ24	1	1	1	1			
EgrVQ25	1	1	1	1	1		
EgrVQ26	1	1	1	1		1	
EgrVQ27	1	1	1	1	1		

[1] cis-acting element involved in the abscisic acid responsiveness; [2] cis-acting regulatory element involved in the MeJA-responsiveness; [3] cis-acting regulatory element involved in the MeJA-responsiveness; [4] cis-acting element involved in salicylic acid responsiveness; [5] MYB binding site involved in drought-inducibility; [6] cis-acting element involved in low-temperature responsiveness; [7] cis-acting element involved in dehydration, low-temperature, and salt stresses.

2.6. Expression Patters of EgrVQ Genes in Different Tissues of E. grandis

In our study, the expression patterns of *EgrVQ* genes were examined in different tissues, including root, xylem, phloem, mature leaves, and young leaves (Figure 6). The majority (88.9%) of *EgrVQ* genes were expressed in all tissues. However, almost all *EgrVQ*s showed differential gene expression. For example, from subfamily I members, *EgrVQ2* and *EgrVQ22* were relative highly expressed when compared to other members, *EgrVQ13* and *EgrVQ14* mainly were expressed in leaves, and *EgrVQ18* was not expressed in the root. Moreover, from subfamily V members, when compared to *EgrVQ20* and *EgrVQ21*, *EgrVQ9* and *EgrVQ11* were relative highly expressed in most tissues. In addition, from subfamily VII members, *EgrVQ1* and *EgrVQ26* were very highly expressed in root and xylem tissues than other tissues. Therefore, *EgrVQ1* and *EgrVQ26* might participate in root development and play vital roles in the xylem development of *E. grandis*.

When considering that 92.6% and 88.9% of *EgrVQ* genes were expressed in young leaves and mature leaves, respectively, we selected leaves for determining the expression profiles of *EgrVQ* genes in response to different treatments.

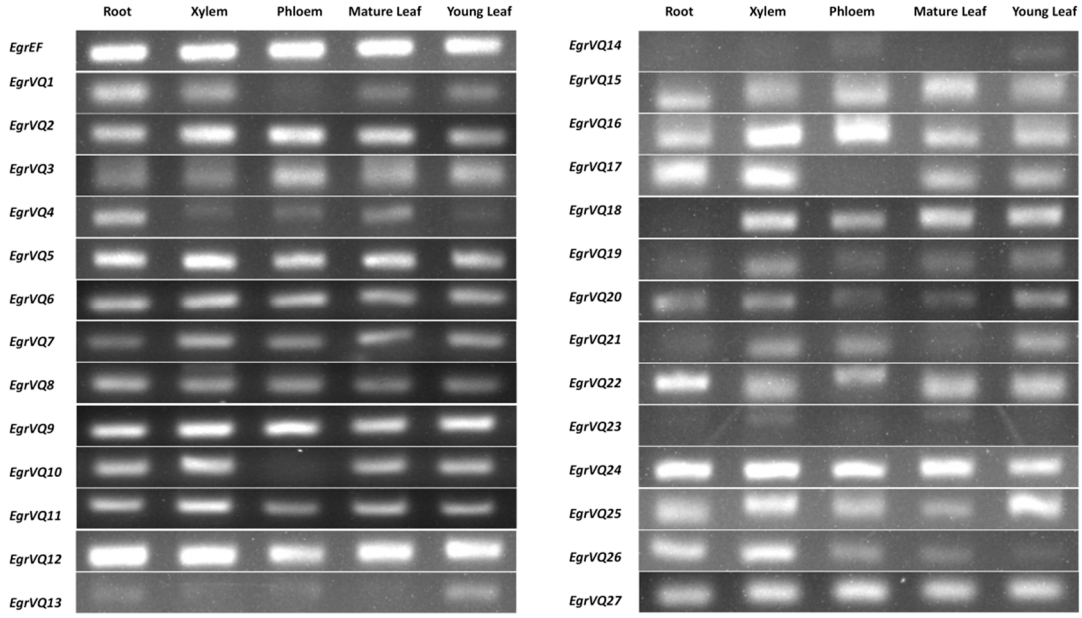


Figure 6. Semi-quantitative Real-Time PCR (RT-PCR) of the expression of *EgrVQ* genes in different tissues. From left to right: root, xylem, phloem, mature leaves, and young leaves. All of these tissues were collected from 6-week-old GL1 plants.

2.7. Expression Profiles of EgrVQ Genes in Response to Hormones

To investigate the function of *EgrVQ* genes in response to biotic stresses, we determined the expression patterns of *EgrVQ* genes under various plant hormone treatments, including BRs, MeJA, SA, and ABA (Figure 7). The results showed that the *EgrVQ* genes were expressed in diverse patterns under different hormone treatments. Most of the *EgrVQ* genes presented differential expression under the different phytohormone treatments. Some differences are extremely significant when compared to the reference gene *EgrEF*.

In the BR treatment, most of *EgrVQ* genes were up-regulated and the *EgrVQs* were clustered into three main groups according to their expression patterns (Figure 7A). Group BR-1 contained three *EgrVQs*, including *EgrVQ19*, *EgrVQ21*, and *EgrVQ27*, which were significantly decreased with the increasing duration of BR treatment. There were three *EgrVQs* in group BR-2 and their expression levels were dramatically increased at 1 h and 168 h after BR treatment.The rest of the *EgrVQ* genes were included in group BR-3, which contained 19 *EgrVQs* and they were continuously up-regulated.

Under SA treatment, the *EgrVQs* were mainly clustered into two groups according to their expression patterns (Figure 7B). The majority of *EgrVQs* (22/27) were clustered into group SA-1, in which the expression levels of *EgrVQs* were greatly up-regulated at 168 h, but they were not significantly altered at 1 h to 24 h after SA treatment. This result indicated that the expression of most *EgrVQs* was affected by long-term treatment with SA. The rest of the *EgrVQ*s were clustered in group SA-2 and their expression was significantly reduced or showed no obvious change.

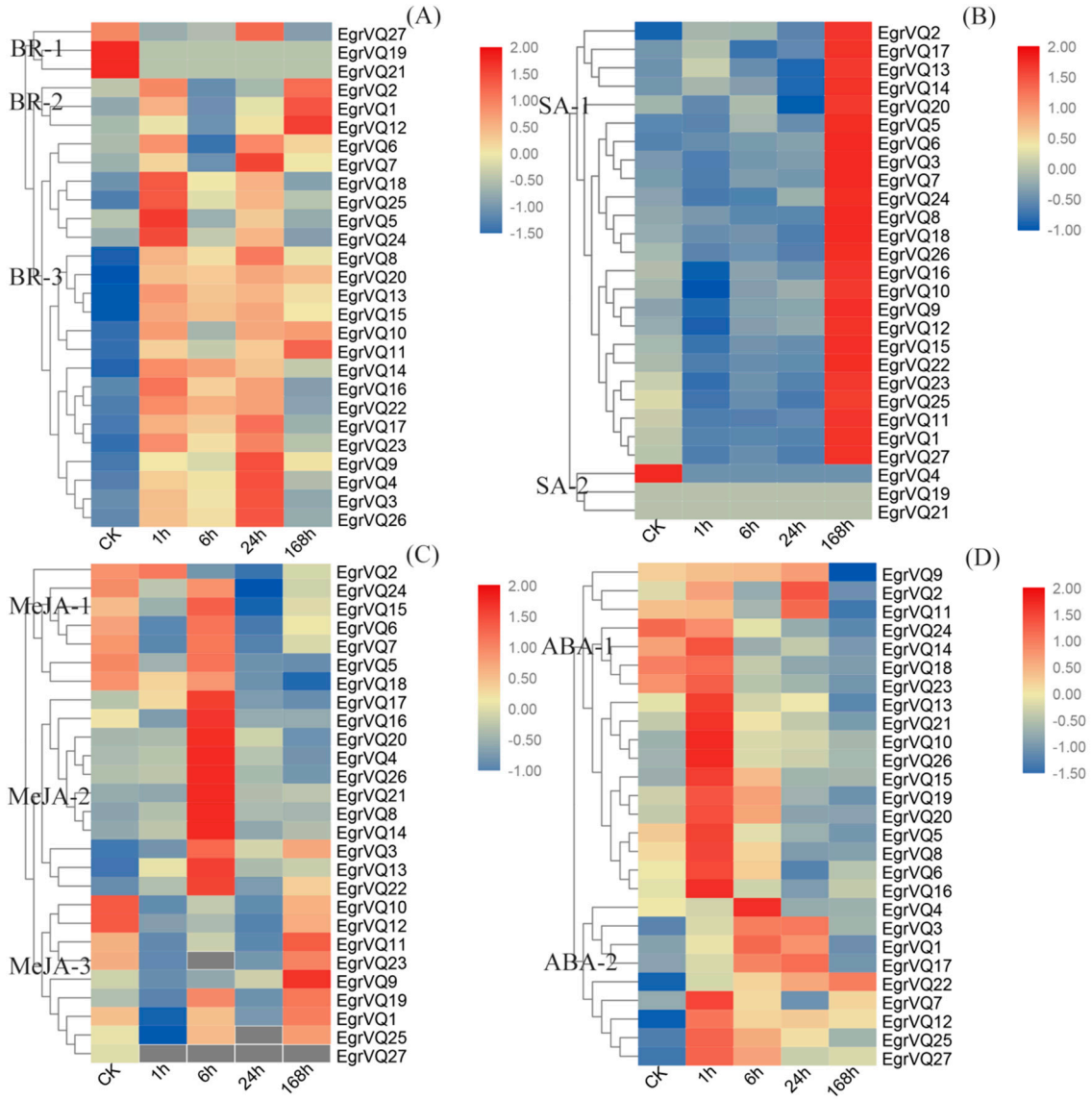


Figure 7. Expression analysis of 25 *EgrVQ* genes in *E. grandis* following brassinosteroid (BR) (**A**), salicylic acid (SA) (**B**), methyl jasmonate (MeJA) (**C**), and abscisic acid (ABA) (**D**) treatment, as determined by Real-Time quantitative PCR (qRT-PCR). The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The heatmap was created using MEV.

In the MeJA treatment, the *EgrVQs* were mainly clustered into three groups according to their expression patterns (Figure 7C). In the MeJA-1 group, there were seven *EgrVQs*, and their expressions were greatly decreased at 1 h, 24 h, and 168 h after MeJA treatment. The MeJA-2 group contained eleven *EgrVQs*, which showed significantly decreased expression at 1 h and then increased the expression at 6 h and reached their maximum expression level and decreased afterward. The rest of *EgrVQ*s were clustered in group MeJA-3, in which the expression levels of *EgrVQs* were greatly down-regulated at

1–24 h, and increased back to the level of the control or increased. However, it was found that *EgrVQ*s were immediately down-regulated under MeJA treatment.

Under ABA treatment, the expression level of the majority of *EgrVQs* was obviously reduced at 168 h and the *EgrVQs* were mainly clustered into two groups (Figure 7D). In group ABA-1, the expression of *EgrVQs* was significantly up-regulated at 1 h and then decreased with increasing treatment time under ABA treatment. In group ABA-2, there were nine *EgrVQs*, and their expression levels increased at 1 h, 6 h, or 24 h.

2.8. Expression Profiles of EgrVQ Genes in Response to Stress Treatments

To research the function of *EgrVQ* genes in response to abiotic stressors, we investigated the expression patterns of *EgrVQ* genes under NaCl, cold, and heat treatments, respectively (Figure 8). Overall, the *EgrVQ* genes showed differential expression under the different abiotic stresses. Some differences are even significant when compared to the reference gene *EgrEF*. Specifically, most of *EgrVQ* genes were up-regulated at 1 h or 6 h and they reached their highest expression levels, but they were then down-regulated at 168 h. These results indicated that *EgrVQ* genes could respond to short-term temperature stress. Interestingly, it was found that similar expression patterns of *EgrVQ* genes occurred under cold and heat treatments.Subsequently, *EgrVQs* were classified into three groups both in the cold and heat treatments, respectively.

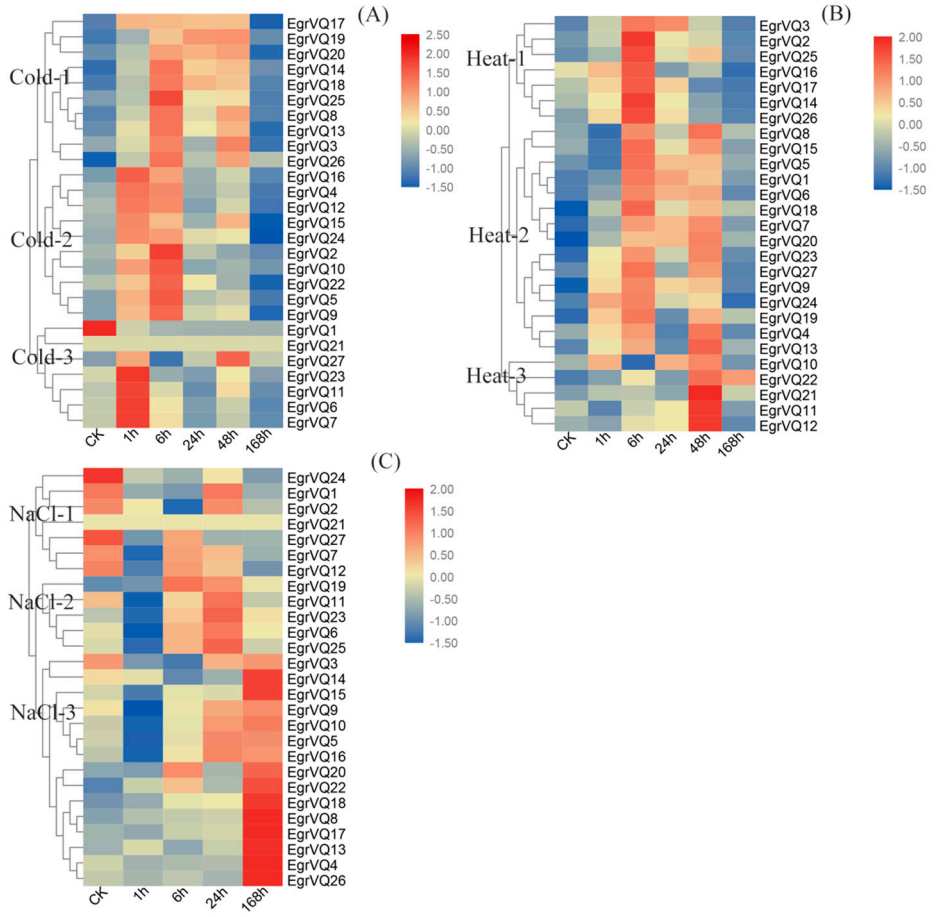


Figure 8. Expression analysis of 25 *EgrVQ* genes in *E. grandis* following cold (**A**), heat (**B**), and NaCl (**C**) treatments, as determined by qRT-PCR. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The heatmap was created using MEV.

In group Cold-1 and group Heat-2, there were 10 and 15 *EgrVQs*, respectively. The expression patterns of these genes were similar, and their expressions were greatly increased at 6-48 h and then

decreased after ABA treatment. Similarly, 10 and seven *EgrVQs* were classified into group Cold-2 and group Heat-1, respectively. Their expressions levels were greatly increased at 1 h and 6 h, but they then decreased after ABA treatment. In group Cold-3, *EgrVQs* were significantly increased at 1 h. In group Heat-3, *EgrVQs* were significantly increased at 48 h.

Under NaCl treatment, the *EgrVQs* were clustered into three groups according to their expression patterns (Figure 8C). In group NaCl-1, the expression of *EgrVQs* showed slight decreases with an increasing treatment time. In group NaCl-2, *EgrVQ*s were down-regulated at 1 h and then up-regulated with increasing NaCl treatment duration at 6 h and 24 h. The remaining of *EgrVQs* were classified to group NaCl-3. In this group, the expression of *EgrVQs* gradually increased and reached their highest expression level at 168 h.

3. Discussion

In our study, 27 *EgrVQ* genes were identified using the genome database in *E. grandis*. Subsequently, phylogenetic analysis, conserved motifs, and analysis of *cis*-elements of *EgrVQs* were performed. All of the *EgrVQs* encoded relatively small proteins of less than 400 amino acids, and most VQ motif-containing proteins contained FxxVQxLTG(20/27), FxxVQxFTG(4/27), and FxxVQxVTG(2/27). Moreover, most *EgrVQs*(25/27) only contained an exon. Our results were consistent with *A. thaliana*, poplar, and rice [7,17,20], which indicated that *VQ* genes were relatively conservative. However, motif FxxVQxLSG(1/27), which was not shown in *A. thaliana*, rice, poplar, or grape, was found in *E. grandis* (Table 3).

Additionally, according to 20 conserved motifs in the analysis of *EgrVQ* genes with Multiple Expectation Maximization for Motif Elicitation (MEME), motif1 was presented in all of the *EgrVQ* genes (Figure 5). This result was consistent with other plants, like poplar [7] and moso bamboo [8,27], implying that the motif 1 imparts specific functions to the VQ protein. Moreover, it was found that several *cis*-elements that related to plant hormones and abiotic stressors were identified in the promoter region of the *EgrVQ* genes (Table 4). Similarly, most of these *cis*-elements occurred in other plants, like rice [17], poplar [7], and bamboo [8,27]. Interestingly, DRE is a *cis*-acting element that is involved in salt stress,cold stress, and dehydration, which was only presented in *EgrVQ20* and *EgrVQ21* (Table 4), and also seldom occurred in other plants.

In a previous study, VQ proteins participated in regulating diverse developmental processes, especially in response to biotic and abiotic stressors [20]. SA and JA are vital defense signaling molecules to response to pathogen infection and other abiotic stress conditions. In our study, the transcription of most *EgrVQ* genes was altered under SA and MeJA treatments. Interestingly, the expression of many *EgrVQ* genes was significantly increased at 168 h of SA treatment, which was not studied in moso bamboo [27], *A. thaliana* [20], and *Vitisvinifera* [28]. It is worth noting that *EgrVQ2*, *EgrVQ18*, and *EgrVQ22*, which were homologs of *AtVQ23* and *AtVQ16*, were highly expressed at 168 h under SA treatment. In *A. thaliana*, the over-expression of *AtVQ23* induced hyper-activate defense-related genes in plants following pathogen infection or SA and MeJA treatments, which enhanced resistance to infection by *Pseudomonassyringae* [29]. Meanwhile, *AtVQ16* could regulate the immune response by regulating *AtWRKY33* and further stimulating the DNA binding activity [30]. It was also found that *EgrVQ*22 expression also increased under MeJA treatment. However, *EgrVQ2* expression was not significantly changed and *EgrVQ18* expression was slightly down-regulated under MeJA treatment. Hence, these results implied that *EgrVQ18* could respond to SA treatment and it was further involved in SA-mediated defense responses by the long-term effect. *EgrVQ22* could also respond to SA and MeJA treatments.

Moreover, *AtVQ22*, which was positively regulated by the COI1 (CORONATINEINSENSITIVE1)-dependent signaling pathway, was a master controller regulating JA-mediated plant response against insects and pathogens [16,31]. *EgrVQ8*, which was homologous with *AtVQ22*, was induced at 6 h under MeJA treatment and 168 h under SA treatment, indicating that *EgrVQ8* has a similar function in

E. grandis. However, *EgrVQ26*, which is another homologous gene with *AtVQ22*, was not changed under MeJA treatment.

Furthermore, ABA and BR were also important in the response of plants to abiotic stressors [32,33]. In our study, most *EgrVQs* were up-regulated at 1 h–24 h when the plants were exposed to ABA, which was similar to moso bamboo *VQ* genes [8]. Meanwhile, only three genes showed up-regulation and multiple *OsVQ* genes showed down-regulation after 12 h under ABA treatment [17]. However, we found that 22 *EgrVQs* showed significantly down-regulation at 168 h under ABA treatment. These results indicated that the VQ genes had special roles in response to ABA treatment in *E. grandis*. It was also found that *EgrVQ* subfamily I members, including *EgrVQ2*, *EgrVQ13*, *EgrVQ14*, *EgrVQ22*,and *EgrVQ18*, were highly expressed under ABA treatment,which was similar with *OsVQ6*, *OsVQ7*, and *OsVQ10* belonging to subfamily. These results implied that these genes have similar functions and they might be related to the ABA signaling pathway. Specifically, *AtVQ15* regulated the osmotic stress tolerance with *WRKY25* and *WRKY51* in *A. thaliana* [20]. Its homologous gene, *EgrVQ27*, was significantly up-regulated after the ABA treatment, suggesting that *EgrVQ27* might affect osmotic stress tolerance in *E. grandis*. In addition, we found that most of the *EgrVQ* genes were highly expressed under the BR treatment (Figure 7A). BRASSINAZOLE-RESISTANT 1(BZR1), which is a key transcription factor (TF) in the BRs signal pathway, can trigger the expression of *FLS2* and *SNC1* to enhance pathogen resistance in plants [34,35].These results implied that *EgrVQ* genes might take part in enhancing the pathogen resistance of the BR signal pathway.

Plants also need to evolve efficient defense systems to protect themselves from abiotic stressors, like extreme temperature, salt, and so on. AtVQ9 acted as a repressor of WRKY8 to maintain a balance in the activation of WRKY8-mediated signaling pathways that are involved in salt stress in *A. thaliana* [22]. In our study, *EgrVQ5* and *EgrVQ9*, both homologous of *AtVQ9*, showed the same expression trends and they were increased at 168 h under NaCl treatment (Figure 8C). This indicated that the *VQ* genes were relatively conservative and were also involved in salt stress. In contrast, another homologous gene, *EgrVQ11*, was decreased at 168 h. *AtVQ15* was induced by dehydration and high salinity, whereas lines that overexpressed this gene showed an increased desensitivity to salt stress during seed germination and seedling growth [20,22]. Interestingly, its homologous gene, *EgrVQ1*, showed obviously decreased expression under NaCl treatment. We speculate that there is a difference in the woody and herbaceous plant.

Furthermore, few previous studies have been conducted on the function of *VQ* genes under heat and cold stresses in the woody plants. Our results showed that 23 and 24 *EgrVQ* genes were up-regulated at 1 h or 6 h, respectively, and then down-regulated at 168 h under cold and heat treatments, respectively (Figure 8A,B), which was consistent with Chinese Cabbage (*Brassica pekinensis*) [36]. It is noteworthy that most of the *EgrVQ* genes showed similar expression patterns between heat and cold treatment, suggesting that *EgrVQs* existed in the same response pathway to heat and cold stress.

Overall, our results implied various stressors motivated or repressed a batch of *EgrVQs*. These results might aid in the selection of available candidate genes in the *EgrVQ* gene family for deeper functional characterization.

4. Materials and Methods

4.1. Identification of VQ Gene Family in E. grandis

We conducted a systematic search on related bioinformatics analysis websites in order to acquire all of the proteins in *Eucalyptus*. Firstly, the sequences of *VQ* genes of *A. thaliana*, rice, and poplar were downloaded. Afterwards, the *VQ* genes were searched in *E. grandis* was using in NCBI (https://www.ncbi.nlm.nih.gov/), UniProt (https://www.uniprot.org), and EucGenIE (https://eucgenie.org/) databases. Next, all of the predicted *VQ* genes were subjected to SMART (http://smart.emblheidelberg.de/) and Pfam (https://pfam.sanger.ac.uk/) to confirm that they contained the VQ motif (PF05678).

Lastly, physical parameters of *VQ* genes, including open reading frame (ORF) length, protein length, isoelectric point (pI), and molecular weight were calculated in ExPASy (http://www.expasy.org/tools) [37]. Subcellular localization was predicted using the WoLF PSORT (http://wolfpsort.org/) [38] and TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) servers [39].

4.2. Chromosomal Location, Gene Duplication, and Identification of Paralogs and Orthologs of EgrVQ Genes

MapChart 2.30 software drew the picture of chromosomal location [40] on the basis of initial position information that was provided in Phytozomev12.1.6 (https://phytozome.jgi.doe.gov/pz/portal.html#). In addition, *EgrVQ* genes, which are located on duplicated chromosomal blocks, were considered to undergo segmental duplication [26]. Next, we used a previously described method for the analysis of paralogs and orthologs [41]. For the elected species, all-against-all nucleotide sequence similarity searches were done among the protein sequences by BLASTN software [42]. Specifically, in one species, a pair of matching sequences was shown as pairs of paralogs, which, when aligned, exceed 300bp and the identity was over 40% [7]. In two different species, the two sequences were defined as orthologs with the reciprocal best hits for each being within >300bp of the aligned sequences from two different species [7,8].

4.3. Promoter Analysis of EgrVQ Genes

The 2,000-bp upstream sequences of the transcriptional start sites of *EgrVQ*genes were submitted to PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [43] to identify the putative *cis*-elements.

4.4. Phylogenetic Analysis and Multiple Sequence Alignment of EgrVQ Proteins

Multiple sequence alignment of 27 VQ full-length protein sequences from *Eucalyptus* was performed using ClustalX2.11 [44]. A phylogenetic tree was constructed with the neighbor-joining (NJ) method in MEGA 7 (https://www.megasoftware.net/home) [45], with 1,000 bootstrap replicates. All VQ protein sequences from *A. thaliana*, poplarand rice were conducted using ClustalX2.11 and MEGA7 software.

4.5. Gene Structure and Conserved Motifs Analysis of EgrVQ Genes

The exon-intron structures of the *EgrVQ* genes were investigated using Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.ch) [46] by aligning cDNA to their corresponding genomic region. In addition, the Multiple Expectation Maximization for Motif Elicitation (MEME) program (http://meme-suite.org/tools/meme) [47] was used to identify the conserved motifs. The parameters were set, as follows: the optimum motif width ranged from 6 to 200, the maximum number of motifs was 20, and other parameters were set at default.

4.6. Plant Material and Treatments

We used approximately 20 cm tall, six-week-old *E. grandis* GL1 clones in hydroponics conditions. The trees were grown in the greenhouse of the Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou, China.

Hormone treatments were carried out by spraying the leaves of individual plants with 100 nM Epi-Brassinosteroid (Epi-BR) and 100 μM ABA, SA, and MeJA, respectively. For the salt stress treatment, the *E. grandis* plants were transferred to 200 mM NaCl solution and cultured for a total of 168 h; leaves were collected after 0, 1, 6, 24, and 168 h. For cold and heat stress treatments, the plants were transferred into 4 °C and 42 °C growth chamber. Then leaves were collected after 0, 1, 6, 24, 48, and 168 h of treatment. All of the samples were immediately frozen in liquid nitrogen and stored at −80 °C for subsequent total RNA extraction. The leaves from at least three plants were collected, and all treatments were performed in triplicate.

4.7. RNA Extraction, Semi-Quantitative Real-Time PCR (RT-PCR), and Real-Time Quantitative PCR (qRT-PCR)

Total RNA was extracted using Aidlab plant RNA kit (Aidlab Biotech, Beijing, China). The concentration and qualification were checked by NanoDrop™ One/OneC (ThermoFisher SCIENTIFIC, Waltham, MA, USA). Total RNA (1.0 μg) was used for first-strand cDNA synthesis using SuperScriptIII (Invitrogen), following the manufacturer's instructions. Gene-specific primers (Supplement Table S1) were designed by using Premier 5.0. qRT-PCR was performed on the Roche LightCycle 96 by using TB Green *Premix Ex Taq*II (TliRNaseH Plus) (RR420Q TaKaRa Biotechnology, Beijing, China), with a total sample volumeof 20 μL. The programs were 95 °C for 30 s, then 40 cycles of 95 °C for 5 s, and 65 °C for 34 s. The relative expression level was calculated while using the $2^{-\Delta\Delta CT}$ method [48]. The *EgrEF* was used as the reference gene [49].

4.8. Statistical Analysis

Statistical analysis was performed using OFFICE Excel and SPSS. The mean values and standard deviations (SDs) were obtained from three biological and three technical replicates; statistical significance was showed by one-way ANOVA, LSD (Least-Significant Difference).

5. Conclusions

Recently, an increasing number of *VQ* genes in different plants were analyzed in genome-wide characterization and evolutionary studies. However, there is little functional information regarding *VQ* genes in the woody plants. Our research showed an overall picture of *EgrVQ* genes, displaying the genome-wide identification, characterization, and expression of *EgrVQ* genes under various stressors. The results verified that the *EgrVQ* genes family either positively or negatively regulated multiple responses, including biotic and abiotic stressors. In addition, this study will provide a theoretical basis for further study of the mechanism of the stress resistance mechanism of the *VQ* gene family and indicated that most of the *VQ* genes participate between the process of biotic and abiotic stress resistance of *E. grandis*.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/7/1765/s1.

Author Contributions: Conceptualization, H.Y.; datacuration, Y.W. and Z.Q.; software, Y.W. and B.H.; supervision, B.Z. and C.F.; writing—original draft, H.Y.; All authors read and approved the final manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 31400554).

Acknowledgments: We'd like to thank Lee Min for his critical comments on this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jing, Y.; Lin, R. The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. *Plant Physiol.* **2015**, *169*, 371–378. [CrossRef] [PubMed]
2. Buscaill, P.; Rivas, S. Transcriptional control of plant defence responses. *Curr. Opin. Plant Biol.* **2014**, *20*, 35–46. [CrossRef]
3. Wang, A.; Garcia, D.; Zhang, H.; Feng, K.; Chaudhury, A.; Berger, F.; Peacock, W.J.; Dennis, E.S.; Luo, M. The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *Plant J.* **2010**, *63*, 670–679. [CrossRef] [PubMed]
4. Pecher, P.; Eschen-Lippold, L.; Herklotz, S.; Kuhle, K.; Naumann, K.; Bethke, G.; Uhrig, J.; Weyhe, M.; Scheel, D.; Lee, J. The Arabidopsis thaliana mitogen-activated protein kinases MPK3 and MPK6 target a subclass of 'VQ-motif'-containing proteins to regulate immune responses. *New Phytol.* **2014**, *203*, 592–606. [CrossRef] [PubMed]
5. Chen, Y.; Zhou, Y.; Yang, Y.; Chi, Y.-J.; Zhou, J.; Chen, J.-Y.; Wang, F.; Fan, B.; Shi, K.; Zhou, Y.-H. Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors. *Plant Physiol.* **2012**, *159*, 810–825. [CrossRef]

6. Li, N.; Li, X.; Xiao, J.; Wang, S. Comprehensive analysis of VQ motif-containing gene expression in rice defense responses to three pathogens. *Plant Cell Rep.* **2014**, *33*, 1493–1505. [CrossRef]
7. Chu, W.; Liu, B.; Wang, Y.; Pan, F.; Chen, Z.; Yan, H.; Yan, X. Genome-wide analysis of poplar VQ gene family and expression profiling under PEG, NaCl, and SA treatments. *Tree Genet. Genomes* **2016**, *12*, 124. [CrossRef]
8. Wang, Y.; Liu, H.; Zhu, D.; Gao, Y.; Yan, H.; Xiang, Y. Genome-wide analysis of VQ motif-containing proteins in Moso bamboo (*Phyllostachys edulis*). *Planta* **2017**, *246*, 165–181. [CrossRef]
9. Guo, J.; Chen, J.; Yang, J.; Yu, Y.; Yang, Y.; Wang, W. Identification, characterization and expression analysis of the VQ motif-containing gene family in tea plant (*Camellia sinensis*). *BMC Genom.* **2018**, *19*, 710. [CrossRef]
10. Liu, X.; Wang, X.; Pang, Y.; Liang, J.; Liu, S.; Sun, X.; Tang, K. Molecular cloning and characterization of a novel WRKY gene from Brassica chinensis. *Mol. Biol.* **2006**, *40*, 816. [CrossRef]
11. Chi, Y.; Yang, Y.; Zhou, Y.; Zhou, J.; Fan, B.; Yu, J.Q.; Chen, Z. Protein-Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol. Plant* **2013**, *6*, 287–300.
12. Xie, T.; Chen, C.; Li, C.; Liu, J.; Liu, C.; He, Y. Genome-wide investigation of WRKY gene family in pineapple: Evolution and expression profiles during development and stress. *BMC Genom.* **2018**, *19*, 490. [CrossRef]
13. Lin, Z.; Li, Y.; Wang, F.; Cheng, Y.; Fan, B.; Yu, J.-Q.; Chen, Z. Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *Plant Cell* **2011**, *23*, 3824–3841.
14. Song, W.; Zhao, H.; Zhang, X.; Lei, L.; Lai, J. Genome-Wide Identification of VQ Motif-Containing Proteins and their Expression Profiles Under Abiotic Stresses in Maize. *Front. Plant Sci.* **2016**, *6*, 1177. [CrossRef]
15. Perruc, E.; Charpenteau, M.; Ramirez, B.C.; Jauneau, A.; Galaud, J.P.; Ranjeva, R.; Ranty, B. A novel calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in Arabidopsis thaliana seedlings. *Plant J.* **2004**, *38*, 410–420. [CrossRef]
16. Hu, P.; Zhou, W.; Cheng, Z.; Fan, M.; Wang, L.; Xie, D. JAV1 Controls Jasmonate-Regulated Plant Defense. *Mol. Cell* **2013**, *50*, 504–515.
17. Kim, D.Y.; Kwon, S.I.; Choi, C.; Lee, H.; Ahn, I.; Park, S.R.; Bae, S.C.; Lee, S.C.; Hwang, D.J. Expression analysis of rice VQ genes in response to biotic and abiotic stresses. *Gene* **2013**, *529*, 208–214.
18. Wang, X.; Zhang, H.; Sun, G.; Jin, Y.; Qiu, L.J. Identification of active VQ motif-containing genes and the expression patterns under low nitrogen treatment in soybean. *Gene* **2014**, *543*, 237–243.
19. Rushton, P.J.; Somssich, I.E.; Ringler, P.; Shen, Q.J. WRKY transcription factors. *Trends Plant Sci.* **2010**, *15*, 247–258. [CrossRef]
20. Viana, V.E.; Busanello, C.; da Maia, L.C.; Pegoraro, C.; de Oliveira, A.C. Activation of rice WRKY transcription factors: an army of stress fighting soldiers? *Curr. Opin. Plant Biol.* **2018**. [CrossRef]
21. Birkenbihl, R.P.; Diezel, C.; Somssich, I.E. Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *Plant Physiol.* **2012**, *159*, 266–285.
22. Hu, Y.; Chen, L.; Wang, H.; Zhang, L.; Wang, F.; Yu, D. Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J.* **2013**, *74*, 730–745. [CrossRef]
23. Batish, D.R.; Singh, H.P.; Kohli, R.K.; Shalinder, K. Eucalyptus essential oil as a natural pesticide. *For. Ecol. Manag.* **2008**, *256*, 2166–2174. [CrossRef]
24. Barton, A. The Oil Mallee Project: A Multifaceted Industrial Ecology Case Study. *J. Ind. Ecol.* **2010**, *3*, 161–176.
25. Popescu, C.-M.; Popescu, M.-C.; Singurel, G.; Vasile, C.; Argyropoulos, D.S.; Willfor, S. Spectral characterization of eucalyptus wood. *Appl. Spectrosc.* **2007**, *61*, 1168–1177.
26. Myburg, A.A.; Grattapaglia, D.; Tuskan, G.A.; Hellsten, U.; Hayes, R.D.; Grimwood, J.; Jenkins, J.; Lindquist, E.; Tice, H.; Bauer, D. The genome of Eucalyptus grandis. *Nature* **2014**, *510*, 356–362.
27. Peng, Z.; Lu, Y.; Li, L.; Zhao, Q.; Feng, Q.; Gao, Z.; Lu, H.; Hu, T.; Yao, N.; Liu, K. The draft genome of the fast-growing non-timber forest species moso bamboo (Phyllostachys heterocycla). *Nat. Genet.* **2013**, *45*, 456–461.
28. Wang, M.; Vannozzi, A.; Wang, G.; Zhong, Y.; Corso, M.; Cavallini, E.; Cheng, Z.M. A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation with WRKY proteins. *Front. Plant Sci.* **2015**, *6*, 417. [CrossRef]
29. Xie, Y.D.; Li, W.; Guo, D.; Dong, J.; Zhang, Q.; Fu, Y.; Ren, D.; Peng, M.; Xia, Y. The Arabidopsis gene SIGMA FACTOR-BINDING PROTEIN 1 plays a role in the salicylate- and jasmonate-mediated defence responses. *Plant Cell Environ.* **2010**, *33*, 828–839.

30. Li, Y.; Jing, Y.; Li, J.; Xu, G.; Lin, R. Arabidopsis VQ MOTIF-CONTAINING PROTEIN29 Represses Seedling Deetiolation by Interacting with PHYTOCHROME-INTERACTING FACTOR1. *Plant Physiol.* **2014**, *164*, 2068–2080. [CrossRef]
31. Ali, M.R.; Uemura, T.; Ramadan, A.; Adachi, K.; Nemoto, K.; Nozawa, A.; Hoshino, R.; Abe, H.; Sawasaki, T.; Arimura, G.-I. The Ring-type E3 Ubiquitin Ligase JUL1 Targets the VQ-motif Protein JAV1 to Coordinate Jasmonate Signaling. *Plant Physiol.* **2019**, *179*, 1273–1284. [CrossRef]
32. Fariduddin, Q.; Mir, B.A.; Yusuf, M.; Ahmad, A. 24-epibrassinolide and/or putrescine trigger physiological and biochemical responses for the salt stress mitigation in *Cucumis sativus* L. *Photosynthetica* **2014**, *52*, 464–474. [CrossRef]
33. Clouse, S.D. Brassinosteroid/Abscisic Acid Antagonism in Balancing Growth and Stress. *Dev. Cell* **2016**, *38*, 118–120.
34. Ali, S.S.; Kumar, G.B.S.; Mojibur, K.; Doohan, F.M. Brassinosteroid enhances resistance to fusarium diseases of barley. *Phytopathology* **2013**, *103*, 1260–1267.
35. Nawaz, F.; Naeem, M.; Zulfiqar, B.; Akram, A.; Ashraf, M.Y.; Raheel, M.; Shabbir, R.N.; Hussain, R.A.; Anwar, I.; Aurangzaib, M. Understanding brassinosteroid-regulated mechanisms to improve stress tolerance in plants: A critical review. *Environ. Sci. Pollut. Res. Int.* **2017**, *24*, 15959–15975.
36. Zhang, G.; Wang, F.; Li, J.; Ding, Q.; Zhang, Y.; Li, H.; Zhang, J.; Gao, J. Genome-Wide Identification and Analysis of the VQ Motif-Containing Protein Family in Chinese Cabbage (*Brassica rapa* L. ssp. Pekinensis). *Int. J. Mol. Sci.* **2015**, *16*, 28683–28704. [CrossRef]
37. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]
38. Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [CrossRef]
39. Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [CrossRef]
40. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [CrossRef]
41. Blanc, G.; Wolfe, K.H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **2004**, *16*, 1667–1678. [CrossRef]
42. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
43. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
44. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Res.* **1997**, *25*, 4876–4882.
45. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
46. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2014**, *31*, 1296–1297. [CrossRef]
47. Timothy, L.B.; Mikael, B.; Fabian, A.B.; Martin, F.; Charles, E.G.; Luca, C.; Jingyuan, R.; Wilfred, W.L.; William, S.N. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208.
48. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [CrossRef]
49. De Oliveira, L.A.; Breton, M.C.; Bastolla, F.M.; Camargo Sda, S.; Margis, R.; Frazzon, J.; Pasquali, G. Reference Genes for the Normalization of Gene Expression in Eucalyptus Species. *Plant Cell Physiol.* **2012**, *53*, 405–422.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

International Journal of Molecular Sciences Editorial Office
E-mail: ijms@mdpi.com
www.mdpi.com/journal/ijms

Academic Open
Access Publishing

www.mdpi.com

ISBN 978-3-0365-8091-3