genes

Recent Advances in Genetics and Breeding of Major Staple Food Crops

Edited by
Joong Hyoun Chin, B.P. Mallikarjuna Swamy and Yeisoo Yu

Printed Edition of the Special Issue Published in *Genes*

www.mdpi.com/journal/genes

MDPI

Recent Advances in Genetics and Breeding of Major Staple Food Crops

Recent Advances in Genetics and Breeding of Major Staple Food Crops

Editors

Joong Hyoun Chin
B.P. Mallikarjuna Swamy
Yeisoo Yu

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

Editors

Joong Hyoun Chin
Sejong University
Korea

B.P. Mallikarjuna Swamy
International Rice Research Institute
Philippines

Yeisoo Yu
DNACare co., Ltd.
Korea

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 *Genes* (ISSN 2073-4425) (available at: https://www.mdpi.com/journal/genes/special_issues/food_crops).

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.

ISBN 978-3-0365-0976-1 (Hbk)
ISBN 978-3-0365-0977-8 (PDF)

Contents

About the Editors

Joong Hyoun Chin (Prof. Dr.) Department of Integrative Biological Sciences and Engineering, Sejong University, Korea. Interests: crop genetics, molecular breeding; plant genomics; abiotic stress; variety development.

B.P. Mallikarjuna Swamy (Dr.) Rice Breeding and Biofortification, The International Rice Research Institute, The Philippines. Interests: crop genetics; breeding; cereals; MAS; genomic selection; QTL; varietal development.

Yeisoo Yu (Dr.) Bioinformaticist and CEO, DNACare Co., Ltd., Korea. Interests: genomics; bioinformatics; molecular marker; genomic selection; data integration.

Preface to "Recent Advances in Genetics and Breeding of Major Staple Food Crops"

Food and nutritional security is of prime importance for achieving zero hunger, improved health, better income and livelihoods of global population. However, the world population is increasing rapidly, which is estimated to reach 10 billion by 2050 and meeting their food and nutritional needs is a major global challenge. As per some recent predictions, we may have to produce 60% more to feed the population, which will be a very tedious task as it needs to be achieved with limited resources such as land, water, labor etc. Moreover, several production constraints, such as the impact of climate change, biotic stresses, abiotic stresses, soil nutrition deficiency problems, etc., have further compounded global food security by affecting the production of food crops, especially major cereals such as rice, wheat and maize—these three crops cover more than 50% of global caloric intake. Hence, there is an urgent need to produce better crop varieties by efficiently using genetic resources and molecular tools. The harnessing of genetic diversity by novel allele mining assisted by recent advances in biotechnological and bioinformatics tools will enhance the utilization of the hidden treasures in the gene bank, and their efficient utilization in crop breeding to produce climate smart, multi-stress tolerant, and nutritious crop varieties. The elucidation of the genetic, physiological, and molecular basis of useful traits are key activities for variety development. High-throughput genotyping systems assisted by bioinformatics and data science provide efficient and easy tools for geneticists and breeders. Recently, new breeding techniques applied in some food crops have become game-changers in the global food crop market.

In this Special Issue, we focus on modern food crop genetics and breeding: plant molecular systems focusing on food crops; olant genetic diversity—QTL and gene identification utilizing high-throughput genotyping systems and their validation; new breeding techniques in food crops—targeted mutagenesis, genome editing, etc.; abiotic and biotic stresses—QTL/gene identification and their molecular physiology; plant nutrition, grain quality improvement, and yield enhancement.

Joong Hyoun Chin, B.P. Mallikarjuna Swamy, Yeisoo Yu
Editors

genes

MDPI

Review

The Use of Genetic and Gene Technologies in Shaping Modern Rapeseed Cultivars (*Brassica napus* L.)

Linh Bao Ton [1], Ting Xiang Neik [2] and Jacqueline Batley [1,*]

[1] School of Biological Science, The University of Western Australia, Perth, WA 6009, Australia; linh.ton@research.uwa.edu.au

[2] Sunway College Kuala Lumpur, No. 2, Jalan Universiti, Bandar Sunway, Selangor 47500, Malaysia; tingxiang@gmail.com

* Correspondence: jacqueline.batley@uwa.edu.au; Tel.: +61-(0)8 6488 5929

Received: 1 September 2020; Accepted: 27 September 2020; Published: 30 September 2020

Abstract: Since their domestication, Brassica oilseed species have undergone progressive transformation allied with the development of breeding and molecular technologies. The canola (*Brassica napus*) crop has rapidly expanded globally in the last 30 years with intensive innovations in canola varieties, providing for a wider range of markets apart from the food industry. The breeding efforts of *B. napus*, the main source of canola oil and canola meal, have been mainly focused on improving seed yield, oil quality, and meal quality along with disease resistance, abiotic stress tolerance, and herbicide resistance. The revolution in genetics and gene technologies, including genetic mapping, molecular markers, genomic tools, and gene technology, especially gene editing tools, has allowed an understanding of the complex genetic makeup and gene functions in the major bioprocesses of the Brassicales, especially Brassica oil crops. Here, we provide an overview on the contributions of these technologies in improving the major traits of *B. napus* and discuss their potential use to accomplish new improvement targets.

Keywords: canola; *Brassica napus*; genetics; gene technology; genomics; disease resistance

1. Introduction

Rapeseed includes the Brassica oilseed crops in which *Brassica napus* is the leading crop globally, while *Brassica juncea*, *Brassica rapa*, *Brassica carinata*, and *Brassica nigra* are cultivated in selected regions of the world [1]. The term "canola" refers to the modified rapeseed low in erucic acid (<2%) and glucosinolate (<30 µmol/g of dried defatted meal) content, which includes cultivars of *B. napus*, *B. rapa*, and *B. juncea* [2]. The allopolyploid oilseed plants were formed following interspecific hybridization events, for example, *B. napus* (AACC, 2n = 38) was formed by hybridization of *B. oleracea* (CC, 2n = 18) and *B. rapa* (AA, 2n = 20) ~7500 years ago [3]. Modern approaches in evolutionary analysis have attempted to consolidate the origin of *B. napus* [4,5]. The use of rapeseed oil in lamps has been recorded since 2000 BC in India and since the 13th century in Europe [2] with additional uses in making food and soap [6]. Rapeseed cultivars were introduced to Canada in 1936 [7] for lubricant production for war ships, and was first commercially grown in Australia in 1969 [2].

Intensive breeding programs in *B. napus* were initiated in many countries around 1970, including Australia [2]. Back-crossing with the low-erucic- acid line, Liho, brought successful release of the first low erucic-acid cultivars of *B. napus* and *B. rapa* in 1968 and 1971, respectively [7]. High oleic and low linolenic (HOLL, > 65% oleic acid and < 3% linolenic acid) rapeseed genotypes were created by chemical mutagenesis using ethyl methanesulfonate (EMS) [8,9], and HOLL varieties have been commercially grown in Canada since 2005 and in Australia since 2006 [10]. The emergence of genomic and genetic tools has facilitated breeding programs [11,12] leading to the release of multiple rapeseed/canola varieties such as high erucic acid rapeseed (HEAR) [13], for non-food purposes, and varieties for

food industries e.g., high lauric acid canola Laurical™, Roundup Ready®(Monsanto, Missouri, US), InVigorTM (BASF), with provisional herbicide tolerance (HT) and resistance against major pathogens. The parental lines for rapeseed breeding programs varied depending on the geographic location, with the progenitors contributing towards the three ecotypes of *B. napus*, spring type which is widely cultivated in Canada, Australia, and northern Europe, winter type *B. napus* which is predominant in Asia and the remaining area of Europe [14], and semi-winter type as the primary rapeseed in China [15]. The increasing human population and sustainable energy initiatives have led to a high demand of canola oil for the food industry, biofuel, and various industrial purposes [16], prompting investments for further improvement of *B. napus*, with emphasis on the major traits relating to oil yield and quality, meal quality, herbicide tolerance, biotic and abiotic stress tolerance [17,18]. Fast-pace development of genetic and gene technologies has facilitated the identification of those breeding targets as well as achieving significant advances in the improvement of canola genomics resources.

During the last 20 years, dramatic innovations in next-generation sequencing (NGS) technologies have led to a significant drop in sequencing cost and offered exciting opportunities to explore plant genomes [19]. Recent rapeseed genomes were assembled using Illumina high-throughput short read sequence technologies and third generation sequencing technologies, notably, single molecule real-time (SMRT) sequencing (PacBio) and Nanopore sequencing technologies (Oxford Nanopore Technologies, Oxford, UK) capable of generating long-read sequences [20,21]. Genome assembly in highly complex genomic regions frequently found in *Brassica* species can be greatly improved using optical mapping [22,23] and chromosome conformation capture (Hi-C) technologies [23].

High-throughput SNP marker identification and genotyping of *Brassica* species, achieved through NGS, allow efficient identification of quantitative trait loci (QTL) and generation of high density genetic maps, which means there is a higher chance of detection of a candidate gene linked to a nearby SNP [24,25]. The SNP markers are useful for genome-wide association studies (GWAS), marker-assisted selection (MAS), genomic selection (GS), and germplasm identification [26]. For example, SNP markers have been widely deployed in mapping QTL conferring resistance to blackleg, Sclerotinia stem rot [27] and clubroot [28] in *Brassica*. In addition, genotyping can be implemented quickly using the *Brassica* 60K Illumina Infinium™ array with the capability to genotype of 52,157 SNPs in *B. napus* [29]. High-throughput genotyping technologies allow large-scale genomic characterisation of the huge *Brassica* germplasm collections worldwide comprising > 74,000 *Brassica* accessions mainly stored in The Netherlands, Norway, Spain, UK, US, and Australia [30] to facilitate improvement and broaden the genetic diversity of *Brassica* varieties.

Limitations in gene pool diversity [31,32], and the time and laborious constraints in canola breeding can now be overcome by using genetic modification [32] and genome editing technologies whereby transgenes or mutations can be directly introduced into plants, providing additional ways to investigate gene functions in biological processes [33–36]. The most commonly applied genome editing tool in eukaryotes in the last 8 years, the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas) system [37,38], is gaining more interest compared to conventional genetic modification methods due to its high efficiency in targeted nucleotide modification and generation of a transgene-free end-product [39]. In a CRISPR/Cas system, a single-guide RNA (sgRNA) binds to the Cas protein and directs it to one or more specific target sequences [40] marked by a short sequence, protospacer-adjacent motif (PAM), located at the 3′ downstream position [37] and varying in accordance with the type of Cas enzyme utilized [40,41]. Recently, a platform for induction of mutagenesis in oilseed rape via the CRISPR/Cas system has been developed [42], facilitating early utilization of this tool in improving agronomic traits of *B. napus* e.g., plant height [43], silique development [44], pod shatter resistance [45], and flowering time [46]. This genome editing tool provides a faster route to interpret the relationship between genes and phenotypic traits controlled by complex genetic structure, biosynthetic pathways, and regulatory elements [26].

Development of breeding techniques and the genetics and genomics associated technologies described above have improved the quality and agronomic traits of canola varieties, which make

the current canola varieties well-suited for expectations of growers, processing sectors, and final users [47,48]. The advent of technologies, including omics technologies, genetic transformation, and genome editing, have facilitated deeper understanding of the physiological and biochemical processes regulating the development of the major phenotypic traits of *Brassica* oilseed crops, which has brought tremendous improvements in the desirable traits of canola in the last decade [49,50]. This review provides a summary of their impressive contributions towards improvement and innovations of canola varieties.

2. Improvement and Innovations of Canola Varieties

2.1. Higher Resolution for Rapeseed Genome Characterization

Genomic data is an integral part in developing breeding strategies and exploiting genetic potential of germplasm. The revolution in sequencing technologies has accelerated genome research, where it used to be an obstacle for polyploid species [21]. Using a combination of NGS and Sanger sequencing technologies, the first genome of *B. napus* was sequenced, from which its evolutionary history was clarified with the linkages to its ancient ancestor [5]. Meanwhile, the combination of PacBio's SMRT and NGS short read sequencing technologies have been deployed to study the various forms and complexity of the gene transcripts responsible for another development in *B. rapa* (Chinese cabbage), providing comprehensive transcriptome data for better accuracy of genome annotation [51]. Version 3 of the *B. rapa* genome was assembled using combinations of SMRT, optical mapping (BioNano), and Hi-C technologies with up ~ 30-fold improvement compared to the previous versions [52]. Other more recent genome assemblies for the *Brassica* species are *B. olerecea* accessions JZS v2 [53], *B. nigra* [54], *B. napus* 'Darmor-bzh [55], and German winter *B. napus* Express 617 [56]. These *Brassica* reference genomes were sequenced and assembled long-read sequencing technologies using SMRT and Oxford Nanopore Technologies (MinION and higher throughput version of the MinION, PromethION) complemented with Illumina data and verified by optical mapping and/or Hi-C data [52,56–58]. Qiao et al. [59] have proven cytoplasmic genomes from mitochondria and cytoplasms would be a highly effective approach for phylogenetic and evolutionary studies on *B. napus* including employing this resource in improved breeding.

2.2. Pan-Genome

Although there is an increasing amount of reference genome data available for *Brassica* species, a high-quality reference genome comprised of genome data of individuals within a species is mandatory for comparison studies and evaluation of genetic diversity, which is essential for improvement of breeding [57]. The original effort in assembling a core genome representing a population of *Streptoccocus agalactiae* [60] introduced the new term "Pan-genome". Pan-genomics exploits the genomic content of all the individual lines of a species under study, including closely related species such as crop wild relatives, differentiating the core genes which are genes present in every individual of the species, from variable genes, which are genes present only in one or some individuals [61,62]. The variations in the genomes can be in many forms including SNP, copy number variation, and presence-absence variation (PAV). These may arise due to many rounds of DNA exchange events happening within or between the (sub)genomes, accumulated within a species during the evolution from the ancestral species and subsequent breeding history with different genetic backgrounds, resulting in high species diversity [63]. The first pan-genome of *B. rapa* was constructed by comparing genomic data of *B. rapa* species including Chinese cabbage cultivar, Chiifu; an oil-like rapid cycling line (RC-144) and a Japanese vegetable turnip [64], which is 283.84 Mb in size and includes 41,052 predicted gene models, of which 90% are the common genes and 9% are unique genes only found in one of the three genomes. Constructed by an iterative assembly approach, the pan-genome of *B. oleracea* revealed the number of variable genes was roughly more than 20% of the total pan-genome gene number; 12,806 (out of 61,379 pan-genome genes) [65], whilst in *B. napus* over one third of the genes were

variable: 35,481 (out of the total 94,013 genes in the pan genome) [66]. Recently, by using a de novo approach, a pan-genome representing 9 *B. napus* genomes was obtained comprising of 58,714 (~56%) core-gene clusters present in at least seven genomes, 44,035 (~42%) dispensable gene clusters and around 2% specific genes [57].

Further dissection of gene functions within pan-genomes across species showed dispensable genes may have higher mutation rates, and are less functionally conserved compared to core genes, of which functions are associated with responsive mechanisms to biotic and abiotic signals [62]. In the pan-genome of *B. oleracea*, Bayer et al. [67] found that the most abundant resistance gene analog (RGA) candidates in the additional pan-genome contigs were leucine rich repeat (NBS-LRR) genes compared to receptor-like kinase (RLK) and receptor-like protein (RLP) genes, and identified 59 RGA candidates linked to known QTL of *Sclerotinia*, clubroot, and *Fusarium* resistance. It was found that 753 out of the total 1749 RGAs in the pan-genome of *B. napus* [68] are variable, 368 of which are not present in the reference genome (cv. Darmor-bzh), suggesting many genes relating to plant resistance mechanisms may be unknown. This is consistent with the fact that the core genes (10584 SNPs identified, 70.97%) contained more SNPs than variable genes (4734 SNPs identified in 299 genes), and 688 SNPs were found in 106 RGAs within QTL (*LepR1*, *LepR2*, *Rlm1*, *Rlm3*, *Rlm4*, *Rlm7*, and *Rlm9*) conferring blackleg resistance in *B. napus* [68]. Transcriptomic analysis coupled with pathotyping could assist functional annotation of RGA candidates [69,70] while their roles in responsive mechanisms could be validated using gene technology or/and CRISPR/Cas tool [71]. Pan-genomes of *Brassica* oilseed crop provide insights on the genetic variations leading to phenotypic variations, differential trait expression of individuals within species which supports finding new candidate genes. Higher coverage of pangenomes allows more precise characterization and prediction of gene content of individuals within a species [63].

3. Breeding for Economically Important Agronomic Traits of *B. napus*

3.1. Oil Content and Specialty as Priority

Canola oil, the main product of *B. napus*, is the most consumed vegetable oil worldwide [72]. Canola food oil possesses a unique fatty acid profile compared to other types of vegetable oil: low in saturated fatty acid (SFAs), typically 7%; high monounsaturated fatty acids (MUFAs) and polyunsaturated fatty acid (PUFAs), comprised of 61% oleic acid, 21% linoleic acid, and 11% alpha-linolenic acid (ALA) [73]. Phytosterols and vitamin E are additional healthy components of canola oil [74]. The health benefits of a canola oil-based diet were clinically proven as reduction of blood glucose levels and risks of coronary heart disease, promoting immunity and prevention of tumor cell growth, contributed by MUFAs, PUFAs, and vitamin E [75]. For cooking purposes and biofuel production, canola oil with high oleic acid (C18:1) and low linolenic acid (C18:3) is preferred through providing high stability and longer shelf-life for the oil [12,76].

There are tailored profiles of canola fatty acids required for specified industries and industrial applications which were achieved through conventional breeding and mutagenesis during the 1990s [76], Figure 1. To meet the demands of the increasing market of cooking oil, priority for improvement of canola varieties is an increase in oleic acid (C18:1 or ω-9) level and decrease in linolenic acid (C18: 3 or ω-3) and SFAs with specific targets varying by country. For example, in Australia, the aim is for up to 67–75% for oleic acid and less than 3% linolenic acid [77]. Expansion of the acreage of high oleic and specialty canola varieties (to 33% of canola acres) is also part of Canada's scope for the canola crop in the period 2015–2025, as well as further reducing SFAs to below 6.8%, and especially palmitic and stearic levels below 4% [76,78]. To maximize canola intake in both biofuel and feed industries, varieties with high oleic acid and extremely low glucosinolate content are targeted [12]. High oil yield (40–45% of the mass) and low production cost are features making canola a potential crop for producing high demand fatty acids [36].

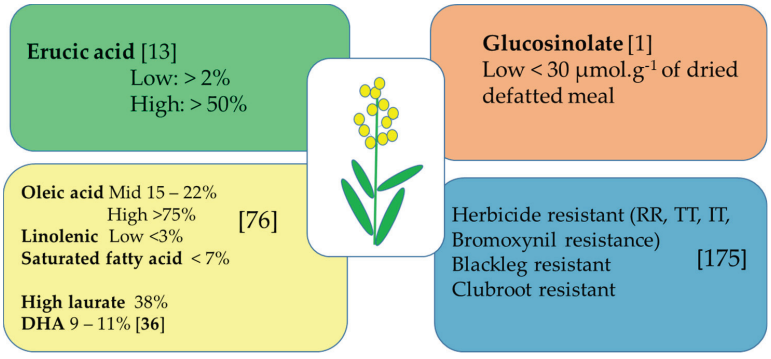


Figure 1. Desired optimal traits in current canola varieties, specifically for oil and meal content, herbicide, and disease resistance.

Understanding biosynthesis of the major C18 unsaturated fatty acids (UFA) in prokaryotes and eukaryotes has been well documented since the 1990s, where synthesis pathways, key fatty acid synthetic enzymes, and key regulating factors have been identified [79]. In *B. napus*, fatty acid desaturase (*FAD*) genes, specifically *FAD2* and *FAD3*, together with stearoyl-acyl carrier protein desaturase (*SAD*) catalyzing desaturation of steric acid (C18:0) to C18:1 are a focus for controlling oil quality [79], however they were also found to be involved in abiotic stress tolerance such as high cadmium (250 μM) and salinity (100 mM) conditions [80]. Elevated oleic levels were reported in lines with mutated *FAD2* genes [35,81]. Targeted mutagenesis of a *FAD2-Aa* allele of *B. napus* using CRISPR/Cas9 produced the heritable mutant, *fad2_Aa* allele with a 4-bp deletion, which was separated from the transgenes by backcrossing [35]. By genome wide analysis of the *FAD* gene family, 84, 45, and 44 *FAD* genes were identified in *B. napus*, *B. rapa*, and *B. oleracea* genomes, respectively, with different distribution, which were assigned into four and six subfamilies in term of soluble and membrane-bound *FAD*s, respectively [82]. Based on 201,187 SNP markers developed from SLAFseq (specific locus amplified fragment sequencing) and GWAS of four important fatty acid content traits (erucic acid, oleic acid, linoleic acid, and linolenic acid), 148 SNP loci were found significantly associated with these traits and 20 orthologs of the candidate genes regulating the fatty acid biosynthesis (fatty acid synthesis, desaturation, elongation, and metabolism) and 14 candidate genes on chromosomes A8 and C3 with potential contributions were identified [83]. Roles of *BnTT8* genes in *B. napus*, the conserved gene complex controlling flavonoid accumulation in plant crops, have been elucidated by using the CRISPR/Cas9 tool to induce targeted mutations at *BnA09.TT8* and *BnC09.TT8b* genes, in which the mutant lines produced seeds with a yellow seed coat, significant increase in seed oil and protein content, and altered FA composition [84].

High demand long chain-unsaturated fatty acids, such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which were once exclusively obtained from fish, are now novel components of transgenic *B. napus* lines [36,85]. In 2005, by transformation using *A. tumefaciens* containing 6 different constructs of desaturase genes from *Thraustochytrium* sp., *Pythium irregular* and *Calendula officinalis*, and fatty acid elongase from *Physcomitrella patens* in *B. juncea*, very long chain PUFAs such as arachidonic acid (AA) and EPA were produced at levels of up to 25% and 15% of total seed fatty acids, however, the DHA biosynthetic pathway needed further optimization [34]. Using a similar approach, Walsh et al. [85] produced transgenic *B. napus* expressing PUFA synthases (*OrfA*, *OrfB*, and hybrid *OrfC*) from *Schizochytrium* sp. ATCC 20888 (Schizo20888). These microalgal PUFA synthases assemble C2 units from malonyl-CoA into long chain PUFAs in cytoplasm with local NADPH supplement. The average DHA contents in T2 seeds from the inbred lines of the selected events were around 2.87–3.43%, and the total content of both DHA and EPA was around 4.4% in field-produced canola oil. In the latest transgenic DHA canola variety by Petrie et al. [36], seven fatty

acid biosynthesis genes from yeast and microalgae were de novo synthesized as a single fragment of 19,750 bp and regulate DHA production through the delta-6-desaturase aerobic long-chain (C20) polyunsaturated fatty acid synthesis pathways in *B. napus*. Levels of DHA from 9 to 11%, similar to those obtained from fish, were obtained from the best transformation event (NS-B50027-4) in open field trials in Australia and Canada. With the fact that the DHA canola is approved for cultivation for human and animal consumption of the oil in Australia and cultivation approval in the US [86,87], production of other valuable fatty acids through genetic engineering tools is quite promising.

Regarding renewable and sustainable resources, erucic acid from canola oil is an excellent raw material in producing polymers used in film manufacture, nylon, lubricant, and emolient industries [13,17]. Furthermore, with the increasing demand of canola oil for food and efforts to promote renewable resources of high erucic acid rapeseed (HEAR, 45–60% erucic acid), varieties have been grown in European countries for biofuel and raw oil supplement (<2% of total weight of crushed seeds) for human consumption [88,89]. By overexpressing the fatty acid elongase gene (*fae1*) simultaneously with the lysophosphatidic acid acyltransferase gene from *Limnanthes douglasii* (Ld-LPAAT), erucic acid synthesis was predominant over PUFA synthesis in competitive elongation to the triaclyglycerol backbone in transgenic lines, which resulted in an increase in erucic acid content of up to 72% and a PUFA content as low as 6% [13]. The roles of LPAAT genes of *B. napus*, BnLPAT2 and BnLPAT5, in regulating oil biosynthesis has been recently confirmed by targeted mutations using Cas9 with single-gRNAs and multi-sgRNAs whereby the resulted *Bnlat2* and *Bnlat5* mutants showed decreased oil content and enlarged oil bodies [33]. Cytoplasmic genomes are also attracting more interest as the new approach for maximizing oil content in canola [90].

3.2. Exploiting Canola Meal Potential

Despite the protein-rich content of canola meal (38%, by meal weight) and wide adoption of low glucosinolate (GSL) varieties, antinutritional compounds, typically, phenolics, sinapine, and phytate, are still challenging the use of by-products of canola oil production as animal feed [16,91]. The expansion of biofuel production from canola [89] has led to significantly additional volumes of canola meal as a by-product and promoted feed industry [16,92]. However, in the study of Skugor et al. [93], the replacement of rapeseed meal for soybean meal as 20% supplement of diet for growing finishing pigs in 3 months, produced elevated expression of major control factors of energy generation, reduced protein synthesis in muscle tissue, and greater expression of oxidative stress, which was caused by higher fiber content and GSL compounds and other secondary metabolites in rapeseed meal. Other studies on animal nutrition also had similar conclusions about antinutritional compounds in canola meal, including high fiber content, lignin, and GSLs [94,95].

Since the release of the first "double-low" canola cultivar with low levels of both erucic acid and glucosinolates (GSL), breeding efforts to further reduce GSL content in canola cultivars including *B. rapa*, *B. napus*, and *B. carinata* are still in progress [1]. While GSL inclusion in diet affects health of livestock e.g., damaging liver, kidney, and thyroid gland and impairing fertility [96], this secondary metabolite could be useful as a cancer prevention and plant protection agent [97,98]. The importance of GSLs in plant defense has been demonstrated [99] and explains the increase in susceptibility to pests and diseases of the "00" (low seed erucic acid and GSL) rapeseed cultivars [100].

Over the at least 15 years, knowledge on metabolism of GSLs in Brassica crops and Arabidopsis, and the underlying genetic architecture has been achieved [97,98,100,101]. To date, there are at least 130 different classes of plant GSL, almost all from the Brassicales order, summarized by Blažević et al. [102] and approximately 40 different GSLs synthesized in Arabidopsis [103]. Understanding the regulation of GSL biosynthesis through molecular genetic tools (SNP marker identification, QTL mapping) and gene technology (gene cloning, gene knock-out) enables controlling in planta GSL profiles and their targeted position for expression [100]. Using an associative Transcriptomics platform comprising 355,536 SNPs and a transcriptome reference, Kittipol et al. [100] found a region on chromosome A3 of *B. napus* had strong association with variation in root aromatic GSL and confirmed

the gene *Bna*.HAG3.A3 as a key regulator for the trait expression. Although studies of genetic factors controlling expression of GSL in vegetative parts and seeds of canola are being performed, they are clearly the keys to enhance pest resistance for canola varieties, and increase nutrition value and use of rapeseed meal. Future canola varieties with differential expression of GSLs in term of types, levels, and position in plant from which the obtained meal may increase in its intake for the feed industry and expand to other markets (as in cancer treatment, herbicide industry).

Among efforts to accommodate higher-value markets for canola meal, breeding towards enhancement of canola protein content is the most achievable and economical solution [16,104]. Canola protein extracted from the meal with a well-balanced amino acid profile could be a new protein source for humans, providing that allergen causing proteins, napin and cruciferin, are controlled along with antinutritional compounds residing in the seed hull [104]. Therefore, improvement targets for protein from canola meal should be protein content, protein types (napin, cruciferin), and reduction in fiber and GSL [104]. To recognize this, recently Canadian Protein Industries announced a major investment in breeding to improve protein and reduce fiber to diversify markets for food and feed applications of canola meal [105].

Major genetic factors controlling seed cellulose are being investigated in *B. napus* and the model plant *Arabidopsis* [106], providing fundamental knowledge about the genetic control of fiber and protein content of seed. Using RNA interference technology, the functions of the two genes encoding cellulose synthase subunits in *Arabidopsis thaliana*; AtCESA1 and AtCESA9, were elucidated by generating transgenic *A. thaliana* containing the hairpin RNA (hpRNAi) constructs which specifically spliced their complementary targets [106].

Yellow-seed rapeseed varieties of *B. napus*, *B. rapa*, and *B. juncea* have been shown to have lower fiber content than brown-seeded canola, thus, showed higher feed efficiency in an in vivo experiment on chickens than brown-seed canola [94]. The characteristic genetic background of yellow-seeded *B. napus* could be differentiated from the black-seeded ones by comparing the correlation between differentially expressed genes (DEGs) from the transcriptomic data and the observed traits in the two *B. napus* types, from which down-regulated DEGs in the yellow seeds were linked to reduction in flavonoid and lignin contents, while other DEGs linked to fatty acid biosynthesis and metabolism [107]. Post-translation regulation of seed coat traits could be an effective approach with less effect on cellulose biosynthesis in other plant parts [106]. There have been at least 2 examples of development of yellow seeded line; through directed mutation, an improved yellow-seeded rapeseed line was produced [108]. In addition, in 2007, a new *B. napus* line was introduced by Canadian scientists with low fiber, yellow canola seeds comprising high oleic, low linolenic oil [109]. However, yellow-seed canola is still yet to be commercially successful.

3.3. Disease Resistance

Common diseases in rapeseed reported worldwide includes blackleg disease, caused by *Leptosphaeria maculans* [110,111] clubroot, caused by *Plasmodiopora brassicae* [112,113] and Sclerotinia stem rot, caused by *Sclerotinia sclerotiorum* [114,115]. The availability of genomic data and genetic maps of *Brassica* species, along with associated bioinformatics tools, allow deeper dissection of the genetic architecture of resistance. An efficient platform for identification of disease resistance genes can be implemented based on resistance gene analogs (RGAs) with the use of pangenomics data and confirmed SNP markers [68]. Taking this approach, Dolatabadian et al. [68] identified a total of 15,318 SNPs associated within 1030 *R*-genes in the *B. napus* pan-genome, which facilitates identification of candidate blackleg *R*-genes. Focusing on *Brassica* improvement for biotic stress, a study on pan-genomics in *B. oleracea* showed that the wild C genome species, *B. macrocarpa* contains the highest number of disease resistance gene candidates (1495) compared to the average RGAs of the cultivars (1450) [67], inferring that the crop wild relatives have high gene diversity with novel alleles which are suitable for introgression into cultivated *Brassica* species to improve varieties of *B. napus*.

3.3.1. Blackleg Resistance

Blackleg disease, caused by *Leptosphaeria maculans*, is still the major constraint for the main canola producers including Australia, Canada, Europe, and the United States, despite blackleg resistant cultivars being adopted roughly 30 years ago and undergoing continuous improvement [27,110,111,116–119]. The host-pathogen interaction during the disease process has been clearly understood based on pathotyping and molecular techniques (PCR, nucleotide sequencing, and gene technology) in which qualitative (major genes) and quantitative (multiple genes) resistance is triggered [120,121]. Molecular and gene technologies allied with fast development of bioinformatics has accelerated the identifications of resistance genes (*R*) in the host plants and avirulence genes (*Avr*) in *L. maculans*. At least 19 major *R* genes against *L. maculans* have been genetically mapped in *B. juncea*, *B. rapa*, and *B. napus* [119,120], with 3 genes cloned [120,122]. By contrast, nine of the 16 *Avr* genes identified in *L. maculans* have been cloned to date [123]. Reliance on a single *R* gene and little or no rotation with less susceptible crops leads to rapid evolution of *L. maculans* populations and resistance breakdown [119,120]. Using a GWAS approach, Fu et al. [116] identified 32 and 13 SNPs from Canada and Chinese accessions tightly linked with blackleg resistance which were distributed on chromosomes A03, A05, A08, A09, C01, C04, C05, and C07, in which potential SNPs located on A8 associated with resistance to 12 *L. maculans* isolates and 25 RGAs were identified within this region. Analysis of quantitative resistance (QR) to *L. maculans* in 177 double haploid lines under three experimental conditions using DArTseq markers revealed 3 SNP associations for QR on chromosomes A3, A4, and A7 [117]. New strategies for identification of genes relating to blackleg resistance is quite promising as providing an efficient way to obtain and evaluate new *R* genes in accordance with race structure of *L. maculans*, which is crucial for development of new rapeseed varieties [110,111].

3.3.2. Clubroot Resistance

Clubroot, caused by the soil-borne biotrophic pathogen *Plasmodiopora brassicae*, has long been an important canola disease and has recently been increasing in Europe, China, India, Nepal, and Australia [124–129]. The yield losses by this disease can be around 60–90% when susceptible canola/rapeseed cultivars are planted in *P. brassicae* infested field [129]. Chemical treatment or sanitation practices for eradication of the pathogen from infested cultivating soil is impractical and impossible [113], which suggests utilization of resistant cultivars is the most effective and sustainable solution [125]. Clubroot resistance (CR) genes and gene loci contributing to the resistance in Brassica oilseed crops have been identified and mapped. Clubroot resistance has been discovered to be controlled by dominant loci in *B. rapa*, with two resistant genes being cloned, [130,131], 13 CR loci mapped on chromosomes A01, A02, A03, A06, A08 and the most recently A05 with CrrA5 [125,132,133]. At least 20 QTL in the C genome of *B. oleracea* [28,134] and more than 30 QTL in the A C genome of *B. napus* conferring CR have also been identified and mapped [134]. The CR gene (*Rcr6*) on chromosome B3 of *B. nigra* (BB, n = 8), was the first CR gene identified and mapped in the B-genome of Brassica species [132]. Utilization of modern genomic approaches e.g., GBS (genotyping by sequencing), GWAS, and RNA sequence analysis has facilitated identification of a number of CR candidate genes corresponding to specific race/pathotype of *P. brassicae* in *B. oleracea*, *B. rapa*, *B. napus*, which were summarized by Zhou et al. [135]. These valuable CR resources need to be further explored and precisely identified by molecular markers for breeding CR and dealing with resistance overcome by novel strains of *P. brassicae* [28,134,136]. By mapping RNA sequences of *B. napus* genotypes with different response to *P. brassicae* pathotype 5X to the *B. napus* reference genome (AST_PRJEB5043_v1), Galindo-González et al. [70] found differential expression in these genotypes whereby the standard defense-related proteins e.g., chitinases and thaumatins, and salicylic acid-mediated responses were up regulated in the resistant genotype coupled with mostly inhibited jasmonic acid-mediated responses. The study also identified major proteins relating to *P. brassicae* resistance and susceptibility e.g., calmodulin binding protein 60g (CBP60g), systemic-acquired resistant deficient 1 (SARD1), and xyloglucan endotransglucosylase/hydrolase (XTH).

3.3.3. *Sclerotinia* Resistance

Sclerotinia sclerotioum is a devastating fungal pathogen infecting more than 400 plant species across 75 families [137]. The typical symptoms caused by *Sclerotinia* pathogens include water-soaked rotting of leaves, stems, or fruit covered by fluffy fungal mycelium which subsequently develop compact resting bodies or sclerotia [115]. For over 15 years, its interaction with host plants has been investigated at the molecular level due to the complication of *S. sclerotiorum* possessing both biotrophic and necrotrophic lifestyles, which involves a variety biological processes including reactive oxygen species, lipid and calcium signaling, and DNA methylation-mediated transcription regulation [31]. This diverse interaction may be caused by a large set of candidate effector proteins, which can be identified through bioinformatic research [138]. Efforts in producing *Sclerotinia*-resistant canola varieties includes mutagenesis [139,140], breeding resistance [141,142], and genetic manipulation [143]. In 2011, a *Sclerotinia*-resistant *B. napus* variety was released by Pioneer Hi Bred International [144], however, there is limited information about quantitative disease resistance to the pathogen and associated QTL. By integrating QTL for the resistance, 26 candidate NBS-LRR genes in *B. napus* were found associated with *S. slerotiorum* resistance [145]. Through a transcriptome analysis approach, Chittem et al. [69] confirmed major pathways underlying communications between *S. sclerotiorum* and *B. napus* and the associated key genes. Based on knowledge on the plant immunity involvement of WRKY transcription factors in *Arabidopsis thaliana*, the roles of the corresponding WRKY genes in *B. napus*, *BnWRKY11*, and *BnWRKY70*, were elucidated using Cas9 enzymes and sgRNAs to induce nucleotide specific mutations in these genes, and the results showed the increased resistance to *S. sclerotiorum* in the mutant lines of *BnWRKY70*, while *Sclerotinia* susceptible phenotype was observed in lines with overexpressed *BnWRKY70* [71].

3.4. Abiotic Stress Tolerance

Brassica tolerance towards key abiotic stresses such as heat, drought, cold, and salinity are critical for *Brassica* breeding programs. The physiological changes of canola and the signaling pathways involved in response to the various abiotic responses are often interconnected across the various abiotic stress types [146]. For example, the stress response genes such as the homeodomain-leucine zipper subfamily transcription factors play a role in drought tolerance of *B. napus* [147] and other abiotic stresses, supported by the findings of 113 homeobox genes being identified in Chinese cabbage as differentially expressed under multiple stresses such as cold and osmotic stress [148]. Other examples of genes related to multiple abiotic stresses include the membrane-bound FAD genes in canola that also play a role in improving oil quality. The FAD genes were found to be differentially expressed under cadmium and salinity stresses using qRT-PCR [80]. The FAD candidate genes identified from Xu et al. [80] are highly promising for breeding cadmium and salt tolerant canola. In addition, phytohormones such as abscisic acid (ABA) are commonly involved in abiotic stress response besides regulating plant growth and development [149]. Targeting the coding and transcription factor genes involved in the ABA-mediated signaling pathway and using these genes for creating transgenics *B. napus* that are tolerant towards drought, cold, osmotic, and salinity stresses is very promising [150]. Regulatory pathways of abiotic stress responses in *B. napus*, typically drought, salinity, extreme temperature, and cadmium toxicity, along with the gene families identified in *B. napus* were reviewed in detail by Lohani et al. [146] which highlighted multigenic engineering approach in coping with multiple abiotic stresses and CRISPR/Cas9 as a versatile tool to achieve this target. Mutant *B. napus* lines of paralogous genes of *Bna.RGA* family, orthologues of Arabidopsis REPRESSOR OF GA1-3 (RGA) genes functioning as repressor in gibberellin signaling, were generated via CRISPR/Cas9 with high efficiency [151] in which *bnaa6.rga-D* (gain-of-function mutant) presented enhanced drought tolerance and its stomata were hypersensitive to abscisic acid (ABA) signal, while *bnarga* (loss-of-function mutant) possessed reduced drought tolerance and less sensitivity to ABA treatment [152]. Using the CRISPR/Cas9 tool will efficiently exploit and elucidate roles of identified major genes/factors regulating stress responses in *B. napus*, providing qualified germplasm for breeding of abiotic stress tolerance.

3.5. Herbicide Tolerance/Resistance

Weed invasion is ubiquitous and a persistent threat to crop yield, especially to major field crops including canola [153], for which chemical treatment is the most effective control and constitutes considerable expenditure, approximately 20–36% of total operating cost of canola production [154,155]. Use of herbicide tolerance (HT) crops in combination with a broad spectrum herbicide as glyphosate has been assumed as revolutionary in weed management [156,157] as evidenced by their benefits to agronomy, economics, and the environment [158,159], providing flexibility in scheduling weed treatment and crop yield enhancement [160].

HT canola varieties have been obtained either through conventional breeding e.g., Clearfield® and triazine-tolerant (TT) canola varieties, or genetic engineering e.g., InVigor® (Bayer CropScience, Leverkusen, Germany) and Roundup Ready (RR, Monsanto) [158]. The first TT canola variety, developed through the incorporation of the TT trait from *B. campetris* and *B. rapa* by backcrossing [161], was quickly adopted in Canada and Australia in the 1990s. In 1995, the first imidazolinone-tolerant *B. napus* was registered, of which the trait was developed through mutagenesis by BASF, and the first transgenic canola variety tolerant to the herbicide glyphosate (Roundup) was developed by Monsanto [2]. Glyphosate tolerance in the initial RR cultivars was conferred by the expression of CP4 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase from *Agrobacterium* sp. strain CP4, the enzyme involved in the shikimate pathway essential for the biosynthesis of aromatic amino acids [162]. This is the most common genetically modified (GM) trait in the current biotech crops, with over 80 transformation events being approved for commercial release [163]. Roundup Ready canola has gained approvals for commercial release in major canola producing countries including Canada, Japan, USA, and Australia within 1994 and 2003, where the obtained oil has been approved for food use in addition to European countries, Mexico, Philippines, and Singapore [164]. At least 34 transformation events have been approved in canola varieties to produce the crops with glyphosate HT alone or stacked with other GM traits [163]. Since their commercial release, HT canola varieties have dominated in canola growing countries [165]. However, the intensive cultivation of HT canola and overuse of herbicides, which are often broad-spectrum herbicides with wider application windows, has increased selection pressure on this trait in weed, which has resulted in HT weeds [157]. An increase of up to 15-fold in glyphosate herbicide use in cultivation compared to 1974 [166] was recorded, contributed by the rapid spread of RR crops globally [156,160]. Overreliance on a few HT crops and single active herbicidal ingredients has caused leaking of HT traits [160,167–169], with cases relating to all of the introduced HT systems being reported worldwide [170].

Recently, negative effects on human and animals from glyphosate exposure in the field and through the food chain have caused restrictions or bans on its applications in Austria, France, Germany, and Vietnam [156]. Therefore, other alternatives for glyphosate applications in agriculture, including weed management, are required for the development of sustainable agriculture with reduced reliance on one active ingredient as glyphosate [169,171]. Due to the lack of novel broad-spectrum herbicides, newly developed HT crops still rely on glyphosate and glufosinate, along with other common types of herbicides such as ALS inhibitors, synthetic auxins, HPPD inhibitors, and ACCase inhibitors [160]. To date, combination or stacking HT traits in one variety is the choice of seed companies, starting with the release of RT technology producing hybrids (Hyola® 525RT, Bayer® 3000TR) with glyphosate (RR) and triazine (TT) tolerance, and followed by "stacked" varieties with Triazine tolerance and Clearfield® (TT+CL), which are tolerant to both triazine and imidazolinone (IMI) herbicides [172]. Strict control of the use of HT varieties and combining other weed management practices is an efficient way to manage herbicide tolerance of weeds.

4. Prospects and Future Directions

Although conventional breeding methods are an integral part of crop improvement, modern molecular genetics and gene technologies have accelerated progress and enabled the incorporation of genetic resources across species and genera in canola. These technologies, especially gene editing,

allow more detailed observations on functions of the genome compartments, as well as major genetic factors controlling biological and biochemical processes in canola, from which the obtained knowledge facilitates effective manipulation of the genetic resources towards enhancement of canola varieties. Contributions of modern genetic and gene technologies towards shaping current canola varieties are summarized in Figure 2.

As the current major market for canola crops are food, feed, and biofuel industries [17], the improvement targets prioritize oil yield and quality [9,36], and further reduction of GSL content and other antinutritional compounds [104], while simultaneously enhancing resistance against the major pathogens and tolerance to abiotic stress and herbicides [31,69,116,117,132,134,148,173–175]. The approval for commercial release of DHA canola [86,87] is the most recent achievement in canola improvement and exploits natural resources in fatty acid biosynthesis, where modern genetic and gene technologies play a crucial role in the success. Glucosinolates, which have been long time avoided in canola meal, are now found useful for diverse bioactivities which facilitate plant defense against non-adapted pathogen and insect pests [132,176], cancer treatment [177], and nutritional quality of canola meal [104]. Knowledge on GSL biosynthesis and their roles in physiological and biological processes of *B. napus* are being made more specific by genomic and transcriptomic approaches [100,177]. Improvement of pathogen resistance or herbicide tolerance need continued attention as pathogens and weeds can overcome these resistances after introduction of the pathogen resistant or herbicide tolerant varieties, with the pace depending on the management and rotation of those varieties in combination with other management practices [156,158,171]. The application of genomics and gene technologies in *Brassica* species have validated hypotheses and incorporated new knowledge on the evolution and the genetic diversity of *B. napus*, and explored the genetic layout of the complex genome of amphidiploid canola [63]. Genomic studies in rapeseed which used to be challenged by the complexity of the genome, are now facilitated by "omic" and gene technologies, enabling more insights about factors controlling 1,trait development and further enhancement of ca [80,100,107,134].

Facing climate change, rapid evolution of phyto-pathogens and weeds and an increasing world population, canola is assumed as an economical effective "multiple-purposes" crop fitting a broad range of markets, industries, and sustainable development policies [36]. The choice of technological approach for developing and enhancing canola varieties, such as GM canola, is affected by local legislation, research investment, and improvement targets. Although current GM canola varieties benefit in enhancement values of canola products and reducing labor costs (HT canola), the transgenes might affect performance of other genomic compartments Therefore, eliminating unessential transgenes from GM canola, environmental impact assessment and obtaining legislative approvals for commercial release of these varieties are requirements to developing a GM variety of which procedure prolongs the time reaching their markets. The world's largest canola importer, EU, has imposed strict regulations on GM organisms, which affects technological choices in improving canola. Taking environmental and legislation perspectives, gene editing tools, such as CRISPR/Cas, are the most promising technology for enhancement of canola for commercial purpose [178]. With the advance of genomics and pan-genomics, the genetic architecture underlying response to the major pathogens and abiotic stresses such as salinity and drought need to be further dissected. Cytoplasmic genomes are attracting more interest as the new source for maximizing oil content in canola [90]. Considering factors shaping modern canola varieties, genetic and gene editing technologies are proven powerful tools for achieving new breeding targets, allowing thorough exploration and exploitation of *B. napus* genes which are largely unknown.

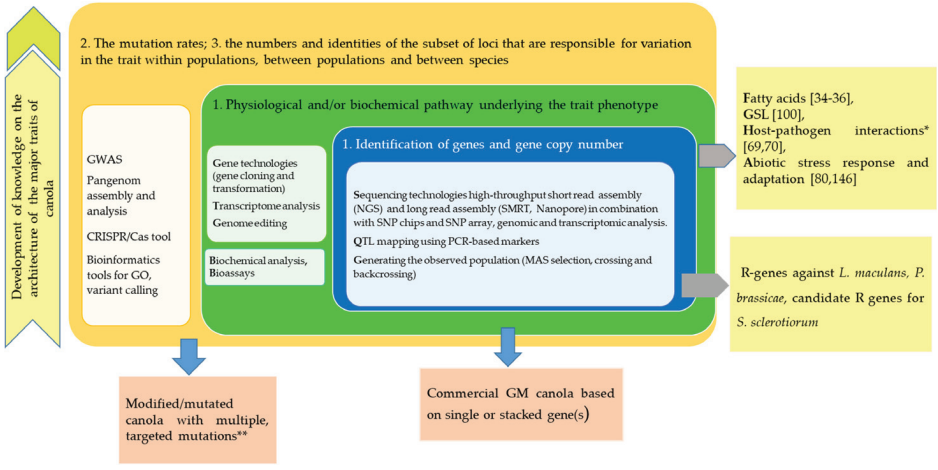


Figure 2. Contributions of the modern genetic and gene technologies to the understanding of the genetic architecture of the major traits of Brassica oilseed crop, adapted from Mackay [179] suggestions, and achievements in canola variety research and development. GWAS, Genome-Wide Association Study; CRISPR/Cas, the clustered regularly interspaced short palindromic repeat; GO, gene ontology; NGS, next-generation sequencing technique; SMRT, single molecule real time; SNP, single nucleotide polymorphism; QTL, quantitative trait loci; MAS, marker-assisted selection; GSL, glucosinolate; GM, genetically modified; (*) molecular biology fundamentals of these mechanisms is being established; (**) products from experimental studies.

Author Contributions: Conceptualization, L.B.T. and T.X.N.; writing—original draft preparation, L.B.T. and T.X.N; writing—reviewing and editing, J.B.; visualization, L.B.T.; funding acquisition, J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Australian Research Council Projects DP1601004497, LP160100030.

Acknowledgments: We acknowledge The University of Western Australia's support through a 'UWA International Fee Scholarship and University Postgraduate Award'.

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

References

1. McVetty, P.B.E.; Duncan, R.W. Canola/Rapeseed: Genetics and Breeding. In *Reference Module in Food Science*; Elsevier: Amsterdam, The Netherlands, 2016. [CrossRef]
2. OGTR. *The Biology of Brassica napus L. (Canola)*; Aging, Department of Health and Aging, Ed.; Office of the Gene Technology Regulator: Canberra, ACT, Australia, 2008.
3. Nagaharu, U. Genome-analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J. Bot.* **1935**, *7*, 389–452.
4. Lu, K.; Wei, L.; Li, X.; Wang, Y.; Wu, J.; Liu, M.; Zhang, C.; Chen, Z.; Xiao, Z.; Jian, H.; et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* **2019**, *10*, 1154. [CrossRef]
5. Chalhoub, B.; Denoeud, F.; Liu, S.; Parkin, I.A.P.; Tang, H.; Wang, X.; Chiquet, J.; Belcram, H.; Tong, C.; Samans, B.; et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **2014**, *345*, 950. [CrossRef]
6. Gómez-Campo, C.; Prakash, S. 2 Origin and domestication. In *Developments in Plant Genetics and Breeding*; Gómez-Campo, C., Ed.; Elsevier: Amsterdam, The Netherlands, 1999; Volume 4, pp. 33–58.
7. Eskin, N.A.M.; Przybylski, R. RAPE SEED OIL/CANOLA. In *Encyclopedia of Food Sciences and Nutrition*, 2nd ed.; Caballero, B., Ed.; Academic Press: Oxford, UK, 2003; pp. 4911–4916. [CrossRef]

8. Scarth, R.; McVetty, P.B.E.; Rimmer, S.R.; Stefansson, B.R. STELLAR LOW LINOLENIC-HIGH LINOLEIC ACID SUMMER RAPE. *Can. J. Plant Sci.* **1988**, *68*, 509–511. [CrossRef]
9. Long, W.; Hu, M.; Gao, J.; Chen, S.; Zhang, J.; Cheng, L.; Pu, H. Identification and Functional Analysis of Two New Mutant BnFAD2 Alleles That Confer Elevated Oleic Acid Content in Rapeseed. *Front. Genet.* **2018**, *9*. [CrossRef]
10. Maher, L.B.W.; Salisbury, P.; Debonte, L.; Deng, X. (Organising Committee of: Wuhan). High Oleic, low linolenic (HOLL) specialty canola development in Australia. In Proceedings of the 12th International Rapeseed Congress, Wuhan, China, 23–30 March 2007; Volume 5, pp. 22–24.
11. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **2018**, *32*, 261–285. [CrossRef]
12. Spasibionek, S.; Mikołajczyk, K.; Ćwiek-Kupczyńska, H.; Piętka, T.; Krótka, K.; Matuszczak, M.; Nowakowska, J.; Michalski, K.; Bartkowiak-Broda, I. Marker assisted selection of new high oleic and low linolenic winter oilseed rape (*Brassica napus* L.) inbred lines revealing good agricultural value. *PLoS ONE* **2020**, *15*, e0233959. [CrossRef]
13. Nath, U.K.; Wilmer, J.A.; Wallington, E.J.; Becker, H.C.; Möllers, C. Increasing erucic acid content through combination of endogenous low polyunsaturated fatty acids alleles with Ld-LPAAT + Bn-fae1 transgenes in rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.* **2009**, *118*. [CrossRef]
14. Hayward, A. Introduction-Oilseed Brassicas. In *Genetics, Genomics and Breeding of Oilseed Brassicas*; Edwards, D., Batley, J., Parkin, I., Kole, C., Eds.; Taylor & Francis Group: Baton Rouge, LA, USA, 2011.
15. Wei, D.; Cui, Y.; He, Y.; Xiong, Q.; Qian, L.; Tong, C.; Lu, G.; Ding, Y.; Li, J.; Jung, C.; et al. A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding. *J. Exp. Bot.* **2017**, *68*, 4791–4801. [CrossRef]
16. Bonnardeaux, J. *Uses for Canola Meal*; Department of Agriculture and Food: Perth, Western Australia, Australia, 2007.
17. Nath, U.K.; Kim, H.-T.; Khatun, K.; Park, J.-I.; Kang, K.-K.; Ill-Sup, N. Modification of Fatty Acid Profiles of Rapeseed (*Brassica napus* L.) Oil for Using as Food, Industrial Feed-Stock and Biodiesel. *Plant Breed. Biotechnol.* **2016**, *4*, 123–134. [CrossRef]
18. Palmer, C.E.D.; Keller, W.A. Future Prospects for Brassica Oilseed Improvement through Genomics. In *Genetics, Genomics and Breeding of Oilseed Brassicas*; Edwards, D., Batley, J., Parkin, I., Kole, C., Eds.; Taylor & Francis Group: Baton Rouge, LA, USA, 2011; pp. 1–13.
19. Li, F.-W.; Harkess, A. A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **2018**, *6*, e1030. [CrossRef]
20. Schiessl, S.V.; Quezada-Martinez, D.; Tebartz, E.; Snowdon, R.J.; Qian, L. The vernalisation regulator FLOWERING LOCUS C is differentially expressed in biennial and annual *Brassica napus*. *Sci. Rep.* **2019**, *9*, 14911. [CrossRef]
21. Kyriakidou, M.; Tai, H.H.; Anglin, N.L.; Ellis, D.; Strömvik, M.V. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* **2018**, *9*, 1660. [CrossRef]
22. Yuan, Y.; Bayer, P.E.; Lee, H.T.; Edwards, D. runBNG: A software package for BioNano genomic analysis on the command line. *Bioinformatics* **2017**, *33*, 3107–3109. [CrossRef]
23. Oluwadare, O.; Highsmith, M.; Cheng, J. An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biol. Proced. Online* **2019**, *21*, 7. [CrossRef]
24. Malmberg, M.M.; Barbulescu, D.M.; Drayton, M.C.; Shinozuka, M.; Thakur, P.; Ogaji, Y.O.; Spangenberg, G.C.; Daetwyler, H.D.; Cogan, N.O.I. Evaluation and Recommendations for Routine Genotyping Using Skim Whole Genome Re-sequencing in Canola. *Front. Plant Sci.* **2018**, *9*. [CrossRef]
25. Qu, C.; Jia, L.; Fu, F.; Zhao, H.; Lu, K.; Wei, L.; Xu, X.; Liang, Y.; Li, S.; Wang, R.; et al. Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers. *BMC Genom.* **2017**, *18*, 232. [CrossRef]
26. Scheben, A.; Batley, J.; Edwards, D. Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. *Plant Biotechnol. J.* **2017**, *15*, 149–161. [CrossRef]

27. Fikere, M.; Barbulescu, D.M.; Malmberg, M.M.; Spangenberg, G.C.; Cogan, N.O.I.; Daetwyler, H.D. Meta-analysis of GWAS in canola blackleg (*Leptosphaeria maculans*) disease traits demonstrates increased power from imputed whole-genome sequence. *Sci. Rep.* **2020**, *10*, 14300. [CrossRef]
28. Farid, M.; Yang, R.C.; Kebede, B.; Rahman, H. Evaluation of Brassica oleracea accessions for resistance to Plasmodiophora brassicae and identification of genomic regions associated with resistance. *Genome* **2020**, *63*, 91–101. [CrossRef]
29. Mason, A.S.; Higgins, E.E.; Snowdon, R.J.; Batley, J.; Stein, A.; Werner, C.; Parkin, I.A. A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. *Theor. Appl. Genet.* **2017**, *130*, 621–633. [CrossRef] [PubMed]
30. Katche, E.; Quezada-Martinez, D.; Katche, E.I.; Vasquez-Teuber, P.; Mason, A.S. Interspecific Hybridization for Brassica Crop Improvement. *Crop Breed. Genet. Genom.* **2019**, *1*, e190007. [CrossRef]
31. Wang, Z.; Ma, L.-Y.; Cao, J.; Li, Y.-L.; Ding, L.-N.; Zhu, K.-M.; Yang, Y.-H.; Tan, X.-L. Recent Advances in Mechanisms of Plant Defense to *Sclerotinia sclerotiorum*. *Front. Plant Sci.* **2019**, *10*. [CrossRef]
32. Cardoza, V.S.C.N. Agrobacterium-Mediated Transformation of Canola. In *Transgenic Crops of the World*; Curtis, I.S., Ed.; Springer: Dordrecht, The Netherlands, 2004.
33. Zhang, K.; Nie, L.; Cheng, Q.; Yin, Y.; Chen, K.; Qi, F.; Zou, D.; Liu, H.; Zhao, W.; Wang, B.; et al. Effective editing for lysophosphatidic acid acyltransferase 2/5 in allotetraploid rapeseed (*Brassica napus* L.) using CRISPR-Cas9 system. *Biotechnol. Biofuels* **2019**, *12*, 225. [CrossRef]
34. Wu, G.; Truksa, M.; Datla, N.; Vrinten, P.; Bauer, J.; Zank, T.; Cirpus, P.; Heinz, E.; Qiu, X. Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.* **2005**, *23*, 1013–1017. [CrossRef]
35. 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]
36. Petrie, J.R.; Zhou, X.-R.; Leonforte, A.; McAllister, J.; Shrestha, P.; Kennedy, Y.; Belide, S.; Buzza, G.; Gororo, N.; Gao, W.; et al. Development of a *Brassica napus* (Canola) Crop Containing Fish Oil-Like Levels of DHA in the Seed Oil. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
37. Puchta, H. Applying CRISPR/Cas for genome engineering in plants: The best is yet to come. *Curr. Opin. Plant Biol.* **2017**, *36*, 1–8. [CrossRef]
38. Burmistrz, M.; Krakowski, K.; Krawczyk-Balska, A. RNA-Targeting CRISPR-Cas Systems and Their Applications. *Int. J. Mol. Sci.* **2020**, *21*, 1122. [CrossRef]
39. European Academies Science Advisory Council. *The Regulation of Genome-Edited Plants in the European Union*; EU: Geneva, Switzerland, 2020.
40. Makarova, K.S.; Koonin, E.V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol. Biol.* **2015**, *1311*, 47–75. [CrossRef]
41. Borrelli, V.M.G.; Brambilla, V.; Rogowsky, P.; Marocco, A.; Lanubile, A. The Enhancement of Plant Disease Resistance Using CRISPR/Cas9 Technology. *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
42. Li, C.; Hao, M.; Wang, W.; Wang, H.; Chen, F.; Chu, W.; Zhang, B.; Mei, D.; Cheng, H.; Hu, Q. An Efficient CRISPR/Cas9 Platform for Rapidly Generating Simultaneous Mutagenesis of Multiple Gene Homoeologs in Allotetraploid Oilseed Rape. *Front. Plant Sci.* **2018**, *9*, 442. [CrossRef] [PubMed]
43. Zheng, M.; Zhang, L.; Tang, M.; Liu, J.; Liu, H.; Yang, H.; Fan, S.; Terzaghi, W.; Wang, H.; Hua, W. Knockout of two BnaMAX1 homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnol. J.* **2020**, *18*, 644–654. [CrossRef] [PubMed]
44. Yang, Y.; Zhu, K.; Li, H.; Han, S.; Meng, Q.; Khan, S.U.; Fan, C.; Xie, K.; Zhou, Y. Precise editing of CLAVATA genes in *Brassica napus* L. regulates multilocular silique development. *Plant Biotechnol. J.* **2018**, *16*, 1322–1335. [CrossRef]
45. Zhai, Y.; Cai, S.; Hu, L.; Yang, Y.; Amoo, O.; Fan, C.; Zhou, Y. CRISPR/Cas9-mediated genome editing reveals differences in the contribution of INDEHISCENT homologues to pod shatter resistance in *Brassica napus* L. *Theor. Appl. Genet.* **2019**, *132*, 2111–2123. [CrossRef]
46. Sriboon, S.; Li, H.; Guo, C.; Senkhamwong, T.; Dai, C.; Liu, K. Knock-out of TERMINAL FLOWER 1 genes altered flowering time and plant architecture in *Brassica napus*. *BMC Genet.* **2020**, *21*, 52. [CrossRef]
47. Jat, R.S.; Singh, V.V.; Sharma, P.; Rai, P.K. Oilseed brassica in India: Demand, supply, policy perspective and future potential. *OCL* **2019**, *26*, 8. [CrossRef]

48. Snowdon, R.J.; Iniguez Luy, F.L. Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breed.* **2012**, *131*, 351–360. [CrossRef]
49. Afzal, M.; Alghamdi, S.S.; Habib Ur Rahman, M.; Ahmad, A.; Farooq, T.; Alam, M.; Khan, I.A.; Ullah, H.; Nasim, W.; Fahad, S. Current status and future possibilities of molecular genetics techniques in Brassica napus. *Biotechnol. Lett.* **2018**, *40*, 479–492. [CrossRef]
50. Witzel, K.; Neugart, S.; Ruppel, S.; Schreiner, M.; Wiesner, M.; Baldermann, S. Recent progress in the use of 'omics technologies in brassicaceous vegetables. *Front. Plant Sci.* **2015**, *6*, 244. [CrossRef]
51. Tan, C.; Liu, H.; Ren, J.; Ye, X.; Feng, H.; Liu, Z. Single-molecule real-time sequencing facilitates the analysis of transcripts and splice isoforms of anthers in Chinese cabbage (*Brassica rapa* L. ssp. pekinensis). *BMC Plant Biol.* **2019**, *19*, 517. [CrossRef] [PubMed]
52. 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] [PubMed]
53. Cai, G.; Yang, Q.; Chen, H.; Yang, Q.; Zhang, C.; Fan, C.; Zhou, Y. Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Rep.* **2016**, *6*, 21625. [CrossRef]
54. Perumal, S.; Koh, C.S.; Jin, L.; Buchwaldt, M.; Higgins, E.E.; Zheng, C.; Sankoff, D.; Robinson, S.J.; Kagale, S.; Navabi, Z.-K.; et al. A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral Brassica genome. *Nat. Plants* **2020**, *6*, 929–941. [CrossRef] [PubMed]
55. Rousseau-Gueutin, M.; Belser, C.; Silva, C.D.; Richard, G.; Istace, B.; Cruaud, C.; Falentin, C.; Boideau, F.; Boutte, J.; Delourme, R.; et al. Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh. *bioRxiv* **2020**. [CrossRef]
56. Lee, H.; Chawla, H.S.; Obermeier, C.; Dreyer, F.; Abbadi, A.; Snowdon, R. Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
57. Song, J.-M.; Guan, Z.; Hu, J.; Guo, C.; Yang, Z.; Wang, S.; Liu, D.; Wang, B.; Lu, S.; Zhou, R.; et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* **2020**, *6*, 34–45. [CrossRef]
58. Wang, W.; Guan, R.; Liu, X.; Zhang, H.; Song, B.; Xu, Q.; Fan, G.; Chen, W.; Wu, X.; Liu, X.; et al. Chromosome level comparative analysis of Brassica genomes. *Plant Mol. Biol.* **2019**, *99*, 237–249. [CrossRef]
59. Qiao, J.; Zhang, X.; Chen, B.; Huang, F.; Xu, K.; Huang, Q.; Huang, Y.; Hu, Q.; Wu, X. Comparison of the cytoplastic genomes by resequencing: Insights into the genetic diversity and the phylogeny of the agriculturally important genus Brassica. *BMC Genom.* **2020**, *21*, 480. [CrossRef]
60. Tettelin, H.; Masignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13950. [CrossRef]
61. Tirnaz, S.; Edwards, D.; Batley, J. The importance of plant pan-genomes in breeding. *Quant. Genet. Genom. Plant Breed.* **2020**, *2*, 27–32.
62. Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and Exploiting Pan-genomics for Crop Improvement. *Mol. Plant* **2019**, *12*, 156–169. [CrossRef] [PubMed]
63. Bayer, P.E.; Golicz, A.A.; Scheben, A.; Batley, J.; Edwards, D. Plant pan-genomes are the new reference. *Nat. Plants* **2020**, *6*, 914–920. [CrossRef] [PubMed]
64. 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] [PubMed]
65. Golicz, A.A.; Batley, J.; Edwards, D. Towards plant pangenomics. *Plant Biotechnol. J.* **2016**, *14*, 1099–1105. [CrossRef] [PubMed]
66. Hurgobin, B.; Golicz, A.A.; Bayer, P.E.; Chan, C.-K.K.; Tirnaz, S.; Dolatabadian, A.; Schiessl, S.V.; Samans, B.; Montenegro, J.D.; Parkin, I.A.P.; et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* **2018**, *16*, 1265–1274. [CrossRef]
67. Bayer, P.E.; Golicz, A.A.; Tirnaz, S.; Chan, C.-K.K.; Edwards, D.; Batley, J. Variation in abundance of predicted resistance genes in the Brassica oleracea pangenome. *Plant Biotechnol. J.* **2019**, *17*, 789–800. [CrossRef]
68. Dolatabadian, A.; Bayer, P.E.; Tirnaz, S.; Hurgobin, B.; Edwards, D.; Batley, J. Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* **2020**, *18*, 969–982. [CrossRef]

69. Chittem, K.; Yajima, W.R.; Goswami, R.S.; del Río Mendoza, L.E. Transcriptome analysis of the plant pathogen Sclerotinia sclerotiorum interaction with resistant and susceptible canola (*Brassica napus*) lines. *PLoS ONE* **2020**, *15*, e0229844. [CrossRef]
70. Galindo-González, L.; Manolii, V.; Hwang, S.-F.; Strelkov, S.E. Response of *Brassica napus* to Plasmodiophora brassicae Involves Salicylic Acid-Mediated Immunity: An RNA-Seq-Based Study. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
71. Sun, Q.; Lin, L.; Liu, D.; Wu, D.; Fang, Y.; Wu, J.; Wang, Y. CRISPR/Cas9-Mediated Multiplex Genome Editing of the BnWRKY11 and BnWRKY70 Genes in *Brassica napus* L. *Int. J. Mol. Sci.* **2018**, *19*, 2716. [CrossRef] [PubMed]
72. Shahbandeh, M. *USDA Foreign Agricultural Service, US Department of Agriculture, Consumption of Vegetable Oils Worldwide from 2013/14 to 2019/2020, by Oil Type (in Million Metric Tons)*; US Department of Agriculture: Washington, DC, USA, 2020.
73. Przybylski, R. Canola/Rapeseed Oil. In *Vegetable Oils in Food Technology*; Gunstone, F.D., Ed.; Wiley-Blackwell: Oxford, UK, 2011. [CrossRef]
74. Aukema, H.; Campbell, L. 9—Oil Nutrition and Utilization. In *Canola*; Daun, J.K., Eskin, N.A.M., Hickling, D., Eds.; AOCS Press: Urbana, IL, USA, 2011. [CrossRef]
75. Lin, L.; Allemekinders, H.; Dansby, A.; Campbell, L.; Durance-Tod, S.; Berger, A.; Jones, P.J. Evidence of health benefits of canola oil. *Nutr. Rev.* **2013**, *71*, 370–385. [CrossRef] [PubMed]
76. Scarth, R.; McVetty, P. Designer oil canola –A review of new food-grade brassica oils with focus on high oleic, low linolenic types. In Proceedings of the 10th International GCIRC Congress, Canberra, Australia, 26–29 September 1999; p. 10.
77. Salisbury, P.A.; Cowling, W.A.; Potter, T.D. Continuing innovation in Australian canola breeding. *Crop Pasture Sci.* **2016**, *67*, 266–272. [CrossRef]
78. Canola Council of Canada. Resilient. 2015. Available online: https://www.canolacouncil.org/media/575447/ccc_2015_resilient.pdf (accessed on 12 July 2020).
79. He, M.; Qin, C.-X.; Wang, X.; Ding, N.-Z. Plant Unsaturated Fatty Acids: Biosynthesis and Regulation. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
80. Xu, L.; Zeng, W.; Li, J.; Liu, H.; Yan, G.; Si, P.; Yang, C.; Shi, Y.; He, Q.; Zhou, W. Characteristics of membrane-bound fatty acid desaturase (FAD) genes in *Brassica napus* L. and their expressions under different cadmium and salinity stresses. *Environ. Exp. Bot.* **2019**, *162*, 144–156. [CrossRef]
81. von Borcke, L. High Oleic Acid Oilseed: Oil Seed Rape FAD2 Mutants with Increased Mono-Unsaturated Fatty Acid Content. Granted Patent AU 2012223022, 2 March 2012. International Publication No. WO 2012/117256, US Publication No. US 2014/0150132.
82. Xue, Y.; Chen, B.; Wang, R.; Win, A.N.; Li, J.; Chai, Y. Genome-Wide Survey and Characterization of Fatty Acid Desaturase Gene Family in *Brassica napus* and Its Parental Species. *Appl. Biochem. Biotechnol.* **2018**, *184*, 582–598. [CrossRef]
83. Zhu, Q.; King, G.J.; Liu, X.; Shan, N.; Borpatragohain, P.; Baten, A.; Wang, P.; Luo, S.; Zhou, Q. Identification of SNP loci and candidate genes related to four important fatty acid composition in *Brassica napus* using genome wide association study. *PLoS ONE* **2019**, *14*, e0221578. [CrossRef]
84. Zhai, Y.; Yu, K.; Cai, S.; Hu, L.; Amoo, O.; Xu, L.; Yang, Y.; Ma, B.; Jiao, Y.; Zhang, C.; et al. Targeted mutagenesis of BnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol. J.* **2020**, *18*, 1153–1168. [CrossRef]
85. Walsh, T.A.; Bevan, S.A.; Gachotte, D.J.; Larsen, C.M.; Moskal, W.A.; Merlo, P.A.O.; Sidorenko, L.V.; Hampton, R.E.; Stoltz, V.; Pareddy, D.; et al. Canola engineered with a microalgal polyketide synthase-like system produces oil enriched in docosahexaenoic acid. *Nat. Biotechnol.* **2016**, *34*, 881–887. [CrossRef]
86. USDA (US Department of Agriculture). *Determination of Nonregulated Status for Nuseed DHA Canola*; USDA, Ed.; USDA: Washington, DC, USA, 2018.
87. OGTR. *Issue of Licence DIR 155 to Nuseed Pty Ltd for the Commercial Release of GM Canola*; Department of Health and Aging, Ed.; OGTR: Canberra, Australia, 2018.
88. Warner, D.J.; Lewis, K.A. Evaluation of the Risks of Contaminating Low Erucic Acid Rapeseed with High Erucic Rapeseed and Identification of Mitigation Strategies. *Agriculture* **2019**, *9*, 190. [CrossRef]
89. Ogunkunle, O.; Ahmed, N.A. A review of global current scenario of biodiesel adoption and combustion in vehicular diesel engines. *Energy Rep.* **2019**, *5*, 1560–1579. [CrossRef]

90. Liu, J.; Hao, W.; Liu, J.; Fan, S.; Zhao, W.; Deng, L.; Wang, X.; Hu, Z.; Hua, W.; Wang, H. A Novel Chimeric Mitochondrial Gene Confers Cytoplasmic Effects on Seed Oil Content in Polyploid Rapeseed (Brassica napus). *Mol. Plant* **2019**, *12*, 582–596. [CrossRef] [PubMed]
91. Wanasundara, J.P.D.; Tan, S.; Alashi, A.M.; Pudel, F.; Blanchard, C. Chapter 18—Proteins From Canola/Rapeseed: Current Status. In *Sustainable Protein Sources*; Nadathur, S.R., Wanasundara, J.P.D., Scanlin, L., Eds.; Academic Press: San Diego, CA, USA, 2017. [CrossRef]
92. Hein, T. Canola Meal in Dairy Rations. 2020. Available online: https://www.allaboutfeed.net/New-Proteins/Articles/2020/7/Canola-meal-in-dairy-rations-618105E/ (accessed on 3 August 2020).
93. Skugor, A.; Kjos, N.P.; Sundaram, A.Y.M.; Mydland, L.T.; Ånestad, R.; Tauson, A.-H.; Øverland, M. Effects of long-term feeding of rapeseed meal on skeletal muscle transcriptome, production efficiency and meat quality traits in Norwegian Landrace growing-finishing pigs. *PLoS ONE* **2019**, *14*, e0220441. [CrossRef] [PubMed]
94. Slominski, B.A. Developments in the breeding of low fibre rapeseed/canola. *J. Anim. Feed Sci.* **1997**, *6*, 303–318. [CrossRef]
95. Jensen, S.K.; Liu, Y.-G.; Eggum, B.O. The influence of seed size and hull content on the composition and digestibility of rapeseeds in rats. *Anim. Feed Sci. Technol.* **1995**, *54*, 9–19. [CrossRef]
96. OGTR. *The Biology of Brassica Napus L. (Canola) and Brassica Juncea (L.) Czern. & Coss. (Indian mustard)*; Health ADO, Ed.; OGTR: Canberra, Australia, 2017.
97. Sønderby, I.E.; Geu-Flores, F.; Halkier, B.A. Biosynthesis of glucosinolates—Gene discovery and beyond. *Trends Plant Sci.* **2010**, *15*, 283–290. [CrossRef]
98. Zang, Y.X.; Kim, H.U.; Kim, J.A.; Lim, M.H.; Jin, M.; Lee, S.C.; Kwon, S.J.; Lee, S.I.; Hong, J.K.; Park, T.H.; et al. Genome-wide identification of glucosinolate synthesis genes in Brassica rapa. *FEBS J.* **2009**, *276*, 3559–3574. [CrossRef]
99. Piślewska-Bednarek, M.; Nakano, R.T.; Hiruma, K.; Pastorczyk, M.; Sanchez-Vallet, A.; Singkaravanit-Ogawa, S.; Ciesiołka, D.; Takano, Y.; Molina, A.; Schulze-Lefert, P.; et al. Glutathione Transferase U13 Functions in Pathogen-Triggered Glucosinolate Metabolism. *Plant Physiol.* **2018**, *176*, 538–551. [CrossRef]
100. Kittipol, V.; He, Z.; Wang, L.; Doheny-Adams, T.; Langer, S.; Bancroft, I. Genetic architecture of glucosinolate variation in Brassica napus. *J. Plant Physiol.* **2019**, *240*, 152988. [CrossRef]
101. Geu-Flores, F.; Nielsen, M.T.; Nafisi, M.; Møldrup, M.E.; Olsen, C.E.; Motawia, M.S.; Halkier, B.A. Glucosinolate engineering identifies a γ-glutamyl peptidase. *Nat. Chem. Biol.* **2009**, *5*, 575–577. [CrossRef]
102. Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C.E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2020**, *169*, 112100. [CrossRef]
103. Burow, M.; Halkier, B.A. How does a plant orchestrate defense in time and space? Using glucosinolates in Arabidopsis as case study. *Curr. Opin. Plant Biol.* **2017**, *38*, 142–147. [CrossRef]
104. Campbell, L.; Rempel, C.B.; Wanasundara, J.P.D. Canola/Rapeseed Protein: Future Opportunities and Directions-Workshop Proceedings of IRC 2015. *Plants* **2016**, *5*, 17. [CrossRef]
105. Knutt, C. Protein Industries Canada Invests to Improve Canola Protein Content. Available online: https://www.proteinindustriescanada.ca/news/protein-industries-canada-invests-to-improve-canola-protein-content (accessed on 27 July 2020).
106. Jayawardhane, K.N.; Singer, S.D.; Ozga, J.A.; Rizvi, S.M.; Weselake, R.J.; Chen, G. Seed-specific down-regulation of Arabidopsis CELLULOSE SYNTHASE 1 or 9 reduces seed cellulose content and differentially affects carbon partitioning. *Plant Cell Rep.* **2020**, *39*, 953–969. [CrossRef]
107. Jiang, J.; Zhu, S.; Yuan, Y.; Wang, Y.; Zeng, L.; Batley, J.; Wang, Y.-P. Transcriptomic comparison between developing seeds of yellow- and black-seeded *Brassica napus* reveals that genes influence seed quality. *BMC Plant Biol.* **2019**, *19*, 203. [CrossRef]
108. Facciotti, D. Production of Improved Rapeseed Exhibiting Yellow-Seed Coat. US Patent 006547711B2, 15 April 2003.
109. Chungu, C.; Rakow, G.; Raney, J.P.; Kubik Thomas, J. Low Fiber Yellow Canola Seeds Comprising High Oleic, Low Linolenic Oil. World Intellectual Property Organization WO2007/016521A2, 8 February 2007.
110. Zhang, X.; Fernando, W.G.D. Insights into fighting against blackleg disease of *Brassica napus* in Canada. *Crop Pasture Sci.* **2017**, *69*, 40–47. [CrossRef]

111. Diaz, C.; Cevallos, F.; Damicone, J. Characterization of the Race Structure of *Leptosphaeria maculans* Causing Blackleg of Winter Canola in Oklahoma and Kansas. *Plant Dis.* **2019**, *103*, 2353–2358. [CrossRef]
112. Hwang, S.F.; Strelkov, S.E.; Ahmed, H.U.; Manolii, V.P.; Zhou, Q.; Fu, H.; Turnbull, G.; Fredua-Agyeman, R.; Feindel, D. Virulence and inoculum density-dependent interactions between clubroot resistant canola (*Brassica napus*) and Plasmodiophora brassicae. *Plant Pathol.* **2017**, *66*, 1318–1328. [CrossRef]
113. Rempel, C.B.; Hutton, S.N.; Jurke, C.J. Clubroot and the importance of canola in Canada. *Can. J. Plant Pathol.* **2014**, *36*, 19–26. [CrossRef]
114. Denton-Giles, M.; Derbyshire, M.C.; Khentry, Y.; Buchwaldt, L.; Kamphuis, L.G. Partial stem resistance in *Brassica napus* to highly aggressive and genetically diverse *Sclerotinia sclerotiorum* isolates from Australia. *Can. J. Plant Pathol.* **2018**, *40*, 551–561. [CrossRef]
115. Aldrich-Wolfe, L.; Travers, S.; Nelson, B.D., Jr. Genetic Variation of Sclerotinia sclerotiorum from Multiple Crops in the North Central United States. *PLoS ONE* **2015**, *10*, e0139188. [CrossRef]
116. Fu, F.; Zhang, X.; Liu, F.; Peng, G.; Yu, F.; Fernando, D. Identification of resistance loci in Chinese and Canadian canola/rapeseed varieties against Leptosphaeria maculans based on genome-wide association studies. *BMC Genom.* **2020**, *21*, 501. [CrossRef]
117. Raman, R.; Diffey, S.; Barbulescu, D.M.; Coombes, N.; Luckett, D.; Salisbury, P.; Cowley, R.; Marcroft, S.; Raman, H. Genetic and physical mapping of loci for resistance to blackleg disease in canola (*Brassica napus* L.). *Sci. Rep.* **2020**, *10*, 4416. [CrossRef]
118. Rong, S.; Feng, J.; Li, Q.; Fei, W.; Ahmed, H.U.; Liang, Y.; Hwang, S.-F.; Strelkov, S.E. Pathogenic variability and prevalence of *Avr* genes in *Leptosphaeria maculans* populations from Alberta, Canada. *J. Plant Dis. Prot.* **2015**, *122*, 161–168. [CrossRef]
119. Zhang, X.; Peng, G.; Kutcher, H.R.; Balesdent, M.-H.; Delourme, R.; Fernando, W.G.D. Breakdown of Rlm3 resistance in the Brassica napus–Leptosphaeria maculans pathosystem in western Canada. *Eur. J. Plant Pathol.* **2016**, *145*, 659–674. [CrossRef]
120. Van De Wouw, A.P.; Marcroft, S.J.; Howlett, B.J. Blackleg disease of canola in Australia. *Crop Pasture Sci.* **2016**, *67*, 273–283. [CrossRef]
121. Raman, H.; Raman, R.; Diffey, S.; Qiu, Y.; McVittie, B.; Barbulescu, D.M.; Salisbury, P.A.; Marcroft, S.; Delourme, R. Stable Quantitative Resistance Loci to Blackleg Disease in Canola (*Brassica napus* L.) Over Continents. *Front. Plant Sci.* **2018**, *9*. [CrossRef]
122. Zhou, T.; Xu, W.; Hirani, A.H.; Liu, Z.; Tuan, P.A.; Ayele, B.T.; Daayf, F.; McVetty, P.B.E.; Duncan, R.W.; Li, G. Transcriptional Insight Into *Brassica napus* Resistance Genes LepR3 and Rlm2-Mediated Defense Response Against the Leptosphaeria maculans Infection. *Front. Plant Sci.* **2019**, *10*. [CrossRef]
123. Zou, Z.; Liu, F.; Selin, C.; Fernando, W.G.D. Generation and Characterization of a Virulent Leptosphaeria maculans Isolate Carrying a Mutated AvrLm7 Gene Using the CRISPR/Cas9 System. *Front. Microbiol.* **2020**, *11*, 1969. [CrossRef]
124. Wallenhammar, A.-C. Prevalence of *Plasmodiophora brassicae* in a spring oilseed rape growing area in central Sweden and factors influencing soil infestation levels. *Plant Pathol.* **1996**, *45*, 710–719. [CrossRef]
125. Hirani, A.H.; Gao, F.; Liu, J.; Fu, G.; Wu, C.; McVetty, P.B.E.; Duncan, R.W.; Li, G. Combinations of Independent Dominant Loci Conferring Clubroot Resistance in All Four Turnip Accessions (*Brassica rapa*) From the European Clubroot Differential Set. *Front. Plant Sci.* **2018**, *9*. [CrossRef]
126. Manzanares-Dauleux, M.J.; Delourme, R.; Baron, F.; Thomas, G. Mapping of one major gene and of QTLs involved in resistance to clubroot in Brassica napus. *Theor. Appl. Genet.* **2000**, *101*, 885–891. [CrossRef]
127. Zhou, L.; Zhang, L.; He, Y.; Liu, F.; Li, M.; Wang, Z.; Ji, G. Isolation and characterization of bacterial isolates for biological control of clubroot on Chinese cabbage. *Eur. J. Plant Pathol.* **2014**, *140*, 159–168. [CrossRef]
128. Sedaghatkish, A.; Gossen, B.D.; Yu, F.; Torkamaneh, D.; McDonald, M.R. Whole-genome DNA similarity and population structure of *Plasmodiophora brassicae* strains from Canada. *BMC Genom.* **2019**, *20*, 744. [CrossRef]
129. Dixon, G. The Occurrence and Economic Impact of *Plasmodiophora brassicae* and Clubroot Disease. *J. Plant Growth Regul.* **2009**, *28*, 194–202. [CrossRef]
130. Hatakeyama, K.; Suwabe, K.; Tomita, R.N.; Kato, T.; Nunome, T.; Fukuoka, H.; Matsumoto, S. Identification and Characterization of Crr1a, a Gene for Resistance to Clubroot Disease (*Plasmodiophora brassicae* Woronin) in Brassica rapa L. *PLoS ONE* **2013**, *8*, e54745. [CrossRef]
131. Ueno, H.; Matsumoto, E.; Aruga, D.; Kitagawa, S.; Matsumura, H.; Hayashida, N. Molecular characterization of the CRa gene conferring clubroot resistance in *Brassica rapa*. *Plant Mol. Biol.* **2012**, *80*, 621–629. [CrossRef]

132. Chang, A.; Lamara, M.; Wei, Y.; Hu, H.; Parkin, I.A.P.; Gossen, B.D.; Peng, G.; Yu, F. Clubroot resistance gene Rcr6 in Brassica nigra resides in a genomic region homologous to chromosome A08 in *B. rapa*. *BMC Plant Biol.* **2019**, *19*, 224. [CrossRef]
133. Nguyen, M.L.; Monakhos, G.F.; Komakhin, R.A.; Monakhos, S.G. The New Clubroot Resistance Locus Is Located on Chromosome A05 in Chinese Cabbage (*Brassica rapa* L.). *Russ. J. Genet.* **2018**, *54*, 296–304. [CrossRef]
134. Hejna, O.; Havlickova, L.; He, Z.; Bancroft, I.; Curn, V. Analysing the genetic architecture of clubroot resistance variation in *Brassica napus* by associative transcriptomics. *Mol. Breed.* **2019**, *39*, 112. [CrossRef]
135. Zhou, Q.; Galindo-González, L.; Hwang, S.-F.; Strelkov, S.E. Application of genomics and transcriptomics to accelerate development of clubroot resistant canola. *Can. J. Plant Pathol.* **2020**. [CrossRef]
136. Strelkov, S.E.; Hwang, S.-F. Clubroot in the Canadian canola crop: 10 years into the outbreak. *Can. J. Plant Pathol.* **2014**, *36*, 27–36. [CrossRef]
137. Boland, G.J.; Hall, R. Index of plant hosts of *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* **1994**, *16*, 93–108. [CrossRef]
138. Guyon, K.; Balagué, C.; Roby, D.; Raffaele, S. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genom.* **2014**, *15*, 336. [CrossRef]
139. Liu, S.; Wang, H.; Zhang, J.; Fitt, B.D.L.; Xu, Z.; Evans, N.; Liu, Y.; Yang, W.; Guo, X. In vitro mutation and selection of doubled-haploid *Brassica napus* lines with improved resistance to *Sclerotinia sclerotiorum*. *Plant Cell Rep.* **2005**, *24*, 133–144. [CrossRef]
140. Mullins, E.; Quinlan, C.; Jones, P. Isolation of Mutants Exhibiting Altered Resistance to Sclerotinia sclerotiorum from Small M2 Populations of an Oilseed Rape (*Brassica napus*) Variety. *Eur. J. Plant Pathol.* **1999**, *105*, 465–475. [CrossRef]
141. Auclair, J.; Boland, G.J.; Cober, E.; Graef, G.L.; Steadman, J.R.; Zilka, J.; Rajcan, I. Development of a new field inoculation technique toassess partial resistance in soybean to Sclerotinia sclerotiorum. *Can. J. Plant Sci.* **2004**, *84*, 57–64. [CrossRef]
142. You, M.P.; Uloth, M.B.; Li, X.X.; Banga, S.S.; Banga, S.K.; Barbetti, M.J. Valuable New Resistances Ensure Improved Management of Sclerotinia Stem Rot (*Sclerotinia sclerotiorum*) in Horticultural and Oilseed Brassica Species. *J. Phytopathol.* **2016**, *164*, 291–299. [CrossRef]
143. Wang, Z.; Wan, L.; Xin, Q.; Chen, Y.; Zhang, X.; Dong, F.; Hong, D.; Yang, G. Overexpression of OsPGIP2 confers Sclerotinia sclerotiorum resistance in *Brassica napus* through increased activation of defense mechanisms. *J. Exp. Bot.* **2018**, *69*, 3141–3155. [CrossRef]
144. Falak, I.; Charne, D.R.; Patel, J.D.; Tulsieram, L. Sclerotinia-Resistant Brassica. US Patent US008558067B2, 15 October 2013.
145. Li, J.; Zhao, Z.; Hayward, A.; Cheng, H.; Fu, D. Integration analysis of quantitative trait loci for resistance to Sclerotinia sclerotiorum in Brassica napus. *Euphytica* **2015**, *205*, 483–489. [CrossRef]
146. Lohani, N.; Jain, D.; Singh, M.B.; Bhalla, P.L. Engineering Multiple Abiotic Stress Tolerance in Canola, Brassica napus. *Front. Plant Sci.* **2020**, *11*. [CrossRef] [PubMed]
147. Khanzada, H.; Wassan, G.M.; He, H.; Mason, A.S.; Keerio, A.A.; Khanzada, S.; Faheem, M.; Solangi, A.M.; Zhou, Q.; Fu, D.; et al. Differentially evolved drought stress indices determine the genetic variation of *Brassica napus* at seedling traits by genome-wide association mapping. *J. Adv. Res.* **2020**, *24*, 447–461. [CrossRef] [PubMed]
148. Khan, M.N.; Zhang, J.; Luo, T.; Liu, J.; Ni, F.; Rizwan, M.; Fahad, S.; Hu, L. Morpho-physiological and biochemical responses of tolerant and sensitive rapeseed cultivars to drought stress during early seedling growth stage. *Acta Physiol. Plant.* **2019**, *41*, 25. [CrossRef]
149. Wani, S.H.; Kumar, V.; Shriram, V.; Sah, S.K. Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *Crop J.* **2016**, *4*, 162–176. [CrossRef]
150. Zhu, X.; Wang, Y.; Liu, Y.; Zhou, W.; Yan, B.; Yang, J.; Shen, Y. Overexpression of BcHsfA1 transcription factor from Brassica campestris improved heat tolerance of transgenic tobacco. *PLoS ONE* **2018**, *13*, e0207277. [CrossRef]
151. Yang, H.; Wu, J.-J.; Tang, T.; Liu, K.-D.; Dai, C. CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus. *Sci. Rep.* **2017**, *7*, 7489. [CrossRef]

152. Wu, J.; Yan, G.; Duan, Z.; Wang, Z.; Kang, C.; Guo, L.; Liu, K.; Tu, J.; Shen, J.; Yi, B.; et al. Roles of the *Brassica napus* DELLA Protein BnaA6.RGA, in modulating drought tolerance by interacting with the ABA signaling component BnaA10.ABF2. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
153. Walsh, M.; Newman, P.; Powles, S. Targeting weed seeds in-crop: A New Weed Control Paradigm for Global Agriculture. *Weed Technol.* **2013**, *27*, 431–436.178. [CrossRef]
154. Manitoba Agriculture and Resource Development Office. *Guidelines: Crop Production Costs*; Manitoba Agriculture and Resource Development, Ed.; Manitoba Agriculture and Resource Development Office: Winnipeg, MB, Canada, 2020.
155. McLeod, R. *Annual Costs of Weeds in Australia*; eSYS Development Pty Limited: Canberra, Australia, 2018.
156. Beckie, H.J.; Flower, K.C.; Ashworth, M.B. Farming without Glyphosate? *Plants* **2020**, *9*. [CrossRef]
157. Duke, S.O.; Powles, S.B. Glyphosate: A once-in-a-century herbicide. *Pest Manag. Sci.* **2008**, *64*, 319–325. [CrossRef]
158. Asaduzzaman, M.; Pratley, J.E.; Luckett, D.; Lemerle, D.; Wu, H. Weed management in canola (*Brassica napus* L): A review of current constraints and future strategies for Australia. *Arch. Agron. Soil Sci.* **2020**, *66*, 427–444. [CrossRef]
159. Brookes, G.; Barfoot, P. Environmental impacts of genetically modified (GM) crop use 1996–2018: Impacts on pesticide use and carbon emissions. *GM Crops Food* **2020**, *11*, 215–241. [CrossRef] [PubMed]
160. Green, J.M. Current state of herbicides in herbicide-resistant crops. *Pest Manag. Sci.* **2014**, *70*, 1351–1357. [CrossRef] [PubMed]
161. Wallace, D.B.; Laima, S.K. Development of Triazine Resistance in Crops by Classical Plant Breeding. *Weed Sci.* **1987**, *35*, 9–11.
162. Funke, T.; Han, H.; Healy-Fried, M.L.; Fischer, M.; Schönbrunn, E. Molecular basis for the herbicide resistance of Roundup Ready crops. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13010–13015. [CrossRef]
163. ISAAA's GM Approval Database. 2020. Available online: http://www.isaaa.org/gmapprovaldatabase/ (accessed on 24 July 2020).
164. ACIL Tasman Pty Ltd. *GM Canola: An Information Package*; Department of Agriculture, Fisheries and Forestry Australia: Canberra, Australia, 2007. Available online: https://www.agriculture.gov.au/sites/default/files/sitecollectiondocuments/ag-food/biotech/gm-canola-info-package.pdf (accessed on 15 August 2020).
165. Kirkegaard, J.A.; Lilley, J.M.; Morrison, M.J. Drivers of trends in Australian canola productivity and future prospects. *Crop Pasture Sci.* **2016**, *67*, i–ix. [CrossRef]
166. Benbrook, C.M. Trends in glyphosate herbicide use in the United States and globally. *Environ. Sci. Eur.* **2016**, *28*, 3. [CrossRef]
167. Fernandez-Cornejo, J.; Wechsler, S.; Milkove, D. The Adoption of Genetically Engineered Alfalfa, Canola, and Sugarbeets in the United States. Agriculture, U. S. Department of Agriculture, Ed.; 2016. Available online: https://www.ers.usda.gov/webdocs/publications/81176/eib-163_summary.pdf?v=0 (accessed on 17 July 2020).
168. Aghazadeh, R.; Zamani, M.; Motallebi, M.; Moradyar, M. Agrobacterium -Mediated Transformation of the Oryza sativa Thaumatin-Like Protein to Canola (R Line Hyola308) for Enhancing Resistance to Sclerotinia sclerotiorum. *Iran. J. Biotechnol.* **2017**, *15*, 201–207. [CrossRef]
169. Sutherland, S. Canola Weed Management. 1999. Available online: http://www.australianoilseeds.com/__data/assets/pdf_file/0012/2712/Chapter_12__Canola_Weed_Management.pdf (accessed on 17 July 2020).
170. Heap, I.; Duke, S.O. Overview of glyphosate-resistant weeds worldwide. *Pest Manag. Sci.* **2018**, *74*, 1040–1049. [CrossRef]
171. Antier, C.K.P.; Reboud, X.; Ulber, L.; Baret, P.V.; Messéan, A. Glyphosate Use in the European Agricultural Sector and a Framework for Its Further Monitoring. *Sustainability* **2020**, *12*, 5682. [CrossRef]
172. Department of Primary Industries and Regional Development (DPIRD). *Canola Variety Sowing Guide for Western Australia*; DPIRD: South Perth, WA, Australia, 2019. Available online: https://www.agric.wa.gov.au/sites/gateway/files/DPIRD%20canola%20variety%20sowing%20guide%202019%20Bulletin%204897.pdf (accessed on 17 July 2020).
173. Idnurm, A.; Urquhart, A.S.; Vummadi, D.R.; Chang, S.; Van de Wouw, A.P.; López-Ruiz, F.J. Spontaneous and CRISPR/Cas9-induced mutation of the osmosensor histidine kinase of the canola pathogen Leptosphaeria maculans. *Fungal Biol. Biotechnol.* **2017**, *4*, 12. [CrossRef] [PubMed]

174. Yu, F.; Zhang, X.; Peng, G.; Falk, K.C.; Strelkov, S.E.; Gossen, B.D. Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of *Plasmodiophora brassicae* in *Brassica rapa*. *Sci. Rep.* **2017**, *7*, 4516. [CrossRef] [PubMed]
175. Fu, F.; Liu, X.; Wang, R.; Zhai, C.; Peng, G.; Yu, F.; Fernando, W.G.D. Fine mapping of *Brassica napus* blackleg resistance gene *Rlm1* through bulked segregant RNA sequencing. *Sci. Rep.* **2019**, *9*, 14600. [CrossRef] [PubMed]
176. Hopkins, R.J.; Dam, N.M.V.; Loon, J.J.A.V. Role of Glucosinolates in Insect-Plant Relationships and Multitrophic Interactions. *Annu. Rev. Entomol.* **2009**, *54*, 57–83. [CrossRef]
177. Bell, L. The Biosynthesis of Glucosinolates: Insights, Inconsistencies, and Unknowns. In *Annual Plant Reviews Online*; Roberts, J.A., Ed.; 2019. [CrossRef]
178. Lin, P.; Qin, S.; Pu, Q.; Wang, Z.; Wu, Q.; Gao, P.; Schettler, J.; Guo, K.; Li, R.; Li, G.; et al. CRISPR-Cas13 Inhibitors Block RNA Editing in Bacteria and Mammalian Cells. *Mol. Cell* **2020**, *78*, 850–861.e855. [CrossRef]
179. Mackay, T.F. The Genetic Architecture of Quantitative Traits. *Annu. Rev. Genet.* **2001**, *35*, 303–339. [CrossRef]

genes

MDPI

Article

Genomic Analysis of Selected Maize Landraces from Sahel and Coastal West Africa Reveals Their Variability and Potential for Genetic Enhancement

Charles Nelimor [1,2,3], Baffour Badu-Apraku [2,*], Ana Luísa Garcia-Oliveira [4], Antonia Tetteh [5], Agre Paterne [2], Assanvo Simon-Pierre N'guetta [3] and Melaku Gedil [2]

1 West African Science Service Centre on Climate Change and Adapted Land-use (WASCAL), Université Felix Houphouët Boigny, Abidjan 22 BP 461, Côte d'Ivoire; nelimor.c@edu.wascal.org
2 International Institute of Tropical Agriculture, Ibadan 200001, Nigeria; P.Agre@cgiar.org (A.P.); m.gedil@cgiar.org (M.G.)
3 Department of Bioscience, Université Felix Houphouët Boigny, Abidjan 22 BP 461, Côte d'Ivoire; nguettaewatty@gmail.com
4 Excellence in Breeding (EiB) Platform, International Maize and Wheat Improvement Center (CIMMYT), ICRAF House, Nairobi 00100, Kenya; a.oliveira@cgiar.org
5 Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology, University Post Office Box PMP, Kumasi 00233, Ghana; aytetteh@gmail.com
* Correspondence: b.badu-apraku@cgiar.org

Received: 10 August 2020; Accepted: 28 August 2020; Published: 7 September 2020

Abstract: Genetic adaptation of maize to the increasingly unpredictable climatic conditions is an essential prerequisite for achievement of food security and sustainable development goals in sub-Saharan Africa. The landraces of maize; which have not served as sources of improved germplasm; are invaluable sources of novel genetic variability crucial for achieving this objective. The overall goal of this study was to assess the genetic diversity and population structure of a maize panel of 208 accessions; comprising landrace gene pools from Burkina Faso (58), Ghana (43), and Togo (89), together with reference populations (18) from the maize improvement program of the International Institute of Tropical Agriculture (IITA). Genotyping the maize panel with 5974 DArTseq-SNP markers revealed immense genetic diversity indicated by average expected heterozygosity (0.36), observed heterozygosity (0.5), and polymorphic information content (0.29). Model-based population structure; neighbor-joining tree; discriminant analysis of principal component; and principal coordinate analyses all separated the maize panel into three major sub-populations; each capable of providing a wide range of allelic variation. Analysis of molecular variance (AMOVA) showed that 86% of the variation was within individuals; while 14% was attributable to differences among gene pools. The Burkinabe gene pool was strongly differentiated from all the others (genetic differentiation values >0.20), with no gene flow (Nm) to the reference populations (Nm = 0.98). Thus; this gene pool could be a target for novel genetic variation for maize improvement. The results of the present study confirmed the potential of this maize panel as an invaluable genetic resource for future design of association mapping studies to speed-up the introgression of this novel variation into the existing breeding pipelines.

Keywords: landraces; genetic diversity; population structure; West Africa; maize improvement; DArTseq markers

1. Introduction

Maize (*Zea mays* L.) is one of the most important cereal crops consumed in sub-Saharan Africa (SSA) and an essential component of livestock feed in the developed as well as developing world. Approximately 10,000 years ago in Central Mexico, maize was domesticated from its ancestor, the

wild grass-teosinte (*Zea mays* spp. *parviglumis*) [1]. Post-domestication of teosinte led to the transfer of beneficial adaptive genomic regions into common maize [2]. Mutations together with recombination events, either due to natural or farmer-mediated selection, then generated novel allele combinations [3,4]. Thus arose the "landraces", traditional varieties selected by farmers for adaptation to local conditions and food preferences [5]. The landraces, though typically low in yield, are invaluable sources of diversity that could be drawn upon to broaden the genetic base of elite maize germplasm, and to further enhance adaptation to changing environments and pathogens [6]. Indeed, earlier breeders identified and composited the most productive landraces into genetically diverse populations, forming the foundation of inbred line development and pedigree breeding [7]. Modern breeding practices, in which a narrow range of inbred lines are included in crossing programs, have narrowed the genetic base of most cultivated crops, which have negatively affected the adaptability of the crop to changing climates, devastating pathogens, and insect-pests [8]. In order to provide a buffer against the possible effects of novel threats, it is essential to broaden the genetic base of breeding populations by introgressing an enlarged pool of beneficial alleles.

The high selection pressure under low input and climatically stressed maize growing environments in Africa is likely to have resulted in local adaptations with potential value for breeding stress tolerant and nutritionally enriched varieties [9]. Indeed, there are examples of successful use of local African genetic resources of maize in the development of varieties such as Katumani in Kenya and Longe-5 from Uganda [9]. However, while the New World maize gene pool is well represented in genebanks [10] and well characterized [3], there are collections of African maize germplasm without adequate data of their genetic make-up [9]. This missing genetic data makes searching for promising landraces within the African maize genebank collections like "finding the proverbial needle in a haystack" [11]. Therefore, to develop improved climate resilient maize varieties for SSA, efficient characterization, identification, and utilization of climatically adapted local African maize germplasm is a crucial prerequisite [9].

Recently, we phenotypically characterized a panel of maize landraces originating from Burkina Faso, Ghana, and Togo, which cover a wide range of climatic conditions classified as Sahel and coastal West Africa (WA) [12]. The study revealed that the maize panel varied considerably in flowering date, plant architecture, yield and yield related traits, and other characteristics. These differences allowed the formation of five distinctive morphological groups [12]. The Sahel gene pool was highly distinct and was considered a valuable resource for future genetic enhancement. However, phenotypic variation can be confounded by the environment and a high degree of plasticity [13]. Given that gene flow from multiple introductions may have shaped the population structure of African maize, diversity analysis of this panel using state-of-the-art genotyping techniques could help provide deep insights into the complexity of its genetic architecture and composition.

The assessment of genetic diversity by genotyping-by-sequencing (GBS) provides robust estimates of diversity and has been increasingly adopted as a fast, high-throughput, and affordable tool for whole-genome genetic diversity analysis in large germplasm sets [14]. The diversity array technology sequence (DArTseq) markers, characterized by high marker coverage, call rates, and scoring reproducibility, has emerged as a useful GBS approach for assessing genetic diversity and population structure in various crops including wheat [15,16], rice [17], watermelon [18], common bean [19], and maize [20,21]. The objective of this study was to examine the genetic diversity and population structure of a maize panel comprising landraces from Sahel and coastal Africa together with a reference population using the GBS-DArTseq approach. The relevance of our results for further exploration and utilization of local genetic resources of maize in Africa is discussed in this manuscript.

2. Materials and Methods

2.1. Plant Material

We analyzed 208 maize accessions obtained from international and national gene banks in Africa (Supplementary Table S1). The maize panel comprised 190 landraces representing gene pools from

Burkina Faso (58), Ghana (43), and Togo (89) (Supplementary Figure S1). The landraces from Burkina Faso and Togo were sourced from the gene bank at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, whereas those from Ghana were provided by the Plant Genetics Resources Institute (PGRI) at Bunso, Ghana. The majority of these landraces were collected from farmers' fields in the 1970s and 1980s. However, the eco-geographical data of the collection sites of the landraces were not available. The study also included a diverse set (18) of drought- and heat-tolerant open pollinated populations (hereafter referred to as reference populations) developed by the Maize Improvement Program at IITA (MIP-IITA), Ibadan, Nigeria.

2.2. DNA Isolation and Genotyping Analysis

For each accession, total genomic DNA was isolated from bulked leaf composites from 15 seedlings at two weeks old according to the DArT protocol (https://www.diversityarrays.com/orderinstructions/plant-dna-extraction-protocol-for-dart/). The quality of each DNA sample was visualized by electrophoresis on 0.8% agarose gel, and the purified DNA was further quantified using a nano-drop spectrophotometer (Thermo Scientific, Wilmington DE, USA). Certified DNA samples were then sent to the Integrated Genomic Service and Support (IGSS) genotyping platform, Nairobi, Kenya, for genotyping. High-throughput genotyping was conducted in 96 plex DArTseq protocol, and SNPs were called using the DArT's proprietary software, DArTSoft, as previously described [22]. Reads and tags found in each sequencing result were aligned to the *Zea mays* L. genome reference, version *AGPV3* (B73 Ref-Gen v4 assembly) [23].

2.3. Data Analysis

A total of 47,441 putative DArTseq markers were generated from the 208 maize panel. Prior to further analysis, the raw data set was filtered to remove markers with call rate <0.8, minor allele frequency (MAF) <0.05, and unmapped SNP markers. Thereafter, markers with no missing rate were retained using the TASSEL software version 5.2.12 [24]. The retained markers were subjected to various genetic diversity analyses including basic diversity statistics such as polymorphic information content (PIC), MAF, observed heterozygosity (H_o), and expected heterozygosity (H_e) using PowerMarker v. 3.2.5 [25]. The population structure of the maize panel was inferred using the Admixture model-based clustering algorithm implemented STRUCTURE 2.3.4 [26]. The *adhoc* number of clusters (k) was varied from 1 to 12, with 10,000 burn-in steps, followed by 10,000 Markov chain Monte Carlo simulations, as previously described [20,27]. For each k, ten independent iterations were implemented. The most likely number of k was determined by the ad hoc Δ k statistics [28] embedded in Structure Harvester [29]. Accessions with membership proportions (Q-value) ≥80% were assigned to groups, while those with membership probabilities less than 80% were designated as admixtures [30]. The population structure of each gene pool (Burkinabe, Ghanaian, Togolese, and reference populations) were also estimated as described above. A discriminant analysis of principal components (DAPC) was carried out on the 208 maize panel using the first 40 principal components using the adegenet R package [31]. Membership probabilities of the individuals for the different gene pools were estimated using the "find cluster" function implemented in adegenet. Further, principal coordinate analysis (PCoA) was conducted to reveal the genetic relationships among the maize accessions using GenAlEx v. 6.503 software [32]. An unrooted neighbor-joining (NJ) tree was constructed by following the procedure of Nei [33] with 1000 bootstrap replicates in PowerMarker v3.25 [25]. The resulting NJ tree was visualized in Molecular Evolutionary Genetics Analysis (MEGA) software version X [34] and edited using Figtree software v.1.4.4 [35]. Genetic relationships within each maize gene pool were elucidated through construction of an unrooted NJ tree, as described above. Analysis of molecular variance (AMOVA) was estimated in GenAlEx v. 6.503 [32] to partition components of genetic variance among and within the populations (k). Calculation of pairwise genetic differentiation statistics (F_{ST}) and haploid number of migrants (Nm) between gene pools was performed using GenAlEx v6.503 [32] with 999 permutations. F_{ST} measures the amount of genetic variance that can be explained by population

structure based on Wright's F-statistics [36], while $Nm = [(1/F_{ST}) - 1]/4$. An Nm value less than 1 indicates limited gene exchange among subpopulations [36].

3. Results

3.1. Analysis of Genetic Diversity Parameters

Out of the total 47,441 putative DArTseq markers, 5974 were retained after filtering. The GBS-DArTseq markers were unequally distributed across the ten chromosomes of the 208 maize panel. Chromosome 1 had the highest number of markers (905), while chromosome 10 had the least (422) (Supplementary Figure S2). H_e, H_o, MAF, and PIC values estimated for the entire maize panel (208 accessions) averaged 0.36, 0.50, 0.28, and 0.29, respectively (Table 1). The average values of H_e, Ho, MAF, and PIC were 0.30, 0.41, 0.23, and 0.24, respectively, for the Burkinabe gene pool, and 0.32, 0.34, 0.23, and 0.26, respectively, for the Ghanaian gene pool. The average H_e, H_o, MAF, and PIC values of the Togolese gene pool were 0.36, 0.50, 0.27, and 0.28, respectively. Furthermore, the He, H_o, MAF, and PIC values of the reference population averaged 0.36, 0.47, 0.28, and 0.29, respectively. For the landraces as a group, H_e, H_o, MAF, and PIC values averaged 0.37, 0.50, 0.28, and 0.29, respectively.

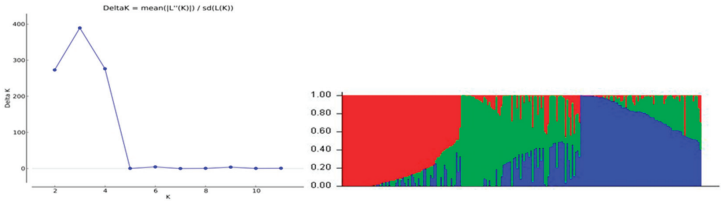
Table 1. Diversity statistics based on 5974 DArTseq markers across 208 maize accessions.

Maize Panel	N	H_e	H_o	MaF	MAF	PIC
Burkina Faso	58	0.30	0.41	0.77	0.23	0.24
Ghana	43	0.32	0.34	0.77	0.23	0.26
Togo	89	0.36	0.50	0.73	0.27	0.28
Reference population	18	0.36	0.47	0.72	0.28	0.29
Landraces	190	0.37	0.50	0.73	0.27	0.29
Entire Panel	208	0.36	0.50	0.72	0.28	0.29

N/number of accessions; H_e/expected heterozygosity; H_o/observed heterozygosity; MaF/major allele frequency; MAF/minor allele frequency; PIC/polymorphic information content.

3.2. Population Structure and Genetic Relationships

The model-based simulation of population structure analysis of the maize panel (208 accessions) showed that the delta K values from the mean log-likelihood probabilities plateaued at K = 3 (389.43), followed by K = 4 (276.33), and K = 2 (273.07) (Figure 1a). At K = 3, the 208 maize panel was divided into three sub-populations (Figure 1b). Using an 80% membership probability threshold, 122 accessions (58.65%) were successfully assigned to the three subpopulations. In comparison, 86 accessions with a probability of associations less than 80% were designated as an admixed population (Supplementary Table S2). Subpopulation 1 was the most uniform (membership coefficient averaged, 90%), and it contained 53 landraces (49 from Burkina Faso, 3 from Togo, and 1 from Ghana). Subpopulations 2 and 3, which constituted 12.5% and 20.67% of the panel, respectively, were admixtures of Ghanaian and Togolese landraces, together with reference populations. Specifically, subpopulation 2 consisted of 26 accessions, 10 reference populations, and 11 and 5 Togolese and Ghanaian landraces, respectively. Subpopulation 3 comprised 24 and 15 landrace accessions from Ghana, and Togo, respectively, and 4 accessions from the reference population. The admixed group contained 60, 13, and 9 landraces from Togo, Ghana, and Burkina Faso, respectively, plus 4 reference populations (Supplementary Table S2). The additional smaller peaks observed at K = 4 (276.33) and K = 2 (273.07) implied the presence of subgroups within the three major groups (Figure 1). Therefore, an independent STRUCTURE run was performed for each gene pool. Sub-clustering of the Burkinabe and Ghanaian gene pools both yielded a sharp peak at K = 2 (Figure 2a,b). Sub-clustering the reference populations and Togolese landraces showed the highest peak at K = 3, and K = 9, respectively (Figure 2c,d). A substantial degree of admixture was observed for each gene pool (Supplementary Table S3).



(**a**) (**b**)

Figure 1. Graphical representation of the population structure of the 208 maize panel. (**a**) Plot of mean likelihood of delta K against the number of K groups. The highest peak observed at K = 3 signifies the grouping of accessions into three groups, while the small peak at K = 2 and 4 signifies further grouping of accessions into two and four groups, respectively. (**b**) Subpopulations at K = 3. The colors represent three subpopulations of 208 accessions. The separation of accessions into subpopulation 1 (red), 2 (green), and 3 (blue) was based on membership coefficient ≥80%.

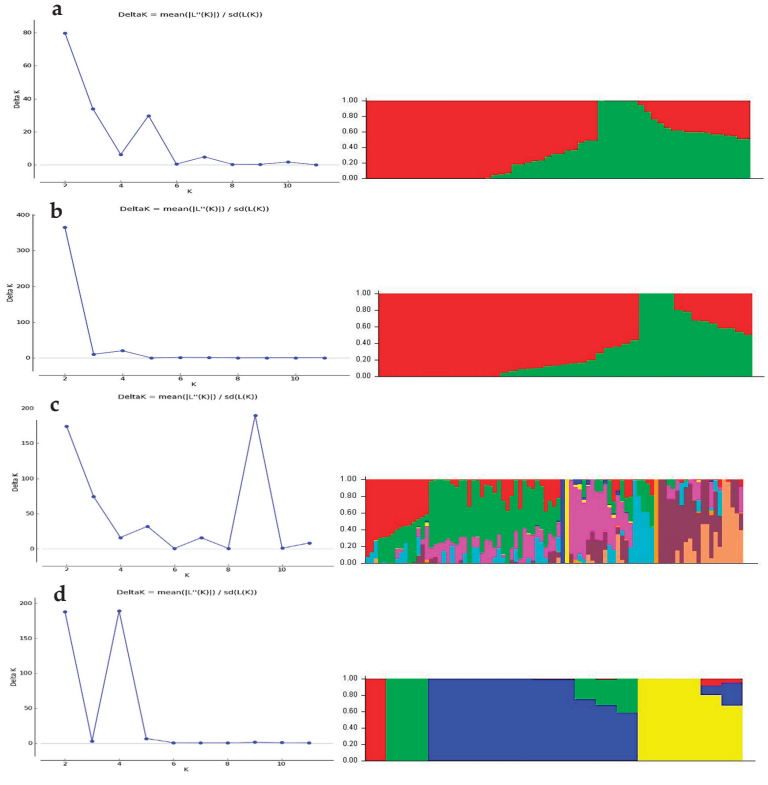


Figure 2. Population structure of Burkinabe (**a**), Ghanaian (**b**), and Togolese (**c**) landrace gene pools, including a reference population (**d**) at K = 2, K = 9, and K = 4, respectively.

Using the Bayesian information criterion (BIC) implemented in DAPC, a maximum of K = 3 was obtained, which corresponded to three groups of maize accessions in the panel (Figure 3). Estimation of the cluster membership revealed that cluster three had the highest number of accessions (94) followed by cluster two with 77 accessions, and cluster one with the smallest number of accessions (37). Of the

94 accessions in cluster three, 58 (61.70%) and 29 (30.85%) were landraces from Togo and Ghana, respectively, including six reference populations and the landraces from Burkina Faso (Supplementary Table S4). All the accessions in cluster two were landraces from Burkina Faso (57), Togo (17), and Ghana (3). Of the 37 accessions in cluster 1, 14 (37.84%) were Togolese landraces, 12 (32.43%) were from the reference populations, while 11 (29.73%) were Ghanaian landraces.

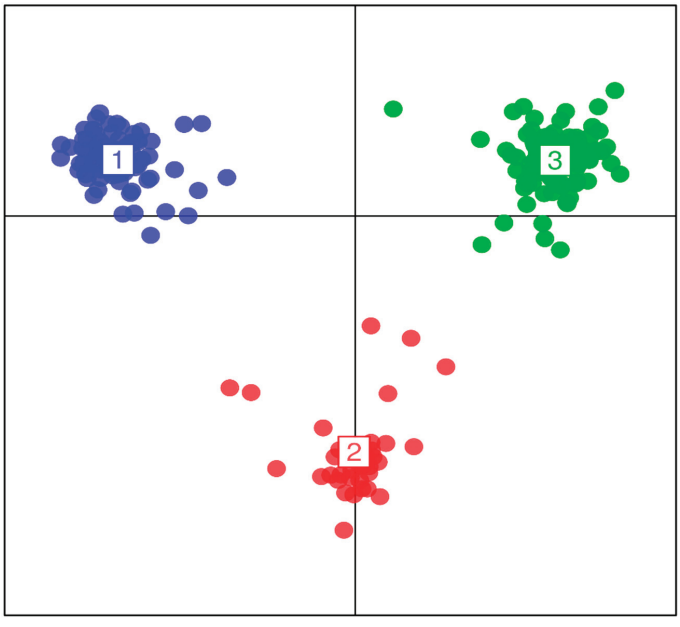


Figure 3. Discriminant analysis of principal components (DAPC) using 5974 DArTseq markers. The axes represent the first two linear discriminants (LD). Each color represents a cluster, while each dot represents an individual. Numbers represent the different subpopulations identified by DAPC analysis.

Further investigation of the genomic structure of the maize panel using the PCoA indicated three subpopulations as per the STRUCTURE simulation and DAPC analyses (Figure 4). The total amount of genetic variation explained by the first two principal coordinates was 57%. The PCoA clearly separated subpopulation 3 (by PCo2), which showed a higher degree of admixture between Ghanaian and Togolese landraces, including six reference populations and a landrace from Burkina Faso. The other two subpopulations appeared to be distributed along PCo1. Although some degree of overlap among landrace gene pools was shown in subpopulation 1, located at the upper extreme of PCo1, ~75% were Burkinabe landraces. Subpopulation 2 distributed along the lower extreme of PCo1 was the most distant of the three, comprising the majority of the reference populations and four Ghanaian landraces.

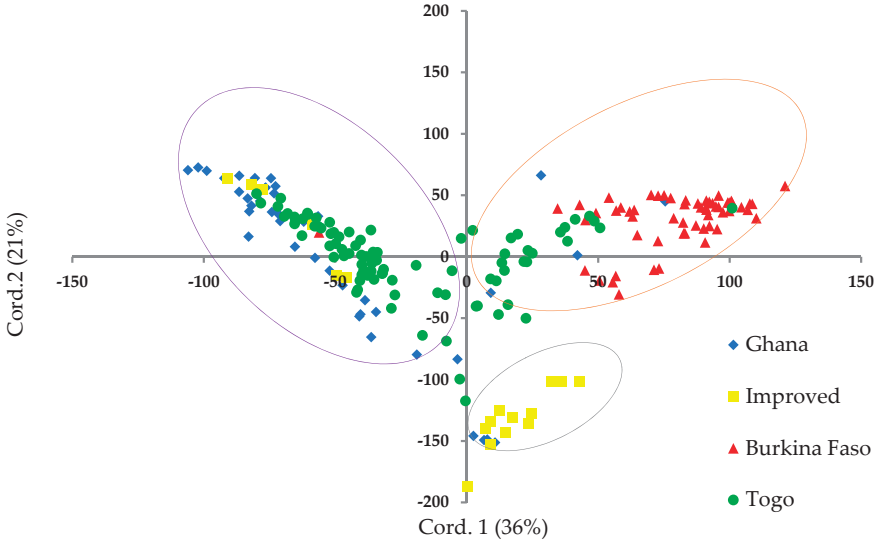


Figure 4. Cluster analysis of 208 maize accessions using principal coordinate analysis (PCoA). Accessions are colored according to origin. The orange, ash, and violet circles represent subpopulation 1, 2, and 3, respectively.

As per the STRUCTURE, DAPC, and PCoA results, the NJ phylogenetic tree also showed three sub-populations with higher degrees of admixture among Ghanaian and Togolese landraces, and reference populations (Figure 5). The neighbor-joining tree performed for each gene pool divided the Burkinabe and Ghanaian gene pools into two main clusters (Figure 6b). The Togolese landraces and the reference populations were grouped into nine and three clusters, respectively (Figure 6b,c).

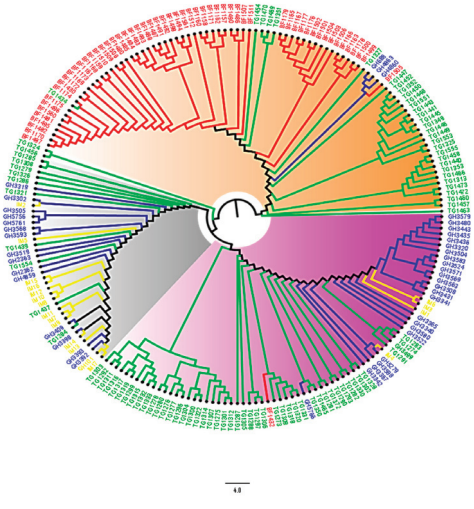


Figure 5. Phylogenetic tree estimated through the neighbor-joining method for 208 maize accessions from West Africa. The green, red, blue, and yellow clades and taxa represent Togolese, Burkinabe, Ghanaian, and reference populations, respectively. The orange, violet, and ash highlights represent subpopulations 1, 2, and 3, respectively.

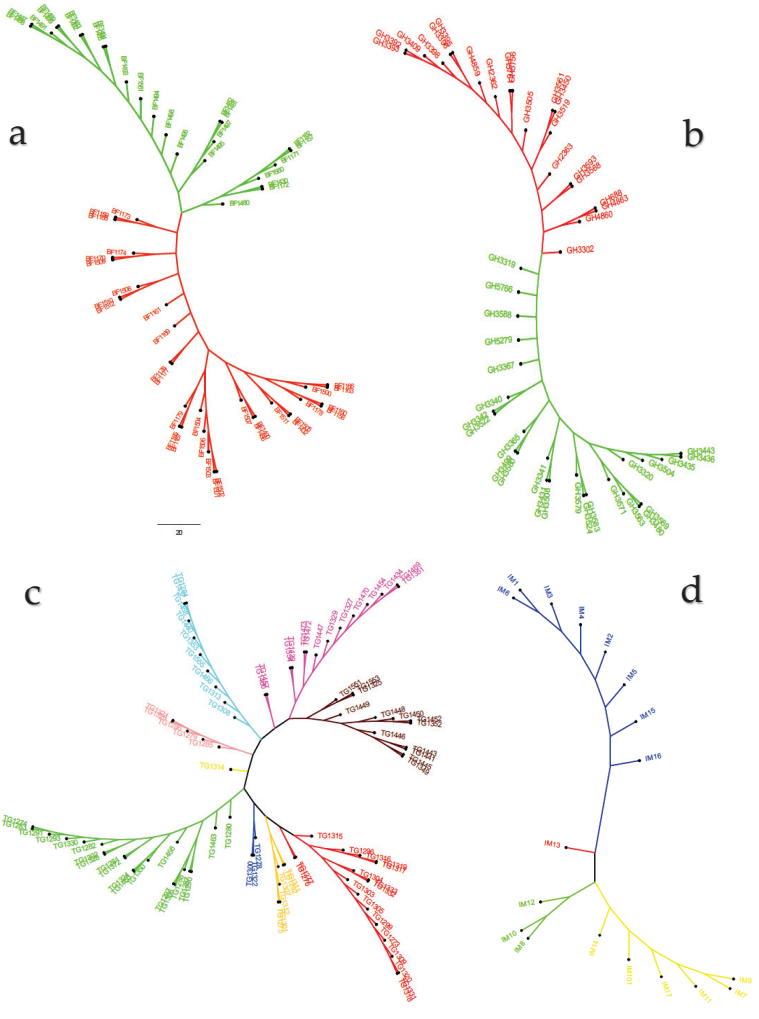


Figure 6. Unrooted neighbor joining tree depicting genetic relationships among Burkinabe (**a**), Ghanaian (**b**), and Togolese (**c**) landrace gene pools, and a reference population (**d**). Colors of each tree represent genetic groups.

3.3. Analyses of Molecular Variance, Genetic Differentiation, and Gene Flow among Gene Pools

The AMOVA revealed that 14% of the total variation was found among gene pools, while the rest (86%) was within gene pools (Table 2). The overall F_{ST} value of the maize panel was 0.21, and the Nm value was 1.58. As shown in Table 3, the Burkinabe gene pool had the highest F_{ST} value (0.28), and the Ghanaian and Togolese gene pools had the lowest (0.18, each). The pairwise F_{ST} values ranged from 0.14 (Ghanaian vs. Togolese) to 0.31 (Burkinabe vs. reference populations). Similarly, Nm values between gene pools varied from 0.98 (Burkinabe vs. reference populations) to 2.83 (Ghanaian vs. reference populations). The Nm value between the Ghanaian and Togolese gene pools was 2.63.

Table 2. Analysis of molecular variance (AMOVA) using 5974 DArTseq markers of the genetic variation among and within four gene pools of 208 maize accessions.

Source	df	SS	MS	Est. Var.	% Var.	*p* Value
Among gene pools	3	54,459.23	18,153.08	168.46	14	0.001
Among individuals	204	420,226.27	2059.93	993.38	80	0.001
Within individuals	208	15,221.5	73.18	73.18	6	0.001
Total	415	489,907.0		1235.02	100	
Fixation index (F_{ST})	0.21					0.001
Gene flow (Nm)	1.58					0.001

df/degree of freedom, SS/Sum of square; MS/Mean sum of square, Est. Var./Estimated variance; Var./Variance.

Table 3. Measure of genetic population differentiation (F_{ST}) (lower diagonal), and estimation of gene flow (Nm) (upper diagonal) within and among the four gene pools of maize accessions.

	Burkina Faso	Ghana	Togo	Reference Population
Burkina Faso	-	1.18	1.31	0.98
Ghana	0.27	-	2.63	2.82
Togo	0.24	0.14	-	2.03
Reference population	0.31	0.14	0.17	-
Average F_{ST}	0.28	0.18	0.18	0.21

4. Discussion

A well-characterized and diverse germplasm is an essential requisite for genetic enhancement of crops. In this study, we applied GBS technology to explore the genetic diversity and population structure of a maize panel comprising landrace gene pools from Burkina Faso, Ghana, and Togo, plus a reference population from IITA-MIP. The results of the estimated diversity indices revealed ample genetic diversity within the maize panel indicated by average H_e (0.36) and H_o (0.5). The He obtained in this study was comparable to the 0.36 reported for provitamin A (PVA) quality protein maize (QPM) germplasm from IITA-MIP [21] but was higher than that reported for maize landraces from Eastern Africa (H_e = 0.25), Western Africa (H_e = 0.18), and Sahel Africa (H_e = 0.24) [9] as well as tropical maize breeding populations (H_e = 0.22) [27] including IITA early-maturing white inbred lines [20]. Characterization of the Burkinabe, Ghanaian, and Togolese maize pools showed different values for the estimated diversity indices. The results indicated that the Togolese gene pool (H_e = 0.36, H_o = 0.50) contained slightly higher diversity than the Burkinabe (H_e = 0.30, Ho = 0.41) and Ghanaian (H_e= 0.32, H_o = 0.34) landrace pools. Further, the low variation in the genetic indices identified between the landraces as a group, and the reference populations showed that the two germplasm sets possessed similar genetic diversity (Table 1). These results agreed with previous findings that tropical maize germplasm is highly diverse with H_e > 0.3 [37–39]. The mean PIC obtained in the present study, 0.29 using 5974 DArTseq SNPs for the 208 maize accessions was higher than the 0.19 and 0.26 reported for tropical early-maturing maize inbred lines using 15,047 [30] and 7224 SNPs for a sample size of 94 and 134, respectively [20,27]. The discrepancies between the results of our study and those of earlier researchers may be due to the use of different genetic materials, the sample sizes, and the number of SNPs used. Nonetheless, the mean PIC value in this study was like the 0.29 recently reported for tropical PVA-QPM maize germplasm using 8171 DArTseq SNP markers [21].

The Evanno criterion employed for the model-based simulation of population structure identified the peak level of ΔK at K = 3 (Figure 1a), which depicted the presence of three genetically distinct subpopulations (Figure 1b). The proportion of admixed accessions (47%) in the maize panel, based on a membership probability threshold of 80%, suggested moderate genetic differentiation and gene flow. The DAPC, PCoA, and NJ phylogenetic analyses results all illustrated the existence of

three subpopulations in the whole set of 208 maize accessions. Comparison of the results of the four complementary clustering methods (STRUCTURE, DAPC, NJ tree, and PCoA) revealed high consistency in the individuals assigned to each group, which reinforced the findings that the identified groups were indeed genetically distinct. The close proximity between Togolese and Ghanaian gene pools suggested high genetic relatedness of the two gene pools. This result was expected due to the geographical proximity of the two countries and the similarity of the climatic conditions. The Burkinabe gene pool largely diverged from all others, suggesting its adaptation to Sahel conditions, which is in agreement with its pattern of phenotypic diversity [12]. Multivariate analyses revealed high affinity of Ghanaian and Togolese landraces with the reference populations (Figures 1 and 3–5). It is likely that some of these accessions are not true landraces but, rather, old improved cultivars that were either recollected or wrongly classified, as farmers usually consider improved varieties cultivated over longer periods in a given area as landraces [40]. The grouping together of some landraces with the reference populations also suggested a pedigree relationship. Hence, it is possible that some of the landraces analyzed in this study were local varieties that were selected by earlier maize breeders in IITA, based on high grain yield, earliness, and resistance to the maize streak virus (MSV), and adaptation to the drought and heat stress as starting materials for the development of inbred lines that were later involved in cross-breeding (see http://r4dreview.iita.org/index.php/tag/maize-improvement/). The additional smaller peaks observed at K = 4 (276.33) and K = 2 (273.07) implied the presence of subgroups within the three major groups (Figure 1). Therefore, an independent STRUCTURE run was performed for each gene pool (Figure 2). The high degree of genetic admixtures within each landrace gene pool observed with ancestry share of <80% probably reflects considerable levels of gene flow or germplasm exchange. Results of previous studies have shown that such an admixture is not unusual in landraces from restricted geographical backgrounds [40].

According to Frankham et al. [41], an F_{ST} value greater than 0.15 can be considered as significantly differentiating populations. Thus, in the present study, the overall F_{ST} value (0.21) supported the presence of significant genetic divergence within the maize panel. Wright [36] reported that an Nm value less than 1 indicated limited gene exchange among populations. In the present study, the overall Nm value of 1.58 (Table 2) indicated that moderate genetic exchange or gene flow may have occurred, leading to the moderate genetic differentiation between gene pools. This observation was consistent with the AMOVA results (Table 2), which indicated that 14% of the total variation was accounted for by gene pool variations. This result is consistent with the findings of previous studies [42]. According to the F_{ST} values, the Burkinabe gene pool was the most differentiated (Table 3), in agreement with its divergence as revealed by the clustering methods (STRUCTURE, PCoA, DAPC, and NJ analyses). The divergence between the reference populations and landraces varied among the different gene pools. In particular, the low affinity of the Burkinabe gene pool with the reference population (F_{ST} = 0.31, Nm = 0.98) suggested little involvement of the original Sahelian gene pool in the development of the modern maize varieties presently grown in the sub region. This observation is biologically and historically meaningful since in West Africa, the reference maize gene pool called Composite Y [43], which was developed through recombination of 145 flint landraces of West Africa savannah zone, contained only 2% each of the genetic materials from Burkina Faso and Niger, as well as 1% of those from Senegal [44]. In the analyses of the isozyme variability in West African maize cultivars, Sanou et al. [45] showed that Burkinabe landraces were distinct, even though some levels of gene flow between them and an elite open pollinated variety (SR 22) developed by IITA in 1984 from CIMMYT Pop 22 and widely adopted in Burkina Faso [46] was observed. Therefore, the Burkinabe gene pool, having been grown and selected by farmers over many generations under warmer and drier conditions, could harbor novel and favorable alleles for improving maize for tolerance to drought and heat stresses. It is notable that in our earlier work on this maize panel, the high degree of tolerance of the Burkinabe landraces to drought, heat, and the combined heat and drought stresses was unrivalled [47,48]. The high genetic similarity observed between the Ghanaian and Togolese landraces was supported by their low F_{ST} (0.14) and high Nm (2.63) values. This result further reflected

the gene flow via seed exchanges and local preferences towards a given agrotype owing to similar climatic conditions. These may have significantly shaped the distribution of the genetic diversity within Ghanaian and Togolese maize landraces, as was previously suggested [12]. The F_{ST} and Nm values (Table 3) suggested that the Ghanaian gene pool was closer to the reference populations, in agreement with PCoA stratification (Figure 4). Indeed, the reference set analyzed in this study included two popular cultivars that are commonly cultivated in Ghana (Aburoheema and Obatanpa GH, coded IM1 and IM6, respectively). The deep knowledge of the genetic diversity and structure of Sahel and coastal West African maize landraces revealed in the present study provides an essential platform for efficient use of these valuable maize gene pools.

5. Conclusions

In the present study, we explored the genetic diversity and relationships within and between a maize panel comprising landrace gene pools from Burkina Faso, Ghana, and Togo and compared each to a reference maize population. The analysis of genetic diversity parameters indicated ample genetic diversity in the maize panel. The four multivariate methods were consistent in dividing the maize panel into three distinct genetic groups, each capable of providing different sources of variation for maize genetic enhancement. The genetic divergence of the Burkinabe gene pool was particularly remarkable. It, therefore, clearly represents an invaluable genetic resource that should be exploited to address the overarching goal of improving maize for adaptation to different environments, ecosystems, and stress situations. Overall, the genetic diversity revealed in this study has provided an invaluable resource for future analyses of candidate genes for local adaptations using robust association mapping experiments.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/9/1054/s1, Figure S1: Map of West Africa showing the countries of origin of the maize landraces analyzed in this study. Figure S2: Distribution of the 5974 DArTseq GBS makers across the 10 chromosomes of the 208 maize accessions. Table S1: Excel file with the description of the accessions analyzed in this study. Table S2: Excel file showing clustering of the maize 208 accessions based on 5974 DArTseq markers. Table S3: Excel file showing clustering of the four gene pools of maize accessions; Sheet 1. Burkinabe landraces; Sheet 2. Ghanaian landraces; Sheet 3. Togolese landraces; Sheet 4. Reference set. Table S4: List of accessions, and categorization into different clusters by DAPC.

Author Contributions: Conceptualization, C.N., B.B.-A., A.L.G.-O., M.G., A.T., and A.S.-P.N.; Supervision: B.B.-A., A.L.G.-O., and A.S.-P.N.; Formal analysis, C.N., A.L.G.-O., A.P., and A.T.; Writing—original draft preparation, C.N.; Writing—review and editing, B.B.-A., A.L.G.-O., M.G., A.T., A.P., C.N., and A.S.-P.N.; Funding acquisition, B.B.-A., M.G., and A.L.G.-O. All authors have read and approved the final version of the manuscript.

Funding: This research was supported by the Bill and Melinda Gates Foundation through the funding support to the Stress Tolerant Maize for Africa (STMA) Project (Grant # OPP1134248). Genotyping was done through the Integrated Genotyping Service and Support (IGSS) platform grant (ref. number PJ-002507). The first author was supported by a PhD fellowship from the German Federal Ministry of Education, through the West African Science Service Centre on Climate Change and Adapted Land-Use (WASCAL).

Acknowledgments: The first author is grateful to WASCAL for the PhD fellowship. We are grateful to the Genetic Resource Centre at IITA, Ibadan, Nigeria, and the Plant Genetics Resources Institute at Bunso, Ghana for providing the maize accessions used in this study. We also thank the IITA bioscience laboratory technicians in Ibadan, Nigeria, for sample collection and DNA extraction.

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

Availability of Data and Materials: The DArTseq datasets used in the present study have been deposited at the IITA repository. CKAN: http://data.iita.org/dataset/genotypic-data-for-216-maize-inbred-lines-for-diversity-studies.

References

1. Mir, C.; Zerjal, T.; Combes, V.; Dumas, F.; Madur, D.; Bedoya, C.; Dreisigacker, S.; Franco, J.; Grudloyma, P.; Hao, P.X.; et al. Out of America: Tracing the genetic footprints of the global diffusion of maize. *Theor. Appl. Genet.* **2013**, *126*, 671–682. [CrossRef] [PubMed]
2. Hufford, M.B.; Lubinksy, P.; Pyhäjärvi, T.; Devengenzo, M.T.; Ellstrand, N.C.; Ross-Ibarra, J. The genomic signature of crop-wild introgression in maize. *PLoS Genet.* **2013**, *9*, e1003477. [CrossRef]

3. Warburton, M.L.; Reif, J.C.; Frisch, M.; Bohn, M.; Bedoya, C.; Xia, X.C.; Crossa, J.; Franco, J.; Hoisington, D.; Pixley, K.; et al. Genetic diversity in CIMMYT non-temperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. *Crop Sci.* **2008**, *48*, 617–624. [CrossRef]
4. Yan, J.; Shah, T.; Warburton, M.L.; Buckler, E.S.; McMullen, M.D.; Crouch, J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* **2009**, *4*, e8451. [CrossRef] [PubMed]
5. Byrne, P.; Richards, C.; Volk, G.M. From wild species to landraces and cultivars. Crop Wild Relatives and their Use in Plant Breeding. In *VolkGM*; Byrne, P., Ed.; Colorado State University: Fort Collins, CO, USA, 2020.
6. Dwivedi, S.L.; Ceccarelli, S.; Blair, M.W.; Upadhyaya, H.D.; Are, A.K.; Ortiz, R. Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. *Trends Plant Sci.* **2016**, *21*, 31–42. [CrossRef]
7. Wu, X.; Wang, A.; Guo, X.; Liu, P.; Zhu, Y.; Li, X.; Chen, Z. Genetic characterization of maize germplasm derived from Suwan population and temperate resources. *Hereditas* **2019**, *156*, 2. [CrossRef]
8. Masuka, B.P.; van Biljon, A.; Cairns, J.E.; Das, B.; Labuschagne, M.; MacRobert, J.; Olsen, M. Genetic diversity among selected elite CIMMYT maize hybrids in East and Southern Africa. *Crop Sci.* **2017**, *57*, 2395–2404. [CrossRef]
9. Westengen, O.T.; Berg, P.R.; Kent, M.P.; Brysting, A.K. Spatial Structure and Climatic Adaptation in African Maize Revealed by Surveying SNP Diversity in Relation to Global Breeding and Landrace Panels. *PLoS ONE* **2012**, *7*, e47832. [CrossRef]
10. Global Crop Diversity Trust (GCDT). Global Strategy for the Ex Situ Conservation and Utilization of Maize Germplasm. 2007. Available online: http://www.croptrust.org/documents/web/MaizeStrategy-FINAL-18Sept07.pdf (accessed on 12 June 2012).
11. Wambugu, P.W.; Ndjiondjop, M.-N.; Henry, R.J. Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Brief. Funct. Genom.* **2018**, *17*, 198–206. [CrossRef]
12. Nelimor, C.; Badu-Apraku, B.; Nguetta, S.P.A.; Tetteh, A.Y.; Garcia-Oliveira, A.L. Phenotypic characterization of maize landraces from Sahel and Coastal West Africa reveals marked diversity and potential for genetic improvement. *J. Crop Improv.* **2019**, *34*, 122–138. [CrossRef]
13. Smith, J.S.C.; Smith, O.S. Fingerprinting crop varieties. *Adv. Agron.* **1992**, *47*, 85–140.
14. Holtz, Y.; Ardisson, M.; Ranwez, V.; Besnard, A.; Leroy, P.; Poux, G.; Roumet, P.; Viader, V.; Santoni, S.; David, J. Genotyping by sequencing using specific allelic capture to build a high density genetic map of durum wheat. *PLoS ONE* **2016**, *11*, e0154609. [CrossRef]
15. Baloch, F.S.; Alsaleh, A.; Shahid, M.Q.; Çiftçi, V.E.; Sáenz de Miera, L.; Aasim, M.; Nadeem, M.A.; Aktas, H.; Özkan, H.; Hatipoglu, R.A. Whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from Central Fertile Crescent. *PLoS ONE* **2017**, *12*, e0167821. [CrossRef] [PubMed]
16. Robbana, C.; Kehel, Z.; Naceur, B.; Sansaloni, C.; Bassi, F.; Amri, A. Genome-wide genetic diversity and population structure of Tunisian durum wheat landraces based on DArTseq technology. *Int. J. Mol. Sci.* **2019**, *20*, 1352. [CrossRef] [PubMed]
17. Ndjiondjop, M.N.; Semagn, K.; Gouda, A.C.; Kpeki, S.B.; Dro-Tia, D.; Sow, M.; Goungoulou, A.; Sie, M.; Perrier, X.; Ghesquiere, A. Genetic variation and population structure of Oryza glaberrima and development of a mini-core collection using DArTseq. *Front. Plant Sci.* **2017**, *8*, 1748. [CrossRef] [PubMed]
18. Yang, H.; Wei, C.L.; Liu, H.W.; Wu, J.L.; Li, Z.G.; Zhang, L.; Jian, J.-B.; Li, Y.-Y.; Tai, Y.-L.; Zhang, J.; et al. Genetic divergence between Camellia sinensis and its wild relatives revealed via genome-wide SNPs from RAD sequencing. *PLoS ONE* **2016**, *11*, e0151424. [CrossRef]
19. Nadeem, M.A.; Habyarimana, E.; Çiftçi, V.; Nawaz, M.A.; Karaköy, T.; Comertpay, G.; Shahid, M.Q.; Hatipoğlu, R.; Yeken, M.Z.; Ali, F. Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicoDArT marker information. *PLoS ONE* **2018**, *13*, e0205363. [CrossRef]
20. Adewale, S.A.; Badu-Apraku, B.; Akinwale, R.O.; Paterne, A.A.; Gedil, M.; Garcia-Oliveira, A.L. Genome-wide association study of Striga resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* **2020**, *20*, 1–16. [CrossRef]
21. Obeng-Bio, E.; Badu-Apraku, B.; Ifie, B.E.; Danquah, A.; Blay, E.T.; Dadzie, M.A.; Noudifoulè, G.T.; Talabi, A.O. Genetic diversity among early provitamin A quality protein maize inbred lines and the performance of derived hybrids under contrasting nitrogen environments. *BMC Genet.* **2020**, *21*, 1–13. [CrossRef]

22. Kilian, A.; Sanewski, G.; Ko, L. The application of DArTseq technology to pineapple. *Acta Hortic.* **2016**, *1111*, 181–188. [CrossRef]
23. 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*, 524–527. [CrossRef] [PubMed]
24. Bradbury, P.J.; Zhang, Z.; 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]
25. Liu, K.; Muse, S.V. Power Marker: Integrated analysis environment for genetic marker data. *Bioinformatics* **2005**, *21*, 2128–2129. [CrossRef]
26. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [PubMed]
27. Boakyewaa, G.A.; Badu-Apraku, B.; Akromah, R.; Garcia-Oliveira, A.L.; Awuku, F.J.; Gedil, M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE* **2019**, *14*, e0214810. [CrossRef] [PubMed]
28. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
29. Earl, D.A.; Vonholdt, B.M. Structure harvester: A website and program for visualizing Structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [CrossRef]
30. Salazar, E.; González, M.; Araya, C.; Mejía, N.; Carrasco, B. Genetic diversity and intra-racial structure of Chilean Choclero corn (*Zea mays* L.) germplasm revealed by simple sequence repeat markers (SSRs). *Sci. Hortic.* **2017**, *225*, 620–629. [CrossRef]
31. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 1–15. [CrossRef]
32. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef]
33. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3321–3323. [CrossRef] [PubMed]
34. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
35. Rambaut, A. FigTree v 1.4.4. 2018. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 25 April 2020).
36. Wright, S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **1965**, *19*, 395–420. [CrossRef]
37. Zhang, X.; Zhang, H.; Li, L.; Lan, H.; Ren, Z.; Liu, D.; Wu, L.; Liu, H.; Jaqueth, J.; Li, B.; et al. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC Genom.* **2016**, *17*, 697. [CrossRef] [PubMed]
38. Aci, M.M.; Lupini, A.; Mauceri, A.; Morsli, A.; Khelifi, L.; Sunseri, F. Genetic variation and structure of maize populations from Saoura and Gourara oasis in Algerian Sahara. *BMC Genet.* **2018**, *19*, 51. [CrossRef] [PubMed]
39. Belalia, N.; Lupini, A.; Djemel, A.; Morsli, A.; Mauceri, A.; Lotti, C.; Sunseri, F. Analysis of genetic diversity and population structure in Saharan maize (*Zea mays* L.) populations using phenotypic traits and SSR markers. *Genet. Resour. Crop Evol.* **2019**, *66*, 243–257. [CrossRef]
40. Mengistu, D.K.; Afeworki, Y.K.; Mario, E.P. Phenotypic diversity in Ethiopian durum wheat (Triticum turgidum var. durum) landraces. *Crop J.* **2015**, *3*, 190–199. [CrossRef]
41. Frankham, R.; Ballou, J.D.; Briscoe, D.A. *Introduction to Conservation Genetics*; Cambridge University Press: Cambridge, UK, 2002. [CrossRef]
42. N'da, D.P.; Louise, A.; Paul, A.K.; Konan, K.C.; Arsène, Z.B.I. Genetic diversity and population structure of maize landraces from Côte d'Ivoire. *Afr. J. Biotech.* **2016**, *15*, 2507–2516. [CrossRef]
43. Marchand, J.L. *Note Sur la Creation D'un Composite de Mais Africain*; CIRAD-IRAT: Paris, France, 1977.
44. Sauvaire, D. *L'opération Régionale Coordonnée 1975–1987*; CIRAD-IRAT: Paris, France, 1987.
45. Sanou, J.; Gouesnard, B.; Charrier, A. Isozyme variability in West African maize cultivars (*Zea mays* L.). *Maydica* **1997**, *42*, 3–12.

46. Efron, Y.; Kim, S.K.; Fajemisin, J.M.; Mareck, J.H.; Tang, C.Y.; Dąbrowski, Z.T.; Buddenhagen, I.W. Breeding for resistance to maize streak virus: A multidisciplinary team approach. *Plant Breed.* **1989**, *103*, 1–36. [CrossRef]
47. Nelimor, C.; Badu-Apraku, B.; Tetteh, A.Y.; N'guetta, A.S.P. Assessment of Genetic Diversity for Drought, Heat and Combined Drought and Heat Stress Tolerance in Early Maturing Maize Landraces. *Plants* **2019**, *8*, 518. [CrossRef] [PubMed]
48. Nelimor, C.; Badu-Apraku, B.; Tetteh, A.Y.; Garcia-Oliveira, A.L.; N'guetta, A.S.P. Assessing the potential of extra-early maturing landraces for improving tolerance to drought, heat, and both combined stresses in maize. *Agronomy* **2020**, *10*, 318. [CrossRef]

genes

MDPI

Article

Characterization of the Common *Japonica*-Originated Genomic Regions in the High-Yielding Varieties Developed from Inter-Subspecific Crosses in Temperate Rice (*Oryza sativa* L.)

Jeonghwan Seo [1], So-Myeong Lee [1,2], Jae-Hyuk Han [3], Na-Hyun Shin [3], Yoon Kyung Lee [1], Backki Kim [1], Joong Hyoun Chin [3,*] and Hee-Jong Koh [1,*]

[1] Department of Plant Science and Research Institute for Agriculture and Life Sciences, and Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Korea; rightseo@hotmail.com (J.S.); olivetti90@korea.kr (S.-M.L.); nknglee2403@snu.ac.kr (Y.K.L.); uptfamily@hanmail.net (B.K.)
[2] Department of Southern Area Crop Science, National Institute of Crop Science, RDA, Miryang 50424, Korea
[3] Department of Integrative Biological Sciences and Industry, Sejong University, 209, Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea; 0724jh@gmail.com (J.-H.H.); skgus1125@gmail.com (N.-H.S.)
* Correspondence: jhchin@sejong.ac.kr (J.H.C.); heejkoh@snu.ac.kr (H.-J.K.)

Received: 23 April 2020; Accepted: 14 May 2020; Published: 18 May 2020

Abstract: The inter-subspecific crossing between *indica* and *japonica* subspecies in rice have been utilized to improve the yield potential of temperate rice. In this study, a comparative study of the genomic regions in the eight high-yielding varieties (HYVs) was conducted with those of the four non-HYVs. The Next-Generation Sequencing (NGS) mapping on the Nipponbare reference genome identified a total of 14 common genomic regions of *japonica*-originated alleles. Interestingly, the HYVs shared *japonica*-originated genomic regions on nine chromosomes, although they were developed through different breeding programs. A panel of 94 varieties was classified into four varietal groups with 38 single nucleotide polymorphism (SNP) markers from 38 genes residing in the *japonica*-originated genomic regions and 16 additional trait-specific SNPs. As expected, the *japonica*-originated genomic regions were only present in the *japonica* (JAP) and HYV groups, except for Chr4-1 and Chr4-2. The *Wx* gene, located within Chr6-1, was present in the HYV and JAP variety groups, while the yield-related genes were conserved as *indica* alleles in HYVs. The *japonica*-originated genomic regions and alleles shared by HYVs can be employed in molecular breeding programs to further develop the HYVs in temperate rice.

Keywords: rice; yield; HYV; Tongil; *indica*; *japonica*; SNP; molecular breeding

1. Introduction

There are two subspecies in cultivated rice (*Oryza sativa*), *indica* and *japonica*. *Indica* rice is known to be adaptable to tropical regions, while *japonica* rice is grown in temperate regions. Therefore, *indica* and *japonica* have different characteristics [1]. In general, *japonica* varieties are known to have relatively low yield potentials, as compared to *indica* varieties. To improve the yield potential of *japonica* rice, inter-subspecific crosses between *indica* and *japonica* have been conducted by conventional rice breeders [2]. As a result of these efforts, several high-yielding varieties (HYVs) have been developed from *indica-japonica* crosses. One of the greatest historical successes of *indica–japonica* crosses was the development of Tongil, an HYV, in Korea. Tongil showed a 30% higher yield than those of the conventional *japonica* varieties. By growing Tongil rice, self-sufficiency in staple food in Korea was possible in 1977 [3]. There were important trade-offs in growing Tongil, such as cold intolerance, pathogen susceptibility, and low eating quality, which were inherited from *indica* parents. A series

of 'Tongil-type' HYVs have been developed from *indica–japonica* crosses to overcome the vulnerable points of Tongil from the late 1970s [2].

HYVs have been developed in Japan using *indica* and *japonica* varieties since the 1980s. Takanari is a Japanese semi-dwarf HYV, developed from the crosses between Milyang 42 and Milyang 25. Takanari shared the ancestry of Tongil [4]. To date, it recorded the highest grain yields for both yield trials (>10 t/ha as brown rice) and individual trials (11.7 t/ha as brown rice) in Japan [5]. Minghui 63, which was derived from the cross between IR 30 and Gui 630, is the male parent of the elite hybrid rice Shanyu 63 from China. IR 30 is a semi-dwarf variety developed in IRRI (International Rice Research Institute) and is a restorer line for WA-CMS A-lines, which have a good plant type, a high resistance to blast, a bacterial blight, and brown planthoppers. Gui 630 is a rice germplasm from Guyana that has a high grain weight, desirable grain quality, and high yield potential [6]. Gui 630 is known as an *indica* restorer variety [7]. Minghui 63 was classified into the *indica II* subpopulation, together with Milyang 23 and some other *indica-japonica* HYVs, by genome sequence analysis [8].

Nipponbare is the *japonica* reference genome of rice and was first sequenced at the high-quality whole-genome level through all the crops [9]. In addition, the whole genome sequences of *indica* rice varieties were reported [10–14]. The genomic difference between *indica* and *japonica* at the sequence level has been extensively studied [15]. At least 384,431 single nucleotide polymorphisms (SNPs) and 24,557 insertion/deletion mutations (InDels) were reported between Nipponbare and 93-11 [16]. With the advent of Next-Generation Sequencing (NGS) technology, numerous genomes of diverse rice germplasm collections have become available. For instance, 3000 rice genomes were sequenced and deployed in genetic and genomic rice studies [17–20]. Recently, more than one hundred high-yielding loci, associated with green revolution phenotypes and derived from the two ancestral *indica* varieties, were identified with the help of pedigree analysis, whole-genome sequencing, and genome editing [21]. Furthermore, most of the quantitative trait loci (QTLs) and genes for high-yielding potential in HYVs originated from *indica* parents in previous studies using HYVs derived from *indica–japonica* crosses [3–5,22,23]. No report exists on the characterization of *japonica* genomic regions in HYVs derived from *indica–japonica* crosses yet.

Previously, we sequenced the whole genomes of Tongil and its parental varieties to analyze the genome composition and genetic factors of Tongil. As a result, the Tongil genome was found to be derived mostly from the *indica* genome, with a small portion of *japonica* genome introgressions [3]. This study was carried out to comparatively analyze the genome structure of eight HYVs and to identify the *japonica*-originated genomic regions that are shared in HYVs, which will be helpful in understanding the role of *japonica* genome in Tongil and other HYVs that are developed in temperate regions for the further development of promising HYVs.

2. Materials and Methods

2.1. Plant DNA Materials

The eight HYVs, including Cheongcheongbyeo, Dasanbyeo, Hanareumbyeo, Milyang 23, Minghui 63, Nampungbyeo, Takanari, and Tongil, and four non-HYVs, including Nipponbare, Yukara, IR 8, and TN1 were used for the whole-genome sequence analysis. Cheongcheongbyeo, Dasanbyeo, Hanareumbyeo, Milyang 23, and Nampungbyeo are Tongil-type HYVs that were developed in Korea. Takanari and Minghui 63 are HYVs from Japan and China, respectively. The pedigree of each of the eight HYVs can be found in Figure S1. A total of 94 rice varieties were used for SNP marker validation (Table S1).

2.2. Whole Genome Sequencing and DNA Variation

Tongil and its three parental varieties (Yukara, IR 8, and TN1) were sequenced in a previous study [3]. The other eight varieties were sequenced in this study using the Illumina Hiseq 1000 and NextSeq 500 platform (Illumina, San Diego, CA, USA) (Table 1). Whole genome sequencing, including the construction of shotgun DNA libraries, was performed according to the methods recommended by

the manufacturer. The Illumina whole-genome shotgun paired-end DNA sequencing data were filtered to obtain high-quality sequence data. Raw sequence reads were subjected to quality trimming using FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/), and the reads with a Phred quality (Q) score <20 were discarded. Adapter trimming was conducted by using Trimmomatic version 0.36 (http://www.usadellab.org/cms/?page=trimmomatic).

The clean reads were mapped on the *japonica* reference Nipponbare genome (Os-Nipponbare-Reference-IRGSP-1.0 [24]) using the Burrows–Wheeler Aligner (BWA) program version 0.7.15 [25]. The alignment results were merged and converted into binary alignment map (BAM) files [26]. The BAM files were used to calculate the sequencing depth and to identify SNPs using the GATK program version 3.4 with default parameters [27].

All the raw sequence data obtained in this study are available in the NCBI Short Read Archive (SRA) database under the following BioProject accession numbers: Nipponbare [PRJNA264254], Milyang 23 [PRJNA264250], Dasan (or Dasanbyeo) [PRJNA222717], Cheongcheong (or Cheongcheongbyeo) [PRJNA616202], Nampung (or Nampungbyeo) [PRJNA616219], Hanareum (or Hanareumbyeo) [PRJNA616209], Takanari [PRJNA616222], and Minghui 63 [PRJNA616216]. The raw sequence data of Tongil, Yukara, IR 8, and TN1 are available in a previous study [3].

2.3. SNP Allele Calling

Genotype calling was performed in order to identify SNPs that originated from the *indica* and *japonica* genomes. There were two types of values calculated in this study: variety value and reference value. The variety value calculated whether it was of *japonica*-type parental allele (Yukara allele) of Tongil or a Tongil-like variety or not. The variety value of SNP was calculated as the sum of the following values: '1' (IR 8 allele); '2' (TN1 allele); '4' (Yukara allele); and '0' (all others). If the SNP is the same as the Nipponbare reference, the value '1' was given to the SNP; otherwise, the value '0' was given. The SNPs showing variety value '4' and reference value '1' were called *japonica*-type SNPs. Then, the total number of *japonica*-type SNPs in each 100 kb block, which is the approximate chromosomal distance for the linkage disequilibrium (LD) decay rate in rice [28], in each chromosome was counted to identify the introgressed regions of *japonica*.

2.4. SNP Marker Development for the Fluidigm Platform

A total of 39 representative genes were selected from the selected regions. The representative genes were well-annotated in the public gene/QTL databases, as of 6 January 2017 (RAP and UniProt). Then, only the genes containing non-synonymous SNPs in the predicted exon and UTR regions between HYVs and non-HYVs were selected. We assumed that these genes might be a part of the candidate genes that are associated with *japonica*-originated traits. Out of the many SNPs in the genes, only one SNP with a substitution polymorphism between *indica* (IR 8 and TN1) and *japonica* (Nipponbare and Yukara) per representative gene was selected for the genomic validation. The SNP markers for the Fluidigm platform (Fluidigm Corporation, San Francisco, CA, USA) were designed using the method by Seo et al. [29]. To design Fluidigm SNP genotyping assays, 60–150 bp sequences, flanking the selected SNPs on either side, were aligned by BLAST. Finally, the selected SNPs and flanking sequences were uploaded on the D3 Assay Design website [30]. After confirming the results, the designed assays were ordered. One Fluidigm SNP assay contained the Allele-Specific Primer 1 (ASP1), ASP2, Locus-Specific Primer (LSP), and Specific Target Amplification (STA) primer.

2.5. DNA Extraction and Fluidigm Genotyping

Young leaves of each plant from all materials used in this study were collected for DNA extraction at the tillering stage. Genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method, as described by Murray and Thompson [31]. The concentration and purity of DNA samples were measured with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies,

Wilmington, DE, USA). DNA samples, showing absorbance ratios above 1.8 at 260/280 nm, were diluted to 50 ng/μL and used for genotyping.

Genotyping was performed using the BioMark™ HD system (Fluidigm) and 96.96 Dynamic Array IFCs (Fluidigm), according to the manufacturer's protocol in National Instrumentation Center for Environmental Management (NICEM), Seoul National University (Pyeongchang, Korea). Specific Target Amplification (STA) was performed prior to SNP genotyping analysis. PCR was performed in a 5 μL reaction containing 50 ng of the DNA sample, according to the manufacturer's protocol. For genotyping, SNPtype assays were performed using STA products, according to the manufacturer's protocol. The genotyping result was acquired using Fluidigm SNP Genotyping Analysis software. All the genotype-calling results were manually checked and any obvious errors in the homozygous or heterozygous clusters were curated.

2.6. Data Analysis and QTL Comparison

We analyzed basic marker statistics, such as major allele frequency (MAF), heterozygosity, and polymorphism information content (PIC) of SNP markers using PowerMarker V3.25 [32]. PowerMarker V3.25 was used to calculate the genetic distance, based on CS Chord [33] and the constructed un-weighted pair group methods with the arithmetic mean algorithm (UPGMA) dendrogram, which was visualized in Molecular Evolutionary Genetics Analysis version 7.0; MEGA7 [34].

We extracted QTL information from the Q-TARO database at 25th August 2017. The physical position in Q-TARO is based on IRGSP build 4 of the Nipponbare genome, while IRGSP 1.0 of the Nipponbare genome was used in this study. Thus, the physical position of the start and end of each QTLs in Q-TARO were converted into the physical positions of IRGSP 1.0. Then, the QTLs overlapped on common *japonica*-originated regions of eight HYVs were selected. After checking if all the selected QTLs were unique or redundant, the redundant QTLs were discarded and only the unique QTLs were left remained. After the filtering step for the QTLs, they were classified into seven categories. The category classification was conducted by checking the characters and the trait names manually.

3. Results

3.1. Whole Genome Sequencing and SNP Calling

To analyze the genomic composition of the HYVs derived from *indica–japonica* crosses, the whole genomes of HYVs and four varieties were sequenced on the Illumina platform. A large number of short reads were mapped onto the reference Nipponbare genome and were then assembled into a consensus sequence. The number of sequence yields and the number of reads and mapping depths were varied. For example, a total of 66,464,246 reads of the Cheongcheongbyeo genome, corresponding to 9,991,040,272 bp (10 Gb), were generated, representing a 21-fold mapping depth. Nipponbare showed the largest number of sequence yields and number of reads and mapping depths (Table 1).

More than one million SNPs were detected in each of the eight HYVs and two *indica* varieties against the Nipponbare (*japonica*) genome. More than 90% of the SNPs were detected in the intergenic region. The smallest number of SNPs were in the 5′ untranslated region (UTR). Two *indica* varieties, IR 8 and TN1, represented a relatively large number of SNPs, as compared to seven HYVs except for Minghui 63. Among the seven HYVs, Milyang 23 showed the smallest number of SNPs (Table 2).

Table 1. Basic sequencing statistics of the varieties used in this study.

Type	Variety Name	Yield (bp)	Read	N (%)	GC (%)	Q30 (%)	Depth (X)	Sequencing Platform
HYV	Cheongcheongbyeo	9,991,040,272	66,464,246	0.07	43.83	83.37	21.76	Illumina NextSeq 500
HYV	Dasanbyeo	5,677,243,407	93,838,734	1.05	39.75	69.57	10.32	Illumina HiSeq 1000
HYV	Hanareumbyeo	8,701,463,207	57,884,660	0.07	43.74	81.74	18.58	Illumina NextSeq 500
HYV	Milyang 23	8,909,120,495	147,258,190	0.31	41.08	76.44	17.79	Illumina HiSeq 1000
HYV	Minghui 63	10,346,919,646	68,794,138	0.07	43.93	83.18	22.48	Illumina NextSeq 500
HYV	Nampungbyeo	10,366,586,498	68,923,446	0.07	43.88	83.52	22.62	Illumina NextSeq 500
HYV	Takanari	9,372,371,998	62,353,794	0.08	43.38	83.10	20.35	Illumina NextSeq 500
HYV	Tongil	13,362,670,165	264,607,330	0.16	42.57	70.41	24.58	Illumina HiSeq 1000
Japonica	Nipponbare	22,212,867,380	439,858,760	0.19	42.10	89.12	51.72	Illumina HiSeq 1000
Japonica	Yukara	9,155,887,048	151,336,976	0.28	41.13	77.37	18.51	Illumina HiSeq 1000
Indica	IR 8	8,287,794,812	136,988,344	0.40	41.14	78.94	17.09	Illumina HiSeq 1000
Indica	TN 1	8,337,247,875	137,805,750	0.31	41.51	77.25	16.83	Illumina HiSeq 1000

Abbreviations are as follow: HYV—high-yielding variety, N (%)—percentage of skipped base, GC (%)—percentage of GC content, Q30 (%)—percentage of bases showing Phred quality score (Q) ≥ 30.

Table 2. Number and location of the SNPs in the varieties against Nipponbare pseudomolecule.

Varieties	Non-Synonymous	Synonymous	Intron	5′ UTR	3′ UTR	Intergenic	Total
Cheongcheongbyeo	19,193	17,439	21,671	7076	29,403	1,019,466	1,114,248
Dasanbyeo	18,534	16,944	21,247	7006	28,598	975,433	1,067,762
Hanareumbyeo	18,059	16,492	20,657	6825	27,766	937,990	1,027,789
Milyang 23	17,836	16,414	20,110	6628	27,271	933,778	1,022,037
Minghui 63	20,117	18,601	22,351	7350	30,731	1,042,153	1,141,303
Nampungbyeo	19,592	17,960	21,588	7195	29,286	1,005,553	1,101,174
Takanari	18,631	16,936	20,881	6887	28,157	966,035	1,057,527
Tongil	18,427	16,837	20,816	6750	27,021	952,421	1,042,272
IR 8	20,578	18,834	22,847	7494	31,152	1,061,383	1,162,288
TN1	20,171	18,179	22,095	7614	30,286	1,041,363	1,139,708

3.2. Evaluation of Japonica-Type SNP Value

To identify the SNPs that originated from the *indica* and *japonica* genomes, the SNP value was evaluated. The SNPs showing a variety value of 4 and a reference value of 1 were *japonica*-originated SNPs. The total number of *japonica*-originated SNPs in each 100 kb block of each chromosome were counted in order to identify introgressed regions from *japonica*. Previously, we discriminated the Tongil genome into segments that originated from *indica* and *japonica* using the sliding window method [3]. In this study, we allowed for two exception from the HYV in order to include Takanari, Cheongcheong, Nampung and Minghui 63. Accordingly, a total of 14 *japonica*-originated genomic regions, which were shared by at least six HYVs on *japonica*-originated segment of Tongil, were detected. The common *japonica*-originated genomic regions were distributed on nine chromosomes, not including chromosomes 8, 10, and 12. There were three *japonica*-type regions on chromosome 2 and two regions on chromosomes 1, 4, and 7. Furthermore, the regions were clustered or closely located on each chromosome. The size of the regions was varied from 0.1 Mb for Chr7-1 and Chr11-1 to 2 Mb for Chr1-2. Out of the 14 regions, seven were common in eight HYVs (Figure 1, Table 3).

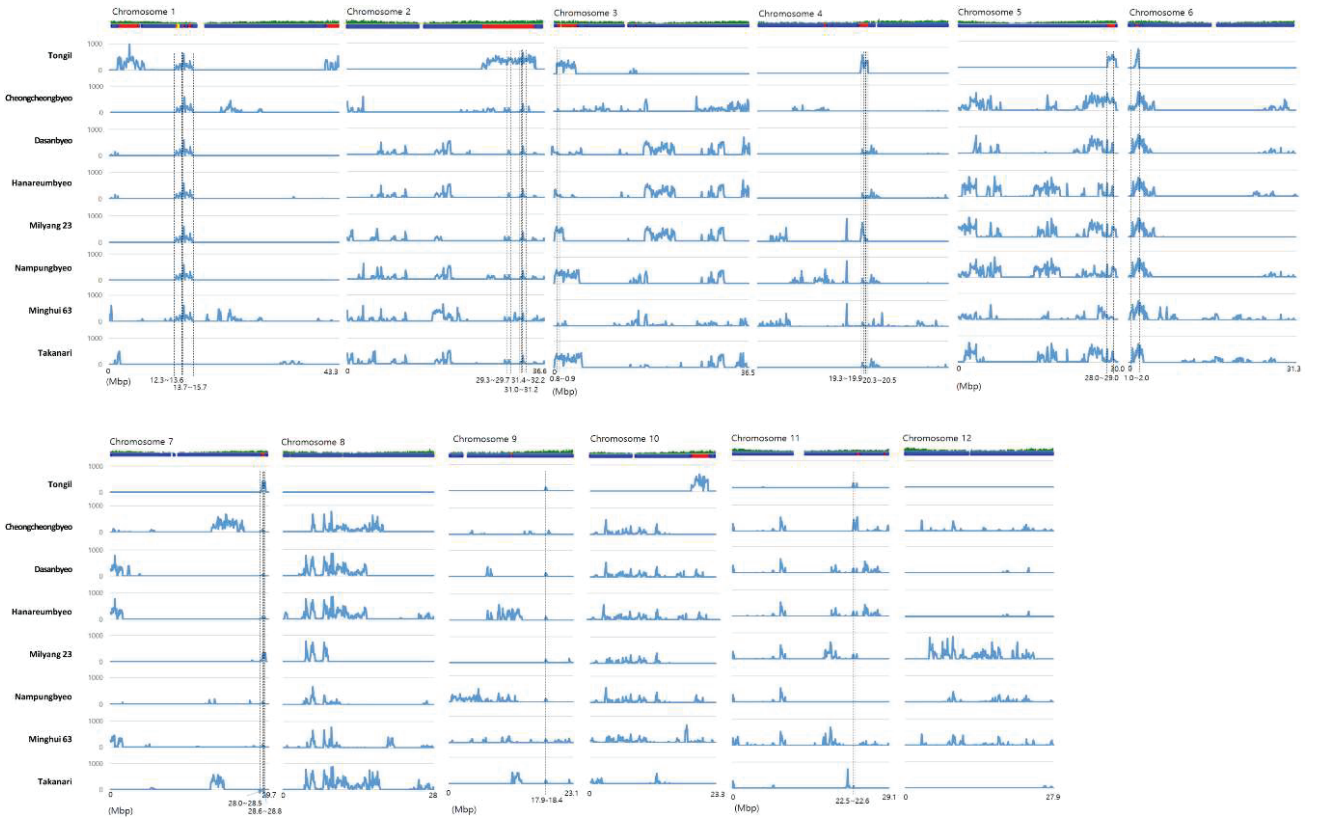


Figure 1. Location of *japonica*-type SNPs on 12 chromosomes of HYVs and their co-location with the *japonica* block of the Tongil genome, shown as a bar above the graphs. The blue and red block of the Tongil genome represents the *indica* and *japonica* blocks, respectively. The blue peak on each graph indicates the number of *japonica*-type SNPs. The position and range for the co-location of *japonica*-type blocks are denoted by vertical black dotted lines.

Table 3. *Japonica*-type SNP frequency (%) of Tongil and the other HYVs at common *japonica*-type regions among the eight HYVs.

Region	Narrowed Range (Mb)	Size (Mb)	Tongil	Cheongcheongbyeo	Dasanbyeo	Hanareumbyeo	Milyang 23	Nampungbyeo	Minghui 63	Takanari	Type
Chr1-1	12.3 ~ 13.6	1.3	23.3	23.3	23.3	23.3	23.3	23.3	23.3	0.0	All (ex. Takanari)
Chr1-2	13.7 ~ 15.7	2	31.4	31.4	31.4	31.4	31.4	31.5	31.5	0.0	All (ex. Takanari)
Chr2-1	29.3 ~ 29.7	0.4	64.1	14.1	14.1	14.1	14.1	14.1	14.4	14.1	All
Chr2-2	31.0 ~ 31.2	0.2	46.5	1.1	1.1	1.1	1.1	1.1	2.1	1.1	All
Chr2-3	31.4 ~ 32.2	0.8	77.4	17.3	17.6	17.6	17.6	17.6	21.6	17.6	All
Chr3-1	0.8 ~ 0.9	0.1	73.9	19.3	16.6	16.6	73.9	73.9	19.2	73.9	All
Chr4-1	19.3 ~ 19.9	0.6	75.9	0.0	8.2	8.2	77.3	8.2	7.0	8.2	All (ex. Cheongcheong)
Chr4-2	20.3 ~ 20.5	0.2	87.9	0.0	12.8	12.8	12.8	5.1	0.3	5.1	All (ex. Cheongcheong)
Chr5-1	28.0 ~ 29.0	1	76.9	82.1	23.9	23.9	23.9	23.9	13.7	23.9	All
Chr6-1	1.0 2.0	1	55.2	51.7	70.1	70.1	70.1	70.1	70.1	70.1	All
Chr7-1	28.5 ~ 28.6	0.1	32.0	6.5	5.0	6.5	4.8	6.5	6.5	0.0	All (ex. Takanari)
Chr7-2	28.6 ~ 28.8	0.2	93.5	9.7	7.7	7.6	33.2	9.7	7.6	0.0	All (ex. Takanari)
Chr9-1	17.9 ~ 18.4	0.5	13.9	0.7	13.9	13.9	13.9	13.9	13.9	13.9	All
Chr11-1	22.5 ~ 22.6	0.1	25.3	65.6	25.3	25.3	25.3	0.0	0.0	12.9	All (ex. Nampung/Minghui63)

Abbreviation is as follows: ex.—except.

3.3. QTL Comparison and Representative Gene Selection in Common japonica-Originated Genomic Regions

To elucidate the function of common *japonica*-type regions in HYVs, we first investigated the reported QTLs in the Q-TARO database [35]. A total of 101 selected QTLs for seven categories were co-located, with 14 common *japonica*-type regions on nine chromosomes. Category classification was carried out by checking the character and trait names of each QTL manually. For instance, the yield-related trait category contained various characters which could affect yield potential, such as source activity- and sink-related morphological traits, and sterility. Only three regions on chromosome 2 were co-located with QTLs for all the seven trait categories. For eating quality, abiotic stress, and the yield-related category, 80 QTLs were identified. The largest number (10) of co-located QTLs was detected in the Chr6-1 region for eating quality. All the regions were co-located with the QTLs for abiotic stress tolerance (Table 4). This information of co-located QTLs, with common *japonica*-type regions, suggests that common genomic regions in HYVs might be mainly associated with quality, yield, and abiotic stress tolerance.

Table 4. Classification of the reported QTLs co-location with common *japonica*-type regions in eight HYVs.

Region	Eating Quality	Abiotic Tolerance	Biotic Resistance	Yield-Related	Root	Flowering	Other	Total
Chr1-1/1-2	0	4	0	2	0	0	1	7
Chr2-1/2-2/2-3	8	6	3	5	1	1	1	25
Chr3-1	1	8	0	0	2	1	0	12
Chr4-1/4-2	0	2	0	4	1	0	1	8
Chr5-1	3	2	0	6	2	0	0	13
Chr6-1	10	1	1	0	1	2	0	15
Chr7-1/7-2	0	1	0	3	0	1	0	5
Chr9-1	0	4	2	7	0	0	0	13
Chr11-1	1	1	0	1	0	0	0	3
Total	23	29	6	28	7	5	3	101

Furthermore, we selected 39 genes containing non-synonymous SNPs, which could affect the molecular function of genes, and are clearly annotated in the databases of 12 common *japonica* chromosomal introgressions. There was no target gene that satisfied the above-mentioned condition in Chr4-2 and Chr7-1. The largest number (13) of selected genes was located on Chr1-2, which is the largest region, spanning 2 Mb. Only one gene was selected from Chr2-2, Chr7-2, and Chr11-1. The size of these three blocks was 0.1–0.2 Mb. The genes annotated from the major criteria of interest were *Os01g0348900*, *Os06g0130000*, *Os06g0130100* (stress tolerance), *Os06g0130400* (eating quality), and *Os01g0367100* (yield potential) (Table 5).

3.4. SNP Marker Development and Genotyping Using Fluidigm Platform

A total of 39 SNP markers were designed in the 39 selected genes from common *japonica*-originated genomic regions, by one marker per one gene. The SNPs for the marker were selected from among the non-synonymous SNPs. Five SNP markers, out of the 39 SNP markers, were designed in 3′ or 5′ UTR. In addition, 14 agronomic traits, related to SNP markers in *indica-japonica* SNP set 2 [29] and four previously developed yield related SNP markers [36], were also used for the genotyping of 94 diverse germplasms. A total of 57 SNP markers were genotyped for 94 germplasms using the Fluidigm system, and consequently, 54 SNPs showed polymorphism and a clear genotype, except one monomorphic SNP marker designed in *Os01g0348900* on the Chr1-2 block and two SNP markers which showed low base call quality, SaF-CT and SLG7-GC, in *indica-japonica* SNP set 2. Therefore, we conducted a further analysis using a total of 54 polymorphic SNP markers (Figure 2, Table S2).

Table 5. Selected 39 genes in the common *japonica*-type regions.

Region	Gene (RAP DB)	Alternative Name	Known Function	
			UniProt	RAP DB
Chr1-1	*Os01g0328500*			Bucentaur or craniofacial development family protein
	Os01g0329800	*IAI1*		YT521-B-like protein family protein
	Os01g0337100	*OsTPS1*		Similar to Sesquiterpene synthase
Chr1-2	*Os01g0347000*	*OsPHS1b, OsPP4, OsSTA14*		Similar to PROPYZAMIDE-HTPERSENSITIVE 1
	Os01g0348900	*salT(sal1), SalT1, sal1, Sal1, SALT, ML, SalT, OsSalT*	Salt stress-induced protein	SalT gene product
	Os01g0351200	*PARP2-A*	Poly [ADP-ribose] polymerase 2-A	Similar to Poly
	Os01g0356800	*OsEnS-6*		Domain of unknown function DUF3406, chloroplast translocase domain containing protein
	Os01g0363900	*OsWAK5*		Similar to HASTY
	Os01g0364400	*OsRLCK35*		Protein kinase, catalytic domain domain containing protein
	Os01g0367100	*PHD1*		Chloroplast-localized UDP-glucose epimerase (UGE), Galactolipid biosynthesis for chloroplast membranes, Photosynthetic capability and carbon assimilate homeostasis (Os01t0367100-01); NAD(P)-binding domain containing protein
	Os01g0369200	*CUL1*	Cullin-like protein	Similar to Cullin-1
	Os01g0370000	*OsOPR9, OsOPR01-2*	Putative 12-oxophytodienoate reductase 9	NADH:flavin oxidoreductase/NADH oxidase, N-terminal domain containing protein
	Os01g0371200	*OsGSTF1, RGSI*	Probable glutathione S-transferase GSTF1	Similar to Glutathione-S-transferase 19E50
	Os01g0371400	*OsGSTF9*		Similar to Glutathione s-transferase gstf2
	Os01g0371500	*OsGSTF10*		Similar to Glutathione-S-transferase 19E50
	Os01g0375100	*OsDjC6*	DNAJ heat shock N-terminal domain-containing protein-like	Similar to DnAJ-like protein slr0093
Chr2-1	*Os02g0713500*	*OsFbox108, Os_F0236*		F-box domain, cyclin-like domain containing protein.
	Os02g0713900	*HMGR I, Hmg1, HMGR1, HMGR 1, OsHMGR1*	3-hydroxy-3-methylglutaryl-coenzyme A reductase 1	Similar to 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1
Chr2-2	*Os02g0743700*			Similar to RING-H2 finger protein ATL1Q
Chr2-3	*Os02g0755200*	*OsHDMA702, HDMA702*	Lysine-specific histone demethylase 1 homolog 1	Similar to amine oxidase family protein
	Os02g0761100	*OsCYP40b, OsCYP-8, OsCYP40-2*		Similar to Cyclophilin-40 (Expressed protein)
Chr3-1	*Os03g0114900*		Mitochondrial import inner membrane translocase subunit Tim17 family protein, expressed	Similar to putaive mitochondrial inner membrane protein
	Os03g0115500			Similar to pyridoxine 5′-phosphate oxidase-related
Chr4-1	*Os04g0395900*			Polynucleotide adenylyltransferase region domain containing protein
	Os04g0400800			Heavy metal transport/detoxification protein domain containing protein
Chr5-1	*Os05g0563400*	*OsARF15, ETT1, OsETT1, ARF3b*	Auxin response factor 15	Similar to Auxin response factor 5
	Os05g0571700	*OsFbox282, Os_F0643*		Cyclin-like F-box domain containing protein
Chr6-1	*Os06g0125132*	*SDH8B*	Succinate dehydrogenase subunit 8B, mitochondrial	Conserved hypothetical protein
	Os06g0128300	*OsABCB23, OsISC32*		Similar to Mitochondrial half-ABC transporter, Similar to STA1 (STARIK 1); ATPase, coupled to transmembrane movement of substances
	Os06g0129900	*Cytochrome P450*	Cytochrome P450	Similar to Cytochrome P450 CYPD
	Os06g0130000	*LMR*		AAA-type ATPase, Defense response, Similar to Tobacco mosaic virus helicase domain-binding protein (Fragment)
	Os06g0130100	*OsSIK1, OsER2, ER2*		Receptor-like kinase (RLK), Drought and salt stress tolerance, Oryza sativa stress-induced protein kinase gene 1
	Os06g0130400	*OsACS6*		ACC synthase, Protein homologous to aminotransferase, Ethylene biosynthesis, Control of starch grain size in rice endosperm
	Os06g0131100	*OsWD40-124*		Similar to guanine nucleotide-binding protein beta subunit-like protein 1, WD40/YVTN repeat-like domain containing protein
	Os06g0136800	*OsClp9, CLP9*	ATP-dependent Clp protease proteolytic subunit	Peptidase S14, ClpP family protein
Chr7-2	*Os07g0677500*	*POX3006, prx114*	Peroxidase	Similar to Peroxidase precursor (EC 1.11.1.7)
Chr9-1	*Os09g0471600*	*OsWAK84*		EGF-like calcium-binding domain containing protein
	Os09g0471800	*OsWAK85, YK10*		Similar to WAK80 - OsWAK receptor-like protein kinase
Chr11-1	*Os11g0592500*		NB-ARC domain containing protein, expressed	Similar to NB-ARC domain containing protein, expressed

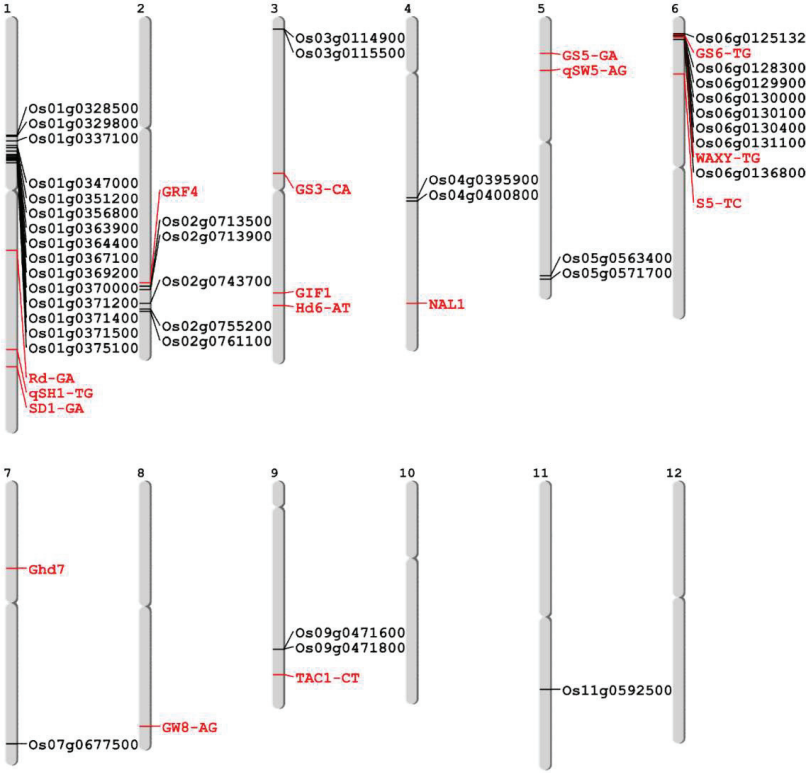


Figure 2. Genomic location of 54 polymorphic SNP markers used in this study. The markers represented by black and red indicate those newly developed on common *japonica* regions and those previously developed for agronomic trait related genes, respectively.

The results of genotyping showed a dividing pattern for 94 varieties. A phylogenetic analysis of 94 varieties was carried out using 54 polymorphic SNP markers. There were four groups, including IND1, IND2, HYV, and JAP, in the phylogenetic tree (Figure 3). All the sequenced HYVs, except Takanari, which possesses *indica* allele of Chr1-1 and Chr1-2, were clustered in the HYV-group with seven Tongil-type and five *indica* varieties. These varieties were developed by inter-subspecific crosses or by inter-varietal crosses including the parents of Tongil-type varieties. For example, Keunseombyeo was derived from a cross between Dasanbyeo and Namyeongbyeo [37]. Taebaekbyeo was used for the development of Hanareumbyeo and Dasanbyeo (Figure S1, F–G). On the other hand, among the five *indica* varieties, four *indica* (IR24, IRBB23, IRBB61, and IRBB66) were developed by the International Rice Research Institute (IRRI). IRBB23, IRBB61, and IRBB66 are near-isogenic line series of IR24 for the improvement of the resistance against bacterial leaf blight [38]. IR24 is a good eating quality and high-yielding variety developed from the inter-subspecific crosses using one tropical *japonica* (CP-SLO) and two *indica* (SIGADIS and IR8) [39]. Furthermore, IR24 was included in the breeding program of six HYVs, which were used for resequencing in this study. (Figure S1, C–H). This implies that the specific genomic regions were conserved in the HYV group by the selections during the conventional breeding programs for HYV development.

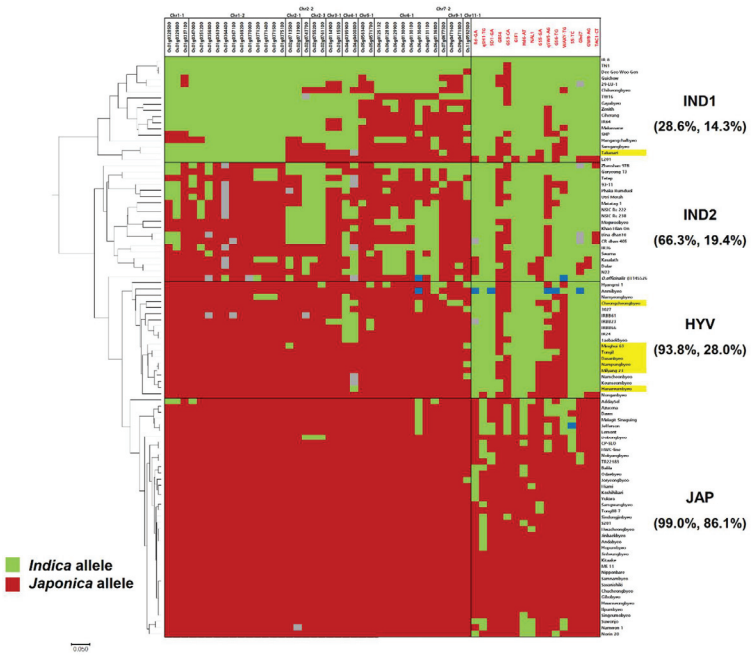


Figure 3. Phylogenetic tree of 94 germplasms based on 54 SNP markers and a genotype heat map. A 38 SNP marker set (left) and 16 SNP marker set (right) were developed on common *japonica* regions and agronomic trait related genes, respectively. The color of the marker ID is the same as in Figure 2. The varieties highlighted with yellow are eight resequenced HYVs. Homozygous alleles, which are identical to Nipponbare, were represented as red, and this is different from Nipponbare, represented as green and heterozygous alleles as blue. Grey indicates a missing genotype. The percentage values in parenthesis under each subgroup represent the percentage for homozygous *indica* allele in the 38 SNP marker set and 16 SNP marker set.

A total of 38 SNP markers were developed in the common Tongil-like *japonica*-like regions and discriminated IND1, IND2, and HYV by the frequency of *japonica*-alleles. The HYV group showed 93.8% for *japonica*-alleles frequency, which is similar to that of JAP. A total of 16 SNP markers were related to yield and some agronomic traits could distinguish JAP from the other three groups. For the 16 SNP set, the JAP group represented 86.1% of *japonica* allele, while the other three groups showed a lower *japonica* allele frequency, which was less than 30%. The HYV group contained *indica* alleles and were informed by the makers that were linked to the genes associated with plant architecture (SD1-GA, NAL1, and TAC-CT), yield potential (GIF1, Hd6-AT, Ghd7, and GW8-AG), and subspecies differentiation (Rd-GA, qSH1-TG, and S5-TC). They contained more than 50% japonica alleles using the markers linked to the genes for grain shape and quality (GRF4, GS3-CA, qSW5-AG, GS6-GT, and WAXY-TG) of HYV-type. Practically, the markers designed in the *japonica*-originated genomic regions and the yield-related markers from *indica* varieties can differentiate HYV from *indica* and *japonica* varieties. In fact, a high proportion of *japonica* alleles on Chr1-1, Chr1-2, and Chr3-1 were found in IND2, which consist of three *aus* varieties, one wild rice relative accession, and one weedy rice variety.

4. Discussion

All eight HYVs used in this study were clustered into the *indica* group, based on *indica–japonica* SNP sets reported in a previous study [29]. Interestingly, some *japonica*-type SNPs were detected within the genomes of HYVs after resequencing analysis. Furthermore, collocations of 14 *japonica*-originated

genomic regions commonly present in Tongil-like HYVs, inherited from *indica–japonica* crosses, were investigated. This suggests that these *japonica*-type genomic segments were commonly and repeatedly selected during independent breeding programs in the temperate rice cultivation area. To investigate the role of these *japonica* segments in HYVs, a comparative study of the reported QTLs and representative gene selection were conducted. Eating quality, stress tolerance, and yield related traits might be main drivers for the selection of the rice HYV breeding program.

A total of 54 SNP markers, including 38 SNP markers developed from 38 selected genes for 14 common *japonica*-type regions and 16 trait-specific SNP markers, were used for the genotyping of 94 diverse varieties across *indica* and *japonica*. For the 38 SNP marker set, 16 SNP markers were located on chromosome 1, which was not identified as a *japonica* genomic segment in Takanari. Consequently, Takanari was clustered into the IND1 subgroup, although it showed a similar genotypic pattern to the other HYVs. The Chr1-1 and Chr1-2 regions were co-located with some abiotic tolerance and yield related QTLs. Furthermore, there were several selected genes that conferred abiotic stress tolerance and yield potential in the blocks. For instance, *Os01g0337100* (*OsTPS1*) reported an association with abiotic stress response and tolerance by knock-out and overexpression [40,41]. *Os01g0367100* (*PHD1*) was shown to be a gene that was involved in galactolipid biosynthesis and affected photosynthetic efficiency [42]. Recently, the effect of the haplotype of *PHD1* on grain yield was also reported using the 3K rice genome panel [18]. Takanari could not have acquired this *japonica* genomic segment from a different natural environment and/or breeder's selection.

In addition, 18 varieties out of the 19 HYV-types, showed *japonica*-type Wx^b allele on the SNP marker WAXY-TG, which was designed on the splicing site in the intron of *Wx* gene in the *japonica*-type region. The *Wx* gene only contains synonymous SNPs, although it is located within Chr6-1; thus, it was not selected in our study. We previously developed a functional SNP marker for the *Wx* gene [29]. The genomic region containing the *Wx* gene is a hotspot for grain quality [43] and has been selected during and after the domestication of rice [44]. The other genomic research using two Tongil-type varieties also showed a *japonica*-type SNP pattern on the common *japonica*-type region on chromosome 6 [45]. Tongil-type varieties showed medium amylose contents, approximately 19–20%, which is similar to that of non-waxy Korean *japonica* varieties [46,47]. Further, *Os06g0130400*, one of the selected genes in Chr6-1, was also reported as the gene controlling starch grain size in endosperm [48]. *Os06g0130000* and *Os06g0130100* were reported for resistance to rice blast and bacterial blight, and tolerance to drought and salt stress, respectively [49,50]. Therefore, Chr6-1, including the *Wx* gene, might be mainly selected for eating quality and latent stress tolerance.

When HYVs were developed by inter-subspecific hybridization, the breeders aimed at not only transferring some of the desirable characteristics, like resistance to lodging, blast, and yield, but also at retaining the ecological adaptability and eating quality of *japonicas* [51]. The *japonica* chromosomal introgression regions identified in this study were regarded as putative temperate region adaptable and improved the eating quality of *indica*. For this reason, the varieties developed by *indica–japonica* crosses could also be considered as 'temperate *indica*'. Recently, new elite rice varieties showing high yield potential and high grain quality were developed by the precise pyramiding of major genes controlling yield and grain quality traits [52]. Furthermore, there was an attempt to develop cold tolerant *indica* using an inter-subspecific cross and marker-assisted selection (MAS) [53]. In other words, breeding *indica* varieties, which are adaptable to the temperate region with high yield potential and good eating quality, can be efficiently achieved through inter-subspecific crosses and marker-assisted selection using the SNP markers developed in this study. Nevertheless, to dissect the exact contribution of *japonica*-type regions in HYVs, a comprehensive genetic and physiological analysis, by applying the molecular markers developed in *japonica*-type regions to the segregation populations derived from cross between HYVs and *indica*, is necessary. In addition, the functional studies of genes in the regions, as well as the selected ones in this study, are also required.

5. Conclusions

Consumer preference in grain shape and quality during conventional breeding procedures, without sacrificing the high-yield potential of *indica*, were revisited by the genomic analysis of HYVs. The 14 *japonica*-originated genomic regions and alleles identified in this study shared by HYVs could be applied in further development of more HYVs through inter-subspecific rice breeding in temperate rice.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/5/562/s1, Figure S1: Pedigrees of eight HYVs used in this study. (A) Tongil, (B) Minghui 63, (C) Cheongcheongbyeo, (D) Milyang 23, (E) Nampungbyeo, (F) Hanareumbyeo, (G) Dasanbyeo, (H) Takanari. Table S1: The list of 94 diverse rice germplasms used in this study. The order of ger1mplasm is identical with order in Figure 3. The varieties highlighted by yellow are eight HYVs used in this study. Table S2: The list of 54 Fluidigm SNP markers used in this study. MAF, Heterozygosity and PIC were calculated in 94 rice varieties.

Author Contributions: Conceptualization, J.H.C. and H.-J.K.; methodology, J.S. and J.H.C.; software, J.S.; validation, B.K. and J.H.C.; formal analysis, J.S. and S.-M.L.; investigation, J.S., S.-M.L., J.-H.H., N.-H.S., and Y.K.L., resources, J.H.C and H.-J.K.; data curation, J.S. and B.K.; writing—original draft preparation, J.S.; writing—review and editing, J.H.C. and H.-J.K.; visualization, J.S., S.-M.L., and J.H.C.; supervision, H.-J.K.; project administration, J.H.C.; funding acquisition, H.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by a grant from the Next-Generation BioGreen 21 Program (No. PJ013165) of the Rural Development Administration, Korea.

Acknowledgments: We appreciate to members of Crop Molecular Breeding Lab., in Seoul National University for their kind advice and assistance for this study.

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

References

1. Oka, H.I. *Origin of Cultivated Rice*; Japan Scientific Societies Press: Tokyo, Japan, 1988.
2. Chung, G.S.; Heu, M.H. Improvement of Tongil-Type Rice Cultivars from *Indica*/*Japonica* Hybridization in Korea. In *Rice*; Bajaj, Y.P.S., Ed.; Springer: Berlin/Heidelberg, Germany, 1991; pp. 105–112. [CrossRef]
3. Kim, B.; Kim, D.-G.; Lee, G.; Seo, J.; Choi, I.-Y.; Choi, B.-S.; Yang, T.-J.; Kim, K.; Lee, J.; Chin, J.; et al. Defining the genome structure of 'Tongil' rice, an important cultivar in the Korean "Green Revolution". *Rice* **2014**, *7*, 22. [CrossRef]
4. Takai, T.; Arai-Sanoh, Y.; Iwasawa, N.; Hayashi, T.; Yoshinaga, S.; Kondo, M. Comparative Mapping Suggests Repeated Selection of the Same Quantitative Trait Locus for High Leaf Photosynthesis Rate in Rice High-Yield Breeding Programs. *Crop Sci.* **2012**, *52*, 2649–2658. [CrossRef]
5. Takai, T.; Ikka, T.; Kondo, K.; Nonoue, Y.; Ono, N.; Arai-Sanoh, Y.; Yoshinaga, S.; Nakano, H.; Yano, M.; Kondo, M.; et al. Genetic mechanisms underlying yield potential in the rice high-yielding cultivar Takanari, based on reciprocal chromosome segment substitution lines. *BMC Plant Biol.* **2014**, *14*, 295. [CrossRef]
6. Xie, F.; Zhang, J. Shanyou 63: An elite mega rice hybrid in China. *Rice* **2018**, *11*, 17. [CrossRef]
7. Zhu, L.; Lu, C.; Li, P.; Shen, L.; Xu, Y.; He, P.; Chen, Y. Using doubled haploid populations of rice for quantitative trait locus mapping. In *Rice Genetics III. Proceedings of the Third International Rice Genetics Symposium, Manila, Philippines, 16–20 October 1995*; Khush, G.S., Ed.; International Rice Research Institute: Los Baños, Philippines, 1996; pp. 631–636.
8. Xie, W.; Wang, G.; Yuan, M.; Yao, W.; Lyu, K.; Zhao, H.; Yang, M.; Li, P.; Zhang, X.; Yuan, J.; et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5411–E5419. [CrossRef]
9. Sasaki, T. The map-based sequence of the rice genome. *Nature* **2005**, *436*, 793–800. [CrossRef]
10. Yu, J.; Hu, S.; Wang, J.; Wong, G.K.-S.; Li, S.; Liu, B.; Deng, Y.; Dai, L.; Zhou, Y.; Zhang, X.; et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. indica). *Science* **2002**, *296*, 79–92. [CrossRef]
11. Schatz, M.C.; Maron, L.G.; Stein, J.C.; Wences, A.H.; Gurtowski, J.; Biggers, E.; Lee, H.; Kramer, M.; Antoniou, E.; Ghiban, E.; et al. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. *Genome Biol.* **2014**, *15*, 1–16. [CrossRef]
12. Sakai, H.; Kanamori, H.; Arai-Kichise, Y.; Shibata-Hatta, M.; Ebana, K.; Oono, Y.; Kurita, K.; Fujisawa, H.; Katagiri, S.; Mukai, Y.; et al. Construction of Pseudomolecule Sequences of the aus Rice Cultivar Kasalath for Comparative Genomics of Asian Cultivated Rice. *DNA Res.* **2014**. [CrossRef]

13. Zhang, J.; Chen, L.-L.; Xing, F.; Kudrna, D.A.; Yao, W.; Copetti, D.; Mu, T.; Li, W.; Song, J.-M.; Xie, W.; et al. Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E5163–E5171. [CrossRef]
14. Du, H.; Yu, Y.; Ma, Y.; Gao, Q.; Cao, Y.; Chen, Z.; Ma, B.; Qi, M.; Li, Y.; Zhao, X.; et al. Sequencing and *de novo* assembly of a near complete *indica* rice genome. *Nat. Commun.* **2017**, *8*, 15324. [CrossRef]
15. Subbaiyan, G.K.; Waters, D.L.; Katiyar, S.K.; Sadananda, A.R.; Vaddadi, S.; Henry, R.J. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J.* **2012**, *10*, 623–634. [CrossRef]
16. Feltus, F.A.; Wan, J.; Schulze, S.R.; Estill, J.C.; Jiang, N.; Paterson, A.H. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.* **2004**, *14*, 1812–1819. [CrossRef]
17. Sun, C.; Hu, Z.; Zheng, T.; Lu, K.; Zhao, Y.; Wang, W.; Shi, J.; Wang, C.; Lu, J.; Zhang, D.; et al. RPAN: Rice pan-genome browser for approximately 3000 rice genomes. *Nucleic Acids Res.* **2017**, *45*, 597–605. [CrossRef]
18. Abbai, R.; Singh, V.K.; Nachimuthu, V.V.; Sinha, P.; Selvaraj, R.; Vipparla, A.K.; Singh, A.K.; Singh, U.M.; Varshney, R.K.; Kumar, A. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* **2019**, *17*, 1612–1622. [CrossRef]
19. Carpentier, M.C.; Manfroi, E.; Wei, F.J.; Wu, H.P.; Lasserre, E.; Llauro, C.; Debladis, E.; Akakpo, R.; Hsing, Y.I.; Panaud, O. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat. Commun.* **2019**, *10*, 24. [CrossRef]
20. Fuentes, R.R.; Chebotarov, D.; Duitama, J.; Smith, S.; De la Hoz, J.F.; Mohiyuddin, M.; Wing, R.A.; McNally, K.L.; Tatarinova, T.; Grigoriev, A.; et al. Structural variants in 3000 rice genomes. *Genome Res.* **2019**, *29*, 870–880. [CrossRef]
21. 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]
22. Adachi, S.; Yamamoto, T.; Nakae, T.; Yamashita, M.; Uchida, M.; Karimata, R.; Ichihara, N.; Soda, K.; Ochiai, T.; Ao, R.; et al. Genetic architecture of leaf photosynthesis in rice revealed by different types of reciprocal mapping populations. *J. Exp. Bot.* **2019**, *70*, 5131–5144. [CrossRef]
23. Takai, T.; Adachi, S.; Taguchi-Shiobara, F.; Sanoh-Arai, Y.; Iwasawa, N.; Yoshinaga, S.; Hirose, S.; Taniguchi, Y.; Yamanouchi, U.; Wu, J.; et al. A natural variant of *NAL1*, selected in high-yield rice breeding programs, pleiotropically increases photosynthesis rate. *Sci. Rep.* **2013**, *3*. [CrossRef]
24. 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*, 4. [CrossRef]
25. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef]
26. Barnett, D.W.; Garrison, E.K.; Quinlan, A.R.; Stromberg, M.P.; Marth, G.T. BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **2011**, *27*, 1691–1692. [CrossRef]
27. 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]
28. Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef]
29. Seo, J.; Lee, G.; Jin, Z.; Kim, B.; Chin, J.H.; Koh, H.-J. Development and application of *indica–japonica* SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding. *Mol. Breed.* **2020**, *40*, 39. [CrossRef]
30. D3 Assay Design-Fluidigm. Available online: https://d3.fluidigm.com (accessed on 25 September 2019).
31. Murray, M.G.; Thompson, W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [CrossRef]
32. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [CrossRef]
33. Cavalli-Sforza, L.L.; Edwards, A.W.F. Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* **1967**, *19*, 233–257.

34. 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]
35. Yonemaru, J.-i.; Yamamoto, T.; Fukuoka, S.; Uga, Y.; Hori, K.; Yano, M. Q-TARO: QTL Annotation Rice Online Database. *Rice* **2010**, *3*, 194–203. [CrossRef]
36. Seo, J.; Lee, S.-M.; Han, J.-H.; Shin, N.-H.; Koh, H.-J.; Chin, J.H. Identification of Yield and Yield-Related Quantitative Trait Loci for the Field High Temperature Condition in Backcross Populations of Rice (*Oryza sativa* L.). *Plant Breed. Biotechnol.* **2019**, *7*, 415–426. [CrossRef]
37. Ha, U.-G.; Song, Y.-C.; Yeo, U.-S.; Cho, J.-H.; Hwang, H.-G.; Kim, Y.-D.; Cho, Y.-H.; Yang, S.-J.; Lee, J.-H.; Oh, B.-G.; et al. A New High Yielding Rice Variety with Multi-Disease Resistance, 'Keunseom'. *Korean J. Breed. Sci.* **2011**, *43*, 576–580.
38. Huang, N.; Angeles, E.R.; Domingo, J.; Magpantay, G.; Singh, S.; Zhang, G.; Kumaravadivel, N.; Bennett, J.; Khush, G.S. Pyramiding of bacterial blight resistance genes in rice: Marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* **1997**, *95*, 313–320. [CrossRef]
39. Khush, G.; Virk, P. *IR Varieties and Their Impact*, 1st ed.; International Rice Research Institute: Los Baños, Philippines, 2005; pp. 46–48.
40. Kim, S.-J.; Jeong, D.-H.; An, G.; Kim, S.-R. Characterization of a drought-responsive gene, *OsTPS1*, identified by the T-DNA Gene-Trap system in rice. *J. Plant Biol.* **2005**, *48*, 371–379. [CrossRef]
41. Li, H.W.; Zang, B.S.; Deng, X.W.; Wang, X.P. Overexpression of the trehalose-6-phosphate synthase gene *OsTPS1* enhances abiotic stress tolerance in rice. *Planta* **2011**, *234*, 1007–1018. [CrossRef]
42. Li, C.; Wang, Y.; Liu, L.; Hu, Y.; Zhang, F.; Mergen, S.; Wang, G.; Schlappi, M.R.; Chu, C. A rice plastidial nucleotide sugar epimerase is involved in galactolipid biosynthesis and improves photosynthetic efficiency. *PLoS Genet.* **2011**, *7*, e1002196. [CrossRef]
43. Sreenivasulu, N.; Butardo, V.M., Jr.; 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]
44. Olsen, K.M.; Caicedo, A.L.; Polato, N.; McClung, A.; McCouch, S.; Purugganan, M.D. Selection Under Domestication: Evidence for a Sweep in the Rice *Waxy* Genomic Region. *Genetics* **2006**, *173*, 975–983. [CrossRef]
45. Ji, H.; Ahn, E.; Seo, B.; Kang, H.; Choi, I.; Kim, K. Genome-wide detection of SNPs between two Korean Tongil-type rice varieties. *Korean J. Breed. Sci.* **2016**, *48*, 460–469. [CrossRef]
46. Kim, H.-Y.; Yang, C.-I.; Choi, Y.-H.; Won, Y.-J.; Lee, Y.-T. Changes of Seed Viability and Physico-Chemical Properties of Milled Rice with Different Ecotypes and Storage Duration. *Korean J. Crop Sci.* **2007**, *52*, 375–379.
47. Kwak, J.; Yoon, M.R.; Lee, J.S.; Lee, J.H.; Ko, S.; Tai, T.H.; Won, Y.J. Morphological and starch characteristics of the *Japonica* rice mutant variety Seolgaeng for dry-milled flour. *Food Sci. Biotechnol.* **2017**, *26*, 43–48. [CrossRef]
48. Matsushima, R.; Maekawa, M.; Kusano, M.; Tomita, K.; Kondo, H.; Nishimura, H.; Crofts, N.; Fujita, N.; Sakamoto, W. Amyloplast Membrane Protein SUBSTANDARD STARCH GRAIN6 Controls Starch Grain Size in Rice Endosperm. *Plant Physiol.* **2016**, *170*, 1445–1459. [CrossRef]
49. Fekih, R.; Tamiru, M.; Kanzaki, H.; Abe, A.; Yoshida, K.; Kanzaki, E.; Saitoh, H.; Takagi, H.; Natsume, S.; Undan, J.R.; et al. The rice (*Oryza sativa* L.) *LESION MIMIC RESEMBLING*, which encodes an AAA-type ATPase, is implicated in defense response. *Mol. Genet. Genom.* **2015**, *290*, 611–622. [CrossRef]
50. Ouyang, S.Q.; Liu, Y.F.; Liu, P.; Lei, G.; He, S.J.; Ma, B.; Zhang, W.K.; Zhang, J.S.; Chen, S.Y. Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants. *Plant J.* **2010**, *62*, 316–329. [CrossRef]
51. Dalrymple, D.G. Development of High-Yielding Rice Varieties. In *Development and Spread of High-Yielding Rice Varieties in Developing Countries*, 2nd ed.; Bureau for Science and Technology Agency for International Development: Washington, DC, USA, 1986; pp. 15–35.
52. Zeng, D.; Tian, Z.; Rao, Y.; Dong, G.; Yang, Y.; Huang, L.; Leng, Y.; Xu, J.; Sun, C.; Zhang, G.; et al. Rational design of high-yield and superior-quality rice. *Nat. Plants* **2017**, *3*, 17031. [CrossRef]
53. Li, L.; Mao, D.; Prasad, M. Deployment of cold tolerance loci from *Oryza sativa* ssp. *Japonica* cv. 'Nipponbare' in a high-yielding *Indica* rice Cultivar '93-11'. *Plant Breed.* **2018**, *137*, 553–560. [CrossRef]

genes

MDPI

Article

Characterization of Quantitative Trait Loci for Germination and Coleoptile Length under Low-Temperature Condition Using Introgression Lines Derived from an Interspecific Cross in Rice

Mirjalol Akhtamov [1], Cheryl Adeva [1], Kyu-Chan Shim [1], Hyun-Sook Lee [1], Sun Ha Kim [1], Yun-A Jeon [1], Ngoc Ha Luong [1], Ju-Won Kang [2], Ji-Yoon Lee [2] and Sang-Nag Ahn [1,*]

[1] Department of Agronomy, Chungnam National University, Daejeon 34134, Korea; miramax90501@gmail.com (M.A.); ccadeva_758@yahoo.com (C.A.); zktnrl@naver.com (K.-C.S.); leehs0107@gmail.com (H.-S.L.); sunha82@cnu.ac.kr (S.H.K.); jya0911@cnu.ac.kr (Y.-AJ.); luongngocha.biotech@gmail.com (N.H.L.)

[2] Department of Southern Area Crop Science, National Institute of Crop Science, RDA, Miryang 50424, Korea; kangjw81@korea.kr (J.-W.K.); minitia@korea.kr (J.-Y.L.)

* Correspondence: ahnsn@cnu.ac.kr; Tel.: +82-42-821-5728

Received: 24 August 2020; Accepted: 9 October 2020; Published: 15 October 2020

Abstract: Previously, five putative quantitative trait loci (QTLs) for low-temperature germination (LTG) have been detected using 96 BC_3F_8 lines derived from an interspecific cross between the Korean *japonica* cultivar "Hwaseong" and *Oryza rufipogon*. In the present study, two introgression lines, CR1517 and CR1518, were used as parents to detect additional QTLs and analyze interactions among QTLs for LTG. The F_2 population (154 plants) along with parental lines, Hwaseong and *O. rufipogon*, were evaluated for LTG and coleoptile length under low-temperature conditions (13 °C). Among five QTLs for LTG, two major QTLs, *qLTG1* and *qLTG3*, were consistently detected at 6 and 7 days after incubation. Three minor QTLs were detected on chromosomes 8 and 10. Two QTLs, *qLTG10.1* and *qLTG10.2*, showing linkage on chromosome 10, exerted opposite effects with the Hwaseong allele at *qLTG10.2* and the *O. rufipogon* allele at *qLTG10.1* respectively, in turn, increasing LTG. Interactions among QTLs were not significant, implying that the QTLs act in an additive manner. Near-isogenic line plants with the combination of favorable alleles from *O. rufipogon* and Hwaseong exhibited higher LTG than two introgression lines. With regard to coleoptile length, three QTLs observed on chromosomes 1, 3, and 8 were colocalized with QTLs for LTG, suggesting the pleiotropy of the single gene at each locus. According to the results, the introgression of favorable *O. rufipogon* alleles could hasten the development of rice with high LTG and high coleoptile elongation in *japonica* cultivars.

Keywords: low-temperature germinability; rice; interspecific cross; QTL; interaction

1. Introduction

Rice is one of the most important crops, feeding more than a third of the global population. Global rice demand is projected to rise from 723 million tons in 2015 to 763 million tons by 2020 and is expected to increase further to 852 million tons by 2035 [1]. To feed the growing population, rice production needs to be improved and stabilized, globally and regionally. Improving potential yield and incorporation of biotic and abiotic stress tolerance mechanisms would facilitate the achievement of such goals. Wild relatives of rice are rich sources of desirable genes, not only with regard to yield but also with regard to disease resistance, stress tolerance, and other traits [2,3]. Exploring wild and exotic rice germplasm for desirable genes and transferring them into cultivars through crossing and

marker-assisted selection (MAS) has been demonstrated to be feasible approaches of increasing rice yield and augmenting stress tolerance [4].

Rice is cultivated in tropical, subtropical, and temperate regions of the globe. Low-temperature is one of the abiotic stress factors that lead to growth retardation and yield loss in rice cultivated at high latitudes and in the northern regions of the globe [5]. Throughout the rice cultivation period, decreasing temperatures to levels lower than the optimum could influence rice development significantly and can result in severe yield loss [6]. During the seed development period, various factors, such as environment and genetic background, influence low-temperature germination [7]. In rice cultivating regions, in areas where rice is irrigated with cold water, weak seedling growth and slow growth reduce yield significantly [8–13].

Along with rapid low-temperature germination (LTG), vigorous coleoptile growth is essential in direct seeding rice, when rice seeds are sown in flooded paddy fields and watered with cold irrigation water [14]. The rice coleoptile is a small and ephemeral organ that emerges first in imbibed rice seeds [15].

To determine the genetic basis of low-temperature germinability and coleoptile elongation in rice, numerous independent studies have been conducted using biparental populations and genome-wide association studies [6,8,9,12,16–21]. A total of 11 QTLs have been associated with low-temperature germination ability in the population derived from a cross between *indica* (N22) and *japonica* (USSR5) rice [12]. Three QTLs associated with low-temperature germination have been detected, and among them, *qLTG-9* was fine mapped to a 72.3-kbp region on chromosome 9 [22]. In addition, Lee et al. (2015) detected two colocalized QTLs for LTG and coleoptile length under a low-temperature condition (13 °C) [6]. Fukuda et al. (2015) also identified two QTLs associated with coleoptile length under 16 °C conditions [23]. Using the genome-wide association (GWAS) method, two QTLs associated with germination under cold tolerance, *qCTGERM8-1* and *qCTGERM8-2*, were identified in 421 accessions from a rice diversity panel (RDP1) [24]. In addition, Fujino et al. (2015) detected *qLTG8* on the RM5647 marker locus in 63 Japanese landraces [18]. Li et al. (2019) reported that Chinese Dongxiang wild rice (*O. rufipogon* Griff.) alleles at all five QTLs resulted in delayed germination rates, and QTL pyramiding line DX71 led to rapid germination and vigorous seedling survival under low-temperature conditions (15 °C and 8 °C) [25]. RNAseq analysis was conducted with two *indica* rice genotypes under cold treatment and detected 1361 differentially expressed transcripts [26].

Such studies indicate that low-temperature germination QTLs are distributed widely throughout the rice genome. However, the stability of such putative QTLs has not been tested in near-isogenic backgrounds, and studies analyzing interactions among QTLs are limited. Characterization of the QTLs underlying LTG and their interactions could provide more insights into the mechanisms of low-temperature germinability in rice and would be potentially more useful for pyramiding QTLs aimed at improving LTG in rice.

In our previous study, five QTLs for low-temperature germinability, *qLTG1*, *qLTG3*, *qLTG4*, *qLTG10*, and *qLTG11* were mapped using a BC_4F_8 population including 96 introgression lines (ILs) derived from an interspecific cross between a Korean elite line Hwaseong and *O. rufipogon* [27]. Recently, *qLTG1* was fine mapped using introgression lines and the new *qLTG3-1* allele of the *O. rufipogon* was identified [28,29]. In the present study, we used two introgression lines, CR1517 and CR1518, which consistently exhibit higher LTG than Hwaseong due to *O. rufipogon* introgression segments in the Hwaseong genetic background, as parents. Using 154 F_2 plants derived from a cross between CR1517 and CR1518, we aimed to: (1) identify and characterize QTLs for low-temperature germinability and coleoptile length, (2) analyze interactions among LTG QTLs, and (3) develop QTL pyramiding lines with enhanced LTG and coleoptile length.

2. Materials and Methods

2.1. Plant Materials

Two introgression lines, CR1517 and CR1518, which were derived from crossing between Hwaseong and *O. rufipogon*, were used as parents. We selected the two parental lines from an interspecific cross population, BC_4F_8 [27]. CR1517 has *O. rufipogon* chromosomal segments on chromosomes 8 and 10, and CR1518 has *O. rufipogon* segments on chromosomes 1, 3, 9, and 10 (Figure 1). Two introgression lines were crossed to produce F_1 seeds. During the winter season of 2018–2019, F_1 seeds were sown in a greenhouse to produce F_2 seeds. In the middle of April 2019, F_2 seeds were sown in a greenhouse and 30-day-old seedlings were transplanted into the experimental paddy field. The F_2 population (154 plants) was grown in a field belonging to Chungnam National University, Daejeon, South Korea. The plants were used for further phenotypic and genotypic analyses. To verify minor QTLs, two F_2 plants (CR8017-4 and CR8020-6) were selected based on the genotypes to verify *qLTG10.1* and *qLTG10.2*, respectively. CR8017-4 is heterozygous at *qLTG10.1* and *qLTG8*, while *O. rufipogon* homozygous for *qLTG3*, and Hwaseong homozygous for *qLTG1* and *qLTG10.2*. CR8020-6 is heterozygous at *qLTG10.2*, while *O. rufipogon* homozygous for *qLTG1* and Hwaseong homozygous for *qLTG3* and *qLTG10.1*. The F_3 plants were grown in the 2019/2020 winter season and dormancy was broken for seeds harvested from each F_3 plant.

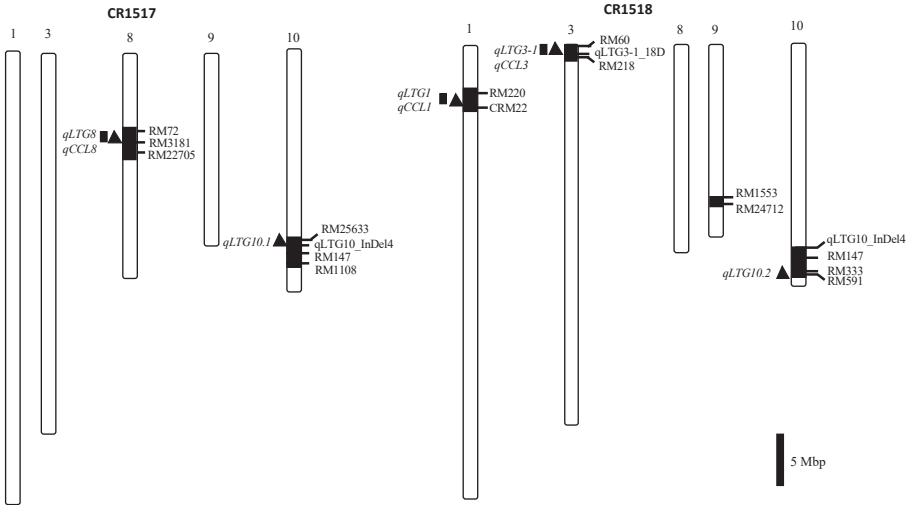


Figure 1. Graphical genotypes of two parental lines (CR1517 and CR1518) with the locations of five and three QTLs for low-temperature (*qLTG* in filled triangle) and coleoptile length (*qCCL* in filled box), respectively, on the left of the chromosomes. Black and white bars represent *O. rufipogon* and Hwaseong chromosome segments, respectively.

2.2. Evaluation of LTG and Coleoptile Length

Seeds from each F_2 plant were harvested five weeks after heading and dried in a greenhouse for two weeks. To break dormancy, seeds were stored in a dry oven at 50 °C for 72 h. For normal temperature and low-temperature germination tests, 20 and 30 seeds were used with three replicates, respectively. Healthy and normal seeds were carefully selected and used in the germination test. The seeds from each plant were placed into 9-cm Petri dishes onto one layer of filter paper and 5 mL of distilled water added. Dormancy breakage was checked at the normal germination temperature (30 °C) for four days. The low-temperature germination test was conducted at 13 °C under the dark condition

for eight days. Germinated seeds of each plant were counted daily and defined as germination rate. The germination rate (%) was calculated as follows: Germination rate (%) for a given day = (number of germinated seeds until the given day/total number of seeds) × 100. The germination tests were repeated two times under similar conditions. The seeds were considered germinated when coleoptile emerged from the seed (Figure S1).

Coleoptile length measurement was conducted based on previous studies [6,14]. F_3 seeds of F_2 plants were used to measure coleoptile length. Twenty normal and healthy seeds per plant were placed on one layer of filter paper in 9-cm Petri dishes and 10 mL of distilled water was poured into each Petri dish. The measurements were performed with two replicates using parental lines under normal (30 °C) and low-temperature (13 °C) conditions. In the F_3 generation, coleoptile length was measured under low-temperature conditions (13 °C) with one replicate. Distilled water was renewed daily and maintained at 30 °C and 13 °C conditions for the normal and low-temperature conditions, respectively. Coleoptile length was measured using a ruler. In the normal temperature treatments, measurements were carried out from 1 to 4 days after incubation (DAI) and, in the low-temperature treatments, measurements were carried out from 5 to 10 DAI. For the normal and low-temperature conditions, coleoptile lengths of the grains were measured at room temperature (26 °C) and at seed storage temperature (10 °C) conditions, respectively, and mean lengths of the coleoptile were used in further analyses.

2.3. DNA Extraction and Marker Analysis

Booting stage plant leaves were gathered for use in extracting genomic DNA, and extraction was performed based on the method of Causse et al. (1994) with some minor modifications [30]. The Cetyl trimethylammonium bromide (CTAB) method was used to extract plant DNA. Leaf fragments from each plant were placed in a 2.0-mL microcentrifuge tube with aluminum beads (4 mm) and 500 µL of CTAB extraction buffer added. The tubes were incubated at 55 °C in a dry oven in CTAB buffer (2% CTAB, 100 mM pH 8.0 Tris-HCl, 20 mM EDTA, 1.4 M NaCl, 0.2% β-mercaptoethanol) for 30 min right after grinding. An equal volume of chloroform was added and the solution was mixed well. The tubes were centrifuged at 4 °C for 15 min at 13,000 rpm. The 200-µL upper supernatant layer was transferred into a new 1.5-mL tube and mixed with 2/3 volume of isopropanol, and after gently mixing, incubated at 4 °C for 15 min at 12,000 rpm. The supernatant was discarded very carefully, and the DNA pellets were washed with 70% ethanol once. After washing the DNA pellets, they were kept at room temperature to dry out the rest of ethanol and then diluted in 100 µL TE buffer (10 mM Tris pH 8, 1 mM EDTA). The concentration of the DNA was checked with a NanoDrop™ 2000 spectrophotometer (Thermo Scientific Inc., Wilmington, DE, USA).

A 15-µL of PCR reaction mixture containing 5 µL DNA template (2~5 ng/µL), 1 µL of forward and reverse primer (10 pmol each), 1.5 µL of 10× PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Gelatin), 1 µL of dNTP (2.5 mM each), 0.1 µL of *Taq* polymerase (5 unit/µL), and 6.4 µL triple distilled water was used. The PCR was performed as described in Panaud et al. (1996) with minor modifications: 94 °C for 5 min, followed by 33–35 cycles of 94 °C for 30–40 s, 52–60 °C (based on annealing temperature of each marker) for 30 s, and 72 °C for 30 s, and a final extension at 5 min at 72 °C [31]. Separation of PCR products was conducted on 3% of metaphor agarose gel stained with Staining Safe Nucleic Acid Gel Stain (RBC, Taiwan). A total of 18 markers, including 15 Simple Sequence Repeat (SSR) and 3 Insertion-Deletion (InDel) markers were used to map QTLs (Table S1). Three InDel markers (qLTG3-1_18D, qLTG10_InDel3, and qLTG10_InDel4) and one SSR marker (CRM22) were designed and used for mapping [27,28].

2.4. Statistical Analysis and QTL Analysis

QTL analyses were conducted in Minitab 16.2.4 (Minitab Inc., State College, PA, USA) using one-way Analysis of Variance and Tukey's test for mean comparisons. QTLs were detected using single-marker analysis, and the QTL was declared when the association between phenotype and genotype of markers

was significant at $p < 0.05$. Gene actions were calculated by the following formula; additive effect (a) = (mean germination rate of *O. rufipogon* homozygote − mean germination rate of Hwaseong homozygote)/2; dominant effect (d) = mean germination rate of heterozygote − (mean germination rate of *O. rufipogon* homozygote + mean germination rate of Hwaseong homozygote)/2; degree of dominance = d/a. To analyze the interaction between QTLs, a general multiple regression model with two QTLs as independent variables and interaction term was employed in Minitab 16.2.4. Nomenclature of the QTLs has been described previously [32].

3. Results

3.1. Low-Temperature Germination Rate of Parental Lines and F_2 Population

Germination rates of four parental lines (Hwaseong, *O. rufipogon*, CR1517, and CR1518) were compared under control conditions (30 °C). The seeds of *O. rufipogon* began germinating at 1 DAI, with a 47% germination rate. At 4 DAI, *O. rufipogon* germination rate reached 100%. Hwaseong, CR1517, and CR1518 began germinating 2 DAI. Subsequently, the germination rate of the lines achieved nearly 100% at 4 DAI. *O. rufipogon* exhibited a significantly higher germination rate than those of other lines from 1 to 2 DAI. There were no significant differences in germination rate among the parental lines from 3 to 4 DAI under the control temperature (Figure S2). However, there were differences in the germination rate between 4 lines under the low-temperature condition (13 °C). *O. rufipogon* began germinating at 3 DAI and attained 100% germination rate at 7 DAI. In contrast, Hwaseong began germinating 4 DAI, and achieved 80% germination 8 DAI. Two introgression parental lines, CR1517 and CR1518, began germinating at 3 DAI, and the lines attained 85% and 94% germination rates, respectively, 7 DAI. They exhibited much more vigorous germinability than that of Hwaseong and lower than that of *O. rufipogon* (Figure S2). Germination rates in F_2 plants ranged from 0 to 100% during the incubation period for 1–8 DAI (Table 1). The results revealed that *O. rufipogon* segments increased LTG in the Hwaseong background.

Table 1. Mean germination rates in *O. rufipogon*, Hwaseong, two introgression lines, and F_2 population.

Germination Rate (%)						
Trait [x]	**Lines (n = 5** [z]**)**				**F_2 Population**	
	O. rufipogon	**Hwaseong**	**CR1517**	**CR1518**	**Mean**	**Range**
4 DAI	60.0 a [y]	3.3 c	30.4 b	34.8 b	15.6 ± 10.1	1.1–48.9
5 DAI	91.9 a	11.1 c	49.3 b	56.7 b	38.3 ± 15.2	4.4–72.2
6 DAI	97.4 a	33.9 d	70.4 c	78.5 b	58.6 ± 17.7	12.2–88.9
7 DAI	100 a	62.2 d	84.2 c	91.9 b	74.9 ± 16.5	25.6–95.6
8 DAI	100 a	80.7 b	99.3 a	100 a	90.7 ± 11.6	51.1–100

[x] DAI: days after incubation. Data are presented as mean germination rate or mean ± standard deviation. [y] Different letters indicate significant differences among lines at each DAI based on Tukey's test ($p < 0.05$). [z] n: number of plants.

3.2. Mapping QTLs for Low-Temperature Germinability

We used low-temperature germination rate of the F_2 population at 6 and 7 DAI for the QTL analysis because we observed larger differences among the parental lines (CR1517 and CR1518) at the periods. The F_2 population exhibited nearly normal distribution of the LTG rate at 6 DAI (skewness = −0.4277) and was skewed to the left at 7 DAI (skewness = −1.0609) (Figure 2). Using QTL-linked SSR markers and one InDel marker, we consistently detected *qLTG1* and *qLTG3* 6 and 7 DAI, respectively. *qLTG1* was detected between RM220 and CRM22 markers, while *qLTG3* was detected on qLTG3-1_18D marker on chromosomes 1 and 3, respectively. The *qLTG1* explained 16.0% and 12.0% phenotypic variation 6 and 7 DAI, respectively, while *qLTG3* explained 23.8% and 22.1% phenotypic variation 6 and 7 DAI, respectively. Based on gene action analysis 6 DAI, the additive effect (a) and dominant effect (d) of *O. rufipogon* allele on *qLTG1* locus (RM220–CRM22) were 10.0 and 0.2, respectively. *O. rufipogon* allele

of *qLTG1* had 7.7 and 2.1% of the additive and dominant effects, respectively, 7 DAI. In the *qLTG3* locus (qLTG3-1_18D), *O. rufipogon* allele accounted for 9.6 and 8.1% of the additive effects 6 and 7 DAI, respectively. The dominant effect was 9.7% on both days. Degrees of dominance were 1.0 and 1.2 at 6 and 7 DAI, respectively (Table 2). In addition, we detected two minor QTLs on chromosome 8 and 10. No QTLs were detected between markers RM72 and RM22705 ($p = 0.086$) on chromosome 8 at 6 DAI. *qLTG8* was detected when the significant difference had a p value of 0.014 at 7 DAI. The phenotypic variation explained by *qLTG8* was 5.45%. In addition, we identified other QTLs on chromosome 10. The QTLs were detected on the most distal end of the short arm of chromosome 10 by RM25633 and RM333-RM591, respectively. The phenotypic variation explained by *qLTG10.1* was 4.6% at 7 DAI. *qLTG10.2* explained 4.2 and 6.6% of the phenotypic variation at 6 and 7 DAI, respectively. In addition, additive and dominant effects of *O. rufipogon* alleles at the *qLTG8* locus (RM72-RM22705) were 4.9 and 2.4% at 7 DAI, respectively. *qLTG10.1* had 2.9 and −1.9% additive and dominant effects, respectively, at 7 DAI. The additive effects of *qLTG10.2* were −4.2 and −5.3% at 6 and 7 DAI, and the dominant effects of the QTL were 5.1 and 5.5% at 6 and 7 DAI, respectively. Although two QTLs, *qLTG10.1* and *qLTG10.2*, were linked on the short arm of chromosome 10, the effect for LTG of *O. rufipogon* allele was different. *O. rufipogon* introgression had a −0.6 degree of dominance at the *qLTG10.1* locus at 7 DAI, while −1.2 and −1.0 degrees of dominance were observed at *qLTG10.2* at 6 and 7 DAI, respectively (Table 2). The gene actions of *O. rufipogon* are different in the five QTLs. The *O. rufipogon* allele is partially dominant at *qLTG8* and *qLTG10.2*, partially recessive at *qLTG10.1*, completely dominant at *qLTG3*, and completely recessive at *qLTG1* with regard to regulating LTG (Table 2).

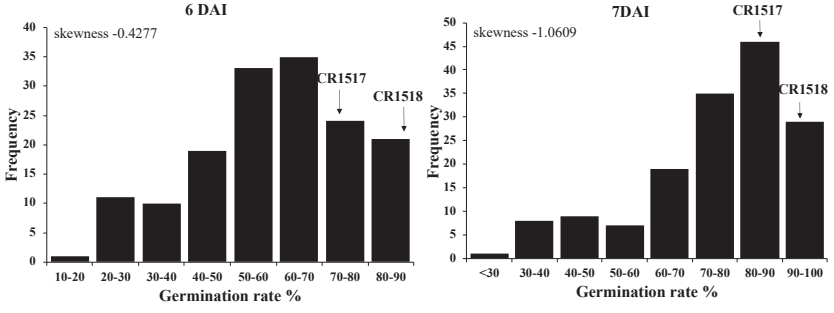


Figure 2. Frequency distribution of germination rate in 154 F_2 plants at 6 and 7 days after incubation (DAI) at 13 °C. Germination rate was measured at 4–8 (DAI). Arrows indicate mean germination rate of CR1517 and CR1518 lines.

Table 2. List of QTLs identified for low-temperature germinability in F_2 population.

Population	Trait [x]	QTL	Chr.	Markers	*F*-Value	*p*-Value	R^2 (%) [y]	Gene Action [z]		
								a	*d*	*d*/*a*
F_2	LTG (6 DAI)	*qLTG1*	1	RM220-CRM22	14.4	0.000	16.0	10.0	0.2	0.0
		qLTG3	3	qLTG3-1_18D	23.1	0.000	23.8	9.6	9.7	1.0
		qLTG10.2	10	RM333-RM591	3.3	0.038	4.2	−4.2	5.1	−1.2
		Total					38.9			
	LTG (7 DAI)	*qLTG1*	1	RM220-CRM22	10.3	0.000	12.0	7.7	2.1	0.3
		qLTG3	3	qLTG3-1_18D	21.0	0.000	22.1	8.1	9.7	1.2
		qLTG8	8	RM72-RM22705	4.4	0.014	5.4	4.9	2.4	0.5
		qLTG10.1	10	RM25633	4.2	0.041	4.6	2.9	−1.9	−0.6
		qLTG10.2	10	RM333-RM591	5.4	0.006	6.6	−5.3	5.5	−1.0
		Total					42.0			

[x] LTG: low-temperature germination. [y] R^2: Coefficient of determination. [z] *a*: Additive effect = (*O. rufipogon* homozygote − Hwaseong homozygote)/2, *d*: Dominant effect = Heterozygote − (*O. rufipogon* homozygote + Hwaseong homozygote)/2, *d*/*a*: degree of dominance.

3.3. Interaction between LTG QTLs

Analysis of interactions among QTLs was carried out using the average low-temperature germination rate of nine genotypic classes at 7 DAI in the F_2 population. Based on the general regression model, no significant interaction was detected between five QTLs. Although two QTL, *qLTG3* and *qLTG8* ($p = 0.145$) showed relatively high QTL interaction at 7 DAI, no significant QTL interaction at $p < 0.05$ was observed. In addition, the presence of *O. rufipogon* allele at the *qLTG8* locus increased LTG in three genotype classes at the *qLTG1* and *qLTG3* loci (Figure S3A,D). The highest LTG scores were observed in the genotype classes with *O. rufipogon* homozygous at *qLTG10.1* and Hwaseong homozygous at *qLTG10.2*. The results suggest that the *O. rufipogon* alleles at *qLTG1*, *qLTG3*, *qLTG8*, and *qLTG10.1* and the Hwaseong allele at *qLTG10.2* increased LTG, and QTLs act in an additive manner to regulate LTG.

3.4. QTL Pyramiding Line for LTG

We compared the mean low-temperature germination rates of the F_2 QTL-pyramiding lines (QTL-PL) with those of the parental lines (*O. rufipogon*, Hwaseong, CR1517, and CR1518) (Table S2). QTL-PL represents *O. rufipogon* alleles at four QTLs, *qLTG1*, *qLTG3*, *qLTG8*, and *qLTG10.1*, and Hwaseong allele at *qLTG10.2* locus. QTL-PL had higher germination rates than CR1517, CR1518, and Hwaseong, and lower germination rates than *O. rufipogon* at 6 and 7 DAI under the 13 °C condition. Although an 87.3% germination rate was observed in QTL-PL, this rate was significantly higher than that of CR1517 but not different from CR1518 at 6 DAI. The results indicate that pyramiding the QTLs with the combination of *O. rufipogon* alleles at *qLTG1*, *qLTG3*, and *qLTG8*, *qLTG10.1* and Hwaseong allele at *qLTG10.2* led to robust germinability under low-temperature conditions (13 °C). The lines could be useful material for developing rice varieties with enhanced LTG capacity (Table S2).

3.5. Verification of LTG QTL Using F_3 Population

Two F_2 plants (CR8017-4 and CR8020-6) were selected based on their genotypes to verify *qLTG10.1* and *qLTG10.2*, respectively. CR8017-4 is heterozygous at *qLTG10.1* and *qLTG8*, while *O. rufipogon* homozygous for *qLTG3* and Hwaseong homozygous for *qLTG1* and *qLTG10.2*. CR8020-6 is heterozygous at *qLTG10.2*, while *O. rufipogon* homozygous for *qLTG1* and Hwaseong homozygous for *qLTG3* and *qLTG10.1*. The F_3 plants were grown in the 2019/2020 winter season and dormancy was broken for seeds harvested from each F_3 plant. We used low-temperature germination rates of two F_3 populations at 6 and 7 DAI for the QTL analysis. QTL analysis indicated that RM25633 explained 21.79% of the phenotypic variance in the CR8017-4 population, confirming the presence of *qLTG10.1* (Figure 3A). However, we failed to detect the effect of *qLTG8* ($p = 0.21$). Additional experiments using larger populations are needed to verify the effect of *qLTG8*. QTL analysis indicated that *qLTG10.2* explained 20.75% of the phenotypic variance in the CR8020-6 population (Figure 3B). The Hwaseong allele at *qLTG10.2* contributed to an increase in LTG. The results indicate that two linked QTLs (*qLTG10.1* and *qLTG10.2*) on chromosome 10 are associated with LTG. Because *O. rufipogon* allele at *qLTG10.1* and Hwaseong at *qLTG10.2* contributed to higher LTG, the *O. rufipogon qLTG10.1* could be selectively introgressed into *japonica* rice using MAS. Two QTLs, *qLTG10.1* and *qLTG10.2*, explained 21.79% and 20.75% of the phenotypic variance, respectively, in the near-isogenic background. The low contribution of the QTL is partly due to the small population size and other environmental factors. The F_3 plants were grown in the 2019/2020 winter season. We recorded the heading date of each plant and harvested them 30 days after heading. However, some late-flowering F_3 plants failed to set adequate seeds to be tested for LTG, leading to the small population size and uneven distribution of three genotypic classes in the F_3 population.

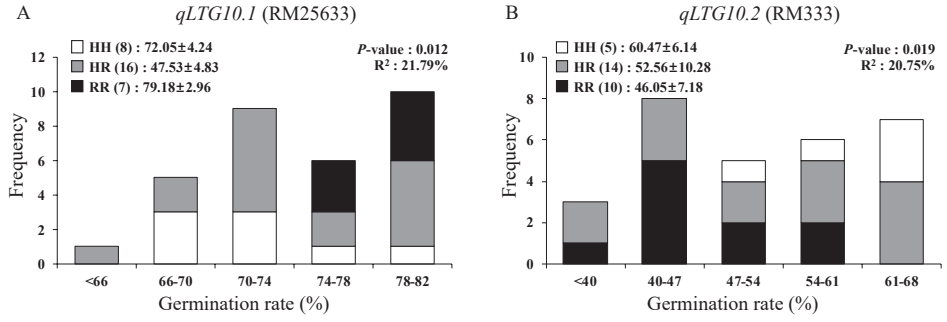


Figure 3. Frequency distribution of LTG at 7 days after incubation in the two F_3 population. Plants in each population are segregating in (**A**) RM25633 for *qLTG10.1* and (**B**) RM333 for *qLTG10.2* in the different genetic background. The numbers in parenthesis indicate plants number followed by mean germination rate ± standard deviation of genotype group. *p*-value and R^2 were determined by one-way ANOVA. HH: Hwaseong homozygoutes, RR: *O. rufipogon* homozygotes, HR: heterozygotes.

3.6. Coleoptile Length in Parental Lines and F_2 Population

O. rufipogon showed significantly higher coleoptile length during 2–3 DAI at optimal condition than Hwaseong, CR1517, and CR1518, while no significant difference was observed in coleoptile length among Hwaseong, CR1517, and CR1518 at 30 °C (Figure S4A). At 1 DAI, mean coleoptile lengths were 0.6 mm and 0.3 mm in *O. rufipogon* and Hwaseong, respectively, while CR1517 and CR1518 were 0.4 mm in length on average. Coleoptile length varied from 12.4 mm to 14.6 mm at 4 DAI among parental lines (Figure S4A).

Coleoptile length varied markedly among parental lines under the 13 °C condition (Figure S4B). The mean length of the coleoptile in *O. rufipogon* was longer than those of Hwaseong, CR1517, and CR1518. The average coleoptile length in *O. rufipogon* at 5 DAI was 1.5 mm, while it reached 9 mm at 10 DAI. In Hwaseong, it was close to zero at 5 DAI and attained 4 mm at 10 DAI. While the average coleoptile length in CR1518 was slightly greater than that in Hwaseong at 5 to 9 DAI, the difference was much greater at 10 DAI. CR1517 had a longer coleoptile than CR1518 and Hwaseong, and a shorter coleoptile than *O. rufipogon*. The differences in coleoptile length between parental lines were more significant at 9 and 10 DAI than at other scoring dates (Figure S4, Table 3).

Table 3. Mean coleoptile lengths (mm) in *O. rufipogon*, Hwaseong, two introgression lines, and F_2 population.

Coleoptile Lengths (mm)						
Trait [x]	**Lines (n = 5** [z]**)**				**F_2 Population**	
	O. rufipogon	**Hwaseong**	**CR1517**	**CR1518**	**Mean**	**Range**
7 DAI	2.6 a [y]	0.7 c	1.6 b	1.1 bc	1.2 ± 0.7	0.2–9.2
8 DAI	4.5 a	1.4 d	2.8 b	2.0 c	3.2 ± 0.5	0.6–9.5
9 DAI	7.0 a	2.8 c	5.0 b	3.5 c	6.6 ± 1.3	2.0–10.0
10 DAI	9.0 a	3.8 d	7.4 b	5.6 c	7.9 ± 1.4	5.0–12.0

[x] DAI: days after incubation. Data are presented as mean germination rate or mean ± standard deviation. [y] Different letters in each row indicate significant difference at $p < 0.05$ based on Tukey's test. [z] n: number of plants.

In the F_2 population, the coleoptile length varied from 0 to 12 mm from 5 to 10 DAI. The mean coleoptile lengths of 149 F_2 plants were 0.4, 0.7, 1.2, and 3.2 mm at 5, 6, 7, and 8 DAI, respectively. At 9 DAI, the average coleoptile length was 6.6 mm, and it reached 7.9 mm at 10 DAI (Table 4, Figure 4). The F_2 population exhibited nearly normally distributed coleoptile elongation at 9 and 10 DAI (Figure 4).

Table 4. List of QTLs identified for coleoptile length under low-temperature condition.

Population	Trait [x]	QTL	Chr.	Markers	*F*-Value	*p*-Value	R^2 (%) [y]	Gene Action [z]		
								a	*d*	*d/a*
F_2	CCL 9 DAI	*qCCL3*	3	qLTG3-1_18D	9.82	0.000	12.0	0.032	0.08	2.6
		qCCL8	8	RM22689-RM22705	4.77	0.001	6.2	0.043	0.02	0.6
	CCL 10 DAI	*qCCL1*	1	RM220-CRM22	4.97	0.008	6.4	0.037	0.02	0.1
		qCCL3	3	qLTG3-1_18D	7.30	0.001	9.2	0.033	0.07	2.3
		qCCL8	8	RM22689-RM22705	5.27	0.006	7.1	0.046	0.03	0.7

[x] CCL: Coleoptile length, [y] Coefficient of determination, [z] *a*: Additive effect = (*O. rufipogon* homozygote-Hwaseong homozygote)/2, *d*: Dominant effect = Heterozygote − (*O. rufipogon* homozygote + Hwaseong homozygote)/2, *d/a*: Degree of dominance.

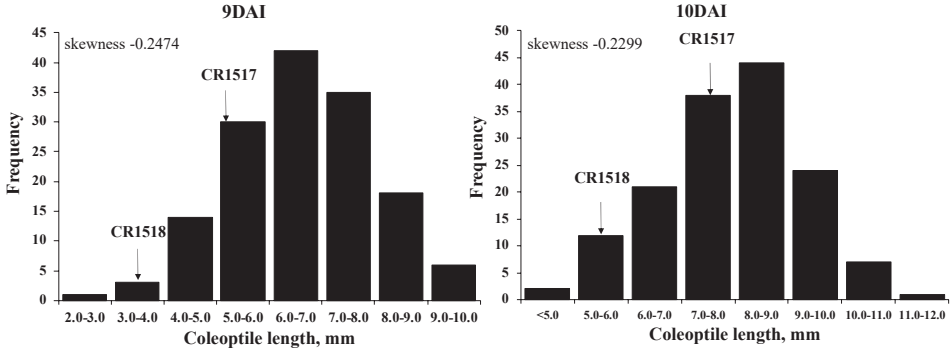


Figure 4. Frequency distribution of coleoptile length in 149 F_2 plants at 13 °C at 9 and 10 DAI. Arrows indicate mean coleoptile length of CR1517 and CR1518 lines.

3.7. QTLs for Coleoptile Length

QTL analysis of coleoptile length was carried out for two scoring dates (9 and 10 DAI) considering that the two parental lines had the greatest difference in coleoptile length (Table 3). A total of three QTLs associated with coleoptile length were detected and they were colocalized with the LTG QTLs (*qLTG1*, *qLTG3*, and *qLTG8*). At 9 DAI, *qCCL3* (qLTG3-1_18D) and *qCCL8* (RM22689-RM22705) were identified and they explained 12.0 and 6.4% of the phenotypic variance, respectively. The *qCCL1* on chromosome 1 between markers RM220 and CRM22 was not significant (p = 0.059) at 9 DAI (data not shown).

At 10 DAI, *qCCL1*, *qCCL3,* and *qCCL8* were detected (Table 4). The QTLs explained 6.4%, 9.2%, and 7.1% of the total phenotypic variance, respectively. No QTL was detected on the other *O. rufipogon* introgression segments on chromosomes 9 and 10 at 9 and 10 DAI. *O. rufipogon* alleles are partially recessive at *qCCL1* and *qCCL8* loci, and dominant at the *qCCL3* locus similar to the case in LTG (Tables 2 and 4). *qLTG8* in this study was colocalized with the *qCCL8* for coleoptile length under cold stress in the study [6] and the allelic relationship between two QTLs remains to be clarified.

4. Discussion

Rice is one of the most important staple food crops globally, being consumed by approximately 50% of the global population [33]. Low-temperature germinability (LTG) is one of the major factors influencing stable crop establishment in the direct seeding method of rice cultivation in tropical and subtropical regions of the world. Along with rapid low-temperature germination, vigorous coleoptile growth is essential in the direct-seeding method of rice when rice seeds are sown in flooded paddy fields and watered with cold irrigation water [14]. Rapid coleoptile elongation after germination is necessary to improve seedling establishment rate [6].

In the present study, we analyzed LTG and coleoptile length in rice cultivated under 13 °C conditions. Using the F_2 population derived from a cross between two introgression lines, CR1517 and CR1518, we detected a total of five and three QTLs for LTG and coleoptile length, respectively, over two scoring dates. QTLs associated with LTG were detected on chromosomes 1, 3, 8, and 10. Among them, two major QTLs, *qLTG1* and *qLTG3*, were detected on chromosomes 1 and 3, respectively. One minor QTL for coleoptile elongation, *qLTG8*, together with *qCCL8*, was detected between RM72 and RM22705 markers on chromosome 8. We failed to detect the QTL using 96 introgression lines in a previous study [27], which could be due to a masking effect of the major QTLs (*qLTG1* and *qLTG3*) and the genetic structure of the population [27]. Several studies have reported the presence of QTLs for LTG and coleoptile elongation on chromosome 8 [5,6,18,34]. For example, Najeeb et al. (2020) detected an LTG QTL, *qLTG(III)8* (497SNP_8_8509144), which is located near *qLTG8* in the present study [34]. The results suggest that the region on chromosome 8 participates in the regulation of germination and coleoptile elongation under low-temperature conditions. Identification of such genes from diverse varieties could enhance our understanding of the roles of the region in germination activities.

Low-temperature germination QTLs are distributed widely throughout the rice genome, including on chromosome 10 [12,17]. We detected two linked QTLs, *qLTG10.1* and *qLTG10.2* for LTG, with an LTG-increasing allele originating from *O. rufipogon* at *qLTG10.1* and Hwaseong at *qLTG10.2*. The results were confirmed in the near-isogenic background in the F_3 population. Notably, the locations of *qLTG10.1* and *qLTG10.2* (20.0~22.9 Mb region) are similar to that of *qGR-10* for low-temperature germination ability detected between C809 and C797 on chromosome 10 (21.0~22.0 Mb region) in a study by Ji et al. (2009) [17]. They mapped *qGR10* using a recombinant inbred line population derived from a cross between Asominori and IR24. At *qGR-10*, the IR24 allele increased germination rate at two scoring dates (8 and 9 DAI), whereas the Asominori allele increased germination rate at later scoring dates (10 and 14 DAI). Similar results have been observed, where the beneficial alleles for LTG originate from two parents, WTR-1 and Haoannong, at two linked loci, $qLTG(I1)_{11}$ and $qLTG(III)_{11}$, on chromosome 11 [34]. Li et al. (2019) also observed that Dongxiang wild rice (*O. rufipogon* Griff.) introgressions at five detected QTLs in 94 BC_1F_7 population delayed germination rates under 15 °C conditions in the background of *indica* variety, Xieqingzao. Among them, two QTLs, *qLTG10-1* and *qLTG10-2*, were identified on chromosome 10 [25]. The results indicate that two linked QTLs potentially act in opposite directions in such QTL regions. Such genetic linkage is common in rice [35]. A linkage between two desirable genes would be advantageous in the selection of improved lines. For example, the tight linkage of two QTLs (*qSPP5* for spikelet no. and *qTGW5* for grain weight) could be valuable for improving rice yield [36]. However, linkage between desirable and undesirable genes is complex in terms of its application in rice breeding [37]. Since two LTG QTLs, *qLTG10.1* and *qLTG10.2*, acting in opposite directions are linked and have minor effects, selection of the high LTG lines with *qLTG10.1* from *O. rufipogon* and *qLTG10.2* could be accomplished using DNA markers. In a previous study, we detected *qLTG10.1* and *not qLTG10.2*, possibly due to the buffering effect of two QTLs and interactions among QTLs in the population [27]. Overall, it appears that the region carries gene(s) with a strong effect on germination performance and represent additional genetic targets for MAS directed development of rice varieties with improved LTG.

The interactions among the four QTLs, *qLTG1*, *qLTG3*, *qLTG8*, *qLTG10.1*, and *qLTG10.2* were examined using general regression models. The plants that harbor the *O. rufipogon* alleles at *qLTG1*, *qLTG3*, *qLTG8*, and *qLTG10.1* exhibited the highest germination rates at 13 °C in the nine groups, and the five QTLs cumulatively explained 42.0% of the phenotypic variance in LTG. The results imply that five QTLs control the LTG in an additive manner. Pyramiding the four QTLs from the *O. rufipogon* into cultivated rice with *qLTG10.2* would facilitate breeding programs aimed at enhancing LTG for direct-seeding production systems. It is also notable that the plants with four *O. rufipogon* alleles at *qLTG1*, *qLTG3*, *qLTG8*, and *qLTG10.1* exhibited lower LTG than *O. rufipogon*, the donor parent at 5–7 DAI, indicating the presence of additional QTLs for LTG in *O. rufipogon*. Further experiments are underway to detect and characterize such unknown QTLs in *O. rufipogon*.

Also, three QTLs for coleoptile length detected on chromosomes 1, 3, and 8 shared their locations with three LTG QTLs, *qLTG1*, *qLTG3*, and *qLTG8,* respectively, and the *O. rufipogon* alleles at all three loci increased the coleoptile length, suggesting a pleiotropy of a single QTL at each locus.

5. Conclusions

In the present study, we performed QTL analysis and identified QTLs for LTG and coleoptile length in the F_2 population. Among the five QTLs for LTG, two major QTLs, *qLTG1* and *qLTG3*, and three minor QTLs were detected on chromosomes 8 and 10. Fine mapping revealed that two QTLs, *qLTG10.1* and *qLTG10.2*, were linked on chromosome 10 and exerted opposite effects with the Hwaseong allele at *qLTG10.2* and the *O. rufipogon* at *qLTG10.1*, respectively, in turn, increasing LTG. Because two LTG QTLs, *qLTG10.1* and *qLTG10.2*, which act in opposite directions are linked, the DNA markers could improve the selection efficiency of the high LTG lines with *qLTG10.1* from *O. rufipogon* and *qLTG10.2* from Hwaseong. Interactions among QTLs were not significant, implying that the QTLs act in an additive manner. NIL plants with combinations of favorable alleles from *O. rufipogon* and Hwaseong exhibited the highest LTG among all groups, supporting the absence of interactions. With regard to coleoptile length, three QTLs observed on chromosomes 1, 3, and 8 were colocalized with QTLs for LTG, suggesting the pleiotropy of the single gene at each locus. According to the results, the introgression of favorable *O. rufipogon* alleles should hasten the breeding of high LTG and coleoptile elongation in *japonica* cultivars.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/10/1200/s1, Figure S1: Measurement of low-temperature germination rate of parental lines (*O. rufipogon*, Hwaseong, CR1517 and CR1518) for 4–8 days after incubation (DAI). Seeds were considered as germinated when coleoptile emerged. Arrows indicate emerged coleoptile, Figure S2: Comparison of germination rates of *O. rufipogon* (Ru), Hwaseong (Hs), CR1517 and CR1518 at normal (30 °C) (A) and low-temperature conditions (13 °C) (B). Different letters at each DAI indicate significant difference at $p < 0.05$ based on Tukey's test. ns: no significant. Error bars indicate standard deviation, Figure S3: Interaction analysis of *qLTG1* and *qLTG3* with other QTL for LTG based on LTG scores of the F_2 population. HH, HR and RR indicate Hwaseong homozygous, heterozygous (Hwaseong/*O. rufipogon*) and *O. rufipogon* homozygous at two QTLs, respectively. Low-temperature germination rates at 7 days after incubation were compared. Error bars indicate standard deviations. Interaction was determined by general linear regression model, Figure S4: Comparison of coleoptile length of parental lines at normal (30 °C) (A) and low-temperature (13 °C) (B) Error bars at each DAI indicate standard deviation. Different letters at each DAI indicate significant difference at $p < 0.05$ based on Tukey's test. ns: no significant. Ru: *O. rufipogon*; Hs: Hwaseong, Table S1: List of primers used in this study, Table S2: Comparison of germination rates among parental lines and pyramiding lines at 6 and 7 DAI.

Author Contributions: M.A. performed experiments and analyzed data. M.A., K.-C.S., C.A., and S.-N.A. conceived the study, designed, and supervised the study. H.-S.L., S.H.K., Y.-AJ., C.A., N.H.L., J.-Y.L. and J.-W.K. investigated the traits. M.A., and K.-C.S. wrote and C.A. and S.-N.A. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01321401)", Rural Development Administration and of Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Golden Seed Project funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (213009-05-3-WT312), Republic of Korea.

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. Brar, D.S.; Khush, G.S. Wild relatives of rice: A valuable genetic resource for genomics and breeding research. In *The Wild Oryza Genomes*; Springer: Cham, Switzerland, 2018; pp. 1–25.
2. Brar, D.; Khush, G. Cytogenetic manipulation and germplasm enhancement of rice (*Oryza sativa* L.). *Genet. Resour. Chromosome Eng. Crop. Improv.* **2006**, *2*, 115–158.
3. Luo, X.D.; Jun, Z.; Dai, L.F.; Zhang, F.T.; Yi, Z.; Yong, W.; Xie, J.K. Linkage map construction and QTL mapping for cold tolerance in *Oryza rufipogon* Griff. at early seedling stage. *J. Integr. Agric.* **2016**, *15*, 2703–2711. [CrossRef]

4. McCouch, S.R.; Sweeney, M.; Li, J.; Jiang, H.; Thomson, M.; Septiningsih, E.; Edwards, J.; Moncada, P.; Xiao, J.; Garris, A. Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica* **2007**, *154*, 317–339. [CrossRef]
5. Wang, Z.F.; Wang, F.H.; Zhou, R.; Wang, J.F.; Zhang, H.S. Identification of quantitative trait loci for cold tolerance during the germination and seedling stages in rice (*Oryza sativa* L.). *Euphytica* **2011**, *181*, 405–413. [CrossRef]
6. Lee, J.; Kwon, S.W. Analysis of quantitative trait loci associated with seed germination and coleoptile length under low temperature condition. *J. Crop Sci. Biotechnol.* **2015**, *18*, 273–278. [CrossRef]
7. Sun, Q.; Wang, J.H.; Sun, B.Q. Advances on seed vigor physiological and genetic mechanisms. *Agric. Sci. China* **2007**, *6*, 1060–1066. [CrossRef]
8. Teng, S.; Zeng, D.L.; Qian, Q.; Yasufumi, K.; Huang, D.I.; Zhu, L.H. QTL analysis of rice low temperature germinability. *Chin. Sci. Bull.* **2001**, *46*, 1800–1804. [CrossRef]
9. Fujino, K.; Sekiguchi, H.; Sato, T.; Kiuchi, H.; Nonoue, Y.; Takeuchi, Y.; Ando, T.; Lin, S.Y.; Yano, M. Mapping of quantitative trait loci controlling low-temperature germinability in rice (*Oryza sativa* L.). *Appl. Genet.* **2004**, *108*, 794–799. [CrossRef]
10. Chen, L.; Lou, Q.; Sun, Z.; Xing, Y.; Yu, X.; Luo, L. QTL mapping of low temperature germinability in rice. *Zhongguo Shuidao Kexue* **2006**, *20*, 159–164.
11. Hou, M.Y.; Wang, C.M.; Jiang, L.; Wan, J.M.; Yasui, H.; Yoshimura, A. Inheritance and QTL mapping of low temperature germinability in rice (*Oryza sativa* L.). *Acta Genet. Sin.* **2004**, *31*, 701–706.
12. Jiang, L.; Liu, S.J.; Hou, M.Y.; Tang, J.Y.; Chen, L.M.; Zhai, H.Q.; Wan, J.M. Analysis of QTLs for seed low temperature germinability and anoxia germinability in rice (*Oryza sativa* L.). *Field Crop. Res.* **2006**, *98*, 68–75. [CrossRef]
13. Satoh, T.; Tezuka, K.; Kawamoto, T.; Matsumoto, S.; Satoh-Nagasawa, N.; Ueda, K.; Sakurai, K.; Watanabe, A.; Takahashi, H.; Akagi, H. Identification of QTLs controlling low-temperature germination of the East European rice (*Oryza sativa* L.) variety Maratteli. *Euphytica* **2016**, *207*, 245–254. [CrossRef]
14. Adachi, Y.; Sugiyama, M.; Sakagami, J.I.; Fukuda, A.; Ohe, M.; Watanabe, H. Seed Germination and Coleoptile Growth of New Rice Lines Adapted to Hypoxic Conditions. *Plant Prod. Sci.* **2015**, *18*, 471–475. [CrossRef]
15. Inada, N.; Sakai, A.; Kuroiwa, H.; Kuroiwa, T. Three-dimensional progression of programmed death in the rice coleoptile. *Int. Rev. Cytol.* **2002**, *218*, 221–258.
16. Miura, K.; Lin, S.Y.; Yano, M.; Nagamine, T. Mapping quantitative trait loci controlling low temperature germinability in rice (*Oryza sativa* L.). *Breed. Sci.* **2001**, *51*, 293–299. [CrossRef]
17. Ji, S.L.; Jiang, L.; Wang, Y.H.; Zhang, W.W.; Liu, X.; Liu, S.J.; Chen, L.M.; Zhai, H.Q.; Wan, J.M. Quantitative trait loci mapping and stability for low temperature germination ability of rice. *Plant Breed.* **2009**, *128*, 387–392. [CrossRef]
18. Fujino, K.; Obara, M.; Shimizu, T.; Koyanagi, K.O.; Ikegaya, T. Genome-wide association mapping focusing on a rice population derived from rice breeding programs in a region. *Breed. Sci.* **2015**, *65*, 403–410. [CrossRef]
19. Ji, S.L.; Ling, J.; Wang, Y.H.; Liu, S.J.; Liu, X.; Zhai, H.Q.; Yoshimura, A.; Wan, J.M. QTL and epistasis for low temperature germinability in rice. *Acta Agron. Sin.* **2008**, *34*, 551–556.
20. Wang, X.; Zou, B.; Shao, Q.; Cui, Y.; Lu, S.; Zhang, Y.; Huang, Q.; Huang, J.; Hua, J. Natural variation reveals that *OsSAP16* controls low-temperature germination in rice. *J. Exp. Bot.* **2018**, *69*, 413–421. [CrossRef]
21. Xie, L.X.; Tan, Z.W.; Zhou, Y.; Xu, R.B.; Feng, L.B.; Xing, Y.Z.; Qi, X.Q. Identification and fine mapping of quantitative trait loci for seed vigor in germination and seedling establishment in rice. *J. Integr. Plant Biol.* **2014**, *56*, 749–759. [CrossRef]
22. Li, L.F.; Liu, X.; Xie, K.; Wang, Y.H.; Liu, F.; Lin, Q.Y.; Wang, W.Y.; Yang, C.Y.; Lu, B.Y.; Liu, S.J.; et al. *qLTG-9*, a stable quantitative trait locus for low-temperature germination in rice (*Oryza sativa* L.). *Appl. Genet.* **2013**, *126*, 2313–2322. [CrossRef] [PubMed]
23. Fukuda, A.; Kataoka, T.; Shiratsuchi, H.; Fukushima, A.; Yamaguchi, H.; Mochida, H.; Ogiwara, H. QTLs for Seedling Growth of Direct Seeded Rice under Submerged and Low Temperature Conditions. *Plant Prod. Sci.* **2014**, *17*, 41–46. [CrossRef]
24. Shakiba, E.; Edwards, J.D.; Jodari, F.; Duke, S.E.; Baldo, A.M.; Korniliev, P.; McCouch, S.R.; Eizenga, G.C. Genetic architecture of cold tolerance in rice (*Oryza sativa*) determined through high resolution genome-wide analysis. *PLoS ONE* **2017**, *12*, e0172133. [CrossRef] [PubMed]

25. Li, L.Y.; Chen, H.P.; Mao, D.H. Pyramiding of rapid germination loci from *Oryza Sativa* cultivar 'Xieqingzao B' and cold tolerance loci from Dongxiang wild rice to increase climate resilience of cultivated rice. *Mol. Breed.* **2019**, *39*, 85. [CrossRef]
26. Dametto, A.; Sperotto, R.A.; Adamski, J.M.; Blasi, E.A.R.; Cargnelutti, D.; de Oliveira, L.F.V.; Ricachenevsky, F.K.; Fregonezi, J.N.; Mariath, J.E.A.; da Cruz, R.P.; et al. Cold tolerance in rice germinating seeds revealed by deep RNAseq analysis of contrasting *indica* genotypes. *Plant Sci.* **2015**, *238*, 1–12. [CrossRef] [PubMed]
27. Nguyen, H.N.; Park, I.K.; Yeo, S.M.; Yun, Y.T.; Ahn, S.N. Mapping quantitative trait loci controlling low-temperature germinability in rice. *Korean J. Agric. Sci.* **2012**, *39*, 477–482. [CrossRef]
28. Shim, K.C.; Kim, S.; Le, A.Q.; Lee, H.S.; Adeva, C.; Jeon, Y.A.; Luong, N.H.; Kim, W.J.; Akhtamov, M.; Ahn, S.N. Fine mapping of a low-temperature Germinability QTL *qLTG1* using introgression lines derived from *Oryza rufipogon*. *Plant Breed. Biotechnol.* **2019**, *7*, 141–150. [CrossRef]
29. Shim, K.C.; Kim, S.H.; Lee, H.S.; Adeva, C.; Jeon, Y.A.; Luong, N.H.; Kim, W.J.; Akhtamov, M.; Park, Y.J.; Ahn, S.N. Characterization of a New *qLTG3-1* Allele for Low-temperature Germinability in Rice from the Wild Species *Oryza rufipogon*. *Rice* **2020**, *13*, 10. [CrossRef]
30. Causse, M.A.; Fulton, T.M.; Cho, Y.G.; Ahn, S.N.; Chunwongse, J.; Wu, K.S.; Xiao, J.H.; Yu, Z.H.; Ronald, P.C.; Harrington, S.E.; et al. Saturated Molecular Map of the Rice Genome Based on an Interspecific Backcross Population. *Genetics* **1994**, *138*, 1251–1274.
31. Panaud, O.; Chen, X.; McCouch, S.R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L). *Mol. Gen. Genet.* **1996**, *252*, 597–607. [CrossRef]
32. McCouch, S.R.; Cooperative, R.G. Gene Nomenclature System for Rice. *Rice* **2008**, *1*, 72–84. [CrossRef]
33. Shinada, H.; Iwata, N.; Sato, T.; Fujino, K. QTL pyramiding for improving of cold tolerance at fertilization stage in rice. *Breed. Sci.* **2014**, *63*, 483–488. [CrossRef]
34. Najeeb, S.; Ali, J.; Mahender, A.; Pang, Y.L.; Zilhas, J.; Murugaiyan, V.; Vemireddy, L.R.; Li, Z. Identification of main-effect quantitative trait loci (QTLs) for low-temperature stress tolerance germination- and early seedling vigor-related traits in rice (*Oryza sativa* L.). *Mol. Breed.* **2020**, *40*, 10. [CrossRef]
35. Xiao, J.; Li, J.; Grandillo, S.; Ahn, S.N.; Yuan, L.; Tanksley, S.D.; McCouch, S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* **1998**, *150*, 899–909. [PubMed]
36. Luo, X.; Ji, S.-D.; Yuan, P.-R.; Lee, H.-S.; Kim, D.-M.; Balkunde, S.; Kang, J.-W.; Ahn, S.-N. QTL mapping reveals a tight linkage between QTLs for grain weight and panicle spikelet number in rice. *Rice* **2013**, *6*, 33. [CrossRef] [PubMed]
37. Fukuoka, S.; Saka, N.; Koga, H.; Ono, K.; Shimizu, T.; Ebana, K.; Hayashi, N.; Takahashi, A.; Hirochika, H.; Okuno, K. Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* **2009**, *325*, 998–1001. [CrossRef] [PubMed]

genes

MDPI

Article

Study of Quantitative Trait Loci (QTLs) Associated with Allelopathic Trait in Rice

Ill-Min Chung [1], Tae-Ho Ham [1], Gi-Won Cho [2], Soon-Wook Kwon [3], Yoonjung Lee [1], Jeonghwan Seo [4], Yeon-Ju An [1], So-Yeon Kim [1], Seung-Hyun Kim [1] and Joohyun Lee [1,*]

1 Department of Crop Science, Konkuk University, Seoul 05029, Korea; imcim@konkuk.ac.kr (I.-M.C.); lion78@daum.net (T.-H.H.); yoon10.lee@gmail.com (Y.L.); ayj3043@konkko.ac.kr (Y.-J.A.); hellosys1@konkuk.ac.kr (S.-Y.K.); kshkim@konkuk.ac.kr (S.-H.K.)
2 Seed Development Team, Agro Division, Orion Corporation, Gangwon 25323, Korea; gwcho2669@orionworld.com
3 Department of Plant Bioscience, Pusan National University, Milyang 50463, Korea; swkwon@pusan.ac.kr
4 Department of Plant Science, Research Institute for Agriculture and Life Sciences and Plant Genomics and Breeding, Seoul National University, Seoul 08826, Korea; jhseo83@snu.ac.kr
* Correspondence: edmund@konkuk.ac.kr

Received: 23 March 2020; Accepted: 23 April 2020; Published: 26 April 2020

Abstract: In rice there are few genetic studies reported for allelopathy traits, which signify the ability of plants to inhibit or stimulate growth of other plants in the environment, by exuding chemicals. QTL analysis for allelopathic traits were conducted with 98 F8 RILs developed from the cross between the high allelopathic parents of 'Sathi' and non-allelopathic parents of 'Nong-an'. The performance of allelopathic traits were evaluated with inhibition rate on root length, shoot length, total length, root weight, shoot weight, and total weight of lettuce as a receiver plant. With 785 polymorphic DNA markers, we constructed a linkage map showing a total of 2489.75 cM genetic length and 3.17 cM of average genetic distance between each adjacent marker. QTL analysis detected on QTL regions on chromosome 8 responsible for the inhibition of shoot length and inhibition of total length. The qISL-8 explained 20.38% of the phenotypic variation for the inhibition on the shoot length. The *qITL-8* explained 14.93% of the phenotypic variation for the inhibition on total length. The physical distance of the detected QTL region was 194 Kbp where 31 genes are located.

Keywords: rice; allelopathy; QTL

1. Introduction

Crops are susceptible to attacks from pests and diseases and competition from weeds, which lead to considerable yield losses. Unlike pests and pathogens, which generally invade the crop from external sources, weeds are active within the same cultivation area as the crop and compete for light, nutrients, and water. In the most severe situations, weed competition can lead to complete loss of the crop [1]. Cultivation practices often exacerbate weed germination and development, and weed control is thus essential in most cropping situations. Substantial economic resources are used by farmers for weed-control practices, such as herbicide application [2,3]. Use of allelopathy may allow weeds to be managed in a more cost-effective manner.

Allelopathy is the ability of plants to inhibit or stimulate growth of plants in the neighboring environment through the activity of exuded bioactive secondary metabolites referred to as allelochemicals [4]. Allelopathic potential in rice was found to be proportional to the amount and type of phytotoxic compounds produced, including phenolic acids and momilactones [5]. Allelochemicals can elicit a wide range of effects, including changes to plant membrane permeability that impact nutrient absorption; suppression of metabolic activities, such as photosynthesis, respiration, and diverse

enzyme functions; and disruption of growth and development through inhibition of cell division and elongation and alterations to submicroscopic structures [6].

Plants with allelopathic potential are termed donor plants, whereas plants affected by the allelopathic compounds from the donor plant are referred to as receiver plants. Donor and receiver plants can affect one another through both allelopathy and competition. The combined effect of these two interactions is termed interference [7]. Allelopathic interactions are complex, and it is difficult to exclude the effects of competition and the environment when assessing allelopathic potential. Although field-based screening is an important component of plant-breeding programs, it is almost impossible to distinguish allelopathic potentials from competition under natural field conditions. A range of plant species can be used as receivers in bioassays, to assess allelopathic activity [8] and several considerations, such as susceptibility and genetic uniformity, are important when selecting a receiver species for testing allelopathic potential. Some standard species, such as lettuce (Lactuca sativa), radish (Raphanus sativa), and duckweed (Lemna minor), are recommended for preliminary testing because of their availability and high sensitivity to allelopathic actions [9].

Although many agronomic traits in rice have been studied at the genetic level, relatively few studies have been conducted on the genetic basis of allelopathy. Most published studies to date have involved screening and evaluation of existing plant materials. One quantitative trait loci (QTL) study examined the allelopathic potential of 150 recombinant inbred lines (RILs) generated from a cross between 'AC1423', a highly allelopathic rice cultivar, and a minimally allelopathic line, 'Aus196'. Allelopathic potential was evaluated against the vigorous weed species Echinochloa crus-galli (L.) in laboratory and greenhouse conditions. QTLs linked to allelopathic traits were found on chromosomes 3–10 and 12 [10]. A separate QTL study examined the allelopathic potential of an F2 population derived from 'PI312777', a highly allelopathic indica cultivar, and 'Rexmont', a minimally allelopathic japonica cultivar. Water-soluble extracts from the F2 seedlings were supplied to lettuce seedlings as receiver plants, and candidate allelopathic QTLs were identified on chromosomes 1, 3, 5, 6, 7, 11, and 12 [11].

Recent advances in molecular breeding technologies, such as the development of high-density DNA markers, DNA chips, and next-generation sequencing (NGS), have facilitated the identification and characterization of many genes associated with quantitative traits. However, genetic studies to uncover the basis of allelopathic traits in rice remain in their infancy. This is due to the complications arising from distinguishing allelopathic potential from the effects of competition and the environment, which necessitate performing large numbers of replicated studies in field and greenhouse conditions, as well as in the laboratory. In this study, QTL analysis to identify candidate regions associated with allelopathic traits in rice was conducted in controlled laboratory conditions.

2. Materials and Methods

2.1. Allelopathic Assay

Allelopathic potential was assessed by using the equal compartment agar method (ECAM) [12], with minor modifications. A total of 98 F8 RILs were produced by single-seed descent from a cross between 'Sathi', an indica cultivar with high allelopathic potential, and 'Nong-an', a non-allelopathic Tong-il cultivar [13,14]. Genetically uniform cultivated lettuce 'Yeolpungjeokchima' (Lactuca sativa, cv.)—lettuce exhibits high sensitivity to low concentrations of allelopathic chemicals [4]—was used as a receiver species. For ECAM, dehulled rice seeds were sterilized for 15 min, using 2% sodium hypochlorite to prevent fungal and bacterial contamination. Seeds were then rinsed seven times with sterilized distilled water before placing on Whatman No. 5 filter paper (Whatman, Maidstone, England) in a Petri dish (SPL life Sciences, Pocheon, Korea) with 7 mL of sterilized distilled water. Rice and lettuce seeds were germinated in a controlled environment chamber for 3 days, at 28 °C, in the dark. Six germinated rice seeds were transplanted into one side of a Magenta box (SPL Life Sciences, Pocheon, Korea) filled with 30 mL of 0.3% nutrient-free water agar. After 1 week, 6 lettuce seeds were transplanted into the other side of the magenta box (Figure 1). The allelopathic

potential of parent rice lines 'Nong-an' and 'Sathi' and 98 derived F8 RILs were evaluated with ECAM, with lettuce as a receiver plant. Lettuce parameters (root length, root weight, shoot length, shoot weight, total length from base of root to apex of shoot, and total weight) were determined after 1-week co-incubation with rice seedlings. Control lettuce plants were grown in the absence of rice, under the same cultivation conditions. Inhibition rates in lettuce (%) were calculated as follows: [(control − treatment)/control] × 100. Two independent experiments were conducted for each RIL, and thus the average from the total of 12 lettuce plants was used to evaluate each RIL.

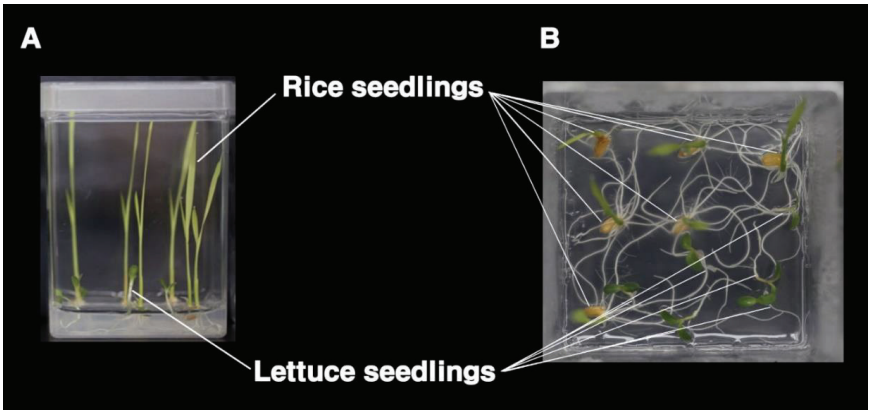


Figure 1. Evacuating method for allelopathic potentials of rice against lettuces seedlings: (**A**) side view of the magenta box where rice seedlings and lettuce seedlings are grown together; (**B**) top view of the magenta box, where rice seedlings and lettuce seedlings are grown together.

2.2. Rice DNA Extraction and High-Throughput SNP Genotyping

DNA was extracted from the parent rice cultivars 'Nong-an' and 'Sathi' and the 98 F8 RILs, using the CTAB method [15]. The extracted DNA was assessed, using a NanoDrop 2000c spectrophotometer with 230, 260, and 280 nm (Thermo Fisher Scientific, Waltham, MA, USA).

All equipment and resources required for the Axiom 2.0 Assay (Thermo Fisher Scientific, Waltham, MA, USA) with automated target preparation were from the Axiom® 2.0 Assay Automated Workflow (Thermo Fisher Scientific, Waltham, MA, USA). Using the Axiom® 2.0 Reagent Kit, ~200 ng of genomic DNA was amplified and randomly fragmented into 25 to 125 bp fragments and then purified and resuspended. The fragments whose size was confirmed from 25 to 125 bp were denatured and transferred to the hybridization tray in the part of GeneTitan® MC Instrument (Thermo Fisher Scientific, Waltham, MA, USA). The hybridization step followed the GeneTitan® Multichannel Instrument User's Manual, using the KNU Axiom Oryza 580K Genotyping Array [16] After ligation, the arrays were stained and imaged on the GeneTitan MC Instrument. The image was then analyzed by using Affymetrix® GeneChip® Command Console® Software (Thermo Fisher Scientific, Waltham, MA, USA). Genotype calls were conducted, using Affymetrix-power-tools. BRLMM-P algorithm was applied, which is a model-based approach which performs 1-dimensional clustering by fitting a Gaussian mixture model (BRLMM-P: a Genotype Calling Method for the SNP 5.0 Array [17]). The part of the KNU Axiom Oryza 580K Genotyping Array, the PolyHighResolution chip, which is comprised of 247,578 SNP markers, was used for high-throughput SNP genotyping. The 247,578 SNP markers were designed from the genomic data of 3494 accessions including the Korean rice core set version 2 (KRICE_CORE v2), including wild rice and 3K IRRI world collections [18].

2.3. QTL Analysis

A linkage map of the 'Nong-an'/'Sathi' 98 RIL population comprising 785 SNP markers that were polymorphic between parents was constructed by QTL IciMapping 4.1 [19]. For mapping, Kosambi's function was used, and the Anchor filter option was applied. QTL analysis was conducted by QTL IciMapping 4.1, using inclusive composite interval mapping of additive and epistatic QTLs. Significant LOD threshold value was calculated for each QTL, using 1000 times permutations at $p = 0.05$. The additive effect and phenotypic variation explained by each QTL for allelopathy traits were consequently calculated during QTL analysis, using QTL IciMapping 4.1.

3. Results

3.1. Screening of Allelopathy in Rice

Allelopathic potentials for each RIL were thus represented by the inhibition rates of the six traits. Clear inhibition was detected for all six traits (Figures 2 and 3). Allelopathic differences between the parent cultivars were more apparent for height traits than weight traits. For root length, 'Sathi' elicited 62% inhibition, whereas 'Nong-an' elicited only 17% inhibition. For shoot length, 'Sathi' elicited a 57% inhibition rate compared to 30% with 'Nong-an'. Overall, total length inhibition was 62% with 'Sathi' and 22% with 'Nong-an'. The highly allelopathic 'Sathi' cultivar inhibited both root and shoot length by approximately 60%. By contrast, the non-allelopathic 'Nong-an' cultivar elicited stronger inhibition of shoot length (30%) than root length (17%). For root weight, 'Sathi' elicited 40% inhibition and 'Nong-an' elicited 32% inhibition. For shoot weight, 'Sathi' elicited 60% inhibition compared with 42% with 'Nong-an'. Overall, total weight inhibition was 58% with 'Sathi' and 41% with 'Nong-an'. Although differences in inhibition between the two parent cultivars were relatively small, RILs exhibited large variations in inhibition for all six traits. Transgressive segregation, where RILs exceeded parental phenotypes, was also observed (Figure 3).

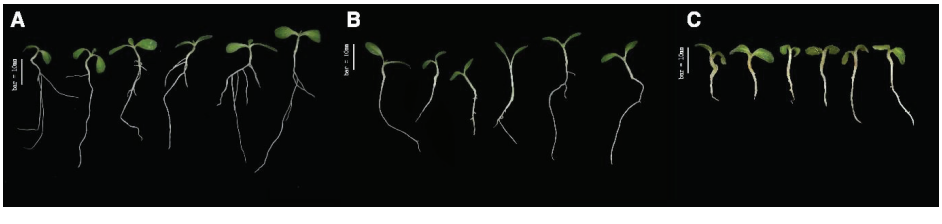


Figure 2. Allelopathic effect on growth inhibition of lettuce: (**A**) lettuces cultivated without rice control; (**B**) lettuces cultivated with rice 'Nong-an'; and (**C**) lettuces cultivated with rice 'Sathi'.

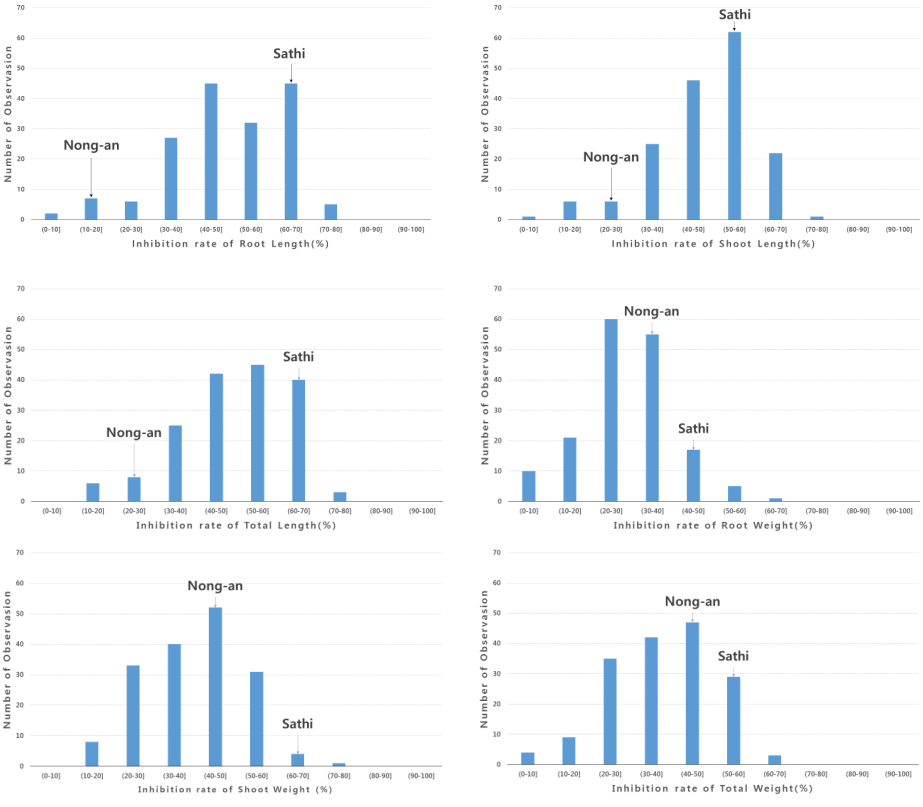


Figure 3. Distribution of growth inhibition rate. (**A**) Inhibition rate of root length, (**B**) inhibition rate of shoot length, (**C**) inhibition rate of total length, (**D**) inhibition rate of root weight, (**E**) inhibition rate of shoot weight, and (**F**) inhibition rate of total weight.

3.2. High-Throughput SNP Genotyping and QTL Analysis

For QTL analysis, 785 polymorphic markers between 'Nong-an' and 'Sathi' were used in constructing a linkage map with 98 RILs (Supplementary Materials Table S1). Among the 247,578 SNP markers used, monomorphic or low-quality markers were eliminated to generate genotypes for 110,770 markers. Additionally, for the generated genotypes for 110,770 markers, binning was carried out for each chromosome, using BIN functionality QTL IciMapping 4.1 to select 2654 markers. Further filtering based on the physical distance (~ 400 kb) was conducted to generate the final 785 markers for the QTL analysis. The markers on the genetic map presented the same ordering presented in the physical map from which the SNP markers were selected, based on the database of IRGSP 1.0. linkage map construct using 98 RIL individuals and 785 markers. Overall, 785 markers were distributed among all 12 rice chromosomes with an average of one marker per 450 kb (Supplementary Materials Figure S1). The number of markers varied from 38 (chromosome 11) to 89 (chromosome 8) with an average of 65.4 per chromosome. The longest length was found in chromosome 1, with 249.5 cM, followed by chromosome 3 with 248.8 cM and by chromosome 12 with 213.2 cM. The linkage map was 2489.75 cM in total length, with the average distance between markers being 3.17 cM. Two main effect-additive QTLs for the allelopathic traits on one chromosome region were detected. One QTL was *qISL-8* (inhibition rate of shoot length), and the other was *qITL-8* (inhibition rate of total length), on chromosome 8, 176.3 cM ~ 177.3 cM (Figure 4). The *qISL-8* showed

an LOD value of 3.38, which explained 20.83% phenotypic variance. The qITL-8 showed an LOD value of 3.24, which explained 14.94% phenotypic variance (Table 1).

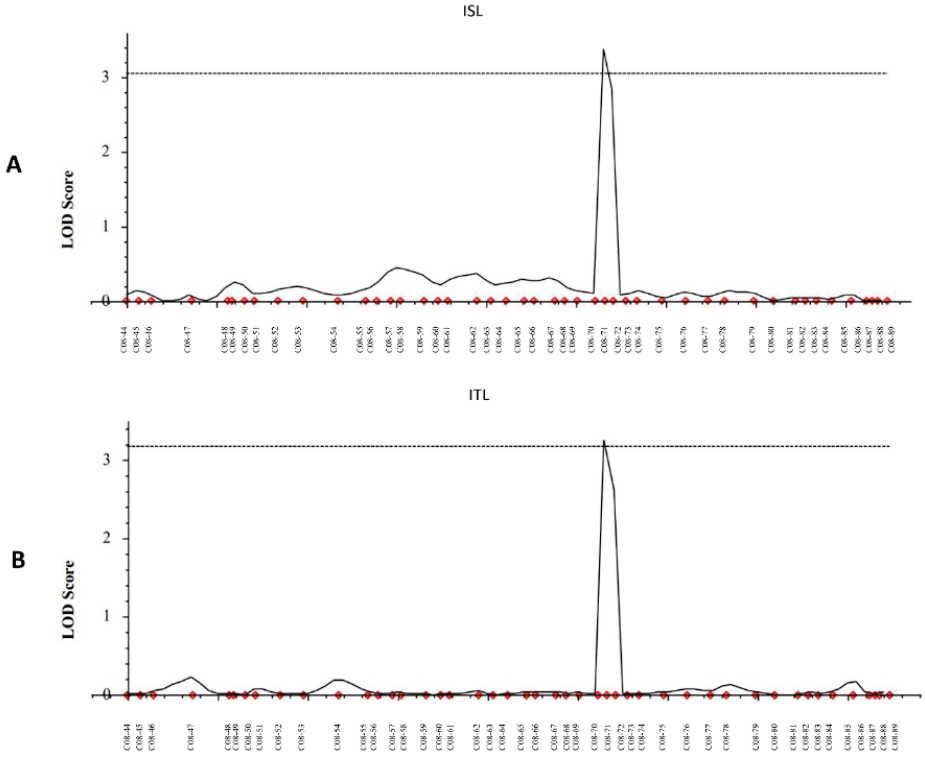


Figure 4. Detected QTLs (qISL and qITL) region associated with inhibition of growth on chromosome 8: (**A**) the LOD score of DNA markers associate with ISL (inhibition rate of shoot length) and position of qISL-8 on chromosome 8; (**B**) the LOD score of DNA markers associate with ITL (inhibition rate of total length) and position of qITL-8 on chromosome 8. Abbreviations are as follows: ISL—inhibition rate of shoot length; ITL—inhibition rate of total length.

Table 1. Quantitative trait loci (QTLs) of allelopathy analyzed by inclusive composite interval mapping.

QTL	Trait	Chromosome	Position (cM)	Left Marker	Right Marker	LOD	PVE (%)	Add	Left CI	Right CI
qISL-8	ISL	8	177	C08-70	C08-71	3.3848	20.8345	−5.29	176.3	177.3
qITL-8	ITL	8	177	C08-70	C08-71	3.2409	14.9362	−5.02	176.3	177.3

Abbreviations are as follows: ISL—inhibition rate of shoot length; ITL— inhibition rate of total length; PVE (%)—phenotypic variance explained; Add—estimated additive effect of Nong—an allele for the QTL; CI—confidence interval calculated by one LOD drop from the estimated QTL position.

The phenotype differences were compared between 'Nong-an' allele and 'Sathi' allele on SNP markers C08-70 and C08-71. For the traits of ISL and ITL, the result of the t-test showed significant differences between homozygous alleles of 'Nong-an' and 'Sathi' (Figure 5).

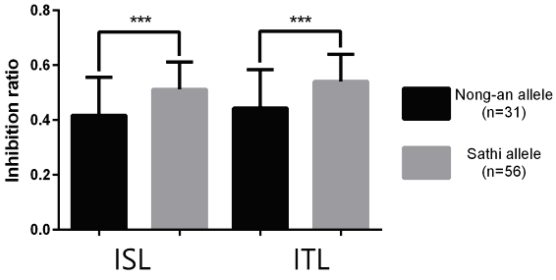


Figure 5. Comparison of inhibition ratio between 'Nong-an' and 'Sathi' alleles on SNP markers C08-70 and C08-71 for ISL (inhibition rate of shoot length) and ITL (inhibition rate of total length). Significance was determined by *t*-test. *** Indicates significance in 0.001 probability level.

Two digenic epistatic QTLs for inhibition rate of shoot weight (ISW) and inhibition rate of total weight (ITW) were also identified on identical genomic regions of chromosomes 1 and 8. The two interacting regions for two digenic epistatic QTLs were located in the interval between markers C01-75 and C01-76 on chromosome 1 and the interval between markers C08-42 and C08-43 on chromosome 8. The two digenic epistatic QTLs showed similar phenotypic variance explained (PVE), which are 23.97% for ISW and 23.29% for ITW. Furthermore, additive-by-additive effects of two digenic epistatic QTLs were −7.06 for ISW and −7.31 for ITW (Table 2).

Table 2. Digenic epistatic QTLs of allelopathy analyzed by inclusive composite interval mapping.

Trait	Chr.	Left Marker	Right Marker	Chr.	Left Marker	Right Marker	LOD	PVE (%)	Add1	Add2	Add by Add
ISW	1	C01-74	C01-75	8	C08-42	C08-43	5.3819	23.9703	−2.39	−3.99	−7.06
ITW	1	C01-74	C01-75	8	C08-42	C08-43	5.2872	23.2899	−2.98	−4.48	−7.31

Abbreviations are as follows: Chr—chromosome; ISW—inhibition rate of shoot weight; ITW—inhibition rate of total weight; PVE(%)—phenotypic variance explained; Add1—estimated additive effect of first QTL; Add2—estimated additive effect of second QTL; Add by Add—additive by additive epistatic effect of QTL at the two scanning points.

4. Discussion

The use of allelopathy for weed control has great potential as a biological control method. Despite this, few genetic studies have examined allelopathy [20] due to the complex challenge of accurately assessing allelopathic interactions in field situations in the presence of natural variability and changing environmental conditions. In this study, an analysis of rice RILs was used to identify QTLs contributing to allelopathic interactions with lettuce, a susceptible receiver plant, in controlled conditions. Six traits were examined, all of which were inhibited in the receiver plant. Inhibition by the different RILs varied widely, and transgressive segregation was observed. Although the cultivation media and growth chamber conditions were well controlled, the conditions in which the receiver plants were cultivated alongside the RILs may produce an inhibitory environment for lettuce growth. Greater inhibition was observed in the non-allelopathic parent ('Nong-an') than in lettuce plants grown in the absence of rice. One possibility is that 'Nong-an' may exhibit low allelopathy, rather than being strictly non-allelopathic [10]. Alternatively, the receiver plant inhibition may have been at least partly due to competition for space. Although it was not possible to completely eliminate the effects of competition in our experimental design, the inhibitory effect of the highly allelopathic parent 'Sathi' was readily apparent, particularly on receiver plant root growth.

Even though we evaluate the allelopathic response with inhibition of weight and length, for the evaluated traits, two main effect additive QTLs for inhibition of root length and inhibition of total length were identified. The locations of both QTLs were identical on the chromosome 8. Because the total length includes the shoot length, this QTL region is mainly for the inhibition of shoot length.

The level of explained phenotypic variation for qISL-8 on chromosome 8 was 20.83%. Other previous reports that showed relatively low phenotypic variation explain the value of individual QTL ranging from 5.0% to 11.1%, in general [11]. This relatively low phenotypic variation of values for the individual QTLs is part of the general nature of the allelopathic trait, representing the difficulty in measuring the allelopathic trait at the individual genotype level. In this study, qISL-8 showed a relatively high value of 20.83%, suggesting it is a possible candidate for further study for cloning genes for the allelopathy. The physical distance of the detected QTL region was 194 Kbp where 31 genes are located (Supplementary Materials Table S2). Among them, 12 genes are unknown or hypothetical proteins, and other proteins were reported to be related with auxin response, dehydration, protein kinase, zinc finger protein, chaperone, peroxidase, and isoamylase. Further study for these candidate genes will be conducted, and the development of near-isogenic lines with each QTL are undergoing. The two digenic epistatic QTLs were detected on identical genomic regions for inhibition rate of shoot length (ISL) and inhibition of total weight (ITW). In addition, PVE and the effect of both digenic epistatic QTLs showed ~ 23%. This is possibly the closest related trait between the shoot weight trait and total weight.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/5/470/s1, Table S1: The 785 SNP markers genotype for 98 RILs. Table S2: The candidate 31 genes located in the detected QTL region on chromosome 8. Figure S1: Linkage map for 12 chromosomes.

Author Contributions: Formal analysis, G.-W.C. and Y.L.; investigation, T.-H.H. and G.-W.C.; methodology, I.-M.C. and S.-W.K.; project administration, I.-M.C. and J.L.; resources, I.-M.C. and J.S.; software, Y.L.; supervision, I.-M.C., S.-W.K., J.S., and J.L.; validation, T.-H.H., S.-W.K., and J.L.; visualization, T.-H.H. and J.S.; writing–original draft, T.-H.H.; writing–review and editing, T.-H.H., Y.-J.A., S.-Y.K., S.-H.K. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2017R1E1A1A01075282) and Konkuk University Researcher Fund in 2018.

Acknowledgments: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2017R1E1A1A01075282) and Konkuk University Researcher Fund in 2018.

Conflicts of Interest: All of the authors declare that they have no conflicts of interest in this publication.

Availability of Data and Materials: The datasets used and/or analyzed during the current study are available from the corresponding author, upon reasonable request.

References

1. Mamun, A. Weeds and their control: A review of weed research in Bangladesh. Agricultural and Rural Development in Bangladesh. Japan Intl. Co-operation Agency, Dhaka, Bangladesh. *JSARD* **1990**, *19*, 45–72.
2. Asaduzzaman, M.; Islam, M.M.; Sultana, S. Allelopathy and allelochemicals in rice weed management. *Bangladesh Res. Publ.* **2010**, *4*, 1–14.
3. Khanh, T. Role of allelochemicals for weed management in rice. *Allelopathy J.* **2007**, *19*, 85–95.
4. Khanh, T.; Xuan, T.; Chung, I. Rice allelopathy and the possibility for weed management. *Ann. Appl. Boil.* **2007**, *151*, 325–339. [CrossRef]
5. Rimando, A.M.; Olofsdotter, M.; Dayan, F.E.; Duke, S.O. Searching for Rice Allelochemicals. *Agron. J.* **2001**, *93*, 16–20. [CrossRef]
6. Li, Z.-H.; Wang, Q.; Ruan, X.; Pan, C.-D.; De-An, J. Phenolics and Plant Allelopathy. *Molecules* **2010**, *15*, 8933–8952. [CrossRef] [PubMed]
7. Wu, H.; Pratley, J.; Lemerle, D.; Haig, T.; An, M. Screening methods for the evaluation of crop allelopathic potential. *Bot. Rev.* **2001**, *67*, 403–415. [CrossRef]
8. Shilling, D.G.; Yoshikawa, F. A Rapid Seedling Bioassay for the Study of Allelopathy. In *Chemistry Student Success: A Field-Tested, Evidence-Based Guide*; American Chemical Society (ACS): Washington, DC, USA, 1987; Volume 330, pp. 334–342.
9. Putnam, A.R.; DeFrank, J.; Barnes, J.P. Exploitation of allelopathy for weed control in annual and perennial cropping systems. *J. Chem. Ecol.* **1983**, *9*, 1001–1010. [CrossRef] [PubMed]

10. Jensen, L.B.; Courtois, B.; Olofsdotter, M. Quantitative Trait Loci Analysis of Allelopathy in Rice. *Crop Sci.* **2008**, *48*, 1459–1469. [CrossRef]
11. Ebana, K.; Yan, W.; Dilday, R.H.; Namai, H.; Okuno, K. Analysis of QTL Associated with the Allelopathic Effect of Rice Using Water-soluble Extracts. *Breed. Sci.* **2001**, *51*, 47–51. [CrossRef]
12. Wu, H.; Pratley, J.; Lemerle, D.; Haig, T. Laboratory screening for allelopathic potential of wheat (Triticum aestivum) accessions against annual ryegrass (Lolium rigidum). *Aust. J. Agric. Res.* **2000**, *51*, 259. [CrossRef]
13. Junaedi, A.; Jung, W.S.; Chung, I.M.; Kim, K.H. Differentially expressed genes of potentially allelopathic rice in response against barnyardgrass. *J. Crop Sci. Biotech.* **2007**, *10*, 231–236.
14. Junaedi, A.; Chozin, M.A.; Lontoh, A.P.; Salam, M.D.A.; Lee, S.B.; Kim, K.H.; Kim, E.H.; Chung, I.M. Promising allelopathic rice lines derived from the recombinant inbred lines of Nongan/sathi cross. *Allelopathy J.* **2012**, *30*, 177–188.
15. Murray, M.; Thompson, W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [CrossRef] [PubMed]
16. Kim, K.-W.; Nam, J.; Chu, S.-H.; Phitaktansakul, R.; Yoo, J.-M.; Kang, J.-S.; Min, M.-H.; Cheng, L.; Cao, Y.; Aung, K.M.; et al. Development of KNU Axiom Oryza 580K Genotyping Array. In Proceedings of the KSBS & SABRAO International Conference on Plant Breeding for Sustainable Development, Gwangju, Korea, 2–5 July 2019; PCS02-58. p. 158.
17. Thermo Fisher Scientific. *BRLMM-p: A Genotype Calling Method for the SNP 5.0 Array*; Thermo Fisher Scientific: Waltham, MA, USA, 2007.
18. 3,000 rice genomes project. The 3000 rice genomes project. *Gigascience* **2014**, *3*, 7. [CrossRef]
19. Meng, L.; Li, H.; Zhang, L.; Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **2015**, *3*, 269–283. [CrossRef]
20. Cheng, F.; Hayat, S. Research Progress on the use of Plant Allelopathy in Agriculture and the Physiological and Ecological Mechanisms of Allelopathy. *Front. Plant Sci.* **2015**, *6*, 347. [CrossRef] [PubMed]

genes

MDPI

Article

Quantitative Trait Loci (QTLs) Associated with Microspore Culture in *Raphanus sativus* L. (Radish)

Kyeongmin Kim [1], Yuna Kang [1], Sol-Ji Lee [1], Se-Hyun Choi [1], Dong-Hyun Jeon [1], Min-Young Park [2], Suhyoung Park [2], Yong Pyo Lim [3] and Changsoo Kim [1,4,*]

1 Department of Crop Science, College of Agricultural and Life Sciences, Chungnam National University, Daejeon 34134, Korea; katty1502@cnu.ac.kr (K.K.); dkwl3120@cnu.ac.kr (Y.K.); solji2m@naver.com (S.-J.L.); chltpgus1996@naver.com (S.-H.C.); jemdong@cnu.ac.kr (D.-H.J.)
2 National Institute of Horticultural & Herbal Science, Rural Development Administration (RDA), Wanju 55365, Korea; endrmfdl99@naver.com (M.-Y.P.); psh@korea.kr (S.P.)
3 Department of Horticultural Science, College of Agricultural and Life Sciences, Chungnam National University, Daejeon 34134, Korea; yplim@cnu.ac.kr
4 Department of Smart Agriculture Systems, College of Agricultural and Life Sciences, Chungnam National University, Daejeon 34134, Korea
* Correspondence: changsookim@cnu.ac.kr; Tel.: +82-42-821-5729

Received: 3 March 2020; Accepted: 20 March 2020; Published: 21 March 2020

Abstract: The radish is a highly self-incompatible plant, and consequently it is difficult to produce homozygous lines. Bud pollination in cross-fertilization plants should be done by opening immature pollen and attaching pollen to mature flowers. It accordingly takes a lot of time and effort to develop lines with fixed alleles. In the current study, a haploid breeding method has been applied to obtain homozygous plants in a short period of time by doubling chromosomes through the induction of a plant body in the haploid cells, in order to shorten the time to breed inbred lines. We constructed genetic maps with an F1 population derived by crossing parents that show a superior and inferior ability to regenerate microspores, respectively. Genetic maps were constructed from the maternal and parental maps, separately, using the two-way pseudo-testcross model. The phenotype of the regeneration rate was examined by microspore cultures and a quantitative trait loci (QTL) analysis was performed based on the regeneration rate. From the results of the culture of microspores in the F1 population, more than half of the group did not regenerate, and only a few showed a high regeneration rate. A total of five significant QTLs were detected in the F1 population, and five candidate genes were found based on the results. These candidate genes are divided into two classes, and appear to be related to either PRC2 subunits or auxin synthesis.

Keywords: radish; microspore culture; regeneration rate; outcrossing; two-way pseudo-testcross model

1. Introduction

With the recent development of life science technology, the importance of molecular breeding technology as well as hybridization and selection breeding is growing rapidly in seed production. In addition, it is necessary to establish a breeding system to maintain international competitiveness in the vegetable seed market by shortening the development time of new varieties through molecular breeding. However, there is a lack of biotechnology-based varieties due to the difficulty in marker development and insufficient information on genetic resources. The embryo development of flowering plants using microspores is a fascinating system in that it can dramatically promote the process of plant breeding.

The Brassicaceae includes important plants, such as *Arabidopsis thaliana* L., *Brassica napus* L., *Brassica carinata* L., *Brassica juncea* L., and *Brassica rapa* L. These plants are being used as model genomes for molecular biology research of Brassicaceae. In particular, comparative genomic studies of

Arabidopsis and cabbage have been actively conducted, and many research papers on cabbage, radishes, and rapeseed have recently been published [1–3]. In the Brassicaceae, the radish has been relatively understudied on the genomics level, and research has focused on traits related to F1 breeding, including male sterility. Genomics research on radishes is currently conducted mainly in Korea, China, Japan, and the United States. The genome size of radishes is known to be approximately 530–573 Mb [4,5]. Recently, 402 Mb genome sequences of the Japanese cultivar 'Aokubi' have been published. A total of 1345 scaffolds were assigned to the linkage map, spanning 116.0 Mb. However, the sequences are draft scaffolds with a small size of N50 (46.3 kb) and only 116 Mb of these sequences been assigned to chromosomes showing incomplete genome assembly [6]. A continuum-level genome draft has been reported for the North American wild radish in the United States [3]. The assembly size is 254 Mb and N50 is 10.1 kb. In Korea, the assembly of pseudomolecule-level assemblies, rather than contigs or scaffolds, has been announced, which is assembled at 426.2 Mb and the size of scaffold N50 is 1.22 Mb. The scaffold mapped to the chromosome in the standard dielectric is 344 Mb, forming nine pseudomolecules [7]. This shows the level of maturity of the recent *B. rapa* and *B. oleracea* standards. The first genetic maps in radish were made using populations created by cross-breeding between *R. sativus* and *R. raphanustrum* to enhance DNA polymorphism, and the RFLP molecular markers identified in *Brassica* spp. were used. Over time, genetic mapping has been done using various types of markers through intraspecific hybridization.

Despite the diversity and importance of radishes, little genetic information has been released about this plant. Most horticultural or vegetable crops have relatively long breeding cycles, from six months to a year or more [8]. Conventionally, only one generation per year of radishes is produced. Research on molecular markers related to shortening the breeding cycle is still ongoing, and the introduction of a breeding system using modern breeding technologies based on molecular markers or genomic information should be developed. For radish breeding and genetic research, it is important to shorten the breeding cycle and then create pure lines in a short time frame. Owing to the high self-incompatibility of the radish, it is necessary to secure homozygous genetic resources with fixed traits for the development of gene functional molecular markers [9]. Because the brassica crops have pronounced self-incompatibility, pollination is difficult [10,11]. For this reason, bud pollination, the process of opening immature pots, and attaching flower pots of mature flowers are utilized [12,13]. In addition, it requires a lot of time and effort to develop fixed lines since at least it takes 6–7 generations of single seed descent. In order to solve this problem, a haploid breeding method to obtain a homozygous plant in a short period of time by doubling chromosomes through the induction of a plant from haploid cells has attracted attention [14,15]. Haploids can quickly produce homozygotes through chromosomal duplications, which can significantly shorten the breeding age through self-pollination [16]. Haploid plants are also used in many fields, as they are advantageous for screening phenotypes of superior recessive traits that are not expressed in diploids, as well as for use in interspecific cross-breeding. Haploid is a general term describing sporophytes with a single set of parental chromosomes. Normal haploids are poorly grown and have low values due to difficulties such as infertility, but the embryoid body grown by doubled haploid lines are characterized by high purity.

In traditional breeding, the selection of these quantitative traits is usually performed by phenotypic testing, mainly by the grower's experience, but these traits are easily influenced by the external environment. Therefore, by creating a genetic map of quantitative trait loci (QTL) with molecular markers that are closely related to the target trait, marker-assisted selection (MAS) can be used to simplify the breeding process and improve efficiency. These molecular markers have become an essential tool for plant genetics and breeding studies. Starting with RFLP (Restriction Fragment Length Polymorphism) markers, various labels, such as RAPD (Random Amplification of Polymorphic DNA), SSR (Simple Sequence Repeat), and SNP (Single Nucleotide Polymorphism), have been developed. Among them, SNPs have been developed. They are molecular markers that are suitable for genome analysis, mapping, and marker-assisted breeding.

In this study, to improve the molecular breeding efficiency of radishes, we generated linkage maps with an F1 population and performed a QTL study associated with the regeneration rate by microspore culture, which is considered to be important in radish. Using an F1 population derived from an intra-specific cross between the high regeneration rate radish lines and low regeneration rate radish lines, 'GX71' and 'GX50', respectively, genetic linkage maps were constructed with SNPs. In order to obtain a large amount of SNP markers, low-cost and high-efficiency genotype-by-sequencing (GBS) was performed, and the SNP markers were obtained by filtering several steps to secure reliable associations between genotypes and phenotypes. The results will be utilized to develop molecular markers associated with the efficiency of microspore culture so that early stage screening is possible for the doubled haploid breeding programs in radishes.

2. Materials and Methods

2.1. Plant Materials and Genomic DNA Extraction

A population of 62 F1 individuals was derived from a cross of 'GX50' × 'GX71', which were maintained in the National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju, Korea. The F1 mapping population consists of 62 progenies from an inter-specific cross between two heterozygous genotypes.

The total genomic DNA of each parent and the progenies were extracted from young and fresh leaf tissues using the modified cetyltrimethylammonium bromide (CTAB) method [17]. The collected leaves were immediately stored in liquid nitrogen, and then transferred to the laboratory and stored at −80 °C until use. The extracted DNA samples were diluted to 20 ng/µL and then used for library preparation. DNA concentration and purity were determined using a Nano stick-S (Scinco, Seoul, Korea).

2.2. Microspore Culture and Regeneration

Microspore cultures were conducted according to Na et al. [18] with some modifications. To isolate the microspores, 30 flower buds of 2–3 mm size before flowering were selected and collected in gauze, and then surface sterilized with 1% sodium hypochlorite (NaOCl) for 15 min on a shaker followed by three rounds of 3-minute washing with sterile water. The surface-sterilized buds were then ground in mortar and pestles in 2–3 mL of NLN [19,20] medium supplemented with 13% sucrose, filtered through a 45-µm sieve, and suspended in 30 mL of NLN medium. Centrifugation was performed at 1000 rpm for 3 min, and when the centrifugation was completed, the supernatant was discarded in a clean bench. The same process was repeated three times by adding a new NLN medium. The separated microspores were suspended in 75 mL of NLN medium containing activated carbon at a density of 40,000 microspores to 1 mL and then divided into 2.5 mL petri dishes (60 mm × 15 mm) for incubation and sealed with parafilm and wraps. On average, the concentration and the volume were adjusted to include one flower bud per petri dish [21]. To the induce embryonic development of the microspores, they were treated for 48 hours in a dark condition of 30 °C and then maintained at 25 °C for about 2 weeks. Afterwards, the embryos were transferred to a light condition of 25 °C and shaken at 75–80 rpm to observe embryogenesis. Microspore-derived embryos were maintained in a shaker incubator up to about 7 mm in size and transferred to a half strength of Murashige-Skoog (MS) [22] solid medium containing 3% sucrose for plant regeneration. These counted embryos were defined as embryos from microspores collected at an early stage. The regeneration rate was calculated using the equation below.

Regeneration rate (%) = [The total number of induced embryos from microspores culture/30(mL) × 100,000 (microspores/mL)] × 100

2.3. Genotyping-by-Sequencing (GBS) Library and Illumina Sequencing

A sequencing library was prepared according to the GBS protocol in [23] using two restriction endonucleases, *Nsi*I-HF and *MseI*, which were successfully used for constructing GBS libraries in

our previous report [24]. For the construction of the GBS library, a set of 64 barcoded adapters was generated from complementary oligonucleotides with the *Nsi*I-HF (New England Biolabs, Ipswich, MA) overhang sequence and unique barcode of length 5–10 bp and they have been tested using two parental lines to determine the optimum conditions for the GBS analysis. The adapter was treated with 2000 U T4 DNA ligase (New England Biolabs, Ipswich, MA) for 2 h at 22 °C and maintained at 65 °C for 20 min to eliminate ligase activity. Barcoded DNA samples were collected into a 50 mL falcon tube and purified using a PCI (phenol-chloroform-isoamyl alcohol) extraction method. It was then size-selected with AMPure XP beads (Beckman Coulter, High Wycombe, UK) to exclude small fragments which are preferentially amplified by polymerase chain reaction (PCR). The product from PCR went through size selection again to be the final product for further processes. The concentrations of the GBS libraries were determined using a Qubit fluorimeter (Thermo Fisher Scientific, Grand Island, NY, USA). Subsequent sequencing was performed on an Illumina HiSeq 2500 platform. The sequencing data are deposited in the NCBI's SRA database (PRJNA610719).

2.4. Sequence Data Analysis and SNP Calling

When this study was completed, there was a reference genome available for aligning Illumina raw reads for SNP discovery. Raw reads were de-multiplexed and the barcode sequences were removed. Any sequences not containing the expected restriction sites for both enzymes were removed. Subsequently, the reads were filtered and trimmed using recommended settings in Trimmomatic-0.39 [25]. Burrows–Wheeler Aligner (BWA) software [26] was then used to assemble and to align the clean reads from each individual against the radish's reference genome. Genomic data (Rs 1.0 chromosome) at the chromosome pseudomolecule level of the Radish Genome Database (http://radish-genome.org/) were used as the reference genome [27]. Alignment files ware converted to bam files using the SAMtools software [28]. A genome Analysis Toolkit (GATK) and Picard tools [29] were used for variant calling. The GATK 'HaplotypeCaller' was used to find all possible SNP and indel sites. In addition, filtering was performed using GATK 'VariantFiltration' (Figure 1).

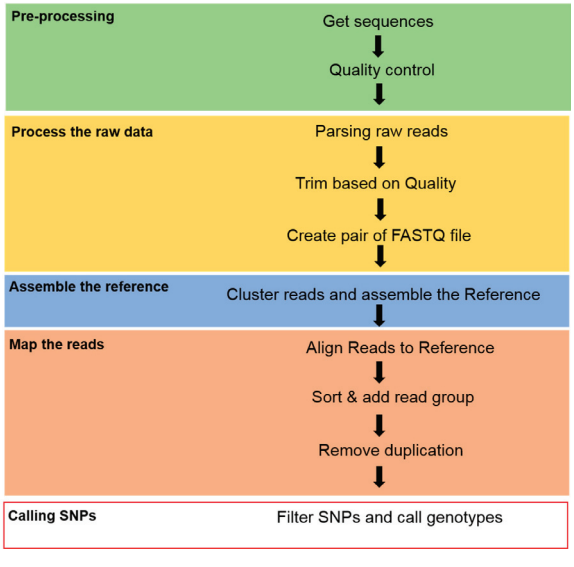


Figure 1. A pipeline for calling single nucleotide polymorphisms (SNPs) from genotype-by-sequencing (GBS) data.

2.5. Linkage Map Construction

The F1 individuals were genotyped based on marker polymorphisms. The marker-segregation data were analyzed with JOINMAP Version 4.1 by treating the segregation data of the markers as a "cross pollination" (CP) population. The significance of each allele was tested by JoinMap, which further filters for independence of segregation using logarithm of odds (LOD) scores [30]. The significance was determined at a LOD threshold of 5.0 and 8.0. SNP markers, which were heterozygous in only one of the parents (testcross markers), were scored either lm × ll or nn × np depending on the parent, and heterozygous markers in both parents with two alleles (intercross markers) were scored as hk × hk. The segregation patterns lm × ll and hk × hk were used for the maternal map construction, while the patterns nn x np and hk × hk were used for the paternal map according to the two-way pseudo-testcross mapping strategy [31]. Chi-square tests were performed to test for deviation from the expected Mendelian segregation ratio for each marker. The testcross markers were tested against a Mendelian segregation ratio of 1:1 using a chi-square test ($P < 0.05$), while those intercross markers were tested against a 1:2:1 ratio ($P < 0.05$). SNP markers with more than 30% missing data were removed from the analysis. After linkage groups (LGs) were computed, their number was assigned according to the chromosome number of the mapped marker. The regression algorithm and Kosambi [32] mapping function were used in the marker distance calculation, expressed in centiMorgans (cM). Maps were viewed using MapChart 2.3 [33].

2.6. QTL Analysis

Inclusive composite interval mapping (ICIM) (http://www.isbreeding.net/) software was used to analyze the LOD profiles with informative markers, as detected by Joinmap. The LOD thresholds for QTL significance were determined by a permutation test (1000 replications) with a genome-wide significance level $\alpha = 0.1$. ICIM tests both additive and epistasis effects and it avoids the possible increase in sampling variance and the complicated background marker selection process, which are common problems in composite interval mapping.

3. Results

3.1. GBS Library Sequencing and SNP Calling

The Hiseq 2500 platform was used to conduct the paired-end (100 bp each) sequencing of the GBS library for the parents and 62 progenies. A total of 422,282,974 reads and 42,650,580,374 bases (42.65 Gb) were generated. A total of 97,772 SNPs and InDels were detected using the GATK's VariantsToTable command. Among these, only 80,283 SNPs were called, and then 22,158 SNPs that have all of the parental genotypes were selected again. SNPs were evenly distributed throughout the genome.

3.2. Linkage Map

A linkage analysis was done in the F1 mapping population consisting of 62 individuals from a cross 'GX71' (high regeneration rate) between 'GX50' (low regeneration rate). Two individuals were excluded from the analyses because they have too many missing genotypes. We keep the missing rate below 30% so that the linkage analysis is properly performed. As a result, a total of 4462 SNPs were selected by first classifying genotypes into lm × ll, nn × np, and hk × hk types (see the Section 2.5 for details). Finally, based on the genotypes of the parents, monomorphic markers were removed. Consequently, the total number of SNPs for genetic mapping was chosen as follows: 1295, 1325, and 1852 segregating SNPs with the classes of lm × ll, hk × hk, and nn × np, respectively (Figure 2). Markers deviating significantly ($P < 0.05$) from the Mendelian segregation ratio for dominant and codominant markers, respectively, were excluded from the analysis. To increase the mapping efficiency, locus pairs or locus groups with identical genotypes were identified and a single marker was selected to represent the group. After removing the redundant SNP markers, the final set of 178 SNPs remained for linkage map construction. The linkage maps were constructed using Joinmap 4.1 software (logarithm

of odds (LOD) ≥5). The linkage maps contained nine linkage groups, which were consistent with the basal chromosome number of the radish. Because of the lack of bi-parental marker types that can connect parental genetic maps in this study, we first identified a set of SNP markers to assign the nine radish chromosomes to nine LGs. For the maternal linkage map, 87 markers were mapped from all three pseudo-testcross segregation types (Table 1, Figure 3). The length of the maternal linkage map was 439.2 cM, with the longest linkage group, LG5 (56.0 cM), and the shortest linkage group, LG2 (36.7 cM). The average interval marker distance was 5.9 cM and the maximum gap size ranged from 7.7 cM (LG9) to 13.4 cM (LG5). For the paternal linkage map, 91 markers were mapped from all three pseudo-testcross segregation types (Table 2, Figure 4). The length of the paternal linkage map was 420.8 cM, with the longest linkage group, LG5 (53.3 cM), and the shortest linkage group, LG8 (40.6 cM). The average interval marker distance was 5.2 cM and the maximum gap size ranged from 3.1 cM (LG1) to 11.3 cM (LG3). Owing to this unambiguous assignment of linkage groups to chromosomes, the linkage groups were designated with the chromosome numbers (e.g., chromosome 1 = LG 1 etc.).

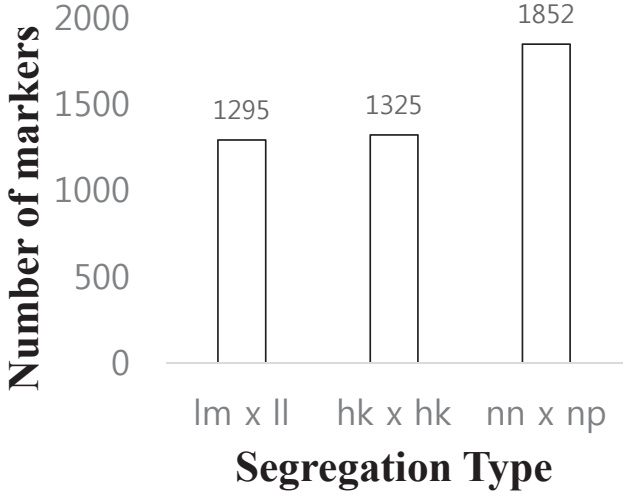


Figure 2. Segregation types of polymorphic SNP markers. The *x-axis* indicates the three segregation types; the *y-axis* indicates the number of markers included in each marker type. lm × ll = 'GX50' heterozygous 'GX71' homozygous. hk × hk = 'GX50' heterozygous 'GX71' heterozygous. nn × np = 'GX50' homozygous 'GX71' heterozygous.

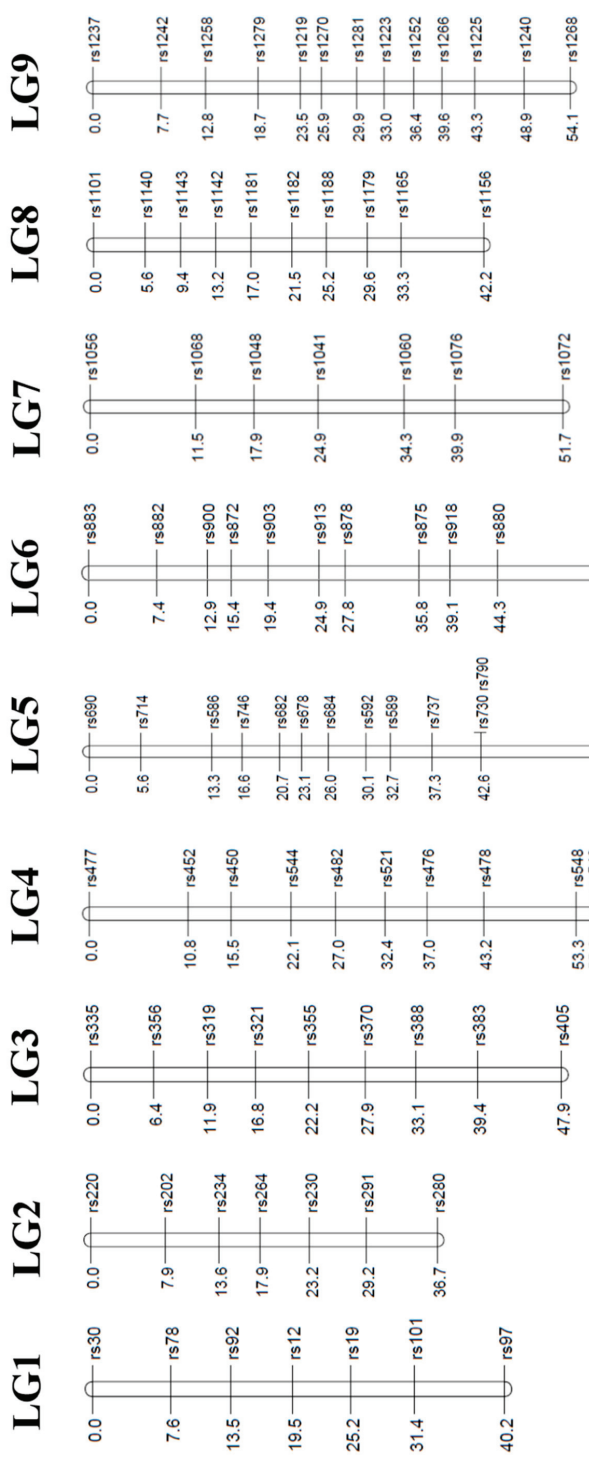


Figure 3. Genetic linkage map of GX50 constructed from an outbred population between GX50 and GX71. The map was constructed from 87 discovered SNPs and genotyped using GBS, including SNPs from all three pseudo-testcross segregation types.

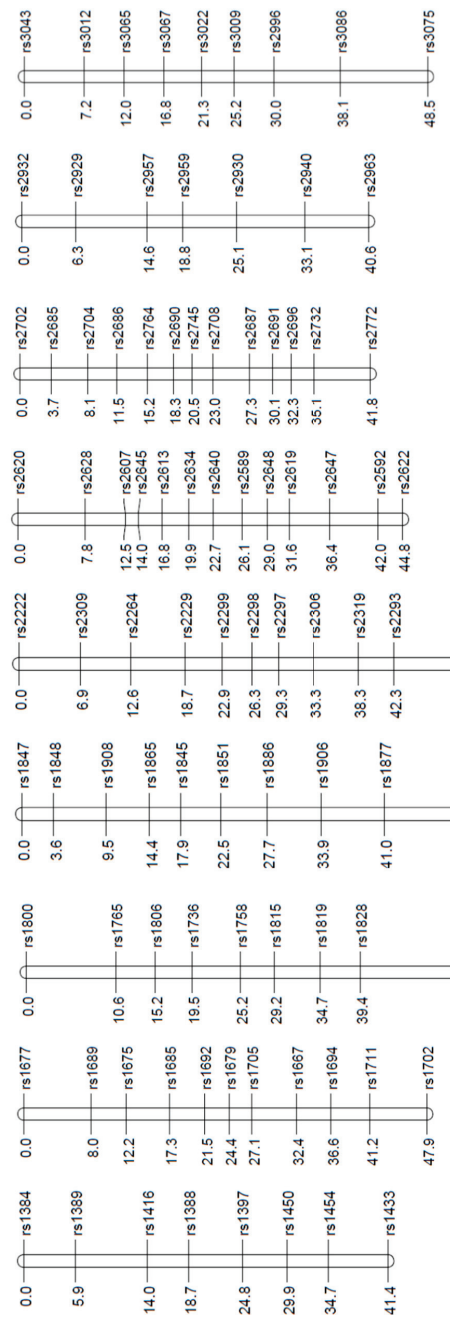


Figure 4. Genetic linkage map of GX71 constructed from an outbred population between GX50 and GX71. The map was constructed from 91 SNPs discovered and genotyped using GBS, including SNPs from all three pseudo-testcross segregation types

Table 1. Length, number of markers, average spacing, and largest intervals of GX50 genetic maps (LG; linkage group, SNP; single nucleotide polymorphism, cM; centi-Morgan).

LGs	Length (cM)	Number of SNPs	Average Interval (cM)	Largest Interval (cM)
1	40.2	7	6.7	8.8
2	36.7	7	6.1	7.9
3	47.9	9	6.0	8.5
4	55.2	10	6.1	10.8
5	56.0	13	4.7	13.4
6	55.2	11	5.5	10.9
7	51.7	7	8.6	11.8
8	42.2	10	4.7	8.9
9	54.1	13	4.5	7.7
Total	439.2	87	5.9	

Table 2. Length, number of markers, average spacing, and largest intervals of GX71 genetic maps.

LGs	Length (cM)	Number of SNPs	Average Interval (cM)	Largest Interval (cM)
1	41.4	8	5.9	3.1
2	47.9	11	4.8	6.7
3	50.7	9	6.3	11.3
4	51.8	10	5.8	10.8
5	53.3	11	4.2	11
6	44.8	13	3.7	7.8
7	41.8	13	3.5	6.7
8	40.6	7	6.8	8.3
9	48.5	9	6.0	10.4
Total	420.8	91	5.2	

3.3. Regeneration Rate from Microspore Culture

The microspore cultures were repeated twice. After completing the microspore culture experiment, each embryonic body was induced for about a month, and the induced embryos were confirmed. The average number of embryos was 3.52 in 62 F1 individuals. Only embryos that were not contaminated were selected and cultured on the MS medium. The average number of embryos cultured on the MS medium was 1.71 (Table 3). The regeneration rate was calculated using the formulas shown in the Materials and Methods. As we can see from the averages, the overall regeneration rate was quite low. The individuals without any embryogenesis totaled 31 of the F1 individuals.

Table 3. The average number of embryos and cultures of the F1 individuals.

	F1 Individuals
Total embryo average [1]	3.52 [3]
Total culture average [2]	1.71

[1] The average number of embryos from the entire 62 F1 individuals with two replications. [2] The average number of culturable embryos (not contaminated) from the total amount of embryos. [3] Rounded to two decimal places.

3.4. QTL Mapping Analysis of Regeneration Rate and the Candidate Genes Related to Microspore Culture

Using SNP markers obtained through the GBS analysis, linkage maps of 439.2 cM and 420.8 cM, respectively, were constructed, and a QTL analysis was performed by ICIM (inclusive composite interval mapping) by inputting individual phenotypic values corresponding to traits related to the regeneration rate. For the QTL peak of the trait, only those above the threshold value of the trait were considered as significant peaks. From the results of the QTL analysis, four QTL peaks in GX50 were observed (one in chromosome 3, one in chromosome 8, and two in chromosome 9). The LOD values for each QTL are 1.86, 1.79, 1.87, and 1.68, respectively. The additive effect appears to be

negative, indicating that those alleles originate from GX71. The phenotypic variations of peaks were 5.81%, 5.54%, 6.25%, and 5.24%, respectively, and all the values are within a 5–7% range and represent small values (Table 4, Figure 5). In the QTL analysis of GX71, only one QTL peak was observed in chromosome 9. The LOD value is 1.75 and the additive effect is positive, indicating that the allele also originates from GX71. In addition, the phenotypic variation was 13.19%, which was somewhat higher than the GX50 (Table 5, Figure 6). A candidate gene approach was used to determine whether specific loci may explain the responses for the regeneration rate observed in two lines selected for a high regeneration rate and a low regeneration rate. At the position of each QTL, we looked up genes corresponding to 1Mb up and down stream sequences based on left and right markers. Genes were identified in the Radish Genome Database (http://radish-genome.org/), and candidate genes for regeneration were selected based on previous studies (Table 6).

Table 4. Effects of the QTLs associated with the regeneration rate detected in F1 populations (GX50).

QTL	LOD	A.E[1]	PVE(%) [2]
LG3_1	1.86	−6.13	5.81
LG8_1	1.79	−5.96	5.54
LG9_1	1.87	−6.38	6.25
LG9_2	1.68	−5.74	5.24

[1] Estimated additive effect of QTL. Negative values indicate effects from GX71; Positive values indicate effects from GX50. [2] Phenotypic variation explained by QTL.

Table 5. Effects of the QTL on the regeneration rate detected in F1 populations (GX71).

QTL	LOD	A.E [1]	PVE(%) [2]
LG9_1	1.75	5.90	13.19

[1] Estimated additive effect of QTL. Negative values indicate effects from GX50; Positive values indicate effects from GX71. [2] Phenotypic variation explained by QTL.

Table 6. List of *Arabidopsis* orthologs associated with plant regeneration near the QTL region.

QTL	Gene ID	A.T ortholog	Gene Description
P1_Chr8_1[1]	Rs426380	AT4G32540	Flavin-binding monooxygenase family protein
P1_Chr8_1	Rs426400	AT4G32540	Flavin-binding monooxygenase family protein
P1_Chr9_1/P1_Chr9_2	Rs465100	AT5G51230	VEFS-Box of polycomb protein
P2_Chr9_1[2]	Rs479580	AT4G02020	SET domain-containing protein
P2_Chr9_1	Rs479680	AT4G02020	SET domain-containing protein

[1] P1 is parent 1, meaning GX50. [2] P2 is parent 2, meaning GX71.

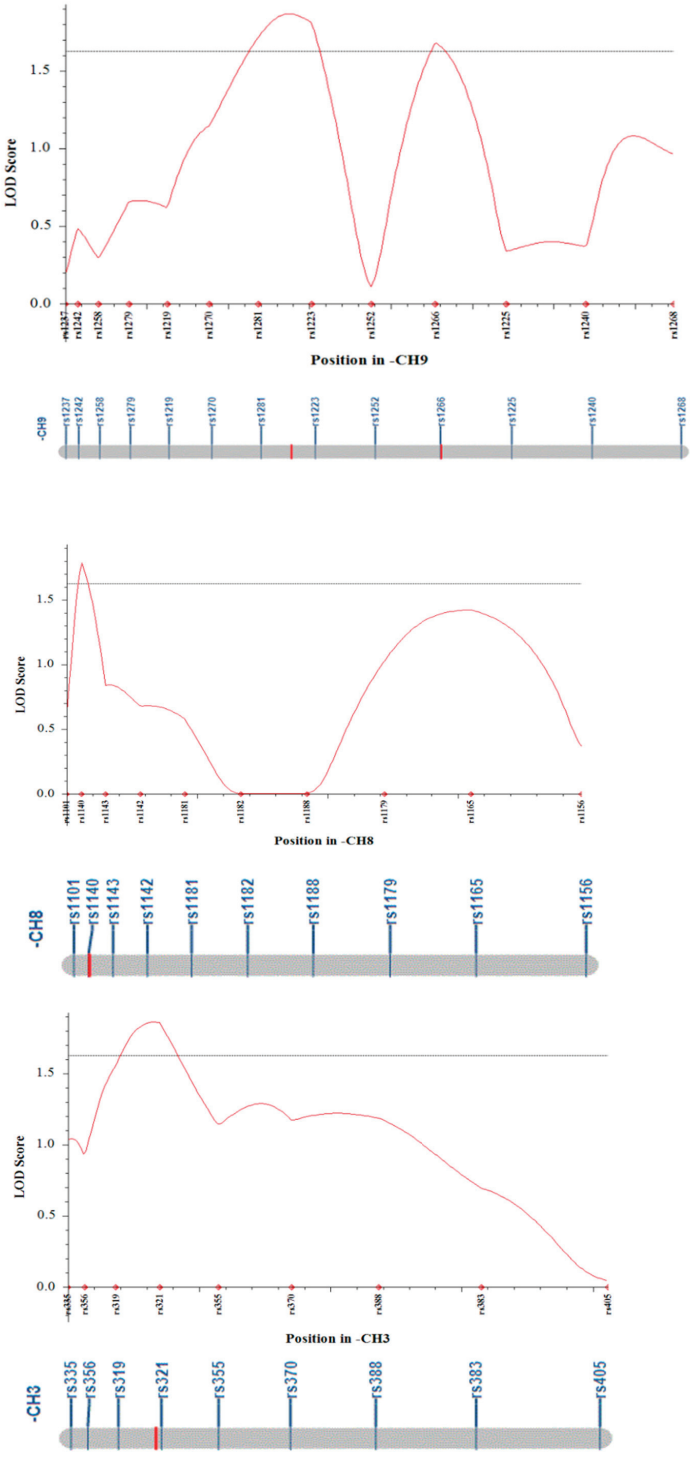


Figure 5. Logarithm of odds (LOD) score plots for chromosomes containing QTL with LOD scores (GX50). The horizontal lines indicate the thresholds for the LOD.

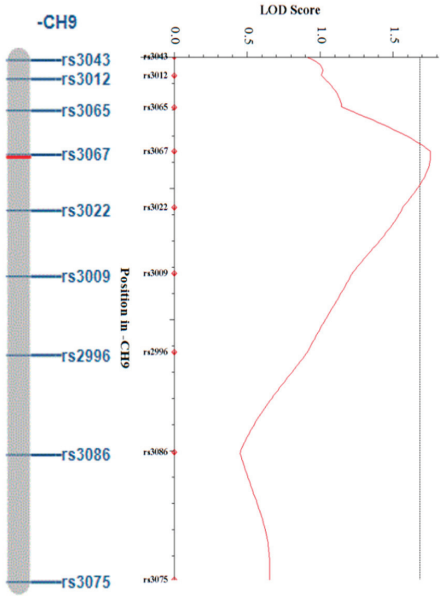


Figure 6. LOD score plots for chromosome 9 containing QTL with LOD score of 1.75 (GX71). The horizontal lines indicate the thresholds for the LOD.

4. Discussion

4.1. Phenotypic Variation of Regeneration Rate in Parent Lines and F1 Population

The radish is a highly self-incompatible plant, and it is difficult to produce homozygous lines due to this characteristic. In previous studies, the parental lines with fixed traits through generation advancement were used for QTL analysis [34,35]. In cross-fertilization plants, bud pollination should be done by opening immature pollen and attaching pollen to mature flowers. Accordingly, it takes a lot of time and effort to develop lines with any traits fixed. In this study, in order to resolve this problem, a haploid breeding method has been attempted to obtain homozygous plants in a short period of time by doubling chromosomes [14,15]. Since the QTL analysis is based on allelic differences between parent lines, it is important to select parents that exhibit a broad range of phenotypic variations for the trait. Based on the comprehensive judgment, GX50 and GX71 were selected for crossing to generate an F1 mapping population, showing differences in terms of regeneration rate. The experiments showed that over 50% of the population did not regenerate, and only a few have high regeneration rates. These regeneration rates are consistent with previous findings [23,36].

4.2. GBS Analysis Using an NGS Platform and Genetic Mapping Using SNP Markers

This study developed markers showing polymorphism between two parental lines, and genotyping for all F1 individuals was completed for a number of polymorphic markers. An next-generation sequencing (NGS) analysis is a sequencing method that can generate large amounts of genomic information at a low cost and is applied for genomic research of various plants. Unlike the costly and time-consuming re-sequencing, GBS analysis is a technique that uses a reduced representation library (RRL) to partially sequence a genome of interest using restriction enzymes [37]. Since GBS does not sequence the entire genome, it can reduce the complexity of the chromosome to allow a complex genome analysis, and large numbers of SNPs can be found for a variety of plant genetic studies, including genetic mapping and population analysis [23,38]. Creating a linkage map plays a

very important role in detecting the location of loci in the chromosome, such as QTL analysis. SNPs are the most frequently occurring genomic variations and can be used as markers for genome analysis, mapping, marker-assisted breeding, etc. The GBS method was adopted to obtain a large amount of SNPs in this study. In the preparation of the GBS library, a double-digestion method using *Nsi*l-HF and *MseI* was applied to reduce the genomic representation so that the sequence coverage for each allele increases. The library was generated with an average size of 344 bp. Considering that the size of the library suitable for NGS analysis is 170 ~ 350 bp, the generated library is considered suitable for sequencing [37]. Through the high filtering, 4462 high quality SNP markers were extracted and filtered again to generate 439.2 cM and 420.8 cM linkage maps using 178 markers, respectively. The average distance of the markers is 5.9 and 5.2 cM, respectively. When considering the length of the entire map, the interval between the markers is decent, but the length per linkage group is short [39,40]. The short linkage groups could be caused by two reasons: (1) the number of F1 individuals is small, and (2) we excluded redundant SNPs (could be reside in the same LD blocks). The genetic maps generated by the GBS have common lengthy linkage groups because some redundant markers overestimate the marker intervals. In this study, we tried to avoid those overestimated intervals by removing redundant SNPs. However, the number of markers mapped may be sufficient for a downstream analysis since we excluded duplicated markers in the same LD blocks. In addition, the SNP markers were mapped based on the locations in the pseudomolecule-level reference genome, and the physical position represented by the marker is clear, which may be useful for QTL mapping and maker-assisted selection (MAS).

4.3. QTLs Associated with Regeneration Rate

Candidate gene approaches have emerged as a way to merge QTL analysis with extensive data on traits involved in the regeneration rate. Plant regeneration is an important step in the success of a plant improvement program using tissue culture technology. The regeneration of plants is achieved through embryogenesis or tissue development [41]. The results of QTL analyses on plant regeneration rate were reported for several crops, such as rice [42,43], wheat [44], and barley [45–47], and a QTL analysis on plant regeneration rate was also reported in *B. oleracea* [48]. However, most of these or other plants are the results of QTL analyses involved in regeneration rates via anther culture [49,50] or tissue culture [51–54]. The QTL analysis of the regeneration rate in *Raphanus sativus* has not been reported to date, but a single marker analysis to find candidate genes involved in the efficiency of microspore culture has recently been reported [23]. There have been many reports on experiments on heat treatment, plant growth regulators, temperature controls, and light conditions to find suitable culture conditions for microspore culture for the regeneration of plants in Brassicaceae [55–60]. In the present paper, based on a review paper published by Ikeuchi [61], the genes related to regeneration were identified as candidate genes. The genes Rs426380 and Rs426400 have the same function as AT4G32540, which is known to function as the flavin-binding monooxygenase family protein and is known to act as an enzyme in auxin biosynthesis [62]. It also plays a key role in *de novo* root formation [63,64]. Auxin biosynthesis makes a contribution to root regeneration from leaf explants as well, because taking root is repressed by the chemical inhibition of auxin biosynthesis or YUC1, YUC2, YUC4, and YUC6 quadruple mutants that are faulty in auxin production [65]. The Rs465100 gene has the same function as AT5G51230, which is known as the PRC2 subunit of VEFS-Box of polycomb protein. In PRC2 mutants, these PRC2-targeted genes are ectopically expressed, resulting in spontaneous somatic dedifferentiation, callus formation, and embryonic development [66]. The genes Rs479580 and Rs479680 have the same functions as Arabidopsis's AT4G02020, which is known as a SET domain-containing protein, also known as the PRC2 subunit. PRC2, one of the polycomb-group proteins, has histone methyltransferase activity and primarily trimethylates histone H3 on lysine 27 (H3K27me3). In Arabidopsis, PRC2 is known to inhibit the process of allowing embryos to mature during plant development [67].

5. Conclusions

In this study, a genetic map related to regeneration rate was developed through microspore culture. The genetic map used a large number of SNP markers generated using the GBS analysis method. Due to the F1 mapping, a maternal linkage map and a paternal linkage map are created separately and contain nine LGs. The phenotype was determined by the regeneration rate of embryos resulting from the microspore culture. QTL analysis was performed from an F1 population created by a cross between high and low regeneration varieties of radish by combining the genetic map with phenotypic data. Among the radish genes in the QTL region, genes known to be related to plant regeneration were designated as candidates. We found a total of five QTLs and predicted five candidate genes. These candidate genes are divided into two classes and appear to be either PRC2 subunits or auxin synthesis. Although the detected QTL information should be confirmed through repeated experiments, it may be useful for genetic research and MAS by developing molecular markers in the near future. Based on these results, it is expected that it will help to establish a system to proactively test the efficiency of the microspore culture. In addition, we are planning to develop PCR-based molecular markers associated with microspore culture so that breeders can speed up selecting radish lines with easier haploid generations. Moreover, we expect that the candidate genes will be further investigated in the near future with reverse genetics approaches to find out exact functions related to microspore culture.

Author Contributions: Conceptualization, C.K. and S.P.; methodology, K.K., S.-J.L., S.-H.C., and D.-H.J.; software, Y.K.; validation, K.K.; formal analysis, M.-Y.P.; investigation, C.K.; resources, S.P.; data curation, Y.K.; writing—original draft preparation, K.K.; writing—review and editing, C.K.; visualization, Y.K.; supervision, Y.P.L.; writing discussion, C.K.; project administration, C.K.; funding acquisition, C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Golden Seed Project, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (213006-05-3-WTH11).

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

References

1. Hall, A.E.; Fiebig, A.; Preuss, D. Beyond the Arabidopsis genome: Opportunities for comparative genomics. *Plant Physiol.* **2002**, *129*, 1439–1447. [CrossRef]
2. Schranz, M.E.; Lysak, M.A.; Mitchell-Olds, T. The ABC's of comparative genomics in the Brassicaceae: Building blocks of crucifer genomes. *Trends Plant Sci.* **2006**, *11*, 535–542. [CrossRef] [PubMed]
3. Moghe, G.D.; Hufnagel, D.E.; Tang, H.; Xiao, Y.; Dworkin, I.; Town, C.D.; Conner, J.K.; Shiu, S.-H. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish Raphanus raphanistrum and three other Brassicaceae species. *Plant Cell* **2014**, *26*, 1925–1937. [CrossRef]
4. Marie, D.; Brown, S.C. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell* **1993**, *78*, 41–51. [CrossRef]
5. Johnston, J.S.; Pepper, A.E.; Hall, A.E.; Chen, Z.J.; Hodnett, G.; Drabek, J.; Lopez, R.; Price, H.J. Evolution of genome size in Brassicaceae. *Ann. Bot.* **2005**, *95*, 229–235. [CrossRef] [PubMed]
6. Kitashiba, H.; Li, F.; Hirakawa, H.; Kawanabe, T.; Zou, Z.; Hasegawa, Y.; Tonosaki, K.; Shirasawa, S.; Fukushima, A.; Yokoi, S. Draft sequences of the radish (Raphanus sativus L.) genome. *DNA Res.* **2014**, *21*, 481–490. [CrossRef] [PubMed]
7. Mun, J.-H.; Chung, H.; Chung, W.-H.; Oh, M.; Jeong, Y.-M.; Kim, N.; Ahn, B.O.; Park, B.-S.; Park, S.; Lim, K.-B. Construction of a reference genetic map of Raphanus sativus based on genotyping by whole-genome resequencing. *Theor. Appl. Genet.* **2015**, *128*, 259–272. [CrossRef] [PubMed]
8. Williams, P.H.; Hill, C.B. Rapid-cycling populations of Brassica. *Science* **1986**, *232*, 1385–1389. [CrossRef]
9. Pink, D.; Bailey, L.; McClement, S.; Hand, P.; Mathas, E.; Buchanan-Wollaston, V.; Astley, D.; King, G.; Teakle, G. Double haploids, markers and QTL analysis in vegetable brassicas. *Euphytica* **2008**, *164*, 509–514. [CrossRef]
10. Kakizaki, Y.; Kasai, T. Bud Pollination in Cabbage and Radish: Some Examples of Conspicuous "Pseudo-fertility" in Normally Self-incompatible Plants. *J. Hered.* **1933**, *24*, 359–360. [CrossRef]

11. Sakamoto, K.; Kusaba, M.; Nishio, T. Single-seed PCR-RFLP analysis for the identification of S haplotypes in commercial F 1 hybrid cultivars of broccoli and cabbage. *Plant Cell Rep.* **2000**, *19*, 400–406. [CrossRef] [PubMed]
12. Tatebe, T. Studies on the behavior of incompatible pollen in the Japanese radish. *J. Jpn. Soc. Hortic. Sci.* **1940**, *11*, 207–234. [CrossRef]
13. Singh, P.K.; Tripathi, S.; Somani, K. Hybrid seed production of radish (Raphanus sativus L.). *J. New Seeds* **2001**, *3*, 51–58. [CrossRef]
14. Cao, M.Q.; Li, Y.; Liu, F.; Doré, C. Embryogenesis and plant regeneration of pakchoi (Brassica rapa L. ssp. chinensis) via in vitro isolated microspore culture. *Plant Cell Rep.* **1994**, *13*, 447–450. [CrossRef] [PubMed]
15. Jo, M.; Ham, I.; Park, M.; Kim, T.; Lim, Y.; Lee, E. Seed production ability of doubled haploid plants through microspore culture in Chinese cabbage (Brassica rapa L. ssp. pekinensis) introduced from China. *Korean J. Hortic. Sci. Technol.* **2012**, *30*, 573–578. [CrossRef]
16. Germana, M.A. Gametic embryogenesis and haploid technology as valuable support to plant breeding. *Plant Cell Rep.* **2011**, *30*, 839–857. [CrossRef]
17. Sambrook, J.; Russell, D. Isolation and quantification of DNA. In *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2001; pp. 1–80.
18. Na, H.; Park, S.; Hwang, G.; Yoon, M.; Chun, C. Medium, AgNO3, activated charcoal and NAA effects on microspore culture in Brassica rapa. *Korean J. Hortic. Sci. Technol.* **2009**, *27*, 657–661.
19. Nitsch, C.; Nitsch, J. The induction of flowering in vitro in stem segments of Plumbago indica L. *Planta* **1967**, *72*, 355–370. [CrossRef]
20. Lichter, R. Induction of haploid plants from isolated pollen of Brassica napus. *Z. Pflanzenphysiol.* **1982**, *105*, 427–434. [CrossRef]
21. Lee, S.; Nam, S. Microspore culture of broccoli (Brassica oleracea var. italica). *J. Korean Soc. Hortic. Sci.* **1995**, *36*, 635–640.
22. Murashige, T.; Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [CrossRef]
23. Chung, Y.S.; Lee, Y.G.; Silva, R.R.; Park, S.; Park, M.Y.; Lim, Y.P.; Choi, S.C.; Kim, C. Potential SNPs related to microspore culture in Raphanus sativus based on a single-marker analysis. *Can. J. Plant Sci.* **2018**, *98*, 1072–1083. [CrossRef]
24. Poland, J.A.; Rife, T.W. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **2012**, *5*, 92–102. [CrossRef]
25. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
26. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
27. Jeong, Y.-M.; Kim, N.; Ahn, B.O.; Oh, M.; Chung, W.-H.; Chung, H.; Jeong, S.; Lim, K.-B.; Hwang, Y.-J.; Kim, G.-B. Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the Brassica genomes. *Theor. Appl. Genet.* **2016**, *129*, 1357–1372. [CrossRef]
28. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
29. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
30. Van Oojien, J.; Van Ooijen, J.; Van Ooijen, J.; Van Ooijen, J.; Kyazmay, B.; Ooijen, J.; Riel, J.; van OOIJEN, J.; Camp, N.; van't Verlaat, J. *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Population*; ScienceOpen, Inc.: Boston, MA, USA, 2006.
31. Grattapaglia, D.; Sederoff, R. Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: Mapping strategy and RAPD markers. *Genetics* **1994**, *137*, 1121–1137.
32. Kosambi, D.D. The estimation of map distances from recombination values. In *DD Kosambi*; Springer: New Delhi, India, 2016; pp. 125–130.
33. Voorrips, R. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [CrossRef]

34. Hashida, T.; Nakatsuji, R.; Budahn, H.; Schrader, O.; Peterka, H.; Fujimura, T.; Kubo, N.; Hirai, M. Construction of a chromosome-assigned, sequence-tagged linkage map for the radish, Raphanus sativus L. and QTL analysis of morphological traits. *Breed. Sci.* **2013**, *63*, 218–226. [CrossRef] [PubMed]
35. Yu, X.; Choi, S.R.; Dhandapani, V.; Rameneni, J.J.; Li, X.; Pang, W.; Lee, J.-Y.; Lim, Y.P. Quantitative trait loci for morphological traits and their association with functional genes in Raphanus sativus. *Front. Plant Sci.* **2016**, *7*, 255. [CrossRef] [PubMed]
36. Palmer, C.E.; Keller, W.A.; Arnison, P.G. Experimental haploidy in Brassica species. In *In Vitro Haploid Production in Higher Plants*; Springer: Dordrecht, The Netherlands, 1996; pp. 143–172.
37. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*. [CrossRef] [PubMed]
38. Kale, S.M.; Jaganathan, D.; Ruperao, P.; Chen, C.; Punna, R.; Kudapa, H.; Thudi, M.; Roorkiwal, M.; Katta, M.A.; Doddamani, D. Prioritization of candidate genes in "QTL-hotspot" region for drought tolerance in chickpea (Cicer arietinum L.). *Sci. Rep.* **2015**, *5*, 15296. [CrossRef] [PubMed]
39. Xu, L.; Wang, L.; Gong, Y.; Dai, W.; Wang, Y.; Zhu, X.; Wen, T.; Liu, L. Genetic linkage map construction and QTL mapping of cadmium accumulation in radish (Raphanus sativus L.). *Theor. Appl. Genet.* **2012**, *125*, 659–670. [CrossRef]
40. Zou, Z.; Ishida, M.; Li, F.; Kakizaki, T.; Suzuki, S.; Kitashiba, H.; Nishio, T. QTL analysis using SNP markers developed by next-generation sequencing for identification of candidate genes controlling 4-methylthio-3-butenyl glucosinolate contents in roots of radish, Raphanus sativus L. *PLoS ONE* **2013**, *8*. [CrossRef]
41. Firoozabady, E.; DeBoer, D. Plant regeneration via somatic embryogenesis in many cultivars of cotton (Gossypium hirsutum L.). *Vitr. Cell. Dev. Boil. Anim.* **1993**, *29*, 166–173. [CrossRef]
42. Takeuchi, Y.; Abe, T.; Sasahara, T. RFLP mapping of QTLs influencing shoot regeneration from mature seed-derived calli in rice. *Crop Sci.* **2000**, *40*, 245–247. [CrossRef]
43. Kwon, Y.-S.; Kim, K.-M.; Eun, M.-Y.; Sohn, J.-K. Quantitative trait loci mapping associated with plant regeneration ability from seed derived calli in rice (Oryza sativa L.). *Mol. Cells* **2001**, *11*, 64–67.
44. Ma, J.; Deng, M.; Lv, S.-Y.; Yang, Q.; Jiang, Q.-T.; Qi, P.-F.; Li, W.; Chen, G.-Y.; Lan, X.-J.; Wei, Y.-M. Identification of QTLs associated with tissue culture response of mature wheat embryos. *SpringerPlus* **2016**, *5*, 1–7. [CrossRef]
45. Komatsuda, T.; Annaka, T.; Oka, S. Genetic mapping of a quantitative trait locus (QTL) that enhances the shoot differentiation rate in Hordeum vulgare L. *Theor. Appl. Genet.* **1993**, *86*, 713–720. [CrossRef] [PubMed]
46. Mano, Y.; Takahashi, H.; Sato, K.; Takeda, K. Mapping genes for callus growth and shoot regeneration in barley (Hordeum vulgare L.). *Jpn. J. Breed.* **1996**, *46*, 137–142. [CrossRef]
47. Bregitzer, P.; Campbell, R.D. Genetic markers associated with green and albino plant regeneration from embryogenic barley callus. *Crop Sci.* **2001**, *41*, 173–179. [CrossRef]
48. Holme, I.; Torp, A.; Hansen, L.; Andersen, S. Quantitative trait loci affecting plant regeneration from protoplasts of Brassica oleracea. *Theor. Appl. Genet.* **2004**, *108*, 1513–1520. [CrossRef]
49. Torp, A.; Hansen, A.; Andersen, S.B. Chromosomal regions associated with green plant regeneration in wheat (Triticum aestivum L.) anther culture. *Euphytica* **2001**, *119*, 377. [CrossRef]
50. Kwon, Y.; Kim, K.; Eun, M.; Sohn, J. QTL mapping and associated marker selection for the efficacy of green plant regeneration in anther culture of rice. *Plant Breed.* **2002**, *121*, 10–16. [CrossRef]
51. Mano, Y.; Komatsuda, T. Identification of QTLs controlling tissue-culture traits in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* **2002**, *105*, 708–715. [CrossRef]
52. Bolibok, H.; Rakoczy-Trojanowska, M. Genetic mapping of QTLs for tissue-culture response in plants. *Euphytica* **2006**, *149*, 73–83. [CrossRef]
53. Jia, H.; Yi, D.; Yu, J.; Xue, S.; Xiang, Y.; Zhang, C.; Zhang, Z.; Zhang, L.; Ma, Z. Mapping QTLs for tissue culture response of mature wheat embryos. *Mol. Cells* **2007**, *23*, 323–330.
54. Tyagi, N.; Dahleen, L.S.; Bregitzer, P. Candidate genes within tissue culture regeneration QTL revisited with a linkage map based on transcript-derived markers. *Crop Sci.* **2010**, *50*, 1697–1707. [CrossRef]
55. Duijs, J.; Voorrips, R.; Visser, D.; Custers, J. Microspore culture is successful in most crop types of Brassica oleracea L. *Euphytica* **1992**, *60*, 45–55.

56. Custers, J.B.; Cordewener, J.H.; Nöllen, Y.; Dons, H.J.; Campagne, M.M.V.L. Temperature controls both gametophytic and sporophytic development in microspore cultures of Brassica napus. *Plant Cell Rep.* **1994**, *13*, 267–271. [CrossRef] [PubMed]
57. Iqbal, M.; Möllers, C.; Röbbelen, G. Increased embryogenesis after colchicine treatment of microspore cultures of Brassica napus L. *J. Plant Physiol.* **1994**, *143*, 222–226. [CrossRef]
58. Hansen, N.; Andersen, S.B. In vitro chromosome doubling potential of colchicine, oryzalin, trifluralin, and APM in Brassica napus microspore culture. *Euphytica* **1996**, *88*, 159–164. [CrossRef]
59. Sato, S.; Katoh, N.; Iwai, S.; Hagimori, M. Effect of low temperature pretreatment of buds or inflorescence on isolated microspore culture in Brassica rapa (syn. B. campestris). *Breed. Sci.* **2002**, *52*, 23–26. [CrossRef]
60. Chun, C.; Park, H.; Na, H. Microspore-derived embryo formation in radish (Raphanus sativus L.) according to nutritional and environmental conditions. *Hortic. Environ. Biotechnol.* **2011**, *52*, 530. [CrossRef]
61. Ikeuchi, M.; Favero, D.S.; Sakamoto, Y.; Iwase, A.; Coleman, D.; Rymen, B.; Sugimoto, K. Molecular mechanisms of plant regeneration. *Annu. Rev. Plant Biol.* **2019**, *70*, 377–406. [CrossRef]
62. Zhao, Y.; Christensen, S.K.; Fankhauser, C.; Cashman, J.R.; Cohen, J.D.; Weigel, D.; Chory, J. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science* **2001**, *291*, 306–309. [CrossRef]
63. Liu, J.; Sheng, L.; Xu, Y.; Li, J.; Yang, Z.; Huang, H.; Xu, L. WOX11 and 12 are involved in the first-step cell fate transition during de novo root organogenesis in Arabidopsis. *Plant Cell* **2014**, *26*, 1081–1093. [CrossRef]
64. Chen, L.; Tong, J.; Xiao, L.; Ruan, Y.; Liu, J.; Zeng, M.; Huang, H.; Wang, J.-W.; Xu, L. YUCCA-mediated auxin biogenesis is required for cell fate transition occurring during de novo root organogenesis in Arabidopsis. *J. Exp. Bot.* **2016**, *67*, 4273–4284. [CrossRef]
65. Cheng, Z.J.; Wang, L.; Sun, W.; Zhang, Y.; Zhou, C.; Su, Y.H.; Li, W.; Sun, T.T.; Zhao, X.Y.; Li, X.G. Pattern of auxin and cytokinin responses for shoot meristem induction results from the regulation of cytokinin biosynthesis by AUXIN RESPONSE FACTOR3. *Plant Physiol.* **2013**, *161*, 240–251. [CrossRef] [PubMed]
66. Ikeuchi, M.; Iwase, A.; Rymen, B.; Harashima, H.; Shibata, M.; Ohnuma, M.; Breuer, C.; Morao, A.K.; de Lucas, M.; De Veylder, L. PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis. *Nat. Plants* **2015**, *1*, 1–7. [CrossRef] [PubMed]
67. Mozgová, I.; Muñoz-Viana, R.; Hennig, L. PRC2 represses hormone-induced somatic embryogenesis in vegetative tissue of Arabidopsis thaliana. *PLoS Genet.* **2017**, *13*, e1006562. [CrossRef] [PubMed]

genes

MDPI

Article

Characterization of QTLs and Candidate Genes for Days to Heading in Rice Recombinant Inbred Lines

Youngjun Mo [1], Jong-Min Jeong [1], Su-Kyung Ha [1], Jinhee Kim [1], Changmin Lee [1], Gung Pyo Lee [2] and Ji-Ung Jeung [1,*]

[1] National Institute of Crop Science, Rural Development Administration, Wanju 55365, Korea; moyj82@korea.kr (Y.M.); jjm0820@korea.kr (J.-M.J.); rocksue193@korea.kr (S.-K.H.); jinhee2723@korea.kr (J.K.); cropas@korea.kr (C.L.)

[2] Department of Integrative Plant Science, Chung-Ang University, Anseong 17546, Korea; gplee@cau.ac.kr

* Correspondence: jrnj@korea.kr; Tel.: +82-63-238-5231

Received: 10 July 2020; Accepted: 17 August 2020; Published: 19 August 2020

Abstract: Understanding the gene mechanisms controlling days to heading (DH) is important in rice breeding for adaption in the target environment. Using a recombinant inbred line population derived from the cross between two *japonica* rice cultivars, Koshihikari and Baegilmi, we identified three consistent quantitative trait loci (QTLs) for DH for two years, *qDH3*, *qDH6*, and *qDH7* on chromosomes 3, 6, and 7, respectively. While Baegilmi contributed the allele for early heading at *qDH6* and *qDH7* with the additive effect of five days each, Koshihikari contributed the allele for early heading at *qDH3* with the additive effect of three days. Notably, pyramiding two or more alleles for early heading at these QTLs accelerated heading effectively. Sequencing of *Hd16*, *Hd1*, and *Ghd7*, the previously known heading date genes underlying *qDH3*, *qDH6*, and *qDH7*, respectively, revealed that Baegilmi and Koshihikari carry different alleles at the three genes. Molecular markers were developed to screen the allelic compositions of the three genes among 295 Korean commercial rice cultivars. The results showed that few cultivars carry alleles for early heading at the three genes, highlighting that DH can be further accelerated and fine-tuned in breeding programs by combining the desirable alleles of *Hd16*, *Hd1*, and *Ghd7*.

Keywords: rice; days to heading; QTL; *Hd1*; *Ghd7*; *Hd16*

1. Introduction

The transition from vegetative to reproductive stage is a critical developmental event in plants for ensuring offspring survival under favorable environments [1]. Understanding the genetic basis of flowering time control is especially important in crop species for breeding cultivars adapted well in the target environment. In rice, flowering (also referred to as heading) time is regulated by the complex genetic mechanisms involving hundreds of quantitative trait loci (QTLs) and at least 14 cloned genes [2,3].

Various environmental cues are integrated to modulate the expression of genes encoding florigen, which is synthesized in leaves and transferred to the apical meristem to initiate flower development [4]. In rice, florigen is encoded by *Heading Date 3a* (*Hd3a*) and *RICE FLOWERING LOCUS T 1* (*RFT1*), the orthologs of Arabidopsis *FT* [5,6]. *Hd3a* expression is mainly regulated by *Hd1*, the ortholog of Arabidopsis *CONSTANS* (*CO*), which upregulates *Hd3a* under short day and downregulates *Hd3a* under long day [7]. *Hd1* expression is upregulated under both short and long days by *OsGI*, the ortholog of Arabidopsis *GIGANTEA* (*GI*) [8]. Unlike bifunctional *Hd1*, *EARLY HEADING DATE 1* (*Ehd1*) can accelerate heading under both short and long days by upregulating *Hd3a* and *RFT1* under short and long days, respectively [9]. *Ehd1* expression is negatively regulated by *Grain number*, *plant height*, *and heading date 7* (*Ghd7*) under long day [10]. While the *OsGI-Hd1-Hd3a* pathway in rice is

orthologous to the *GI-CO-FT* pathway in Arabidopsis, the *Ghd7-Ehd1-Hd3a/RFT1* pathway is unique in rice without clear orthologs in Arabidopsis [1]. Recent studies revealed more complex rice-specific gene networks regulating the *Ghd7-Ehd1-Hd3a/RFT1* pathway (e.g., *Hd16* and *Hd17* upregulating *Ghd7* under long day) reviewed in [2,3].

To fine-tune days to heading (DH) in breeding programs and maximize yield and grain quality under the target environment, it is essential to characterize the effects of major genes controlling DH, their allelic variation, epistasis, and interaction with environmental factors including daylength and temperature. In Korea, developing early heading rice cultivars is especially important for boosting farmers' income by enabling diverse double cropping patterns in the rice paddies, e.g., late-planting of rice after harvesting winter crops such as cabbage, barley, and wheat, or early-planting of rice followed by cash crops such as garlic and onion [11,12]. The utilization of early heading rice cultivars can be also useful for reducing cropping duration in order to minimize damages from erratic weather events. Although over 80 rice cultivars classified as the early heading group have been released in Korea, they occupy less than 10% of the rice cultivation area in Korea as many rice growers generally prefer mid-late heading cultivars because of their higher yield and grain quality compared to the early heading cultivars [13].

Baegilmi is an extremely early heading rice cultivar recently released in Korea exhibiting high yield in the mid-north plain (milled rice yield 5.01 MT/ha) and north-east coastal (5.27 MT/ha) regions in Korea [14]. However, the genes conferring early heading in Baegilmi have remained unknown. In this study, we used a recombinant inbred line (RIL) population derived from the cross between Koshihikari and Baegilmi to identify the chromosomal regions harboring genes controlling DH. Candidate genes underlying the major DH QTLs were sequenced in Koshihikari and Baegilmi and their allelic compositions were screened among commercial rice cultivars. Molecular breeding strategy using the allelic variations in major DH genes was discussed to facilitate the fine-tuning of DH in breeding programs.

2. Materials and Methods

2.1. Plant Materials and Phenotype Evaluation

Two *japonica* rice cultivars, Koshihikari with high eating quality [15] and Baegilmi with early maturity [14], were used in this study. Days to heading (DH) and grain filling rates of the two cultivars were evaluated at the experimental field of the National Institute of Crop Science (NICS), Suwon, Korea (37°27′ N 126°99′ E) in 2014. The seeds of each cultivar were sown on 25 April and transplanted on 25 May under a randomized complete block design (RCBD) with three replications. Each plot comprised of eight 4.5 m rows, with 30 hills per row and three plants per hill. The hills within a row were spaced by 15 cm and the rows were spaced by 30 cm. DH was determined by counting the number of days from sowing to heading when the panicles emerged in 40% of the plants in a plot. To evaluate the rate of grain filling of Baegilmi and Koshihikari, changes in grain weight of the two cultivars were monitored by measuring 1000 grain weight every three to four days during 19–50 days after heading.

To map QTLs for DH, a RIL population (n = 142) was constructed from the cross between Koshihikari and Baegilmi by the single seed descent method. The RIL population (F_6 and F_7 generation in 2016 and 2017, respectively) and its parents were grown at the experimental field of NICS, Wanju, Korea (35°84′ N 127°05′ E) in 2016 and 2017. The seeds were sown on 9 May and 10 May in 2016 and 2017, respectively, and the seedlings were transplanted four weeks after sowing. Each RIL was transplanted in a 4.5 m row with the individual plants spaced by 15 cm (30 plants per each line) and the rows spaced by 30 cm. DH of each RIL was determined by counting the number of days from sowing to heading when the panicles emerged in 40% of the plants in a row.

Days to heading of 295 commercial rice cultivars released by NICS, Rural Development Administration, Korea, were evaluated under the optimum and early planting conditions at the

experimental field of NICS, Wanju, Korea. Sowing and transplanting dates for optimum planting were 9 May and 1 Jun, respectively in 2018, and those for early planting were 10 Apr and 9 May, respectively in 2019. Planting density and DH evaluation were as described above for the RIL population.

2.2. Sequencing Library Construction and Genotyping

Genomic DNA was extracted from the fresh young leaves of the Koshihikari × Baegilmi RILs using the CTAB (cetyl trimethylammonium bromide) method [16] with minor modifications. The quality and quantity of the extracted DNA were checked using the DeNovix DS-11 spectrophotometer (DeNovix, Wilmington, DE, USA) and the Quant-iT^{TM} dsDNA assay kit (Thermo Fisher Scientific, Waltham, MA, USA). The genotyping-by-sequencing (GBS) library was constructed according to [17]. Briefly, each DNA sample was digested with the *ApeKI* restriction enzyme (New England Biolabs, Ipswitch, MA, USA), ligated with barcode adapters, pooled, and purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). The pooled libraries were amplified by PCR with an adapter-specific primer set, analyzed for the target length of 170–350 bp using BioAnalyzer 2100 (Agilent, Santa Clara, CA, USA), and sequenced using the Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA). The sequencing reads were mapped to the Nipponbare IRGSP-1.0 reference using the BWA aligner [18] and single nucleotide polymorphism (SNPs) were extracted in the format of a VCF (variant calling format) file. The SNP filtering and genotype calling of each RIL were carried out using TASSEL-GBS v2 [19] with the filtering conditions (mapping quality ≥30, base quality ≥20, and coverage ≥7) as described in [20]. Genotype calling for each SNP was conducted by defining a homozygous genotype when the same sequence rate is over 0.90 and a heterozygous genotype when the alternative sequence rate is 0.25–0.27.

2.3. Linkage Mapping and Statistical Analysis

Of 893 SNPs segregating in the RIL population, 128 high quality SNPs were selected for the linkage map construction after excluding SNPs with missing rate over 10%, significant segregation distortion (Chi-square test p–value < 0.05), or overlapping genetic positions. The composite interval mapping was implemented using QTL IciMapping version 4.1 [21] with the threshold LOD (logarithm of the odds) score of 3.0. The same program was also used to estimate the additive effects and the phenotypic variation explained by each QTL at the peak LOD. The mean comparisons of DH between the RILs with different allelic combinations of the identified QTLs were conducted by the Duncan's multiple range test using SAS version 9.4 (Cary, NC, USA). Three-way factorial ANOVAs were conducted to study the main effects of the three DH QTLs and their two-way and three-way interactions using SAS version 9.4.

2.4. Sequencing and Marker Analysis for Hd16, Hd1, and Ghd7

The coding regions of *Hd16*, *Hd1*, and *Ghd7*, the candidate genes underlying the three DH QTLs from the Koshihikari × Baegilmi RIL population, were sequenced to search polymorphisms between Koshihikari and Baegilmi. The tenth exon of *Hd16* was amplified by polymerase chain reaction (PCR) using the primers and conditions described in Supplementary Table S1. The two exons of *Hd1* and the two exons of *Ghd7* were amplified by PCR as described in [22]. Sequencing of the PCR products was performed by Macrogen (Daejeon, Korea). To search polymorphism in the promoter region of *Ghd7*, the 2 kb upstream region of *Ghd7* from Koshihikari and Baegilmi was sequenced using the customized sequencing service at Macrogen (Daejeon, Korea). Molecular markers to differentiate the three main polymorphisms in *Hd16*, *Hd1*, and *Ghd7* between Koshihikari and Baegilmi were developed (Supplementary Table S1) and used to screen 295 Korean rice cultivars released in 1979–2017 by NICS, Rural Development Administration.

3. Results

3.1. Days to Heading and Grain Filling Rate of Baegilmi in Comparison with Koshihikari

The DH of Baegilmi under the optimum planting (i.e., sowing and transplanting on 25 April and 25 May, respectively, in Suwon, Korea) was 82 days, which was 21 days earlier than that of Koshihikari (Figure 1a). To compare the rate of grain filling of Baegilmi and Koshihikari, the grain weight change was monitored by measuring 1000 grain weight (TGW) every 3–4 days from 19 days after heading (DAH) to 50 DAH (Figure 1b). While the TGW of Baegilmi increased faster than that of Koshihikari during the early grain filling stage, reaching 19.7 g at 26 DAH (14.5 g in Koshihikari), the TGW of Koshihikari increased faster in the late grain filling stage, reaching 23.1g at 40 DAH which was similar to Baegilmi (23.2 g). A similar pattern was observed when the changes in grain weight were monitored as the ratio of grain weight to the final grain weight according to the cumulative temperature after heading (Figure 1c). Our results indicated that Baegilmi undergoes rapid grain weight increase in the early stage of grain filling as well as early vegetative-to-reproductive transition compared to Koshihikari.

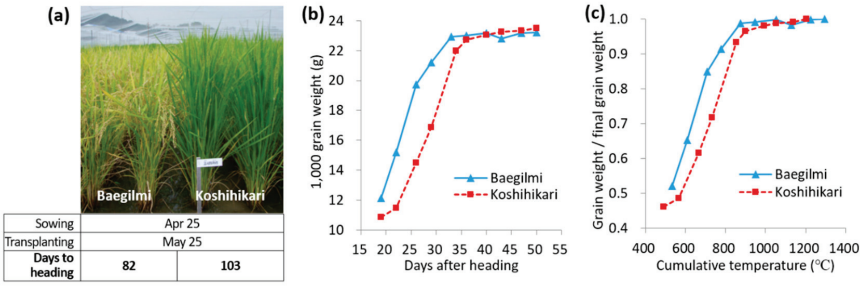


Figure 1. Maturity of Baegilmi in comparison with Koshihikari. (**a**) Representative phenotype of Baegilmi and Koshihikari in the field at 112 days after sowing (15 August). Days to heading was determined as the number of days from sowing to heading; (**b**) Changes in grain weight during 19–50 days after heading; (**c**) Changes in the proportion of grain weight to the final grain weight according to the cumulative temperature after heading.

3.2. Mapping of qDH3, qDH6, and qDH7 for Days to Heading

The Koshihikari × Baegilmi RIL population (n = 142) was used to map QTLs for DH. A GBS experiment detected 893 SNPs segregating in the mapping population. After removing SNPs with low genotyping quality and overlapping genetic positions, a linkage map spanning a total length of 1293 cM was constructed with the 128 selected SNPs (Supplementary Figure S1). The average number of markers per chromosome was 10.7, ranging from six on chromosome 3 to 17 on chromosome 6. The average interval between two adjacent markers was 11.1 cM.

The Koshihikari × Baegilmi RILs showed continuous DH variation (64–105 days in 2016 and 64–107 days in 2017) with positively skewed distribution (Figure 2a,b). Baegilmi headed 19 and 18 days earlier than Koshihikari in 2016 and 2017, respectively. We detected three major QTLs for DH designated *qDH3*, *qDH6*, and *qDH7* on chromosomes 3, 6, and 7, respectively, in both 2016 and 2017 (Figure 3a, Table 1). At *qDH3*, Koshihikari contributed the allele for early heading with the additive effects of −3.1–−3.4 days and the LOD scores of 5.1–7.5, explaining 8.4–10.3% of the DH variation. On the other hand, Baegilmi contributed the allele for early heading at *qDH6* and *qDH7* (additive effects of 5.2–5.4 days and 5.0–5.3 days, respectively), which showed higher LOD scores (12.6–15.1 and 12.7–13.7, respectively) explaining higher levels of the DH variation (24.7–26.2% and 22.8–25.2%, respectively).

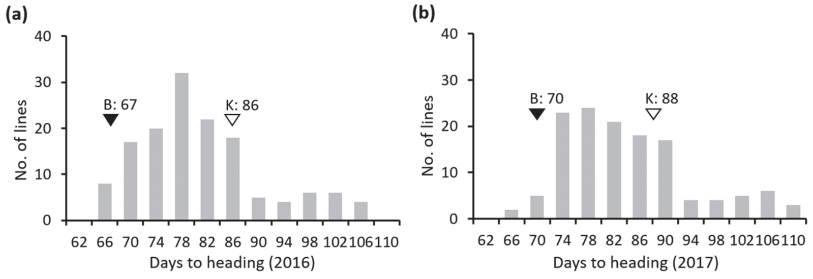


Figure 2. Frequency distribution of days to heading in the Koshihikari × Baegilmi RIL population evaluated in 2016 (**a**) and 2017 (**b**). The values of Baegilmi (B) and Koshihikari (K) are indicated above filled and unfilled triangles, respectively.

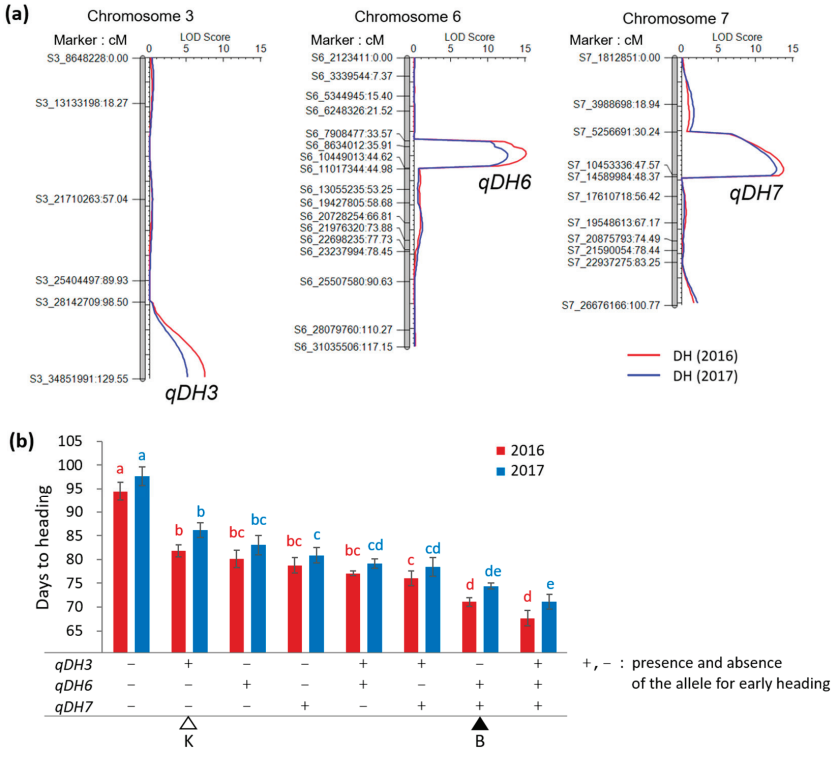


Figure 3. QTLs for days to heading from the Koshihikari × Baegilmi RIL population. (**a**) Mapping of *qDH3*, *qDH6*, and *qDH7* on chromosomes 3, 6, and 7, respectively. Days to heading of the RILs observed in 2016 and 2017 were used for mapping. The number after the letter 'S' in the marker name indicates the chromosome number followed by the physical position according to the IRGSP-1.0 reference; (**b**) Days to heading of the RILs with different allelic combinations of *qDH3*, *qDH6*, and *qDH7*. The markers S3_34851991, S6_8634012, and S7_10453336 were used to represent *qDH3*, *qDH6*, and *qDH7*, respectively. + and − indicate the presence and absence of the allele for early heading, respectively. The genotypes of Baegilmi (B) and Koshihikari (K) are indicated by filled and unfilled triangles, respectively. Different letters above the bars indicate significant difference according to the Duncan's multiple range test at $p < 0.05$. Error bars indicate standard errors.

Table 1. Quantitative trait loci (QTLs) for days to heading from the Koshihikari × Baegilmi recombinant inbred line (RIL) population.

QTL	Chr [a]	Flanking Markers [b]		Year	Peak (cM)	LOD [c]	PVE [d] (%)	Add [e]
		Left	Right					
qDH3	3	S3_28142709	S3_34851991	2016	129	7.5	10.3	−3.4
				2017	129	5.1	8.4	−3.1
qDH6	6	S6_8634012	S6_10449013	2016	39	15.1	24.7	5.2
				2017	40	12.6	26.2	5.4
qDH7	7	S7_5256691	S7_10453336	2016	45	13.7	22.8	5.0
				2017	46	12.7	25.2	5.3

[a] Chromosome number. [b] The number after the letter 'S' in the marker name indicates the chromosome number, followed by the physical location of SNP according to the IRGSP-1.0 reference. [c] Logarithm of the odds. [d] Phenotypic variation explained (%). [e] Additive effect of QTL estimated at the peak position. Positive additive effect indicates that the trait value is increased by the Koshihikari allele.

To analyze the main effects of the three QTLs and their two-way and three-way interactions, 2 × 2 × 2 factorial ANOVAs were conducted for DH of the Koshihikari × Baegilmi RILs (Table 2). The main effects of *qDH3*, *qDH6*, and *qDH7* were highly significant ($3.8 \times 10^{-15} < p < 2.6 \times 10^{-5}$) in both 2016 and 2017. On average, the RILs carrying the Koshihikari alleles at *qDH3* headed 5.1–5.2 days earlier than those carrying the Baegilmi alleles. On the other hand, the RILs carrying the Baegilmi alleles at *qDH6* or *qDH7* headed 9.5–9.9 or 10.7–11.4 days earlier than those with the Koshihikari alleles, respectively. The RILs with different allele combinations of *qDH3*, *qDH6*, and *qDH7* showed that accumulating two or more alleles for early heading of theses QTLs can shorten DH effectively, with the pyramiding effect of *qDH6* and *qDH7* being greater than that of *qDH3* (Figure 3b).

Table 2. Three-way ANOVAs of *qDH3*, *qDH6*, and *qDH7* for days to heading.

Year	Three-Way ANOVA					Mean DH [c]	
	Effect [a]		*F* Value	*p* Value	PVE [b]	K	B
2016	Main	*qDH3*	24.5	2.5×10^{-6}	7.6	76.4	81.6
		qDH6	63.7	1.2×10^{-12}	19.8	83.4	73.9
		qDH7	82.3	3.8×10^{-15}	25.5	84.3	73.6
	Interaction	*qDH3* × *qDH6*	4.1	0.0453	1.3		
		qDH6 × *qDH7*	0.5	ns	-		
		qDH3 × *qDH7*	4.8	0.0301	1.5		
		qDH3 × *qDH6* × *qDH7*	5.5	0.0206	1.7		
2017	Main	*qDH3*	19.4	2.6×10^{-5}	6.6	79.6	84.7
		qDH6	54.9	3.3×10^{-11}	18.8	86.9	76.9
		qDH7	73.4	9.7×10^{-14}	25.2	87.7	76.3
	Interaction	*qDH3* × *qDH6*	2.0	ns	-		
		qDH6 × *qDH7*	2.8	ns	-		
		qDH3 × *qDH7*	4.2	0.0433	1.4		
		qDH3 × *qDH6* × *qDH7*	3.1	ns	-		

[a] *qDH3*, *qDH6*, and *qDH7* were represented by the markers S3_34851991, S6_8634012, and S7_10453336, respectively. [b] Phenotypic variation explained (%) is shown only for the significant (p 0.05) effects. [c] Mean days to heading of the RILs carrying the Koshihikari (K) allele and the Baegilmi (B) allele at each of the three QTL.

The two-way and three-way interactions among *qDH3*, *qDH6*, and *qDH7* were marginally significant ($0.02 < p < 0.05$) and explained little phenotypic variation (1.3–1.7%) (Table 2). The *qDH3* × *qDH6* interaction was significant only in 2016, where the effect of *qDH6* accelerating heading was greater under the absence of the allele for early heading at *qDH3* (Figure 4a). Similarly, the effect of *qDH7* accelerating heading was greater under the absence of the allele for early heading at *qDH3*, and this was significant in both 2016 and 2017 (Figure 4b,c). The *qDH3* × *qDH6* × *qDH7* interaction was significant

only in 2016. Under the absence of the allele for early heading at *qDH3*, the *qDH6* × *qDH7* interaction pattern was similar to the *qDH3* × *qDH6* and *qDH3* × *qDH7* interactions (Figure 4d), i.e., the effect of one QTL accelerating heading was greater under the absence of the allele for early heading at the other QTL. However, under the presence of the allele for early heading at *qDH3*, the *qDH6* × *qDH7* interaction pattern was reversed (Figure 4e), i.e., the effect of one QTL accelerating heading was greater under the presence of the allele for early heading at the other QTL.

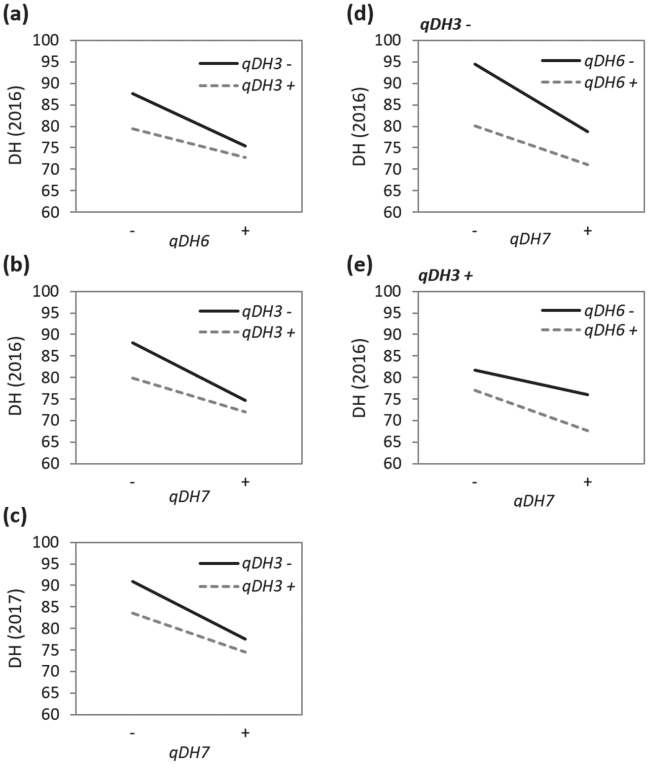


Figure 4. Interactions among the three QTLs for days to heading. (**a**) *qDH3* × *qDH6* interaction in 2016; (**b**) *qDH3* × *qDH7* interaction in 2016; (**c**) *qDH3* × *qDH7* interaction in 2017; (**d**) *qDH6* × *qDH7* interaction under the absence of the allele for early heading at *qDH3* in 2016; (**e**) *qDH6* × *qDH7* under the presence of the allele for early heading at *qDH3* in 2016. Only the significant ($p < 0.05$) interactions are plotted (see Table 2).

3.3. Hd16, Hd1, and Ghd7 Underlying the Days to Heading QTLs

As the three QTLs for DH encompass previously isolated heading date genes—*Hd16* (*Os03g0793500*) at *qDH3*, *Hd1* (*Os06g0275000*) at *qDH6*, and *Ghd7* (*Os07g0261200*) at *qDH7* [7,10,23]—we sequenced the coding regions of the three genes from Koshihikari and Baegilmi to search sequence polymorphisms.

It was previously shown that relative to the functional *Hd16* allele of Nipponbare, Koshihikari carries a G-to-A (alanine-to-threonine) mutation in the 10th exon of *Hd16* that is responsible for early heading under long day [23]. Sequence analysis of the 10th exon of *Hd16* revealed that the sequence of Baegilmi is identical to Nipponbare (Figure 5a). This was consistent with our QTL analysis where Koshihikari provided the allele for early heading and Baegilmi provided the allele for late heading at *qDH3* (Table 1), suggesting *Hd16* as a strong candidate gene for *qDH3*.

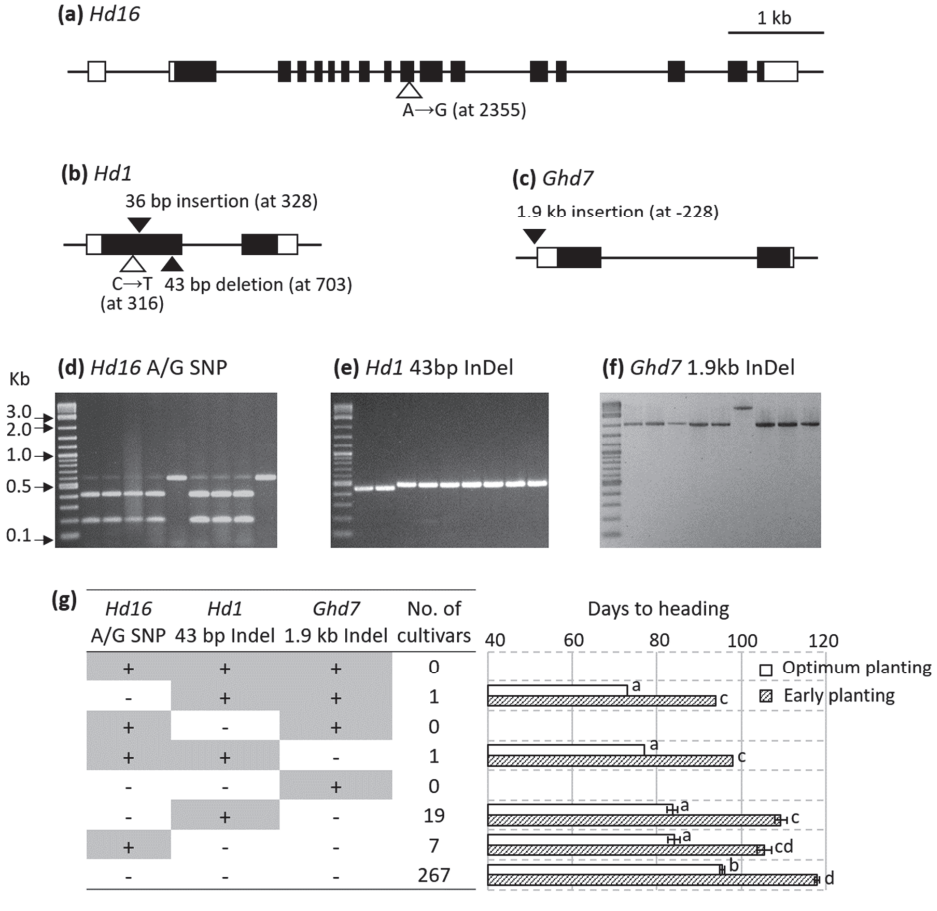


Figure 5. Sequence polymorphisms of *Hd16* (*Os03g0793500*), *Hd1* (*Os06g0275000*), and *Ghd7* (*Os07g0261200*) in Baegilmi relative to Koshihikari (**a**–**c**) and their genotyping among Korean rice cultivars (**d**–**g**). Coding region and untranslated region are depicted as filled and unfilled squares, respectively, while introns are depicted as black lines (**a**–**c**). Unfilled and filled triangles indicate SNP and insertion/deletion of Baegilmi in comparison with Koshihikari, respectively (**a**–**c**). The *Hd16* A SNP (Koshihikari allele) and the G SNP (Baegilmi allele) are visualized as intact (579 bp) and digested (393 bp + 186 bp) bands, respectively (**d**). The *Hd1* 43 bp deletion (Baegilmi allele) is visualized as a lower band (**e**). The *Ghd7* 1.9 kb insertion (Baegilmi allele) is visualized as an upper band (**f**). + and − indicate the presence and absence of the allele for early heading at each gene among 295 Korean rice cultivars, respectively, and the error bars in the bar graph indicate standard errors (**g**). Sowing and transplanting dates for optimum planting were 9 May and 1 June, respectively in 2018, and those for early planting were 10 April and 9 May, respectively in 2019, in Wanju, Korea. Different letters next to the bars indicate significant difference ($p < 0.05$) according to the Duncan's multiple range test in each year.

Sequence analysis of the *Hd1* coding region revealed that Koshihikari carries the functional *Hd1* allele identical to Nipponbare as previously reported [22], while Baegilmi carries a non-functional allele with three sequence polymorphisms relative to Koshihikari—a C-to-T SNP, a 36-bp insertion, and a 43-bp deletion in the first exon of *Hd1* (Figure 5b). The *Hd1* polymorphisms in Baegilmi are

identical to those reported in HS66 (GeneBank ID AB041841), the γ adiation mutant of the Japanese cultivar Ginbouzu [7].

While the sequence analysis found no polymorphism in the coding region of *Ghd7* between Koshihikari and Baegilmi, we identified a 1901 bp insertion at the −228 bp position from the start codon of *Ghd7* in Baegilmi (Figure 5c). The position of the 1901 bp insertion and its sequence were identical to those of the putative retrotransposon inserted in the promoter region of *Ghd7* reported in the Japanese cultivar Sorachi (GenBank ID LC472532) [24].

3.4. Allelic Composition of Hd16, Hd1, and Ghd7 among Commercial Rice Cultivars

To study the allelic composition of the three DH genes in commercial rice cultivars, we screened 295 Korean rice cultivars released in 1979–2017 using the molecular markers designed to genotype the sequence polymorphisms in *Hd16* (A/G SNP), *Hd1* (43 bp Indel), and *Ghd7* (1.9 kb Indel) (Supplementary Table S1, Figure 5d–f). Of the eight (2 × 2 × 2) possible allelic combinations, five were observed among the 295 cultivars (Supplementary Table S2, Figure 5g). The majority (>90%) carried the alleles for late heading for all three polymorphisms, namely, the *Hd16* G SNP, the *Hd1* 43-bp insertion, and the *Ghd7* 1.9-kb deletion. Among 26 cultivars carrying an allele for early heading in one of three DH genes, 19 carried the *Hd1* 43-bp deletion and seven carried the *Hd16* A SNP. Only two out of the 295 Korean rice cultivars carried alleles for early heading in two of the three DH genes—Jopum carried the alleles for early heading at *Hd16* and *Hd1*, while Baegilmi carried the alleles for early heading at *Hd1* and *Ghd7*. None of the 295 cultivars carried the alleles for early heading at all three genes. The average DH of the five cultivar groups according to the allelic combinations of *Hd16* (A/G SNP), *Hd1* (43 bp Indel), and *Ghd7* (1.9 kb Indel) indicated that pyramiding the alleles for early heading at the three genes would accelerate DH effectively in the genetic background of commercial Korean rice cultivars (Figure 5g).

4. Discussion

Baegilmi is an extremely early heading Korean rice cultivar that can be incorporated in diverse double cropping systems to improve land use efficiency [14]. To genetically dissect the early heading characteristics of Baegilmi, we used a RIL population derived from the cross between Koshihikari and Baegilmi and identified the chromosomal locations harboring the genes controlling DH. We identified three major QTLs for DH, of which the allele for early heading was contributed by Baegilmi at *qDH6* and *qDH7*, and by Koshihikari at *qDH3*. As pyramiding of the alleles for early heading at each QTL was effective in accelerating heading, the Koshihikari × Baegilmi RIL population provides useful breeding lines with different allelic combinations of the three QTLs conferring varying DH, ranging from 71 days to 98 days under the natural long day condition in Wanju (36° N), Korea (Figure 3b). Therefore, the RIL population would be useful for selecting promising breeding lines potentially inheriting the good eating quality of Koshihikari with various heading dates that can be adapted in different environments and cropping patterns. To support this idea, we are currently evaluating the RILs with different allelic combinations of the three DH QTLs in terms of important agronomic traits such as yield performance and grain quality.

All three DH QTLs identified in this study harbor previously cloned heading date genes, i.e., *Hd1* [7], *Ghd7* [10], and *Hd16* [23] at *qDH6*, *qDH7*, and *qDH3*, respectively. *Hd1*, the ortholog of Arabidopsis *CO*, is the first heading date gene cloned in rice [7]. While *Hd1* accelerates heading under short day by upregulating *Hd3a*, it downregulates *Hd3a* under long day and represses heading [6–8,25]. Due to this dual function of *Hd1* depending on daylength, nonfunctional *hd1* alleles confer early heading phenotype under long day as the loss-of-function of *Hd1* de-represses *Hd3a* and promotes heading [26,27]. The nonfunctional *hd1* alleles have played an important role in expanding rice cultivation in high latitude areas with natural long day conditions where early heading is critical for rice plants to mature before the cold winter [27]. *Hd1* sequencing revealed that Baegimli carries a nonfunctional *hd1* allele due to the 43 bp deletion in exon 1 causing a frameshift (Figure 5b). The 43 bp deletion in *Hd1* was first reported in the γ ray mutant cultivar HS66 [7], and has been observed in many

rice cultivars grown in high latitude areas such as Jilin (44° N) and Liaoning (41° N) of China [28] and Italy (35–47° N) [27]. Among the 295 Korean rice cultivars screened in this study, 21 including Baegilmi carried the nonfunctional *hd1* allele due to the 43 bp deletion and showed earlier heading phenotype compared to those without the 43 bp deletion (Figure 5g). As there are many other nonfunctional *hd1* alleles arising from frameshift or premature stop codon at different genic positions [22,26–28], further work is underway to define additional allelic variation in *Hd1* among the Korean rice cultivars.

Similar to *Hd1*, *Ghd7* encoding a CCT (CO, CO-LIKE, and TIMING OF CAB1) domain protein delays heading under long day by downregulating *Hd3a* [10]. In addition, similar to nonfunctional *hd1*, nonfunctional *ghd7* alleles such as *ghd7-0* (full gene deletion) and *ghd7-0a* (premature stop codon) accelerate heading under long day and contribute to the adaption of rice in high latitude areas [10,28]. Relative to Koshihikari, Baegilmi has a 1901 bp insertion in the promoter region (−228 bp from the start codon) of *Ghd7* (Figure 5c). The 1901 bp insertion is identical to that reported in *Ghd7* from a Japanese cultivar Sorachi (GenBank ID LC472533) with 449 bp long terminal repeats at each end and 5 bp (AGGTA) target site duplication (Fujino and Yamanouch 2020). This allele has been mainly found in rice cultivars bred in Hokkaido (42–45° N) of Japan, providing an important genetic source for rice adaption in high latitude [24,29]. Among the 295 Korean rice cultivars screened in this study, Baegilmi was the only cultivar carrying the loss-of-function *ghd7* allele due to the 1901 bp retrotransposon insertion (Figure 5g), suggesting that this allele would provide a valuable genetic source to accelerate heading in breeding programs. Interestingly, the same retrotransposon insertion has been also found in exon 2 of *Hd1* from a Taiwanese landrace Muteka (GenBank ID KR230393), creating a loss-of-function *hd1* allele [30]. This illustrates that transposable elements can play important roles in functionally diversifying the heading date genes.

Hd16 encoding casein kinase I delays heading under long day by activating *Ghd7* through phosphorylation, and the missense mutation in exon 10 of *Hd16* from Koshihikari decreases the kinase activity of Hd16, thus accelerating heading under long day [23]. The screening of over 300 worldwide rice cultivars and 30 wild rice accessions revealed that the Koshihikari-type *Hd16* allele is found only in Japanese *japonica* cultivars [23]. Among the 295 Korean rice cultivars, only eight (i.e., Manan, Boseok, Jinkwang, Pungmi, Pungmi 1, Cheonga, Jungsaenggold, and Jopum) carried the Koshihikari-type *Hd16* and showed earlier heading phenotype compared to those carrying the Nipponbare-type *Hd16* (Figure 5g). These cultivars would provide useful genetic sources for developing early heading rice cultivars.

We previously reported that Baegilmi was developed by chemically mutagenizing Koshihikari [14]. However, it is extremely unlikely that the EMS mutagenesis would have created the above-mentioned sequence polymorphisms in *Hd1*, *Ghd7*, and *Hd16* that had been reported previously. This indicates that in the initial stages of breeding Baegilmi, the Koshihikari seed stock used for mutagenesis might have been contaminated during seed handling (e.g., threshing) with a different germplasm which may carry all or some of the previously known alleles, i.e., the HS66-type *Hd1*, the Sorachi-type *Ghd7*, and the Nipponbare-type *Hd16*. We currently do not know the exact source of the potential contamination. Although lacking a clear parentage, the practical value of Baegilmi as a cultivar is unaffected because Baegilmi carries a unique allelic composition of the three heading date genes and exhibits the second earliest heading date among the 295 Korean rice cultivars released in 1979–2017 [31]. Our study also suggests that developing cultivars with earlier heading would be possible by pyramiding the alleles for early heading at *Hd1*, *Ghd7*, and *Hd16*, as none of the 295 cultivars screened in this study carried the alleles for early heading at all three genes.

5. Conclusions

Baegilmi is a *japonica* rice cultivar with extremely early heading that is useful for diversifying cropping patterns in Korea. From the Koshihikari × Baegilmi RIL population, we detected three QTLs for days to heading, *qDH3*, *qDH6*, and *qDH7*. Different allelic combinations of the three QTLs in the RIL population provided breeding lines that can be useful for developing high quality *japonica* rice

cultivars inheriting Koshihikari's high eating quality with varying DH. Screening Korean rice cultivars for the allelic compositions of *Hd16* (A/G SNP), *Hd1* (43 bp Indel), and *Ghd7* (1.9 kb Indel) underlying *qDH3*, *qDH6*, and *qDH7*, respectively, showed that few cultivars carry one or more alleles for early heading at the three genes, suggesting that DH can be further accelerated and fine-tuned in breeding programs by using the allelic variations in *Hd16*, *Hd1*, and *Ghd7*.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/9/957/s1, Supplementary Figure S1: Linkage map from the Koshihikari × Baegilmi RIL population (n = 142) using 128 SNP markers, Supplementary Table S1: Molecular markers designed to genotype sequence polymorphisms in *Hd16*, *Hd1*, and *Ghd7*, Supplementary Table S2: Allelic distributions of the polymorphisms in *Hd16*, *Hd1*, and *Ghd7* in 295 Korean commercial rice cultivars.

Author Contributions: Conceptualization by Y.M. and J.-U.J.; phenotype evaluation and data curation by J.-M.J., S.-K.H., J.K., and C.L.; genotyping analysis by G.P.L.; statistical analysis by Y.M. and J.-U.J.; writing—original draft preparation by Y.M.; writing—review and editing by J.-U.J.; project administration by Y.M. and J.-U.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Rural Development Administration (RDA), Republic of Korea, grant number PJ01357205.

Acknowledgments: We thank Sunghee Kim and Kyoungran Yoo for providing excellent technical supports.

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

References

1. Shrestha, R.; Gómez-Ariza, J.; Brambilla, V.; Fornara, F. Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals. *Ann. Bot.* **2014**, *114*, 1445–1458. [CrossRef] [PubMed]
2. Hori, K.; Matsubara, K.; Yano, M. Genetic control of flowering time in rice: Integration of Mendelian genetics and genomics. *Theor. Appl. Genet.* **2016**, *129*, 2241–2252. [CrossRef] [PubMed]
3. Matsubara, K.; Yano, M. Genetic and molecular dissection of flowering time control in rice. In *Rice Genomics, Genetics and Breeding*; Sasaki, T., Ashikari, M., Eds.; Springer: Singapore, 2018; pp. 177–190.
4. Blümel, M.; Dally, N.; Jung, C. Flowering time regulation in crops—What did we learn from Arabidopsis? *Curr. Opin. Biotech.* **2015**, *32*, 121–129. [CrossRef] [PubMed]
5. Komiya, R.; Yokoi, S.; Shimamoto, K. A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. *Development* **2009**, *136*, 3443–3450. [CrossRef] [PubMed]
6. Tamaki, S.; Matsuo, S.; Hann, L.W.; Yokoi, S.; Shimamoto, K. Hd3a protein is a mobile flowering signal in rice. *Science* **2007**, *316*, 1033–1036. [CrossRef] [PubMed]
7. Yano, M.; Katayose, Y.; Ashikari, M.; Yamanouchi, U.; Monna, L.; Fuse, T.; Baba, T.; Yamamoto, K.; Umehara, Y.; Nagamura, Y.; et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell* **2000**, *12*, 2473–2483. [CrossRef] [PubMed]
8. Hayama, R.; Yokoi, S.; Tamaki, S.; Yano, M.; Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **2003**, *422*, 719–722. [CrossRef]
9. Doi, K.; Izawa, T.; Fuse, T.; Yamanouchi, U.; Kubo, T.; Shimatani, Z.; Yano, M.; Yoshimura, A. Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes Dev.* **2004**, *18*, 926–936. [CrossRef]
10. Xue, W.; Xing, Y.; Weng, X.; Zhao, Y.; Tang, W.; Wang, L.; Zhou, H.; Yu, S.; Xu, C.; Li, X.; et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **2008**, *40*, 761–767. [CrossRef]
11. Kim, S.Y.; Seo, J.H.; Bae, H.K.; Hwang, C.D.; Ko, J.M. Rice cultivars adaptable for rice based cropping systems in a paddy field in the Yeongnam plain area of Korea. *Korean J. Agri. Sci.* **2018**, *45*, 355–363. [CrossRef]
12. Kim, Y.D.; Kang, S.G.; Ku, B.I.; Choi, M.K.; Park, H.K.; Park, T.S.; Back, N.H.; Kim, S.J.; Ko, J.K. Selection of suitable rice cultivars for silage barley-rice double cropping in Honam plain area. *Korean J. Int. Agric.* **2010**, *22*, 337–340. [CrossRef] [PubMed]
13. Park, H.; Jeung, J.; Cho, Y.; Kim, B. Development of early maturing rice lines with genes conferring resistance to bacterial blight and rice stripe virus for enhancing the adaptability in plain area. *Korean J. Breed. Sci* **2015**, *47*, 118–127. [CrossRef]

14. Mo, Y.; Jeong, J.-M.; Kim, W.-J.; Kim, B.-K.; Jeung, J.-U. 'Baegilmi', an extremely early maturing blast resistant rice with good grain appearance. *Korean J. Breed. Sci.* **2019**, *51*, 151–159. [CrossRef]
15. Kobayashi, A.; Hori, K.; Yamamoto, T.; Yano, M. Koshihikari: A premium short-grain rice cultivar—Its expansion and breeding in Japan. *Rice* **2018**, *11*, 15. [CrossRef] [PubMed]
16. Murray, M.G.; Thompson, W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [CrossRef]
17. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef]
18. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
19. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, e90346. [CrossRef]
20. Jang, Y.J.; Seo, M.; Hersh, C.P.; Rhee, S.J.; Kim, Y.; Lee, G.P. An evolutionarily conserved non-synonymous SNP in a leucine-rich repeat domain determines anthracnose resistance in watermelon. *Theor. Appl. Genet.* **2019**, *132*, 473–488. [CrossRef]
21. Meng, L.; Li, H.; Zhang, L.; Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop. J.* **2015**, *3*, 269–283. [CrossRef]
22. Kim, S.R.; Torollo, G.; Yoon, M.R.; Kwak, J.; Lee, C.K.; Prahalada, G.D.; Choi, I.R.; Yeo, U.S.; Jeong, O.Y.; Jena, K.K.; et al. Loss-of-function alleles of *Heading date 1* (*Hd1*) are associated with adaptation of temperate *japonica* rice plants to the tropical region. *Front. Plant Sci.* **2018**, *9*, 1827. [CrossRef]
23. Hori, K.; Ogiso-Tanaka, E.; Matsubara, K.; Yamanouchi, U.; Ebana, K.; Yano, M. *Hd16*, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J.* **2013**, *76*, 36–46. [CrossRef]
24. Fujino, K.; Yamanouchi, U. Genetic effect of a new allele for the flowering time locus *Ghd7* in rice. *Breed. Sci.* **2020**, *70*, 342–346. [CrossRef]
25. Kojima, S.; Takahashi, Y.; Kobayashi, Y.; Monna, L.; Sasaki, T.; Araki, T.; Yano, M. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol.* **2002**, *43*, 1096–1105. [CrossRef] [PubMed]
26. Takahashi, Y.; Teshima, K.M.; Yokoi, S.; Innan, H.; Shimamoto, K. Variations in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4555–4560. [CrossRef] [PubMed]
27. Gómez-Ariza, J.; Galbiati, F.; Goretti, D.; Brambilla, V.; Shrestha, R.; Pappolla, A.; Courtois, B.; Fornara, F. Loss of floral repressor function adapts rice to higher latitudes in Europe. *J. Exp. Bot.* **2015**, *66*, 2027–2039. [CrossRef] [PubMed]
28. Zhang, J.; Zhou, X.; Yan, W.; Zhang, Z.; Lu, L.; Han, Z.; Zhao, H.; Liu, H.; Song, P.; Hu, Y.; et al. Combinations of the *Ghd7*, *Ghd8* and *Hd1* genes largely define the ecogeographical adaptation and yield potential of cultivated rice. *New Phytol.* **2015**, *208*, 1056–1066. [CrossRef] [PubMed]
29. Saito, H.; Okumoto, Y.; Tsukiyama, T.; Xu, C.; Teraishi, M.; Tanisaka, T. Allelic differentiation at the *E1*/*Ghd7* locus has allowed expansion of rice cultivation area. *Plants* **2019**, *8*, 550. [CrossRef]
30. Wei, F.J.; Tsai, Y.C.; Wu, H.P.; Huang, L.T.; Chen, Y.C.; Chen, Y.F.; Wu, C.C.; Tseng, Y.T.; Hsing, Y.C. Both *Hd1* and *Ehd1* are important for artificial selection of flowering time in cultivated rice. *Plant Sci.* **2015**, *242*, 187–194. [CrossRef] [PubMed]
31. Lee, C.M.; Park, H.S.; Baek, M.K.; Suh, J.P.; Kim, C.S.; Lee, K.M.; Park, S.G.; Cho, Y.C. Characterization of grain-related traits of 300 Korean rice varieties. In Proceedings of the 2019 Korean Society of Breeding Science (KSBS) & Society for the Advancement of Breeding Research in Asia and Oceania (SABRAO), Gwangju, Korea, 2–5 July 2019.

genes

MDPI

Article

Drought-Tolerance Gene Identification Using Genome Comparison and Co-Expression Network Analysis of Chromosome Substitution Lines in Rice

Chutarat Punchkhon [1,2], Kitiporn Plaimas [3,4], Teerapong Buaboocha [4,5], Jonaliza L. Siangliw [6], Theerayut Toojinda [6], Luca Comai [7], Nuria De Diego [8], Lukáš Spíchal [8] and Supachitra Chadchawan [2,5,*]

1 Program in Biotechnology, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand; chutarat_tom@hotmail.com
2 Center of Excellence in Environment and Plant Physiology, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand
3 Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand; kitiporn.p@chula.ac.th
4 Omics Science and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand; Teerapong.b@chula.ac.th
5 Molecular Crop Research Unit, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand
6 National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, 113 Phahonyothin Rd. Khlong Nueng, Khlong Luang, Pathumthani 12120, Thailand; jsiangliw@gmail.com (J.L.S.); theerayut@biotec.or.th (T.T.)
7 Genome Center and Department of Plant Biology, UC Davis Genome Center, UC Davis, Davis, CA 95616, USA; lcomai@ucdavis.edu
8 Department of Chemical Biology and Genetics, Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University, Šlechtitelů 27, CZ-783 71 Olomouc, Czech Republic; nuria.de@upol.cz (N.D.D.); lukas.spichal@upol.cz (L.S.)
* Correspondence: Supachitra.c@chula.ac.th or s_chadchawan@hotmail.com

Received: 25 August 2020; Accepted: 9 October 2020; Published: 14 October 2020

Abstract: Drought stress limits plant growth and productivity. It triggers many responses by inducing changes in plant morphology and physiology. KDML105 rice is a key rice variety in Thailand and is normally grown in the northeastern part of the country. The chromosome segment substitution lines (CSSLs) were developed by transferring putative drought tolerance loci (QTLs) on chromosome 1, 3, 4, 8, or 9 into the KDML105 rice genome. CSSL104 is a drought-tolerant line with higher net photosynthesis and leaf water potential than KDML105 rice. The analysis of CSSL104 gene regulation identified the loci associated with these traits via gene co-expression network analysis. Most of the predicted genes are involved in the photosynthesis process. These genes are also conserved in *Arabidopsis thaliana*. Seven genes encoding chloroplast proteins were selected for further analysis through characterization of *Arabidopsis* tagged mutants. The response of these mutants to drought stress was analyzed daily for seven days after treatment by scoring green tissue areas via the PlantScreen™ XYZ system. Mutation of these genes affected green areas of the plant and stability index under drought stress, suggesting their involvement in drought tolerance.

Keywords: CSSLs; drought stress; 'KDML105' rice; co-expression network

1. Introduction

Rice (*Oryza sativa* L.) is one of the important cereal crops of the world [1]. In Thailand, rice is the major agricultural export, especially Khao Dawk Mali 105 (KDML105) rice. The cooked kernels of

KDML105 rice have a highly prized scent and texture [2]. KDML105 rice is normally grown in the northeast of Thailand, based on rain with limited irrigation [3]. Therefore, it is always affected by drought stress, leading to the reduction in growth and yield.

Drought stress affects plant morphology, physiology, and molecular mechanisms. Upon drought stress, cell turgor pressure is decreased due to low water potential in cells. This causes a decrease in the relative water content, leaf water potential, stomatal conductance, and transpiration rate [4]. Cell expansion and elongation are inhibited, resulting in the reduction of plant height, leaf area, growth, and yield [5]. Photosynthesis is one of the important physiological mechanisms affected by drought stress. The decrease in leaf expansion, leaf area, and stomatal conductance limits CO_2 uptake [6]. The photosynthetic pigments (chlorophyll *a*, chlorophyll *b*, and carotenoids) can also be damaged by drought stress, resulting in their degradation and decreased light absorption and maximum photosynthetic rate [7]. During drought stress, phosphoenolpyruvate carboxylase, nicotinamide adenine dinucleotide phosphate-malic enzyme, Rubisco, fructose-1,6-bisphosphatase, and pyruvate orthophosphate dikinase activities are decreased, which can reduce the photosynthetic and electron transport rate [8]. The physiological responses to drought tolerance include osmotic adjustment, osmoprotection, antioxidation, scavenging defense, and photorespiration [9,10].

Kanjoo et al. [11] developed chromosome segment substitution lines (CSSLs) in the background of variety KDML105. CT9993, a variety with a good rooting system, and IR62266, a variety with high osmotic adjustment ability, were hybridized, and their F1 was used to generate double haploid lines. The double haploid lines were evaluated for yield, yield components, and morpho-physiological characters under drought-stress conditions, defining drought-tolerant quantitative trait loci (DT-QTLs) on chromosomes 1, 3, 4, 8 and 9. The doubled haploid line DH212 carries CT9993 alleles on in all chromosomes, while DH103 has IR62266 alleles on chromosome 8. These lines were selected as donor lines for CSSL development. Repeated crossing to KDML105 resulted in CSSLs with the putative drought-tolerant genes from chromosome 8, donated by DH103, and the CSSLs with the DT-QTL from chromosome, 1, 3, 4 and 9, donated by DH212 [3].

CSSL104 is a drought-tolerant KDML105 CSSL carrying the chromosome 8 introgression from inbred DH103 [11]. Compared to KDML105, CSSL104 had higher relative water content, higher chlorophyll fluorescence (F_v/F_m), and lower leaf drying score under 50% field capacity drought-stress conditions [12]. This implied that the region introgressed from DH103 carried the putative drought-tolerant genes.

To investigate the mechanisms affected by the introgression of DH103 genes, the physiological responses to drought stress of CSSL104 were evaluated relative to KDML105 and the DH103 donor line. Then, drought-tolerance genes were predicted based on genomic sequence comparison and co-expression network analysis. Finally, the putative drought-tolerance genes were validated using the corresponding *Arabidopsis* mutants. This study will be beneficial to the future development of drought-tolerant rice.

2. Materials and Methods

2.1. Plant Materials

We used CSSLs with the genetic background of KDML105 rice and containing a putative drought-tolerance segment of chromosome 8 from DH103 between markers RM5353 and RM3480 [11]. These are CSSL97, CSSL104, CSSL106, CSSL107, and parental lines (KDML105 and DH103). They were used to study drought-stress responses. All rice seeds were provided by the Innovative Plant Biotechnology and Precision Agriculture Research Team (APBT) at the National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand.

2.2. Evaluation of Physiological Responses at Vegetative Stage under Drought-Stress Conditions

2.2.1. Rice Growth Condition

KDML105 rice, DH103, and CSSL seeds were incubated at 60 °C for 48 h before planting. The seeds were then germinated by soaking in distilled water for seven days in plastic cups. Rice seedlings were transferred to a plastic tray and continuously grown in WP nutrient solution [13]. Twenty-eight days after germination, rice plants were drought-stressed for three days by the addition of 10% polyethylene glycol 6000 (PEG6000). This condition was previously shown to cause drought stress in rice [14,15]. In order to induce the stronger drought-stress condition, after treatment with WP nutrient solution with 10% PEG6000 for three days, the solution was then changed to WP nutrient solution with 15% PEG. Plants grown in WP nutrient solution without PEG6000 were used as controls. A complete randomized design (CRD) with four replicates was used for physiological evaluation in each parameter.

2.2.2. Net Photosynthesis Rate and Leaf Water Status Detection

The net photosynthesis rate (Pn) of twenty-eight-day-old rice plants was determined with a LI-6400 XT portable photosynthesis system (LI-COR, Lincoln, NE, USA). The measurement was taken at the middle part of the youngest fully expanded leaves between 9 am and 2 pm, with the following conditions: the molar flow of air per unit leaf area was 500 mmol l^{-1} m^{-2} s^{-1}, the photosynthetically active radiation (PAR) at the leaf surface was 1200 μmol m^{-2} s^{-1}, the leaf temperature ranged from 30.0 to 37.0 °C, with a CO_2 concentration of 380.0 mol mol^{-1}. The leaf water potential (LWP) was measured in the youngest fully expanded leaves using plant water status console model 3005 (Soilmoisture, Goleta, CA, USA).

2.3. Identification of Drought-Tolerance Gene

2.3.1. Whole Genome Sequencing

The aboveground parts of KDML105, DH103, and CSSL104 rice plants were collected at fourteen days after germination. Rice genomic DNA was extracted using a Genomic DNA Mini Kit, 'Plant' (Geneaid, New Taipei City, Taiwan). Genomic DNA libraries were prepared for sequencing by using an Illumina genome analyzer (Illumina, San Diego, CA, USA) with Illumina HiSeq3000's protocol. For genome analyses, the sequence reads were classified into specific categories using the pipeline developed by Missirian et al. [16]. The rice genomic sequence from the Rice Genome Annotation Project database [17] was used as a reference genome [18] to map the sequence reads. The raw reads were submitted to GenBank at the NCBI under BioProject no. PRJNA659381. Bioinformatic tools were used to compare the genome of CSSL104 with the KDML105 rice genome to identify loci containing different single nucleotide polymorphisms (SNPs). These loci may contribute to the drought-tolerance phenotypes of CSSL104. The genome comparison first started by discarding all SNPs shared by both CSSL104 and KDML105. The remaining differential SNPs were counted within a sliding window of 5000 background nucleotides. To visualize the chromosome plots with the marks of different SNPs' loci, the window region containing more than 100 SNPs in CSSL104 with different nucleotides from KDML105 was marked as a blue line on the chromosome plots. The analysis of the locations of SNPs in the candidate genes was analyzed.

2.3.2. Gene Co-Expression Network Analysis

The rice loci containing the different SNPs were used for a gene co-expression network analysis. To predict the important loci involved in drought tolerance, a rice oligonucleotide array database was used with abiotic stress-induced gene expression data with a correlation coefficient cut-off of 0.95 [19]. The predicted loci were searched for gene ontology and expression patterns from the rice expression database [20].

2.4. Identification of Drought-Tolerance Gene Function in Arabidopsis

2.4.1. *Arabidopsis* Homologous Gene

The best candidate genes were used to search for the homologous gene in *Arabidopsis* from the Rice Genome Annotation Project database [17] and The *Arabidopsis* Information Resource [19]. *Arabidopsis* mutant lines with T-DNA insertion in the selected gene were ordered from the Arabidopsis Biological Resource Center (ABRC). The *Arabidopsis* seeds were screened for homozygous mutant lines via specific primers, LP and RP, to the gene of interest; the LB primer was used for a specific T-DNA region.

2.4.2. *Arabidopsis* Growth Condition

Four days after germination, *Arabidopsis thaliana* ecotype Col-0 seeds and seven mutant lines [21] including *at1g74880*, *at5g54270*, *at3g63190*, *at4g11960*, *at4g22890*, *at2g27680*, and *at4g34830* were sowed and transferred to 48-well plates, containing Murashige and Skoog (MS) agar media for normal conditions, MS agar media supplemented with 75 mM mannitol for mild drought stress, or MS agar media supplemented with 150 mM mannitol for severe drought stress. The plants were then grown in a growth chamber at 22 °C/20 °C, 16/8 h light/dark cycle, 120 µmol photons of PAR $m^{-2} \cdot s^{-1}$, and 60% humidity. RGB imaging was used to collect the green area of plants twice a day via a PlantScreenTM XYZ system (Photon Systems Instruments, Drásov, Czech Republic) [22].

2.5. Statistical Analysis

Analysis of variance (ANOVA) and Duncan's Multiple Range Test (DMRT) were used for data analysis by SPSS Statistics program version 22 (IBM, Armonk, NY, USA). The images from the PlantScreenTM XYZ system were analyzed using MATLAB (R2015; MathWorks, Inc., Natick, MA, USA), and the data were analyzed by independent *t*-tests using SPSS.

3. Results

3.1. Evaluation of Physiological Responses of CSSLs of KDML105 under Drought-Stress Conditions

Selected CSSLs, namely CSSL97, CSSL104, CSSL106, and CSSL107, were evaluated for drought tolerance by growing the seedlings in soil with 100% or 50% field capacity. In normal growth conditions (100% field capacity), all of the lines were similar (Figure 1A,C), but they differed under 50% field capacity (Figure 1B,D). CSSL104 displayed the most drought-tolerant phenotype, with the lowest leaf dying score and the highest photosystem II (PSII) efficiency (F_v/F_m). This was similar to the performance of DH103 (the drought-tolerant parental line). The highest leaf-death score was detected in CSSL106, while CSSL97 had the lowest PSII efficiency under drought stress. These data suggest that CSSL104 is the most drought-tolerant line, while CSSL97 and CSSL106 are the most susceptible. Therefore, CSSL104 was selected for further characterization.

CSSL104 and its parental lines, KDML105 and DH103 rice, were grown in nutrient solutions corresponding to normal growth conditions and in a solution supplemented with 15% PEG for the drought-stress treatment, which caused about a 50–180% reduction in leaf water potential (Table 1). After nine days of drought stress, we measured the highest reduction of leaf water potential for all lines. Under these conditions, CSSL104 had the lowest leaf water potential at −7.75 MPa, which was about threefold lower than the LWP of the plants grown in normal conditions. The parental lines, KDML105 and DH103, had about a twofold reduction in LWP. Drought stress reduced all parameters of photosynthesis, including net photosynthesis rate, stomatal conductance, transpiration rate, intercellular CO_2 concentration, ΦPSII, and electron transport rate in all lines. However, after six days under drought stress, CSSL104 had a significantly higher photosynthetic rate than the KDML105 parent. In addition, CSSL104 had a greater tendency than KDML105 toward higher values for all photosynthetic parameters after nine days of drought treatment (Table 1).

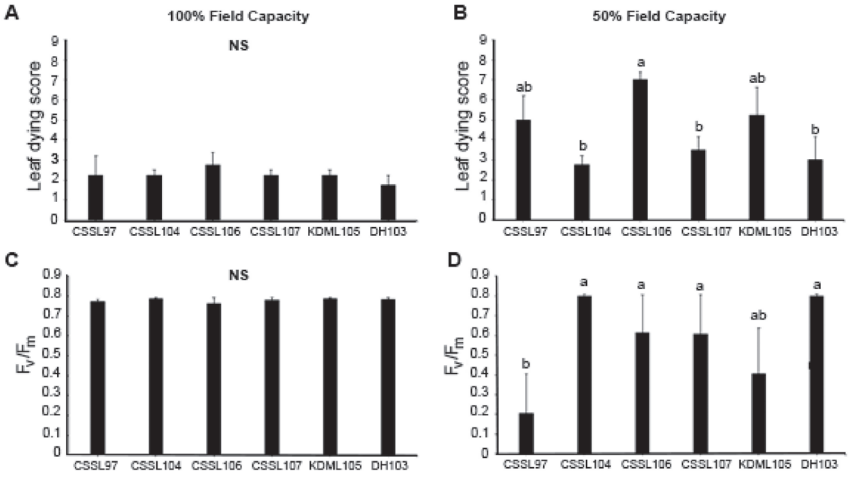


Figure 1. Response to drought stress in chromosome segment substitution lines (CSSLs) and parents. CSSLs, CSSL97, CSSL104, CSSL106, and CSSL107, and their parental lines KDML105 and DH103, were compared for leaf death (leaf dying score) and photosystem II efficiency (F_v/F_m) under (**A**,**C**) normal (100% field capacity) and (**B**,**D**) drought-stress (50% field capacity) conditions. The mean + 1 standard error (SE) was derived from four replicates. Means with a different lowercase letter above them are significantly different ($p < 0.05$). NS demonstrates no significant difference among lines.

Table 1. Photosynthetic performance of CSSL104 and parent lines. Net photosynthesis rate, transpiration rate, stomatal conductance, ΦPSII, electron transport rate, intercellular CO_2 concentration, F_v'/F_m', relative chlorophyll content (determined by portable chlorophyll meter SPAD-501) and leaf water potential of rice at vegetative stage in normal and drought stress conditions. ANOVA and Duncan's Multiple Range Test (DMRT) were used for statistical analysis. The data show mean ± SE. Different superscript letters show the significant difference among lines at p value ≤ 0.05.

Conditions	**Normal (0% PEG)**				**Drought Stress (15% PEG)**			
Timing (Days after Stress)	**0**	**3**	**6**	**9**	**0**	**3**	**6**	**9**
Net photosynthesis rate ($\mu mol \cdot m^{-2} \cdot s^{-1}$)								
KDML105	19.44 ± 4.46	14.50 ± 0.81	17.85 ± 1.85	20.17 ± 0.88	19.44 ± 4.46	10.30 ± 1.74	5.51 ± 0.42 ^{b}	8.64 ± 1.11
DH103	14.47 ± 2.31	17.45 ± 1.21	18.92 ± 1.53	21.35 ± 2.02	14.47 ± 2.31	10.67 ± 0.38	10.20 ± 1.23 ^{a}	12.35 ± 0.59
CSSL104	15.12 ± 1.28	16.74 ± 1.10	18.47 ± 1.61	15.76 ± 0.87	15.12 ± 1.28	11.31 ± 1.88	10.31 ± 1.33 ^{a}	9.38 ± 1.79
Transpiration rate ($mmol \cdot m^{-2} \cdot s^{-1}$)								
KDML105	5.26 ± 0.56	4.86 ± 0.44	0.23 ± 0.01	5.10 ± 0.39 ^{b}	5.26 ± 0.56	2.42 ± 0.14	0.18 ± 0.02	1.87 ± 0.26 ^{b}
DH103	4.99 ±0.48	5.87 ± 0.12	0.26 ± 0.02	6.77 ± 0.71 ^{a}	4.99 ±0.48	2.65 ± 0.24	0.17 ± 0.02	3.33 ± 0.10 ^{a}
CSSL104	4.54 ± 0.28	4.99 ± 0.34	0.25 ± 0.01	4.48 ± 0.27 ^{b}	4.54 ± 0.28	2.57 ± 0.32	0.17 ± 0.02	2.16 ± 0.34 ^{b}
Stomatal conductance ($mmol \cdot m^{-2} \cdot s^{-1}$)								
KDML105	0.35 ± 0.09	0.32 ± 0.05	0.32 ± 0.06 ^{b}	0.33 ± 0.04 ^{b}	0.35 ± 0.09	0.12 ± 0.01	0.08 ± 0.01	0.09 ± 0.01 ^{b}
DH103	0.43 ± 0.05	0.44 ± 0.02	0.44 ± 0.03 ^{a}	0.53 ± 0.07 ^{a}	0.43 ± 0.05	0.12 ± 0.01	0.19 ± 0.04	0.17 ± 0.01 ^{a}
CSSL104	0.36 ± 0.03	0.33 ± 0.06	0.33 ± 0.06 ^{b}	0.28 ± 0.02 ^{b}	0.36 ± 0.03	0.15 ± 0.02	0.11 ± 0.02	0.10 ± 0.02 ^{b}
ΦPSII								
KDML105	0.25 ± 0.00	0.23 ± 0.01	0.23 ± 0.01	0.23 ± 0.01 ^{b}	0.25 ± 0.00	0.20 ± 0.01	0.18 ± 0.02	0.15 ± 0.01
DH103	0.22 ± 0.02	0.25 ± 0.02	0.26 ± 0.02	0.29 ± 0.01 ^{a}	0.22 ± 0.02	0.21 ± 0.03	0.17 ± 0.02	0.18 ± 0.02
CSSL104	0.24 ± 0.01	0.21 ± 0.01	0.25 ± 0.01	0.22 ± 0.00 ^{b}	0.24 ± 0.01	0.21 ± 0.01	0.17 ± 0.02	0.17 ± 0.01
Electron transport rate								
KDML105	163.02 ± 2.37	149.08 ± 9.83	153.51 ± 4.96	152.38 ± 3.91 ^{b}	163.02 ± 2.37	130.73 ± 4.36	117.25 ± 10.71	100.65 ± 8.13
DH103	143.72 ± 10.16	163.29 ± 11.08	168.02 ± 11.52	188.68 ± 4.99 ^{a}	143.72 ± 10.16	139.94 ± 16.65	110.58 ± 11.43	118.43 ± 10.95
CSSL104	155.86 ± 6.09	141.86 ± 7.08	161.10 ± 6.45	144.28 ± 2.79 ^{b}	155.86 ± 6.09	133.31 ± 11.19	113.12 ± 12.52	112.18 ± 8.52
Intercellular CO_2 concentration ($\mu mol \cdot mol^{-1}$)								
KDML105	327.27 ± 5.37	305.04 ± 11.48	282.47 ± 14.96 ^{b}	282.98 ± 10.67	327.27 ± 5.37	255.39 ± 18.30	269.74 ± 18.42	215.42 ± 21.55 ^{a}
DH103	330.97 ± 5.82	318.58 ± 3.45	312.79 ± 8.84 ^{a}	313.36 ± 8.12	330.97 ± 5.82	260.64 ± 19.39	286.68 ± 18.62	265.87 ± 8.18 ^{b}
CSSL104	317.29 ± 2.12	294.48 ± 11.18	281.92 ± 13.68 ^{b}	289.50 ± 9.46	317.29 ± 2.12	241.03 ± 14.35	227.89 ± 18.71	237.02 ± 13.81 ^{ab}

Table 1. *Cont.*

Conditions	Normal (0% PEG)				Drought Stress (15% PEG)			
Timing (Days after Stress)	**0**	**3**	**6**	**9**	**0**	**3**	**6**	**9**
F_v'/F_m'								
KDML105	0.54 ± 0.01 [b]	0.53 ± 0.01	0.50 ± 0.01	0.50 ± 0.02	0.54 ± 0.01 [b]	0.50 ± 0.01	0.48 ± 0.01	0.45 ± 0.02
DH103	0.63 ± 0.02 [a]	0.60 ± 0.03	0.55 ± 0.02	0.52 ± 0.01	0.63 ± 0.02 [a]	0.53 ± 0.04	0.53 ± 0.08	0.47 ± 0.03
CSSL104	0.53 ± 0.01 [b]	0.54 ± 0.01	0.53 ± 0.03	0.52 ± 0.01	0.53 ± 0.01 [b]	0.52 ± 0.03	0.48 ± 0.01	0.43 ± 0.02
Relative chlorophyll content								
KDML105	35.93 ± 0.41	36.63 ± 0.95	36.40 ± 0.81	38.30 ± 1.05	35.93 ± 0.41	37.58 ± 0.40	37.10 ± 0.57	34.20 ± 1.02
DH103	34.43 ± 0.43	35.78 ± 0.21	36.55 ± 0.34	38.15 ± 0.95	34.43 ± 0.43	33.25 ± 1.12	32.58 ± 2.18	34.38 ± 0.97
CSSL104	34.90 ± 0.65	35.70 ± 0.39	38.65 ± 1.70	39.00 ± 0.38	34.90 ± 0.65	35.93 ± 1.06	36.25 ± 0.71	34.48 ± 1.21
Leaf water potential (MPa)								
KDML105	−1.20 ± 0.14	−2.70 ± 0.77	−5.15 ± 1.08	−2.90 ± 0.26	−1.60 ± 0.08	−4.05 ± 0.48	−9.00 ± 1.00	−6.85 ± 0.43
DH103	−1.35 ± 0.10	−2.60 ± 0.54	−3.90 ± 0.33	−3.25 ± 0.10	−2.05 ± 0.32	−4.60 ± 0.50	−7.00 ± 0.38	−6.95 ± 1.03
CSSL104	−1.20 ± 0.14	−1.95 ± 0.15	−4.20 ± 1.29	−2.70 ± 0.13	−1.65 ± 0.05	−4.08 ± 0.31	−7.10 ± 1.04	−7.75 ± 1.50

3.2. Whole Genome Sequence Comparison between CSSL104 and KDML10′ and Co-Expression Network Analysis Revealed that Major Hub Genes Have a Role in Photosynthesis

We compared whole genome sequences of CSSL104 and its drought susceptible parental line KDML105 to define the genes responsible for drought tolerance in CSSL104. A total of 101,950 SNPs located on 3440 genes were detected. The regions with a high density of SNPs were on chromosomes 1, 8, 9, and 11 (Figure 2).

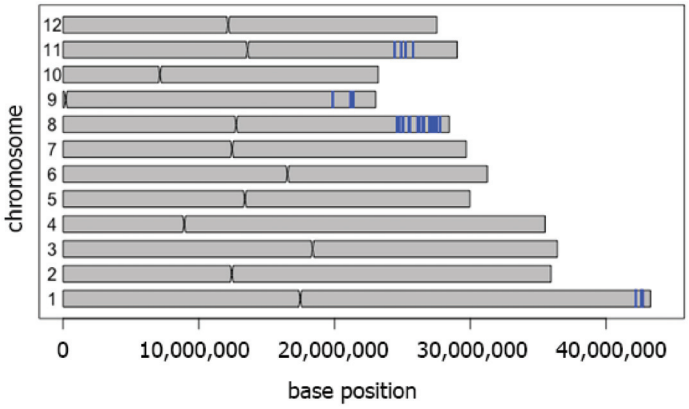


Figure 2. Genetic regions introgressed into the KDML105 genome. Single nucleotide polymorphisms (SNPs) between CSSL104 and KDML105 rice. The blue lines show 100 SNPs within 5000 background nucleotides. All loci containing SNPs were subjected to a co-expression network analysis. The results, shown in Figure 3A, revealed 18 major nodes with a high connection to other genes. The gene ontology of these 18 genes is listed in Table 2. The map position of these genes is shown in Figure 3B. Based on quantitative trait loci (QTL) data from the Qtaro database [23], six loci are located in QTL regions for drought-stress tolerance on chromosomes 1, 3, and 8 (Figure 3B). The high-density SNP on chromosome 1 was consistent with the location of the drought-tolerant (DT)-QTL, which is flanked by markers RZ14 and R117. In this QTL, co-expression network analysis identified two genes, *LOC_Os01g72800* and *LOC_Os01g72950*, as the major nodes. Chromosome 3 did not display high-density SNPs. On this chromosome, *LOC_Os03g02590* and *LOC_Os03g03910* were located in two drought-tolerance QTLs mapped between markers RM7332, RM545, and RG104, RZ329. Another node gene, *LOC_Os03g52460,* is located between markers C136 and R1618, corresponding to another drought-tolerance QTL. Chromosome 8 displays several major nodes: *LOC_Os08g16570* is located between markers RM72 and RM331, while *LOC_Os08g41040* and *LOC_Os08g41460* are located between RM5353 and RM 3480. This region on chromosome 8 also displayed high-density SNPs between CSSL104 and KDML105 (Figures 2 and 3B).

The eighteen rice genes reported here have homologs in *Arabidopsis* for which tagged mutants are available (Table 2). Nine of them (*CPFTSY*, *NDH-O*, *SOQ1*, *LHCB3*, *RRF*, *PGRL1B*, *HCF244*, *NAD(P)-linked oxidoreductase,* and *MRL1*) were annotated to be involved in the photosynthesis process [24–32]. Moreover, the homolog of *LOC_Os11g43600* is *CPRF1,* an *Arabidopsis* gene required for chloroplast development [33]. These findings suggest that these rice genes are involved in photosynthesis adaptation during drought stress.

Therefore, we obtained seven homozygous, T-DNA tagged *Arabidopsis* mutant lines corresponding to *ndhO* (*at1g74880*), *lhcb3* (*at5g54270*), *rrf* (*at3g63190*), *pgrl1b* (*at4g11960*), *pgrl1a* (*at4g22890*), *at2g27680*, and *mrl1* (*at4g34830*). These lines were drought stressed by growing them in MS medium supplemented with 0 mM, 75 mM, or 150 mM mannitol. Their growth response was assessed by measuring the green pixel area per plant and compared to the wild type (WT).

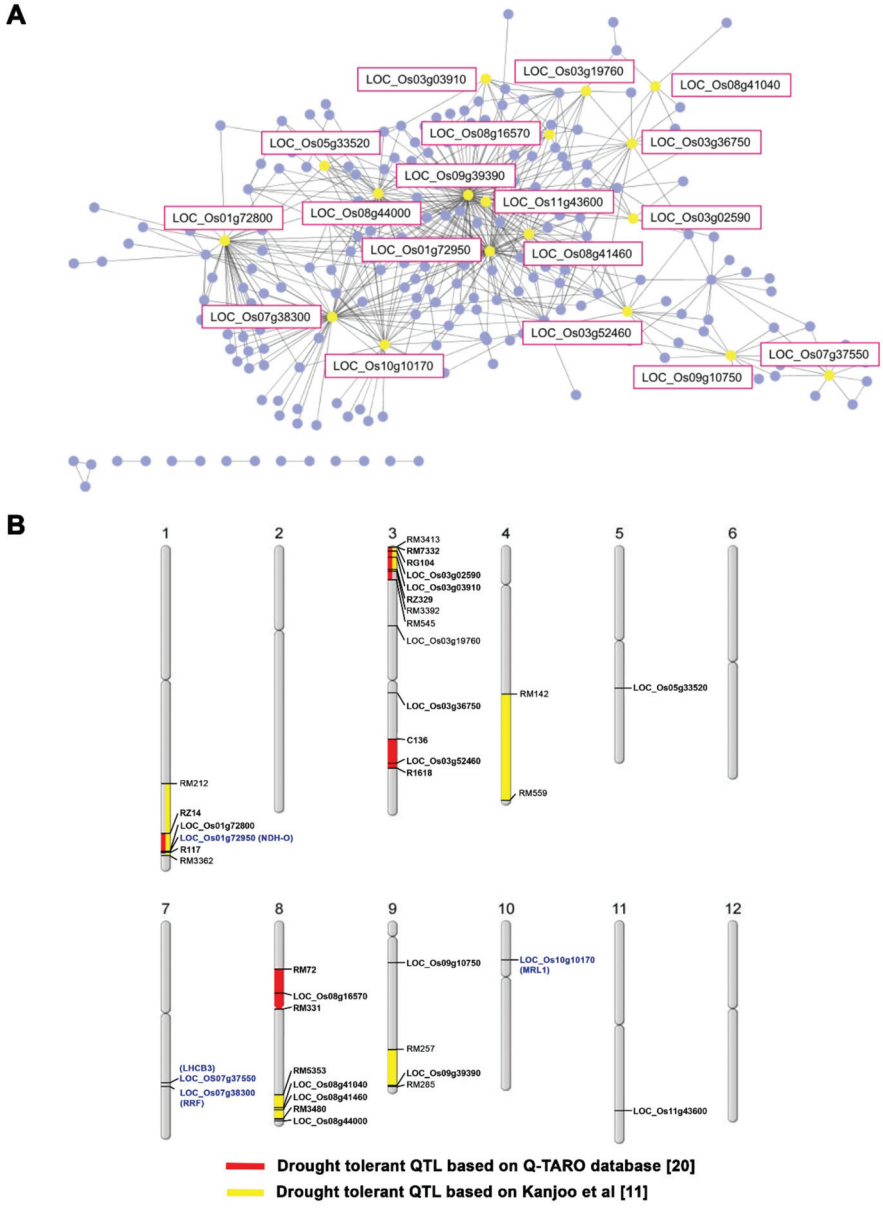


Figure 3. Candidate genes for drought tolerance. Gene co-expression network was analyzed by using the Rice Oligonucleotide Array Database [19], showing the major node genes with yellow dots (**A**), while DT-QTLs from the Q-TARO database [23] and Kanjoo et al. [11] are shown in red and yellow boxes on the chromosome, respectively. (**B**) Loci written in blue letters indicate the proposed drought-tolerance loci based on this study.

Table 2. Rice gene candidates for drought tolerance, their *Arabidopsis* homologs, and the inferred function of the genes in rice.

Rice Locus ID	*Arabidopsis* Locus ID	Gene Description	Mutant Stock	Homozygous	Involved in Photosynthesis
LOC_Os01g72800	AT2G45770	Chloroplast signal recognition particle (SRP) receptor homolog, alpha subunit CPFTSY. Required for light-harvesting chlorophyll *a/b*-binding protein (LHCP) integration into isolated thylakoids.	SALK_070410C	✓	✓
LOC_Os01g72950	AT1G74880	NDH-O, encoding subunit NDH-O of NAD(P)H: plastoquinone dehydrogenase complex (Ndh complex) present in the thylakoid membrane of chloroplasts. This subunit is thought to be required for Ndh complex assembly.	SALK_097351C	✓	✓
LOC_Os03g02590	AT1G01820	PEROXIN11C, member of the peroxin11 (PEX11) gene family, integral to peroxisome membrane, controls peroxisome proliferation	SALK_057358C	✓	
LOC_Os03g03910	AT4G35090	CAT2	SALK_076998	✓	
LOC_Os03g19760	AT1G56500	SOQ1 (Suppressor of quenching 1) prevents the formation of a slowly reversible form of antenna quenching, thereby maintaining the efficiency of light harvesting.	SALK_097577		✓
LOC_Os03g36750	AT3G48420	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	SALK_025204	✓	
LOC_Os03g52460	AT5G19220	APL1, the large subunit of ADP-glucose pyrophosphorylase, which catalyzes the first and rate-limiting step in starch biosynthesis.	CS478981	✓	
LOC_Os05g33520	AT2G48070	RPH1 is a chloroplast protein RPH1 (resistance to Phytophthora 1) involved in immune response to *Phytophthora brassicae*	SALK_102558C	✓	
LOC_Os07g37550	AT5G54270	LHCB3 is a component of the main light-harvesting chlorophyll a/b-protein complex of Photosystem II (LHC II)	SALK_020314C	✓	✓

Table 2. *Cont.*

Rice Locus ID	*Arabidopsis* Locus ID	Gene Description	Mutant Stock	Homozygous	Involved in Photosynthesis
LOC_Os07g38300	AT3G63190	RRF, encoding a chloroplast ribosome recycling factor homolog.	SALK_015954C	✓	✓
LOC_Os08g16570	AT1G16080	Nuclear protein	SALK_007790C	✓	
LOC_Os08g41040	AT4G31115	DUF1997 family protein	SALK_010690C	✓	
LOC_Os08g41460	AT4G11960	PGRL1B—a transmembrane protein present in thylakoids. Plants lacking PGRL1 show perturbation of cyclic electron flow.	SALK_059238C	✓	✓
	AT4G22890		SALK_133856C	✓	✓
LOC_Os09g10750	AT2G42220	Rhodanese/cell cycle control phosphatase superfamily protein	SALK_045769	✓	
LOC_Os09g39390	AT2G27680	NAD(P)-linked oxidoreductase superfamily protein	SALK_073120C	✓	✓
LOC_Os10g10170	AT4G34830	MRL1 (a conserved pentatricopeptide repeat protein) required for stabilization of *rbcL* mRNA	SALK_060806C	✓	✓
LOC_Os11g43600	AT3G62910	Chloroplast ribosome release factor 1, CPRF1, encoding a plastid-localized ribosome release factor 1 that is essential in chloroplast development	SALK_117765C		

Under normal conditions, all mutant lines displayed a significantly lower number of green pixels than WT, suggesting lower growth than WT (Figure 4A). Both drought-stress treatments decreased growth in all lines, with the 150 mM causing the more severe reduction. At 75 mM mannitol, *pgrl1b* had a significantly lower number of green pixels than WT, while *pgrl1a* showed similar growth to WT. Other mutant lines showed better growth than WT (Figure 4B). Under the severe drought-stress conditions induced with 150 mM mannitol, the growth of *lhcb3*, *at2g27680*, *mrl1,* and WT were similar. Mutants *pgrl1b* and *pgrl1a* had a significantly lower growth than WT, while *rrf* had a lower growth at the beginning of the treatment but displayed better growth than WT after 5 days of the treatment. However, similar growth between WT and *rrf* was found after seven days of drought stress. Among the mutant lines, *ndhO* was the only mutant line that had significantly better growth than WT under severe drought stress (Figure 4C).

The stability indexes of *Arabidopsis* mutants and WT were calculated to compare drought tolerance after six days of drought stress. After the intermediate drought stress (75 mM mannitol), all mutant lines except *pgrl1a* displayed significantly higher stability than WT, suggesting the contribution of *NDH-0*, *LHCB3*, *RRF*, *PGRL1b, at2g27680,* and *MRL1* to drought-tolerance adaptation (Figure 5A). Under severe drought (150 mM mannitol), significantly higher stability than WTs was detected for the *ndh-o*, *rrf*, *pgrl1b, at2g27680,* and *mrl1* mutants (Figure 5B). The *rrf* mutant line displayed the highest stability under both intermediate and severe drought-stress conditions.

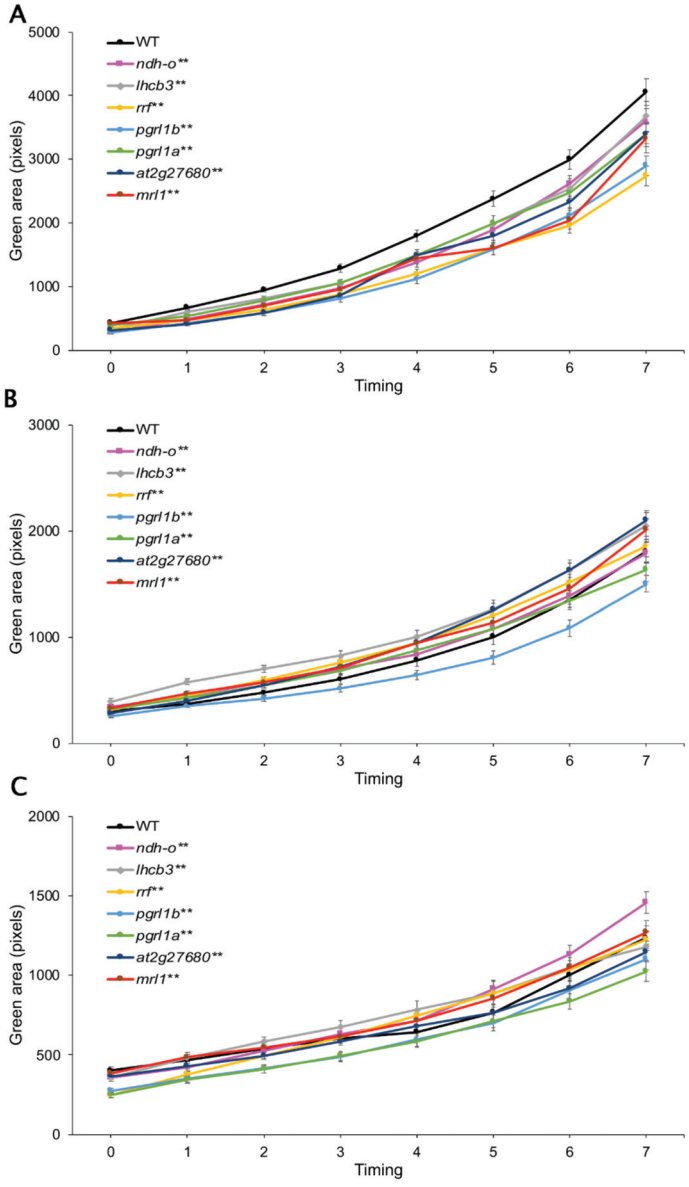


Figure 4. Growth response of seven mutant *Arabidopsis* lines to drought stress. Comparison of growth (green pixels per plant) of wild type (WT) and the T-DNA insertion mutant lines *ndhO* (*at1g74880*), *lhcb3* (*at5g54270*), *rrf* (*at3g63190*), *pgrl1b* (*at4g11960*), *pgrl1a* (*at4g22890*), *at2g27680*, and *mrl1*(*at4g34830*) grown in (**A**) normal Murashige and Skoog (MS) medium, (**B**) under intermediate drought stress (MS medium supplemented with 75 mM mannitol), and (**C**) under severe drought stress (MS medium supplemented with 150 mM mannitol). Statistical analysis was by *t*-test. * and ** above the name of the mutant line represent significant difference ($p < 0.05$) and highly significant difference ($p < 0.01$) between WT and mutant, respectively.

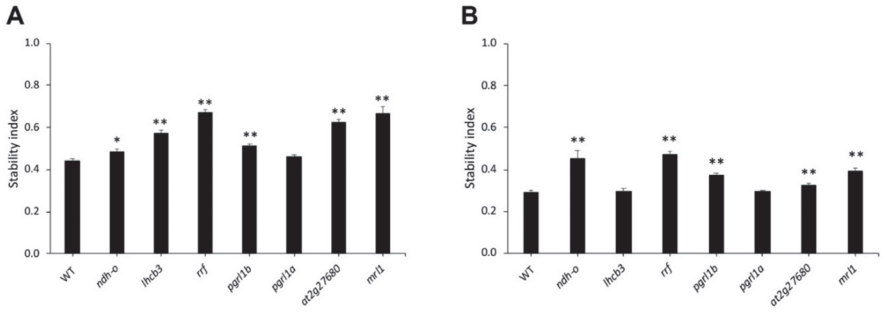


Figure 5. Stability during drought stress. The stability index, which is the ratio between the values from stressed plants and normal growth plants, is shown in (**A**) 75 mM mannitol and (**B**) 150 mM. This figure represents the mean ± SE of WT and mutant lines. The * shows a significant difference at p values ≤ 0.05 and ** indicates p value ≤ 0.01.

4. Discussion

We investigated the effect of a drought-stress QTL introgressed into the elite rice line KDML105. Using leaf water potential under drought stress, we found that introgression line CSSL104 manifested drought tolerance similar to that of the parent line DH103. Both tended to have a better ability to maintain water status in the first fully expanded leaves. After six days of drought stress, both lines had about a 22% higher leaf water potential than the KDML105 parent. Drought stress limits water uptake from the rice root and reduces water availability in the cells, which is critical for survival under drought stress. Water depletion can compromise photosynthesis and cell growth [34], and three main maintenance mechanisms are used by plants to offset water loss: leaf rolling, stomatal closure, and osmoregulation [35,36]. Evidence for the drought tolerance of CSSL104 was also provided by the lower leaf-death score and higher F_v/F_m displayed by CSSL104 compared to KDML105 (Figure 1).

Drought stress resulted in decreased stomatal conductance in all lines (KDML105, DH103, and CSSL104; Table 1). This was a water-preservation mechanism that also resulted in a decline in the net photosynthesis rate. After three days of drought stress, the photosynthesis parameters, net photosynthesis rate, stomatal conductance, transpiration rate, ΦPSII, electron transport rate, intercellular CO_2 concentration, and F_v'/F_m', were similar among all lines. After six days under drought stress, the net photosynthesis rate of DH103 and CSSL104 were about twofold higher than the net photosynthesis rate of KDML105. In comparison with normal plants, KDML105 rice had a nearly 70% reduction in photosynthesis rate, while DH103 and CSSL104 had only a 46 and 44% reduction, respectively. Interestingly, the ΦPSII and electron transport rate of all lines were similar, while the stomatal conductance of KDML105 was 58% lower than DH103. These findings suggest that stomatal closure could be one of the major factors contributing to the decline in the net photosynthesis rate of KDML105. Although the stomatal conductance of CSSL104 was lower than DH103 by 42%, CSSL104 could maintain a net photosynthesis rate (Table 1). These indicated that this CSSL is better adapted than its drought-tolerant parental line. It is possible that a KDML105 locus contributed to maintenance of the photosynthesis process through an epistatic interaction with the introgressed DH103 region.

Using whole genome sequence comparison and co-expression network analysis, we characterized the molecular fingerprint of the introgression. The first detected DNA segments introgressed, while the second identified genes connected with the drought response. During this stress, 18 genes were highly co-expressed with other genes (Figure 3A and Table 2). Nine of them (*CPFTSY*, *NDH-O*, *SOQ1*, *LHCB3*, *RRF*, *PGRL1*, *HCF244*, *NAD(P)-linked oxidoreductase*, and *MRL1*) are involved in the photosynthesis process, and CPRF1 is essential for chloroplast development. These findings indicate that the drought-adaptation QTL affects photosynthetic genes whose modulation maintains the net photosynthesis rate of CSSL104.

The function of the identified genes is as follows. CPFTSY (chloroplast FtsY, i.e., chloroplast signal-recognition particle) is required for light-harvesting chlorophyll *a/b*-binding protein (LHCP) integration into thylakoids [25], and NDH-O is the subunit required for the NADH dehydrogenase-like (NDH) complex assembly that functions in cyclic electron flow [25,37]. SOQ1 is required to maintain light harvesting efficiency especially during nonphotochemical quenching (NPQ) recovery [26]. Light-harvesting chlorophyll (LHC) functions as a light receptor to capture light energy and deliver it to photosystems. The *Lhcb3* gene product regulates the rate of state transition by changing the excitation energy transfer and charge separation [38]. RRF is a ribosome recycling factor in chloroplast [39]. *RRF* is required to maintain photosystem II efficiency (F_v/F_m) and proper stacking of the internal membranes of chloroplast. Loss of these functions led to a lower growth rate for the *rrf* mutant compared to WT [28], which is consistent with the phenotype documented in this study (Figure 4A). *PGRL1A* (*AT4G22890*) and *PGRL1B* (*AT4G11960*) are paralogous genes whose products switch linear electron flow to cyclic electron flow. PGRL1 is the elusive ferredoxin-plastoquinone reductase (FQR) [29].

During drought stress, plants shift the electron transfer route from linear to cyclic to balance the energy flow from light reaction to Calvin cycle and photorespiration [37,40,41]; this change is due to CO_2 limitations caused by stomatal closure. It was recently shown that tomato with co-silencing of the *PGR5/PGRL1A* gene was more susceptible to cold stress [42]. This is consistent with our finding that the *pgrl1* mutant had a significantly lower growth rate than WT in both intermediate and severe drought stress (Figure 5B,C). HCF244 is required for the biogenesis of photosystem II (PSII), and specifically for the synthesis of the reaction center proteins [30–32]. The NAD(P)-linked oxidoreductase gene is involved in redox reactions, but the details are still unclear. *MRL1* is the only gene in this study whose product is involved in the Calvin cycle. It is involved either in the processing or in the stabilization of the large subunit (LS) of *RuBisCO* transcripts [43].

Not surprisingly, impairment of these photosynthesis genes significantly reduced growth under normal conditions (Figure 4A). Moreover, we could not obtain homozygous lines of the *soq1*, *hcf244*, and *cprf1* mutation, probably because the homozygous mutants are lethal. However, under drought-stress conditions, some of the mutants in the genes involved in the light reaction process (*ndhO* (*at1g74880*), *lhcb3* (*at5g54270*), *rrf* (*at3g63190*), and *at2g27680* mutants) displayed significantly higher growth than WT. These responses suggested that the decrease in light energy harvest during drought stress could prevent damage to chloroplasts and prolong survival of photosynthetic tissues. The *mrl1* mutant was expected to lack of the ability to stabilize the *rbcL* mRNA, which could result in a decline in Calvin cycle activity. The higher stability index of the *mrl1* mutant line under intermediate and severe drought stress indicated that a slower rate of the Calvin cycle may help plants cope with drought stress.

Based on the growth phenotype under drought stress, the *ndhO*, *lhcb3*, *rrf*, and *mrl1* mutants showed a higher growth rate than WT (Figure 4B,C). The homologs in rice of these genes are *LOC_Os01g72800*, *LOC_Os07g37550*, *LOC_Os07g38300*, and *LOC_Os10g10170*, respectively. Therefore, we would like to propose that these genes contribute to drought-tolerance regulation in rice by mediating adaptation in the photosynthesis process. *LOC_Os01g72800* is located in the previously reported drought-tolerance QTL between RZ14 and R117 [21], while the other three genes are not. Collectively, our results indicate that the combination of SNP analysis with co-expression network analysis is a suitable method for drought-tolerance gene prediction. This approach will help future exploration to identify the candidate genes for abiotic stress tolerance.

5. Conclusions

The KDML105 chromosome substitution line CSSL104 displayed a drought-tolerance phenotype based on photosynthetic maintenance ability. Identification of SNPs between KDML105 and the tolerant CSSL, together with co-expression network analysis, predicted 18 candidate drought-tolerance genes—ten of which were involved in photosynthesis or chloroplast development. Seven of them were selected for the characterization by using *Arabidopsis* mutant lines for the homologous genes.

Four out of seven mutants showed a higher growth rate than WT under drought stress. Therefore, *LOC_Os01g72800*, *LOC_Os07g37550*, *LOC_Os07g38300*, and *LOC_Os10g10170* are proposed to be the drought-tolerance genes in CSSL104 rice.

Author Contributions: Conceptualization, S.C. and L.C.; methodology, K.P., T.P., L.C., N.D.D., and L.S.; software, K.P., T.P., N.D.D., and L.S.; validation, C.P.; formal analysis, C.P.; investigation, C.P.; resources, J.L.S. and T.T.; data curation, C.P. and T.B.; writing—original draft preparation, C.P.; writing—review and editing, S.C., T.B., and L.C.; visualization, C.P., S.C.; supervision, S.C., N.D.D., and L.S.; project administration, S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: C.P. was supported by the 100th Anniversary of Chulalongkorn University Doctoral Scholarship, and this study was supported by the 90th Anniversary of Chulalongkorn University, Rachadapisek Sompote Fund, grant no. GCUGR1125614034D. The collaboration between CEEPP and CRH was supported by the graduate school and the Rachadapisek Sompote Fund, Chulalongkorn University, Thailand. L.S. and N.D.D. acknowledge the ERDF project entitled "Development of Pre-Applied Research in Nanotechnology and Biotechnology" (No. CZ.02.1.01/0.0/0.0/17_048/0007323).

Acknowledgments: We thank Noppakhun Khunpolwattana for rice DNA extraction and Panita Chutimanukul and Thammaporn Kojonna for their kind help.

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. Saikumar, S.; Gouda, P.K.; Saiharini, A.; Varma, C.M.K.; Vineesha, G.; Padmavathi, G.; Shenoy, V.V. Major QTL for enhancing rice grain yield under lowland reproductive drought stress identify using *O. sativa*/*O. glaberrima* introgression line. *Field Crops Res.* **2014**, *163*, 119–131. [CrossRef]
2. Tulyathan, V.; Leeharatanaluk, B. Change in quality of rice (*Oryza sativa* L.) cv. Khao Dawk Mali 105 during storage. *J. Food Biochem.* **2007**, *31*, 415–425. [CrossRef]
3. Siangliw, J.; Jongdee, B.; Pantuwan, G.; Toojinda, T. Developing KDML105 backcross introgression lines using marker-assisted selection for QTLs associated with drought tolerance in rice. *Sci. Asia* **2007**, *33*, 207–214. [CrossRef]
4. Siddique, M.R.B.; Hamid, A.; Islam, M.S. Drought stress effects on water relations of wheat. *Bot. Bull. Acad. Sin.* **2001**, *41*, 35–39.
5. Hussain, M.; Malik, M.A.; Farooq, M.; Ashraf, M.Y.; Cheema, M.A. Improving drought tolerance by exogenous application of glycine-betaine and salicylic acid in sunflower. *J. Agron. Crop Sci.* **2008**, *194*, 193. [CrossRef]
6. Cornic, G.; Massacci, A. Leaf photosynthesis under drought. In *Photosynthesis and the Environment*; Baker, N.R., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996; pp. 347–366.
7. Anjum, F.; Yaseen, M.; Rasul, E.; Wahid, A.; Anjum, S. Water stress in barley (*Hordeum valgare* L.). I. effect on chemical composition and chlorophyll content. *Pak. J. Agron. Sci.* **2003**, *40*, 45–49.
8. Du, Y.C.; Kawamitsu, Y.; Nose, A.; Hiyane, S.; Murayama, S.; Wasano, K.; Uchida, Y. Effect on water stress on carbon exchange rate and activities of photosynthetic enzymes in leaves of sugarcane (*Saccharum* sp.). *Aust. J. Plant Physiol.* **1996**, *23*, 719–726. [CrossRef]
9. Farooq, M.; Wahid, A.; Kobayashi, N.; Fujita, D.; Basra, S.M.A. Plant drought stress: Effects, mechanisms and management. *Agron. Sustain. Dev.* **2009**, *29*, 185–212. [CrossRef]
10. Zargar, S.M.; Gupta, N.; Nazir, M.; Mahajan, R.; Malik, F.A.; Sofi, N.R.; Shikari, A.B.; Salgotra, R.K. Impact of drought on photosynthesis: Molecular perspective. *Plant Gene* **2017**, *11*, 154–159. [CrossRef]
11. Kanjoo, V.; Punyawaew, K.; Siangliw, J.L.; Jearakongman, S.; Vanavichit, A.; Toojinda, T. Evaluation of agronomic traits in chromosome segment substitution lines of KDML105 containing drought tolerance QTL under drought stress. *Rice Sci.* **2012**, *19*, 117–124. [CrossRef]
12. Punchkhon, C.; Kasetranunt, W.; Kositsup, B.; Siengliw, J.; Chadchawan, S. Evaluation of drought tolerance ability at seedling stage of CSSL rice populations with drought tolerant genes on chromosome 8. In Proceedings of the 9th Botanical Conference of Thailand, Bangkok, Thailand, 3–5 June 2015; pp. 61–71.
13. Vajrabhaya, M.; Vajrabhaya, T. Somaclonal variation of salt tolerance in rice. In *Biotechnology Agriculture Forestry*; Bajaj, Y.P.S., Ed.; Springer: Berlin/Heidelberg, Germany, 1991; pp. 368–382.

14. Pongprayoon, W.; Roytrakul, S.; Pichayangkura, R.; Chadchawan, S. The role of hydrogen peroxide in chitosan-induced resistance to osmotic stress in rice (*Oryza sativa* L.). *Plant Growth Regul.* **2013**, *70*, 159–173. [CrossRef]
15. Chintakovid, N.; Maipoka, M.; Phaonakrop, N.; Mickelbart, M.V.; Roytrakul, S.; Chadchawan, S. Proteomic analysis of drought-responsive proteins in rice reveals photosynthesis-related adaptations to drought stress. *Acta Physiol. Plant.* **2017**, *39*, 240. [CrossRef]
16. Missirian, V.; Comai, L.; Filkov, V. Statistical mutation calling from sequenced overlapping DNA pools in TILLING experiments. *BMC Bioinf.* **2011**, *12*, 287. [CrossRef] [PubMed]
17. Ouyang, S.; Zhu, L.; Hamilton, J.; Lin, H.; Campbell, M.; Childs, K.; Thibaud-Nissen, F.; Malek, R.L.; Lee, Y.; Zheng, L.; et al. The TIGR rice genome annotation resource: Improvements and new features. *Nucleic Acids Res.* **2007**, *35*, 883–887. [CrossRef] [PubMed]
18. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Zhou, S.; Childs, K.L.; et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]
19. Cao, P.; Jung, K.; Choi, D.; Hwang, D.; Zhu, J.; Ronald, P. The rice oligonucleotide array database: An atlas of rice gene expression. *Rice* **2012**, *5*, 17. [CrossRef] [PubMed]
20. Xia, L.; Zou, D.; Sang, J.; Xu, X.J.; Yin, H.Y.; Li, M.W.; Wu, S.Y.; Hu, S.N.; Hao, L.L.; Zhang, Z. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genom.* **2017**, *44*, 235–241. [CrossRef] [PubMed]
21. 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] [PubMed]
22. De Diego, N.; Fürst, T.; Humplik, J.; Ugena, L.; Podlešáková, K.; Spíchal, L. An automated method for high-throughput screening of Arabidopsis rosette growth in multi-well plates and its validation stress condition. *Front. Plan Sci.* **2017**, *8*, 1702. [CrossRef]
23. Yonemaru, J.; Yamamoto, T.; Fukuoka, S.; Uga, Y.; Hori, K.; Yano, M. Q-TARO:QTL Annotation rice online database. *Rice* **2010**, *3*, 194. [CrossRef]
24. Tu, C.J.; Schuenemann, D.; Hoffman, N.E. Chloroplast FtsY, chloroplast signal recognition particle, and GTP are required to reconstitute the soluble phase of light-harvesting chlorophyll protein transport into thylakoid membranes. *J. Biochem. Chem.* **1999**, *274*, 27219–27224. [CrossRef]
25. Rumeau, D.; Bécuwe-Linka, N.; Beyly, A.; Louwagie, M.; Garin, J.; Peltier, G. New subunit NDH-M, -N, -O, encoded by nuclear genes, are essential for plastid Ndh complex functioning in higher plants. *Plant Cell* **2005**, *17*, 219–232. [CrossRef] [PubMed]
26. Brooks, M.D.; Sylak-Glassman, E.J.; Fleming, G.R.; Niyogi, K.K. A thioredoxin-like/β-propeller protein maintains the efficiency of light harvesting in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E2733–E2740. [CrossRef]
27. Adamiec, M.; Gibasiewicz, K.; Luciński, R.; Giera, W.; Chełminiak, P.; Szewczyk, S.; Sipińska, W.; Grondelle, R.; Jackowski, G. Excitation energy transfer and charge separation are affected in *Arabidopsis thaliana* mutants lacking light-harvesting chlorophyll *a/b* binding protein Lhcb3. *J Photochem Photobiol. B Biol.* **2015**, *153*, 423–428. [CrossRef] [PubMed]
28. Guan, Z.; Wang, W.; Yu, X.; Lin, W.; Maio, Y. Comparative proteomic analysis of coregulation of CIPK14 and WHIRLY1/3 mediated pale yellowing of leaves in *Arabidopsis*. *Int. J. Mol. Sci.* **2018**, *19*, 2331. [CrossRef] [PubMed]
29. Hertle, A.P.; Blunder, T.; Wunder, T.; Pesaresi, P.; Pribil, M.; Armbruster, U.; Leister, D. PGRL1 is the elusive ferredoxin-plastoquinone reductase in photosynthetic cyclic electron flow. *Mol. Cell* **2013**, *49*, 511–523. [CrossRef]
30. Hey, D.; Grimm, B. ONE-HELIX PROTEIN2 (OHP2) is required for the stability of OHP1 and assembly factor HCF244 and is functionally linked to PSII biogenesis. *Plant Physiol.* **2018**, *177*, 1453–1472. [CrossRef] [PubMed]
31. Li, Y.; Liu, B.; Zhang, J.; Kong, F.; Zhang, L.; Meng, H.; Li, W.; Rochaix, J.D.; Li, D.; Peng, L. OHP1, OHP2, and HCF244 form a transient functional complex with the photosystem II reaction center. *Plant Physiol.* **2019**, *179*, 195–208. [CrossRef] [PubMed]

32. Chotewutmontri, P.; Williams-Carrier, R.; Barkan, A. Exploring the link between photosystem II assembly and translation of the chloroplast *psbA* mRNA. *Plants* **2020**, *9*, 152. [CrossRef]
33. Motohashi, R.; Yamazaki, T.; Myouga, F.; Ito, T.; Ito, K.; Satou, M.; Kobayashi, M.; Nagata, N.; Yoshida, S.; Nagashima, A.; et al. Chloroplast ribosome release factor 1 (AtcpRF1) is essential for chloroplast development. *Plant Mol. Biol.* **2007**, *646*, 481–497. [CrossRef]
34. Tezara, W.; Mitchell, V.J.; Driscoll, S.D.; Lawlor, D.W. Water stress inhibits plant photosynthesis by decreaseing coupling factor and ATP. *Nature* **1999**, *401*, 914–917. [CrossRef]
35. Teulat, B.; Monneveux, P.; Wery, J.; Borries, C.; Souyris, I.; Charrier, A.; This, D. Relationships between relative water content and growth parameters under water stress in barley: A QTL study. *New Phytol.* **1997**, *137*, 99–107. [CrossRef]
36. González, L.; González-Vilar, M. Determination of relative water content. In *Handbook of Plant Ecophysiology Techniques*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 207–212.
37. Nikkanen, L.; Toivola, J.; Trotta, A.; Diaz, M.G.; Tikkanen, M.; Aro, E.; Rintamäki, E. Regulation of cyclic electron flow by chloroplast NADPH-dependent thioredoxin system. *Plant Direct* **2018**, *2*, e00093. [CrossRef] [PubMed]
38. Damkjaer, J.T.; Kereïche, S.; Johnson, M.; Kovacs, L.; Kiss, A.Z.; Boekema, E.J.; Ruban, A.V.; Horton, P.; Jansson, S. The photosystem II light-harvesting protein Lhcb3 affects the macrostructure of photosystem II and the rate of state transitions in *Arabidopsis*. *Plant Cell* **2009**, *21*, 3245–3256. [CrossRef]
39. Wang, L.; Ouyang, M.; Li, Q.; Zou, M.; Gou, J.; Ma, J.; Lu, C.; Zhang, L. The *Arabidopsis* chloroplast ribosome recycling factor is essential for embryogenesis and chloroplast biogenesis. *Plant Mol. Biol.* **2010**, *74*, 47–59. [CrossRef]
40. DalCorso, G.; Pesaresi, P.; Masiero, S.; Aseeva, E.; Schünemann, D.; Finazzi, G.; Joliot, P.; Barbato, R.; Leister, D. A complex containing PGRL1 and PGRL5 is involved in the switch between linear and cyclic electron flow in *Arabidopsis*. *Cell* **2008**, *132*, 273–285. [CrossRef] [PubMed]
41. Suorsa, M.; Rossi, F.; Tadini, L.; Labs, M.; Colombo, M.; Jahns, P.; Kater, M.M.; Leister, D.; Finazzi, G.; Aro, E.; et al. PGRL5-PGRL1-dependent cyclic electron transport modulates linear electron transport rate in *Arabidopsis thaliana*. *Mol. Plant* **2016**, *9*, 271–288. [CrossRef] [PubMed]
42. Wang, F.; Yan, J.; Ahammed, G.J.; Wang, X.; Bu, X.; Xiang, H.; Li, Y.; Lu, J.; Liu, Y.; Qi, H.; et al. PGR5/PGRL1 and NDH mediate far-red light-induced photoprotection in response to chilling stress in tomato. *Front. Plant Sci.* **2020**, *11*, 669. [CrossRef] [PubMed]
43. Johnson, X.; Wostrikoff, K.; Finazzi, G.; Kuras, R.; Schwarz, C.; Bujaldon, S.; Nickelsen, J.; Stern, D.B.; Wollman, F.; Vallon, O. MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomas* and *Arabidopsis*. *Plant Cell* **2010**, *22*, 234–248. [CrossRef]

genes

MDPI

Article

Development of Chromosome Segment Substitution Lines (CSSLs) Derived from Guangxi Wild Rice (*Oryza rufipogon* Griff.) under Rice (*Oryza sativa* L.) Background and the Identification of QTLs for Plant Architecture, Agronomic Traits and Cold Tolerance

Ruizhi Yuan †, Neng Zhao †, Babar Usman †, Liang Luo, Shanyue Liao, Yufen Qin, Gul Nawaz and Rongbai Li *

State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Agriculture, Guangxi University, Nanning 530004, China; 1717303012@st.gxu.edu.cn (R.Y.); nengzhao@st.gxu.edu.cn (N.Z.); babarusman119@gmail.com (B.U.); 1717303001@st.gxu.edu.cn (L.L.); 1817303014@st.gxu.edu.cn (S.L.); qyf@st.gxu.edu.cn (Y.Q.); gulnawazmalik@yahoo.com (G.N.)

* Correspondence: lirongbai@gxu.edu.cn or lirongbai@126.com; Tel.: +86-136-0009-4135

† These authors contributed equally to this work.

Received: 12 July 2020; Accepted: 21 August 2020; Published: 22 August 2020

Abstract: Common wild rice contains valuable resources of novel alleles for rice improvement. It is well known that genetic populations provide the basis for a wide range of genetic and genomic studies. In particular, chromosome segment substitution lines (CSSLs) ais a powerful tool for fine mapping of quantitative traits, new gene discovery and marker-assisted breeding. In this study, 132 CSSLs were developed from a cultivated rice (*Oryza sativa*) cultivar (93-11) and common wild rice (*Oryza rufipogon* Griff. DP30) by selfing-crossing, backcrossing and marker-assisted selection (MAS). Based on the high-throughput sequencing of the 93-11 and DP30, 285 pairs of Insertion-deletions (InDel) markers were selected with an average distance of 1.23 Mb. The length of this DP30-CSSLs library was 536.4 cM. The coverage rate of substitution lines cumulatively overlapping the whole genome of DP30 was about 91.55%. DP30-CSSLs were used to analyze the variation for 17 traits leading to the detection of 36 quantitative trait loci (QTLs) with significant phenotypic effects. A cold-tolerant line (RZ) was selected to construct a secondary mapping F_2 population, which revealed that *qCT2.1* is in the 1.7 Mb region of chromosome 2. These CSSLs may, therefore, provide powerful tools for genome wide large-scale gene discovery in wild rice. This research will also facilitate fine mapping and cloning of QTLs and genome-wide study of wild rice. Moreover, these CSSLs will provide a foundation for rice variety improvement.

Keywords: rice; chromosome segment substitution lines (CSSLs); quantitative trait locus (QTL); marker-assisted selection (MAS); cold tolerance (CT)

1. Introduction

Wild rice (*Oryza rufipogon* Griff.) contains many novel and useful alleles that control tiller number, shattering, dormancy, pericarp color, mating type, panicle architecture and grain size and number [1]. Therefore, the potentially beneficial genes in wild rice are an important goal to improve cultivated rice (*Oryza sativa* L.) [2]. Although many quantitative trait loci (QTLs), for plant architecture, agronomic traits and cold tolerance (CT) have been identified in rice [3–5], however, there are few reports on those traits that were discovered in same chromosome segment substitution lines (CSSLs). The hybridization between *Oryza sativa* and wild rice and use of marker-assisted selection (MAS) of

individuals retaining a part of the wild chromosome in the background revealed the transfer desirable genes to cultivated rice [6]. QTL identification through CSSLs is advantageous because it completely removes the genetic background interference and provides QTL visualization as a single Mendelian factor [7]. The development of CSSLs is laborious and time-consuming, but it is useful to scientists and plant breeders [8–10]. To date, several CSSLs in rice have been developed and many QTLs for traits of biologic and economic interest have been mapped [8]. The potential of the rice progenitor as a genetic resource has been explored for improving O. *sativa* with 33 chromosome segment substitution lines (CSSLs) of O. *rufipogon* (W0106) [9]. Single segment substitution lines (SSSLs) libraries also has been utilized to detect several QTLs related to plant height, heading date, seed setting rate and 1000-grain weight [10]. These achievements have undoubtedly enhanced the understanding of complex traits and encouraged plant genomic studies.

Some QTLs and genes for plant architecture traits and agronomic traits have also been detected from CSSLs [11,12]. *GL4* can control the grain length and seed shattering ability of African cultivated rice by regulating the longitudinal cell elongation of the outer glume and inner glume [11]. *RLS3* plays an important role in regulating chloroplast degradation and the normal growth of rice [12]. The additive effect, main effect and epistatic effect of QTLs were also studied based on CSSLs. In a previous study, eight SSSLs as experimental material were utilized to estimate the additive and dominant effects of six QTLs (*Hd1, Ehd1, Hd3a or RFT1, EH3, OsMADS50* and *DTH8*) and their epistatic effects among dual QTLs [13].

Cold injuries have been observed at many stages of growth and may result in the failure of germination, retarded seedling growth, stunting, discoloration, panicle tip degeneration, a prolonged duration of cultivation, sterility and irregular maturity [14]. Previous studies identified strong interactions between the cold tolerant QTLs and their environments [15,16]. Therefore, the screening of cold-tolerant germplasm and identifying the QTLs related to cold tolerance will help reduce the losses caused by low temperatures and improve rice production in marginal lands.

In this study, a broad population of DP30-CSSLs was constructed via the backcrossing, selfing and masker-assisted selection (MAS) of cultivated rice (93-11) and Guangxi wild rice (*Oryza rufipogon* Griff. DP30). By performing high-throughput whole genome sequencing, we designed 285 InDel molecular markers that were evenly distributed across the twelve chromosomes of rice. These molecular markers were utilized to select 132 substitution lines through MAS. The rate of coverage of the substitution segments to the whole genome of DP30 wild rice was 91.55%. Meanwhile, we investigated eighteen traits found in the CSSLs. thirty-six QTLs, some of which have been found in previous studies while others are new. Furthermore, we selected a cold-tolerant line (RZ34) to construct a secondary population and *qCT2.1* was located in a 1.7 Mb region on chromosome 2.

2. Materials and Methods

2.1. Plant Materials

The DP30-CSSLs were developed by using the elite rice cultivar 93-11 as the recurrent parent. Guangxi common wild rice DP30 was collected from Nanning, Guangxi Province and used as the donor parent. The original habitat of DP30 was conserved during the collection of the wild rice germplasm. Refer to previous studies to reduce the impact of environment factors on QTL, materials including receptor 93-11 and CSSLs were grown in both seasons of 2018 (fall and spring) [13].

2.2. Phenotyping

The phenotypic data were recorded under natural conditions in the experimental area of Guangxi University (Guangxi, China; 22°38′ N, 108°13′ E) in spring (from March to July) and fall (from July to November). A randomized complete block design (RCBD) was used with three replications for CSSLs, CSSL combinations and their parents. The parent plants and 20 CSSL plants were selected from each replication for data collection. The data for 17 traits were recorded, including leaf sheath color (LSC),

leaf margin color (LMC), tiller angle (TA), heading date (HD), plant height (PH), grain shattering (SH), apiculus color (AC), stigma color (SC), glume color (GC), number of grains per panicle (NGPP), 1000-grain weight (GWT), grain length (GL), grain width (GW), grain length to width ratio (GLWR), awn length (AL), seed coat color (SCC) and cold tolerance (CT). The data for TA was recorded on a scale from 1 to 9 representing the angles of 0–10°, 11–20°, 21–30°, 31–40°, 41–50°, 51–60°, 61–70°, 71–80° and 81–90°, respectively [17–20]. The HD was recorded when the first panicle to emerge reached about 2-cm-long and the number of days from sowing to heading were scored for each plant. PH was measured for each plant at the mature stage from the base of the stem to the tip of the higher panicle. NGPP, GL and GW were recorded according to the previously established methods [19,20]. The GLWR and 100-GWT of the filled grains were investigated after the rice was harvested at the mature and naturally dried stage [10].

2.3. Evaluation of Cold Tolerance at Seedling Stage

To preliminarily evaluate the cold tolerance variation of CSSLs, the experiment was conducted in controlled conditions at seedling stage. Thirty seedlings of each CSSL were planted in soil. Plants were grown in a controlled environment at a day temperature of 25 °C and night temperature of 19 °C till three leaf stage. At three leaf stage, seedlings were exposed to cold stress according to a previously established method [21]. The depth of the water was about 5 cm measured from the surface of the soil in the trey. The cold stress treatment lasted 5 days and the conditions alternated between 10 °C for 10 h during the day and 8 °C for 14 h at night. After the cold treatment, the seedlings were subjected to natural standard growth conditions at 26 °C and the survival rate was investigated. After 5 days of treatment, cold tolerance was evaluated on the basis survival rate and injury level. The experiment was repeated three times under the same cold stress treatment. The average data of three replications were used. The data for CT were recorded on a scale from 0 to 9, representing the survival rates of 0–9%, 10–19%, 20–29%, 30–39%, 40–49%, 50–59%, 60–69%, 70–79%, 80–89% and 90–100%, respectively [21]. The secondary mapping population used to fine-map the major QTL for cold tolerance was developed by backcrossing a CSSL(RZ34) with the recipient parent (93-11). Three hundred and eleven F_2 plants of that cross were genotyped using five InDel markers and fifty-seven cold tolerant and twenty-one cold sensitive plants selected for phenotypic evaluation under cold stress treatment.

2.4. Construction of CSSLs and Genome Sequencing and Development of InDel Markers

CSSLs of common wild rice were constructed by hybridization, backcrossing and marker-assisted selection (MAS) according to the previously described method [22,23]. The genomic DNA of DP30 and 93-11 was prepared and whole-genome re-sequencing (WGRS) was performed on an Illumina HiSeq2500™ by Novo Generation Company, Beijing, China. The standard Illumina protocol was followed for sample preparation and sequencing. The quality trimming (phred quality score, $<Q_{20}$) was carried out by using FastQC [24] and the Cutadapt software was used for adapter trimming with the parameters of −O 5 and −m 32 [25]. The Burrows–Wheeler Aligner (BWA) software was used to map clean reads to the 93-11-reference genome [26,27]. InDel polymorphisms were detected by the GATK tool software with the defined length of insertions and deletions between 1 bp and 10 bp [28]. The larger sized (≥2 bp) InDel regions and high sequencing depths (DP, ≥5-fold) were extracted to design the InDel markers. The primer pairs were designed based on parental sequence differences by the DNAMAN v6.0 software and screened using the NCBI (https://www.ncbi.nlm.nih.gov/) database.

2.5. DNA Isolation and PCR Amplification

Genomic DNA was extracted from fresh leaf tissues using the CTAB method described by the previously established protocol [29]. The PCR amplification, separation of PCR products and confirmation and genotyping of the electrophoretic bands of the PCR products were performed according to previously established methods [30–32].

2.6. QTL Mapping and Data Analysis

The lengths of the substituted segments in CSSLs were assayed according to the previously established method [10]. A chromosomal segment flanked by two donor markers was considered to be 100% donor type, and the length of the segment was considered to be the minimum length of a substituted segment (L-min). A chromosome segment flanked by two recipient markers was considered as 0% donor type, and the length of the segment was considered to be the maximum length of a substituted segment (L-max). A chromosome segment flanked by one marker of donor type and one marker of recipient type was recognized as 50% donor type. The length of each substituted segment in the CSSLs was calculated as the estimated length (L), which is the average of L-min and L-max.

Genotypic graphics and chromosome genetic maps of the CSSLs were generated using the Graphic Geno-Types 2 software (GGT2.0) [10]. A putative QTL was declared at the significance level of $p \leq 0.001$ in a CSSL. If several CSSLs with overlapping substituted segments shared the same QTL, a substitution mapping approach was employed to localize the QTL to a smaller genomic interval [33]. QTL nomenclature was performed according to McCouch and CGSNL: each QTL name is italicized and starts with a lowercase letter "q" to indicate that it is a QTL, followed by a two to five letter standardized "trait name", a number designating the rice chromosome on which it occurs (1–12), a period (".") and a unique identifier to differentiate individual QTLs for the same trait that resides on the same chromosome [34]. The linkage map of QTLs was constructed using MapChart 2.2 [35]. The statistical analysis of the phenotypic and genotypic data of the DP30-CSSL populations was performed by QTL IciMapping 4.1.0 software [36]. Base on the permutation test to set LOD value ≥2.5 as the thresholds for QTL analysis. The chromosomal genetic map was constructed using MapChart 2.3 software.

The genetic effect of different QTLs according to the previously established following formulas [13]:

(1) Additive effect (a) = phenotypic value of CSSL (CSSL) − phenotypic value of 93-11(93-11);
(2) Dominant effect (d) = 93-11 × CSSL − 93-11.

3. Results

3.1. Phenotypic Variation of Plant Architecture, Agronomic Traits and Cold Tolerance in DP30-CSSLs

We evaluated the variation in plant architecture, agronomic and cold tolerance traits (DP30, DP30-CSSLs and 93-11) and calculated the phenotypic values of these traits during two seasons (fall and spring). The DP30-CSSLs showed significant variation in all of these traits. Furthermore, this variation in HD, GWT, GL, GW, GLWR and AL were normally distributed, while TA, PH, NGPP, SH and CT followed biased distribution (Figure S1). First, we speculated that the biased distribution was due to continuous backcrossing, that the CSSLs without related trait substitution segments were similar to 93-11. Second, we counted 132 lines, so an insufficient population may have led to biased distribution. Third, these distributions were frequently skewed because of interactions between the alleles, nonallelic genes or other environmental factors.

3.2. Genome Re-Sequencing and Selection of InDel Markers

The DP30 genome was sequenced by Illumina high-throughput sequencing technology, and the resulting sequence was assembled using IRGSP-1.0, with the 93-11 genome as a reference (Figure 1A). The DP30 genome is displayed according to the input reads (Table S1). The sequencing results reveal that 1,894,103 bp across all 12 chromosomes were different in DP30 and 93-11, representing 6.1% of the genome (Figure 1B). InDel molecular markers were designed such that the base mutation exceeded 20 bp. Based on the lengths of the PCR products (150–200 bp) and the results of gel electrophoresis, we selected 285 InDel markers with an average distance of 1.2 Mb (Table 1). The sequences of the 285 InDel markers (Table S2).

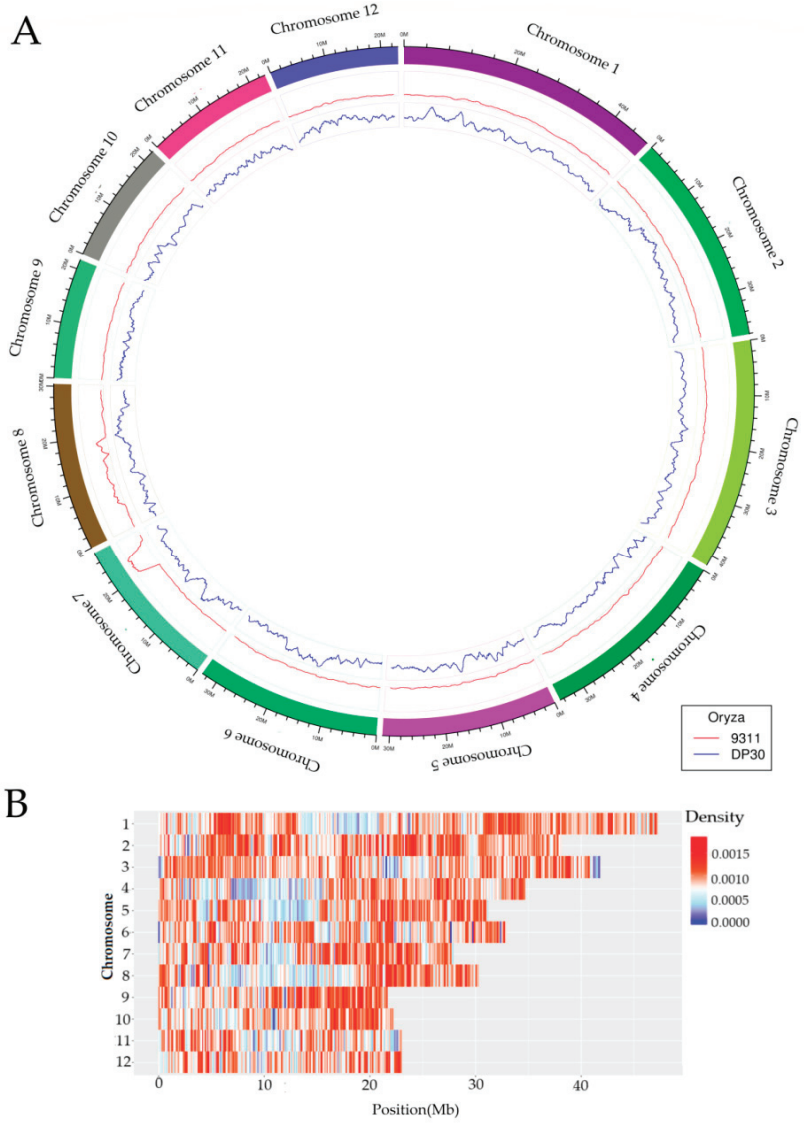


Figure 1. Whole-genome sequencing of DP30 and genome-wide sequence difference. (**A**) Circos plot for SNPs and InDels detected in 93-11 and DP30. The outermost circle depicts the ideogram of the 12 chromosomes in the rice genome. The innermost lines represent the frequency SNPs and InDel markers identified between DP30 (blue) and 93-11 (red); (**B**) genome-wide distribution of sequence differences between 93-11 and DP30. Red and blue colors represent the degree of variation between DP30 and 93-11. Depth of the color indicates the level of variation frequency.

Table 1. Number of single-nucleotide polymorphism (SNPs) detected in wild rice (DP30).

Chromosome	SNP No. (bp)	Chr. Length (bp)	SNP Percentage	Length (Mb)	Marker No.
1	232,372	47,283,185	0.49%	1.32	38
2	210,389	38,103,930	0.55%	1.15	30
3	194,967	41,884,883	0.47%	1.33	32
4	155,761	34,718,618	0.45%	1.21	25
5	164,450	31,240,961	0.53%	1.34	24
6	167,985	32,913,967	0.51%	1.22	25
7	153,895	27,957,088	0.55%	1.1	19
8	154,296	30,396,518	0.51%	1.2	21
9	119,958	21,757,032	0.55%	1.26	18
10	113,121	22,204,031	0.51%	1.21	17
11	107,638	23,035,369	0.47%	1.24	18
12	119,271	23,049,917	0.52%	1.18	18
Average	157,842	31,212,125	0.51%	1.23	24
Total	1,894,103	374,545,499	6.10%		285

SNP-No—number of SNPs detected in DP30; Chr.-length (bp)—chromosome length; Length (Mb)—average length between adjacent markers in each chromosome; SNP Percentage—percentage of SNP; Marker No.—the number of molecular markers on each chromosome; Mb—million bp.

3.3. Development and Substitution Segment Analysis of DP30-CSSLs

3.3.1. Development of the DP30-CSSLs

The procedure we developed the DP30-CSSLs is shown in Figure 2. The F_1 plants were obtained by crossing DP30 and 93-11. Molecular MAS began using the BC_4F_1, resulting in 176 BC_5 F_3 lines, 32 BC_6 F_3 lines and 22 BC_7 F_3 lines containing the target segments. Finally, we selected 132 CSSLs from these 230 lines (candidate plants) for genetic and phenotypic analysis (Figure 2, Table S3).

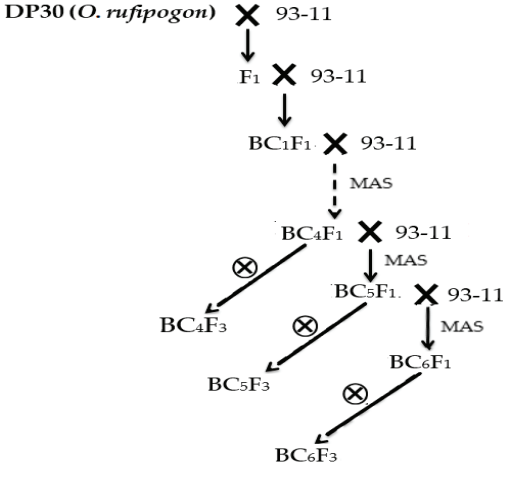


Figure 2. Schematic of the process used to develop DP30-CSSLs in this study. MAS—molecular marker-assisted selection; ×—hybrid; ⊗—self-crossing.

3.3.2. Substitution Segments of DP30-CSSLs

In total, we identified 80 BC_4 F_2 plants from 171 CSSLs and named as DP30-CSSLs. Then, 24 BC_4 F_2 plants were selected and 44 BC_5 F_2 plants were obtained after continuous backcrossing. Finally, 18 BC_6 F_2 were obtained with a relatively complex genetic background to continue backcrossing.

The 132 CSSLs formed a DP30-CSSLs population, and the target replacement segments of these CSSLs accumulated a total length of 536.4 Mb, covering 91.42% of the DP30 genome. The CSSLs were arranged according to the position of their target substitution segments. The 99 substitution lines contained only one substitution segment from DP30. The total length of these CSSLs substitution segments was 359.66 Mb and the coverage rate was 61.36% (Figure 3).

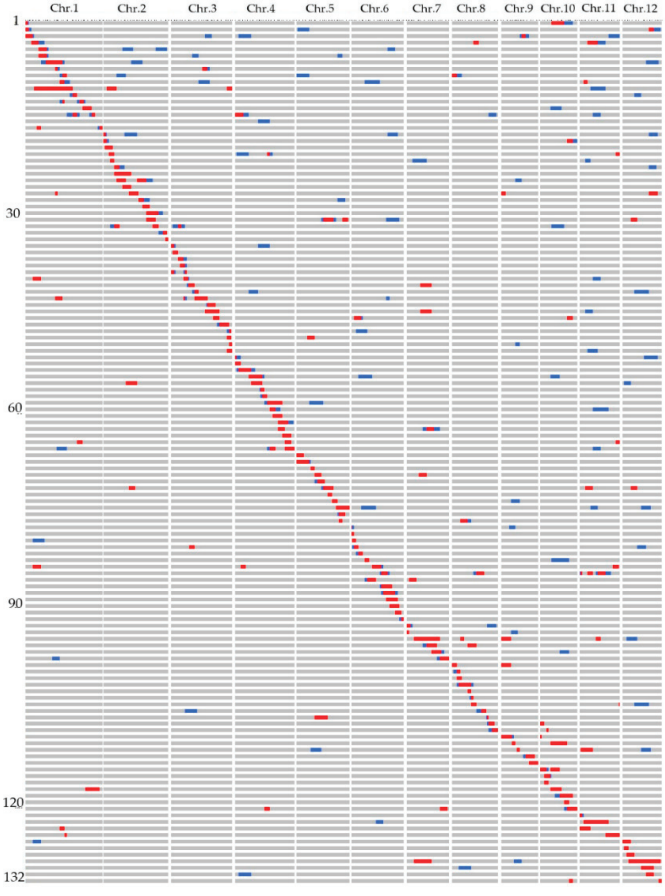


Figure 3. Graphic genotypes of the 132 DP30-CSSLs developed in this study. Red bars indicate homozygous substituted segments derived from DP30. Blue bars indicate heterozygous substituted segments derived from DP30. Gray bars indicate the genetic background of the recipient parent 93-11, "Chr."—chromosome.

Chromosomes with more than 90% coverage of the DP30 genome by chromosome substitution segments include chromosomes 1, 2, 3, 5, 7, 11 and 12. Among them, chromosome 5 had the highest coverage rate (97%), while the lowest coverage rate belonged to chromosome 8 (Table 2).

The average length of DP30-CSSLs population substitution target segment was 4.06 Mb, which ranges from 0.5–22 Mb. The length of the substitution segment of 25 CSSLs was less than 2 Mb; The length range of substitution segments of 55 substitution lines was 2–4 Mb and 30 CSSLs were 4–6 Mb and the length of the substitution segment of 6 CSSLs was larger than 8 Mb.

Table 2. Substitution segment analysis in DP30-CSSL.

Chromosome	No. of CSSLs	Total Length of Target Segment (Mb)	Length of Chromosome (Mb)	Coverage Length (Mb)	Coverage Rate (%)
1	17	72	47	44.5	94.7
2	17	61	38	36.5	96.1
3	17	52.5	41	38.1	93.3
4	15	66.6	35	31.9	91.2
5	11	41.9	32	29.4	92.3
6	15	56.8	33	28.7	87.9
7	6	31.9	28	26.6	95.5
8	11	29.6	30	24.9	83.3
9	5	21.7	22	19.1	87.5
10	7	31.6	22	19.4	88.6
11	4	26.2	23	21.4	93.1
12	7	24.9	23	21.7	93.6

CSSLs No.—number of substitution lines on chromosomes; Chr.-Length (Mb)—chromosome length; Coverage length (Mb)—covering length of the substitution segment to the chromosome; Percentage (%)—coverage of the substitution segment.

3.4. Wild rice QTLs in the DP30-CSSLs

We detected the genotypes of all CSSLs in DP30-CSSLs (Table S3). In total, 36 QTLs were identified in the CSSLs.

3.4.1. Tiller Angle (TA) and Heading Date (HD)

Four QTLs (*qTA1.1*, *qTA7.1*, *qTA9.1* and *qTA2.1*) related to TA were detected in the CSSLs. Seven CSSLs, (RZ18, RZ20, RZ38, RZ104, RZ129, RZ130 and RZ168) had different TAs those of 93-11 (Figure S2A,B). *qTA1.1* was located in the overlapping segment of RZ18 and RZ20, which belong to C1-26–C1-27 in chromosome 1. *qTA7.1* was detected in an overlapping segment of RZ104 and RZ168 that located in C7-2 on chromosome 7, the *PROG1* was also in the same position of chromosome [17]. *qTA9.1* and *TAC1* were detected in overlapping segments of RZ129 and RZ130, which was in the C9-14–C9-15 region of chromosome 9 [20] (Table 3). The TAs of the RZ38 plants were between 0° to 10°, while 93-11 plants had TAs of 30° to 15°, in both seasons (Figure S1A,B). The QTL *qTA2.1* in the segment of RZ38 was located near the molecular marker C2-25 on chromosome 2 (Table 3; Figure S6).

Three CSSLs (RZ142, RZ8 and RZ13) showed significant differences in HD when compared with 93-11 (Figure S2C). These three lines had an overlapping segment near C11-4 on chromosome 11, so we identified a new QTL (*qHD11.1*) in this region that has negative additive effects on HD (Table 3; Figure S6).

Table 3. List of 36 QTLs identified from DP30-CSSLs.

Traits	QTL	LOD	Position	Spring		Fall		Cloned Gene
				a	ap (%)	a	ap (%)	
TA	*qTA1.1*	*2.4*	C1-26–C1-27	−1.6	9.7	0.7	9.7	
TA	*qTA7.1*	*2.5*	C7-2	−2.4	17.2	−1.2	14	*PROG1* [17]
TA	*qTA9.1*	*2.7*	C9-14–C9-15	−1.4	16.5	−1.4	16.5	*TAC1* [20]
TA	*qTA2.1*	*2.6*	C2-25	0.4	4.3	0.6	7.3	
HD	*qHD11.1*	*3*	C11-4	−5	4.1	−3.72	2.9	
PH	*qPH1.1*	*3.1*	JM1-5	31.8	17.6	30.7	28.8	*Sd1* [37]
PH	*qPH6.1*	*3*	C6-8	28.1	24.6	27.3	25.6	
PH	*qPH7.1*	*2.6*	C7-5	25.9	19.5	26.7	25	
NGPP	*qNGPP3.1*	*2.8*	C3-19	34.2	19.8	33.1	20.1	
NGPP	*qNGPP4.1*	*5.2*	C4-12	25.2	14.6	25	15.2	
NGPP	*qNGPP12.1*	*2.9*	C12-2	40.3	23.3	37.3	22.6	
GWT	*qGWT1.1*	*3.1*	C1-19	−3	10.2	−0.3	11.1	*OsAGPL2* [38]
GWT	*qGWT2.1*	*4.2*	C2-19	−0.2	8	−0.2	7.8	
GWT	*qGWT3.1*	*4.4*	C3-25	−0.3	8.3	−0.3	8.3	
GWT	*qGWT4.1*	*3.4*	C4-15	−0.3	10	−0.3	9.9	
GWT	*qGWT5.1*	*4*	C5-23	−0.3	10	−0.3	10.1	
GL	*qGL3.1*	*3.8*	C3-19	−0.9	9	1	10	
GL	*qGL9.1*	*3.4*	C9-12	−0.9	8.5	−0.9	8.7	*SG1* [39]

Table 3. *Cont.*

Traits	QTL	LOD	Position	Spring		Fall		Cloned Gene
				a	ap (%)	a	ap (%)	
GW	*qGW9.1*	*3.2*	C9-8	−0.1	3.2	−0.1	2.7	
GW	*qGW10.1*	*3.1*	C10-3	−0.1	3.4	−0.1	3.1	
GLWR	*qGLWR3.1*	*3.5*	C3-26	0.3	11.4	0.3	11.1	
GLWR	*qGLWR8.1*	*3.9*	C8-2	0.3	10.9	0.3	11.2	
AL	*qAL1.1*	*3*	C1-7	5.6	26.9	5.6	32.7	
AL	*qAL4.1*	*2.7*	C4-8–C4-10	9.8	47.6	10.1	59.3	*An-1* [40]
AL	*qAL4.2*	*4.3*	C4-19	10.3	49.6	10.3	60.8	*An-2* [41]
AL	*qAL8.1*	*2.6*	C8-11	6.8	33.1	6.4	37.7	
AL	*qAL11.1*	*3.9*	C11-2	6.3	30.6	6.9	40.5	
SH	*qSH4.1*	*3.4*	C4-22	−12.5	−8.3	−12	−9	*SH4* [42]
SH	*qSH11.1*	*2*	C11-5~C11-8	−40	−26.7	−42	−27.7	
CT	*qCT1.1*	*2.6*	C1-16	1.3	48.5	1.2	53.3	*OsRAN1* [43]
CT	*qCT2.1*	*2.6*	C2-19	1.2	46.2	0.9	42.9	
CT	*qCT3.1*	*2.9*	C3-2	1.2	45.8	1.1	48.1	
CT	*qCT5.1*	*2.1*	C5-20	1.8	67.3	1.4	65.2	*OsiSAP8* [44]
CT	*qCT6.1*	*2.4*	C6-20	1.8	68.4	1.4	61.4	*OsPYL9* [45]
CT	*qCT10.1*	*3*	C10-3	1.4	51.9	1.4	64.8	
CT	*qCT12.1*	*2.2*	C12-11	1.2	46.8	1	46.5	

List of wild rice QTLs identified in this study. Position—physical position of molecular markers on chromosomes (Mb); length—length of the overlapping part of the lines; a—additive effect; Gene— cloned genes in the overlapping segments. TA— tiller angle; HD—heading date; PH—plant height; NGPP—number of grains per panicle; GWT—1000 grain weight; GL—grain length; GW—grain width; GLWR—grain length to width ratio; AL—awn length; SH—grain shattering; and CT—cold tolerance.

3.4.2. Plant Height (PH)

We found three QTLs (*qPH1.1, qPH6.1* and *qPH7.1*) related to PH in the CSSLs, which were distributed on chromosome 1, 6 and 7, respectively. Five CSSLs (RZ14, RZ16, RZ17, RZ92, RZ105, RZ107 and RZ164) had PH values of 170 ± 9 cm, while 93-11 plants were about 110 ± 5 cm (Figure S2D). *qPH1.1* was identified in the overlapping segment of RZ14, RZ16, RZ17 and the *Sd1* was in the same position [37]. The RZ92 and RZ164 segments overlapped near the molecular marker C6-8 on chromosome 6 and the *qPH6.1* was identified in this region. The *qPH7.1* was identified in the overlapping segments of RZ105 and RZ107 near the molecular marker C7-5 on chromosome 7 (Table 3; Figure S6).

3.4.3. Number of Grains per Panicle (NGPP) and 100 Grain Weight (GWT)

We identified three QTLs (*qNGPP3.1, qNGPP4.1* and *qNGPP12.1*) related to NGPP that were distributed on chromosome 3, 4 and 12. Seven CSSLs (RZ51, RZ50, RZ52, RZ64, RZ20, RZ144 and RZ145) had more than 210 NGPP while 93-11 had less than 180 NGPP in both seasons (Figure 4A). We identified QTL *qNGPP3.1* in C3-19 and *qNGPP4.1* in C4-12. An overlapping segment in C12-2 contained a QTL named *qNGPP12.1* (Table 3; Figure S6).

3.4.4. Grain Length (GL), Grain Width (GW) and Grain Length to Width Ratio (GLWR)

Two QTLs (*qGL3.1* and *qGL9.1*) related to GL were detected in the CSSLs on chromosome 3 and 9. Three CSSLs (RZ3, RZ49 and RZ50) had GLs of 8.40 ± 0.06 mm, while 93-11 had a GL of 10.03 ± 0.13 mm (Figure 4B). The RZ49 and RZ50 segments overlapped near the molecular marker C3-19 on chromosome 3 and this new QTL was named *qGL3.1*. The QTLs *qGL9.1* and *SG1* were detected in the segment of RZ3 near the molecular marker C9-12 on chromosome 9 [39] (Table 3; Figure S6).

Two QTLs (*qGW9.1* and *qGW10.1*) related to GW were detected on chromosome 9 and 10. Three CSSLs, (RZ133, RZ3 and RZ134) had GWs of less than 2.85 mm, while 93-11 had a GW of more than 2.98 mm (Figure 4B). The RZ3 segment was located near the molecular marker C9-8 on chromosome 9 and named *qGW9.1*. We identified *qGW10.1* in the overlapping segment of RZ133 and RZ134, near the molecular marker C10-3 on chromosome 10 (Table 3; Figure S6).

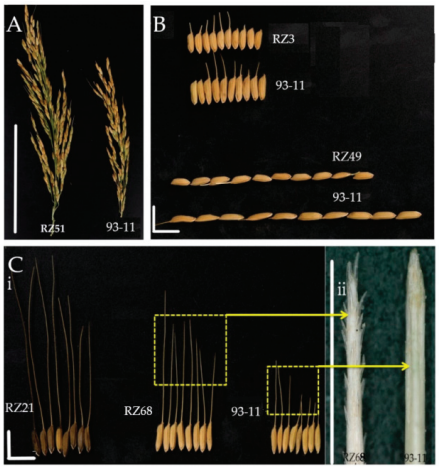


Figure 4. Phenotypic variation in the appearance of grains and awns in CSSLs and 93-11. (**A**) Grain number per ear for RZ50 and 93-11 (bar = 10 cm); (**B**) length of grains of RZ3 and 93-11 (bar = 1 cm); (**C**) Phenotypes of the awns of RZ21, RZ68 and 93-11: (**A**) awn length (bar = 1 cm) and (**C**) awn tip surfaces of RZ68 and 93-11 (bar = 1 mm).

Two QTLs (*qGLWR3.1* and *qGLWR8.1*) related to GLWR were detected on chromosome 3 and 8, respectively. Three CSSLs, (RZ56, RZ111 and RZ167) had GLWRs of 3.5 while 93-11 had a GLWR of 3.3. The overlapping segments of RZ56 and RZ167 were found near the molecular marker C3-26 on chromosome 3 and named *qGLWR3.1*. We detected *qGLWR8.1* in the segments of RZ112, near the molecular marker C8-2 on chromosome 8 (Table 3; Figure S6).

3.4.5. Awn Length (AL)

Five QTLs (*qAL1.1*, *qAL4.1*, *qAL4.2*, *qAL8.1* and *qAL11.1*) related to awn length (AL) were detected on chromosome 1, 4, 8 and 11. Twelve CSSLs (RZ21, RZ6, RZ4, RZ62, RZ63, RZ70, RZ68, RZ119, RZ4, RZ141, RZ88 and RZ142) had ALs of 30 ± 4 mm, while the AL of 93-11 was 14 ± 3 mm (Figure 4C). The overlapping segments of RZ21, RZ6 and RZ4 were found near the molecular marker C1-7 on chromosome 1 and named *qAL1.1*. The QTLs *qAL4.1* and *An-1*/allele were detected in RZ62 and RZ63 near the molecular marker C4-8–C4-10 on chromosome 4. [40] Two other QTLs, *qAL4.2* and *An-2*/allele were detected in the overlapping segments of RZ70 and RZ68 near the molecular marker C4-19 on chromosome 4 [41]. The grains of common wild rice have long and spiny awns. In contrast, cultivated rice species have no awns or short awns with smooth surfaces (Figure S4C). The overlapping segments of RZ135 and RZ4 were located near the molecular marker C8-11 on chromosome 8 and named *qAL8.1*. We found *qAL11.1* in the overlapping segments of RZ142, RZ88 and RZ141 near the molecular marker C11-2 on chromosome 11 (Table 3; Figure S6).

3.4.6. Grain Shattering (SH) and Cold Tolerance (CT)

The grains of 93-11 did not shatter in either season, but the grains of five CSSLs (RZ74, R75, RZ76, RZ8 and RZ88) did shatter (Figure 5). We found *qSH4.1* in the overlapping segments of RZ74, R75 and RZ76 near the molecular marker C4-22 on chromosome 4, the *SH4* was also in the same chromosomal position [42]. *qSH11.1* in the overlapping segments of RZ8 and RZ88 (Table 3; Figure S6).

Seven QTLs (*qCT1.1*, *qCT2.1*, *qCT3.1*, *qCT5.1*, *qCT6.1*, *qCT10.1* and *qCT12.1*) related to cold tolerance (CT) were detected on chromosome 1, 2, 3, 5, 6, 10 and 12. Seventeen CSSLs (RZ10, RZ11, RZ12, RZ34, RZ40, RZ41, RZ85, RZ86, RZ87, RZ99, RZ100, RZ101, RZ136, RZ137, RZ158, RZ162 and RZ163) were more tolerant to cold stress compared with 93-11 (Figure S3). We identified *qCT1.1* in the overlapping

segments of RZ10, RZ11 and RZ12 near the marker C1-16 on chromosome 1, which also contained the gene *OsRAN1* [43]. *qCT2.1* was found in the overlapping segments of RZ34, RZ40, RZ137, RZ158 and RZ170 near the marker C2-19. *qCT3.1* was detected in the overlapping segments of RZ40 and RZ41 near marker C3-2. *Qc T5*.1 was identified in the overlapping segments of RZ85, RZ86, RZ87 and RZ170 near the marker C5-20 on chromosome 5, the *OsiSAP8* was also in the same chromosomal position [44]. The QTLs *qCT6.1* was found in the overlapping segments of RZ99, RZ100, RZ101 and RZ171 near the marker C6-20 on chromosome 6, the *OsPYL9* was also in the same chromosomal position [45]. chromosome 10 contained the QTL *qCT10.1*, which was identified in the overlapping segments of RZ136, RZ137 and RZ171 near the marker C10-3. The overlapping segments of RZ158, RZ162 and RZ163 were located near the marker C12-11 on chromosome 12 and named *qCT12.1* (Table 3).

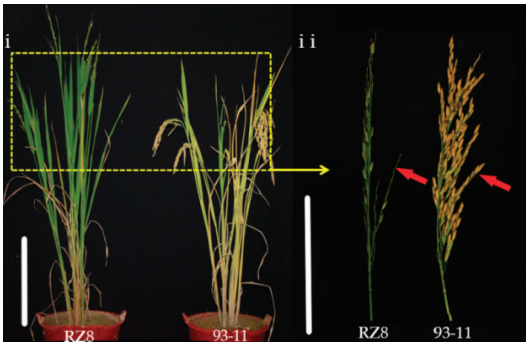


Figure 5. Phenotypic variation of grain shattering in RZ8 and 93-11; Whole plant phenotype showing grain shattering (bar = 50 cm), grain shattering (bar = 10 cm, red arrows).

3.4.7. Identification of Loci Related to Quality Traits

In five CSSLs (RZ5, RZ112, RZ19, RZ75 and RZ106), significant variation with 93-11 in LSC, LMC, AC, SC, GC and SCC (Figures S4 and S5). Segments of these CSSLs are located in chromosomes 1, 4, 7 and 11. Six related trait genes (*LSC11.1*, *LMC11.1*, *AC1.1*, *SC1.1*, *GC4.1* and *SCC7.1*) were identified; *AC1.1* and *SC1.1* may be the same gene. *AC1.1* and *SC1.1* may be the allele gene of *A. GC4.1*, SCC7.1 and *Bh4, Rc* were in the same chromosome position [46–48].

3.5. Construction of a Secondary Population and Mapping of qCT2.1

Our results showed that there was no significant difference in the appearance of RZ34 and 93-11 seedlings before cold treatment (Figure 6A). RZ34 had a lesser degree of wilting after cold treatment compared with 93-11 (Figure 6B). Furthermore, 93-11 died three days after the cold treatment, while RZ34 survived and continued to grow (Figure 6C). The results show that RZ34 seedlings were more tolerant to cold stress than 93-11. All the F_1 individuals in the RZ34 line exhibited a cold tolerant phenotype. We selected some of the F_1 plants and the related molecular markers showed that these plants were heterozygous for the tested genotypes. We obtained a total of 311 plants in the F_2 population, among them 223 plants were cold tolerant, and 88 plants were susceptible to cold stress. This represented a 3:1 segregation ratio for cold tolerance which consistent with Mendelian's rules ($\chi^2 = 1.630 < \chi^2_{0.05,1} = 3.84$), demonstrating that this cold tolerance trait was controlled by a single dominant QTL.

We identified the QTL *qCT2.1* in the segment of RZ34 (Figure 7A) between markers JM2-3 (21.7 Mb) and C2-22 (29.18 Mb). Based on the tracking and overlap of three CSSLs (RZ37, RZ38 and RZ39) segments, the range of *qCT2.1* was reduced to 4.47 Mb (Figure 7A). Five InDel markers (dxw-3, dxw-9, dxw-8, dxw-5 and dxw-4) were developed between C2-18 and C2-20 (Table S4). In order to reduce phenotypic identification error and improve the accuracy of QTL mapping, we selected individual

plants with extreme phenotypic values to detect genotypes. We identified the phenotypes of 311 plants in the F_2 population. Among them, the genotypes of 56 cold-tolerant plants (scale 9–7) and 22 cold-sensitive plants (scale 0) were detected for linkage analysis. Five recombinant plants (R1, R2, R3, R4 and R5) were selected from the constructed F_2 population (Table S5). These plants confirmed that *qCT2.1* is located in a 1.7 Mb region between molecular markers dxw-4 and dxw-9 (Figure 7B) and that the LOD value of *qCT2.1* was 8 (Figure 7C).

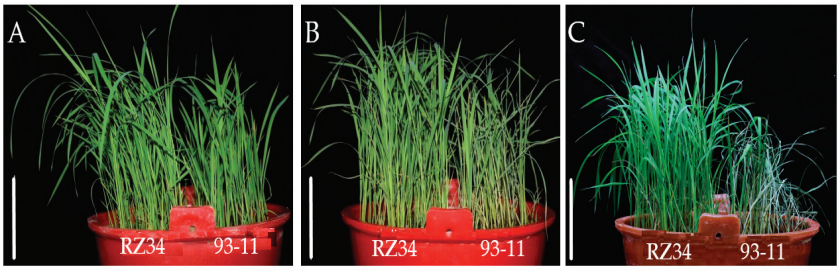


Figure 6. Cold stress treatment to RZ34 and 93-11 at seedling stage. (**A**) Before cold stress (bar = 10 cm); (**B**) after cold stress (bar = 10 cm) and (**C**) after three days of recovery (bar = 10 cm).

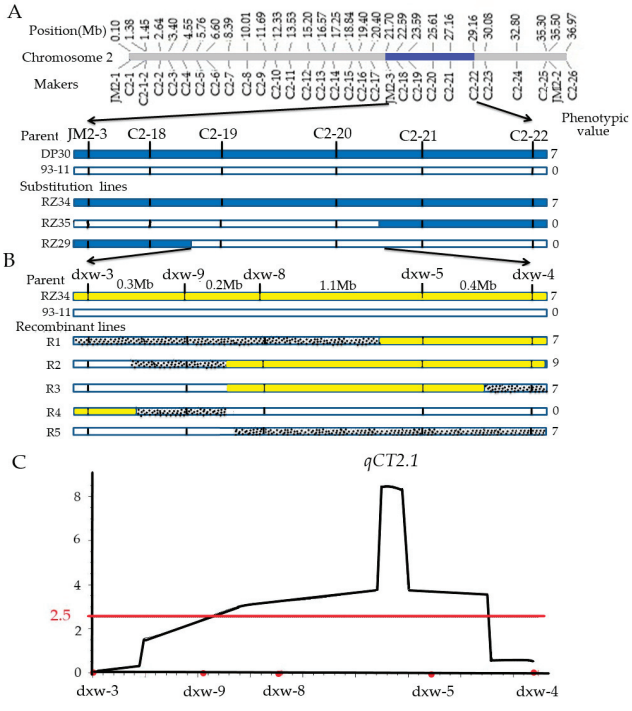


Figure 7. Mapping of grain-shattering QTL *qCT2.1*, (**A**) QTL analysis by overlapping segments of recombinants RZ34, RZ35 and RZ29, blue intervals indicate DP30 genotype; blank intervals indicate: 93-11 genotype; shadow intervals indicate heterozygous genotype; (**B**) linkage analysis of *qCT2.1* yellow intervals indicate RZ34 genotype, blank intervals indicate 93-11 genotype, shadow intervals indicate heterozygous genotype; (**C**) LOD value of *qCT2.1*. dxw-3, dxw-8, dxw-9, dxw-5 and dxw-4 are InDel markers.

4. Discussion

*4.1. Constructing CSSLs in Guangxi Wild Rice (*O. rufipogon *Griff.) DP30*

Since QTL detection is based on the natural allelic differences between parental cultivars, it is important to select parental cultivars that show large phenotypic variation in the target traits [49]. There is rich polymorphism between Guangxi wild rice DP30 (*O. rufipogon* Griff.) and cultivated rice 93-11 (*O. sativa* L.) due to their distant genetic bases [10]. In this study, we performed MAS using Guangxi wild rice DP30 (*O. rufipogon* Griff) as a donor parent to develop 132 DP30-CSSLs. The coverage rate of the DP30 wild rice was 91.55%. The average length of each replacement segment was 0.5–22 Mb. Compared with the previously reported CSSLs in wild rice, the DP30-CSSLs had a higher substitution segment coverage rate and more polymorphic primers for MAS [50,51]. Clearly, the genomic constitution quality of the CSSLs population is important. The chromosome position and genetic effect of QTLs locating on dp30-CSSLs will be more accurately assessed.

4.2. Identification of Quantitative Trait Loci and Measurement of Various Traits in Fall and Spring

Previous reports have shown that the CSSLs of wild rice are effective in the mining and transferring of wild alleles into cultivated rice [10]. Furthermore, several studies have reported the development of CSSLs and the identification of several agronomic and plant architecture traits [17,35–41]. Ma et al. (2019) detected eighteen QTLs were two known grain length- and width-related genes and four novel QTLs. In addition, two QTLs were verified, and two novel QTLs were identified, for panicle neck length, a domestication-related trait [49]. Tan et al. (2004) identified quantitative trait loci (QTLs) associated with plant height and the days to heading in the BC_3 F_2 population. Putative QTLs derived from O. *rufipogon* were detected for plant height on chromosome 1 and identified 6 QTLs for days to heading on chromosomes 1, 3, 7, 8 and 11 [50].

In the present work, we compared the phenotypes of DP30-CSSLs to the phenotypes of known genes and used related molecular markers to confirm whether any of these genes/allele genes were present in the segments. We identified five QTLs related to TA, *qTA7.1* and *PROG1* were in the same position on the chromosome; *qTA9.1* may be the allelic form of *TAC1* (Table 3). *PROG1* (LOC_Os07g05900) controls the creeping growth habits of common wild rice [17]. *TAC1* (LOC_Os09g35980) is a recently discovered gene that controls the TA of rice corresponding to a major QTL [20]. We found that these five QTLs had positive additive effects on TA (Table 3). A QTL named *qHD11.1* had negative additive effects on HD. Three of the QTLs identified in this study (Table 3) had positive additive effects for phenotypic variations in PH in both fall and spring. *qPH1.1* and *Sd1* were in the same position on the chromosome (Table 3). The *Sd1* (LOC_Os01g66100) gene, which controls gibberellin biosynthesis, was among these QTLs [36]. Three QTLs showed negative additive effects on NGPP and *qGWT1.1* may be the allele of *OsAGPL2* (Table 3). *OsAGPL2* (LOC_Os01g44220) is a member of the *OsPYL* gene family that regulates the filling rate of grains, leading to lower final grain weight and yields [37]. Except for the QTLs locus near the *OsAGPL2* gene, we also identified eight other QTLs. All of them showed negative additive effects on 100GWT in both seasons (Table 3). The *SG1* (LOC_Os09g28520) gene has shorter grains than the wild type and a dwarf phenotype [39]. Similar to the *SG1 gene,* one new QTL we identified, which showed negative additive effects on GW (Table 3). We identified five AL-related QTL, *qAL4.1* and *qAL4.2* as alleles of *An-1* and *An-2*, respectively (Table 3). *An-1* (LOC_Os04g28280) regulates the formation of awn primordia, promoting awn elongation and increasing the length of grains [40]. *An-2* (LOC_Os04g43840) also increases the length of awns and makes them spinier (Figure S4) [41]. The QTL *qSH4.2*, which is related to grain shattering, the *SH4* (LOC_Os04g57530) was in the same chromosome position [42] (Table 3). Cold tolerance in seedlings is one of the important traits for the stable production of rice [43–45]. Here, we identified seven QTLs related to cold tolerance including these loci in a similar region with these three previously cloned genes. *qCT1.1*, *qCT5.1* and *qCT6.1* may be the allelic form of *OsRAN1*, *OsiSAP8* and *OsPYL9*, respectively (Table 3). *OsRAN1* (LOC_Os010611100) participates in cell division and the cell cycle and promotes the formation of intact nuclear membranes, thus improving

the cold tolerance of rice [43]. *OsiSAP8* (LOC_Os06g41010) is a zinc finger protein gene that enhances salt, drought and cold stress tolerance in rice [44]. *OsPYL9* (LOC_Os06g33690) is a member of the *OsPYL* gene family and is a possible abscisic acid (ABA) receptor [45]. In addition, there are six quality trait loci, which contain three alleles [46–48]. The alleles of eleven cloned genes showed the reliability of DP30-CSSLs. No cloned genes were found in other QTL, which may contain new genes.

4.3. Construction of Secondary Population and Mapping of qCT2.1

Map-based cloning and mapping of cold tolerance genes in rice have always been a classical method for cold tolerance research in rice.

Previous studies used different populations to obtain some cold tolerance genes of rice [51,52]. According to the published data, more than 250 QTL of low-temperature tolerance has been found on 12 chromosomes of rice. In DP30-CSSLs, *qSCT-3-1* was identified in the RM15031-RM3400 region of the long arm of chromosome three near to the centromere, and the genetic distance between the linkage markers was found to be 1.8 cM [53]. In this study, F_2 populations were constructed by Guangxi common wild rice seedling cold-tolerant segment replacement line RZ34 and cold-sensitive recurrent parent 93-11. Through map-based cloning, it was found that the main cold tolerance QTL *qCT2.1* of rice at the seedling stage was located on chromosome 2 and located in the range of 1.7 Mb between molecular marker dxw-4 and dxw-9. To date, there is no cloned cold tolerance gene at the seedling stage in this interval. *qCT2.1* could enhance cold tolerance at the seedling stage, which has a strong dominant effect, so it is expected to be used in rice breeding.

Rice breeding entered to a new era with the utilization of MAS and whole-genome sequencing to link genotypes with phenotypes. The introduction of wild rice CSSLs promoted gene QTLs mapping and genomic research. This study also suggests using Guangxi common wild rice accessions will provide a broad platform for genomic research and may lead to the discovery of new QTLs that will benefit rice breeding.

5. Conclusions

This study aims to use wild rice to develop CSSLs for cultivated rice and use these CSSLs for comparative mapping of traits related to plant architecture and yield. We focused on germplasm innovation in rice through the identification and transfer of beneficial genes/QTLs from the wild species. We introduced the DP30-CSSL library platform to facilitate pre-designed breeding of cultivated rice to utilize favorable alleles dispersed in Guangxi wild rice resources. The QTLs presented here are expected to provide further clues to identifying underlying mechanisms involved in plant architecture and improved grain. Our ongoing experiments are aimed at confirming the genomic regions and narrowing down of number of genes reported within the QTLs in the present study through comprehensive studies involving high-resolution linkage mapping via high-throughput genotyping by sequencing of advanced generation progenies.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/9/980/s1, Figure S1: Frequency distributions of quantitative traits in DP30-CSSLs during fall and spring, Figure S2: Phenotypic variation in plant-architecture-related traits and heading date in CSSLs and 93-11, Figure S3: Cold tolerance phenotype of some CSSLs and 93-11 at seedling stages (bar = 10 cm), Figure S4: Phenotypic variation of leaf traits in CSSLs and 93-11, Figure S5: Phenotypic variation of in apiculi, glumes and seed coats in CSSLs and 93-11, Figure S6: The distribution of the 36 QTLs in DP30-CSSLs. The molecular markers are shown on the left and the QTL sites are shown on the right. A QTL overlap indicates that there are two QTL in the same region. Table S1: Whole-genome re-sequencing analysis of wild rice DP30. Table S2: The sequence of DP30-CSSLs molecular markers, Table S3: Substitution segments of DP30-CSSLs population, Table S4: The markers sequences of InDel molecular markers for mapping QTL qCT2.1, Table S5: Genotypes and phenotypes of secondary F_2 populations.

Author Contributions: Conceptualization, R.Y. and N.Z.; Data curation, B.U., S.L. and G.N.; Formal analysis, R.Y., N.Z. and Y.Q.; Funding acquisition, R.L.; Investigation, R.Y., B.U., L.L., Y.Q. and G.N.; Methodology, R.Y., L.L., S.L. and G.N.; Project administration, R.L.; Resources, R.L.; Software, N.Z.; validation, R.L.; Visualization, B.U.; Writing—Original draft, R.Y., N.Z., B.U. and G.N.; Writing–review & editing, B.U. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Technology Research and Development Program Guike, Guangxi (Guike AB16380066; Guike AB16380093).

Acknowledgments: We would like to thank Mohsin Niaz, Umair Khalid, Mudassar Abbas, Baoxiang Qin and Fang Liu for the helpful discussion and invaluable comments to make this research meaningful.

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

References

1. Doebley, J.F.; Gaut, B.S. The molecular genetics of crop domestication. *Cell* **2006**, *127*, 1309–1321. [CrossRef] [PubMed]
2. Yano, M.; Sasaki, T. Genetic and molecular dissection of quantitative traits in rice. In *Oryza: From Molecule to Plant*; Springer: Dordrecht, The Netherlands, 1997; pp. 145–153.
3. Li, Z.; Pinson, S. Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (Oryza sativa L.). *Theor. Appl. Genet.* **1995**, *91*, 374–381. [CrossRef] [PubMed]
4. Ray, J.; Yu, L.; McCouch, S. Mapping quantitative trait loci associated with root penetration ability in rice (Oryza sativa L.). *Theor. Appl. Genet.* **1996**, *92*, 627–636. [CrossRef] [PubMed]
5. Redona, E.; Mackill, D. Mapping quantitative trait loci for seedling vigor in rice using RFLPs. *Theor. Appl. Genet.* **1996**, *92*, 395–402. [CrossRef] [PubMed]
6. Yano, M. Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **2001**, *4*, 130–135. [CrossRef]
7. Fujita, D.; Ebron, L.A. Fine mapping of a gene for low-tiller number, Ltn, in japonica rice (Oryza sativa L.) variety Aikawa 1. *Theor. Appl. Genet.* **2010**, *120*, 1233–1240. [CrossRef]
8. Eshed, Y.; Zamir, D. Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* **1996**, *143*, 1807–1817. [CrossRef]
9. Furuta, T.; Uehara, K. Development and evaluation of chromosome segment substitution lines (cssls) carrying chromosome segments derived from oryza rufipogon in the genetic background of Oryza Sativa L. *Breed. Sci.* **2014**, *63*, 468–475. [CrossRef]
10. Zhao, H.; Sun, L. Genetic characterization of the chromosome single-segment substitution lines of O. glumaepatula and O. barthii and identification of QTLs for yield-related traits. *Mol. Breed.* **2019**, *39*, 51. [CrossRef]
11. Wu, W.; Liu, X. A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. *Nat. Plants* **2017**, *3*, 17064. [CrossRef]
12. Lin, Y.; Tan, L. RLS3, a protein with AAAt domain localized in chloroplast, sustains leaf longevity in rice. *J. Integr. Plant Biol.* **2016**, *58*, 971–982. [CrossRef] [PubMed]
13. Yang, Z.; Jin, L. Analysis of epistasis among QTLs on heading date based on single segment substitution lines in rice. *Sci. Rep.* **2018**, *8*, 2071. [CrossRef] [PubMed]
14. Kaneda, C.; Beachell, H.M. Response of indica-japonica rice hybrids to low temperatures. *SABRAO J.* **1973**, *6*, 17–32.
15. Xiao, N.; Gao, Y.; Qian, H.; Gao, Q.; Wu, Y.; Zhang, D.; Zhang, X.; Yu, L.; Li, Y.; Pan, C. Identification of genes related to cold tolerance and a functional allele that confers cold tolerance. *Plant Physiol.* **2018**, *177*, 1108–1123. [CrossRef] [PubMed]
16. Yiting, S.; Zhizhong, G. One SNP in cold1 determines cold tolerance during rice domestication. *J. Genet. Genom.* **2015**, *42*, 133–134. [CrossRef]
17. Tan, L.; Li, X. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **2008**, *40*, 1360–1364. [CrossRef]
18. Wang, J.; Wang, J. Proteomic response of hybrid wild rice to cold stress at the seedling stage. *PLoS ONE* **2018**, *13*, e0198675. [CrossRef]
19. Ando, T.; Yamamoto, T. Genetic dissection and pyramiding of quantitative traits for panicle architecture by using chromosomal segment substitution lines in rice. *Theor. Appl. Genet.* **2008**, *116*, 881–890. [CrossRef]
20. Jiang, J.; Tan, L. Molecular evolution of the TAC1 gene from rice (Oryza sativa L.). *J. Genet. Genom.* **2012**, *39*, 551–560. [CrossRef]
21. Saito, K.; Hayano-Saito, Y. Physical mapping and putative candidate gene identification of a quantitative trait locus Ctb1 for cold tolerance at the booting stage of rice. *Theor. Appl. Genet.* **2004**, *109*, 515–522. [CrossRef]

22. Zhu, H.; Liu, Z. Detection and characterization of epistasis between qtls on plant height in rice using single segment substitution lines. *Breed. Sci.* **2015**, *65*, 192–200. [CrossRef]
23. Ookawa, T.; Aoba, R. Precise estimation of genomic regions controlling lodging resistance using a set of reciprocal chromosome segment substitution lines in rice. *Sci. Rep.* **2016**, *6*, 30572. [CrossRef]
24. Wingett, S.W.; Andrews, S. FastQ screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338. [CrossRef]
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]
26. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows—Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef] [PubMed]
27. Kawahara, Y.; de la Bastide, M. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]
28. McKenna, A.; Hanna, M. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]
29. Rogers, S.O.; Bendich, A.J. Extraction of DNA from plant tissues. In *Plant Molecular Biology Manual*; Springer: Dordrecht, The Netherlands, 1989; pp. 73–83.
30. Qiao, W.; Qi, L. Development and characterization of chromosome segment substitution lines derived from Oryza rufipogon in the genetic background of O. sativa spp. indica cultivar 93-11. *BMC. Genom.* **2016**, *17*, 580. [CrossRef]
31. Panaud, O.; Chen, X. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (Oryza sativa L.). *Mol. Gen. Genet.* **1996**, *252*, 597–607. [CrossRef]
32. Young, N.; Tanksley, S. Restriction segment length polymorphism maps and the concept of graphical genotypes. *Theor. Appl. Genet.* **1989**, *77*, 95–101. [CrossRef]
33. Wan, X.; Weng, J. Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele gw-5 in a recombination hotspot region on chromosome5. *Genet* **2008**, *179*, 2239–2252. [CrossRef] [PubMed]
34. McCouch, S.R. Gene nomenclature system for rice. *Rice* **2008**, *1*, 72–84. [CrossRef]
35. Voorrips, R. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [CrossRef] [PubMed]
36. Meng, L.; Li, H.; Zhang, L.; Wang, J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **2015**, *3*, 269–283. [CrossRef]
37. Ye, H.; Feng, J. Map-based cloning of seed dormancy1-2 identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. *Plant Physiol.* **2015**, *169*, 2152–2165. [CrossRef] [PubMed]
38. Wei, X.; Jiao, G. GRAIN INCOMPLETE FILLING 2 regulates grain filling and starch synthesis during rice caryopsis development. *J. Integr. Plant Biol.* **2017**, *59*, 134–153. [CrossRef] [PubMed]
39. Nakagawa, H.; Tanaka, A. Short grain1 decreases organ elongation and brassinosteroid response in rice. *Plant. Physiol.* **2012**, *158*, 1208–1219. [CrossRef]
40. Luo, J.; Liu, H. An-1 encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant. Cell* **2013**, *25*, 3360–3376. [CrossRef]
41. Gu, B.; Zhou, T. An-2 encodes a cytokinin synthesis enzyme that regulates awn length and grain production in rice. *Mol. Plant* **2015**, *8*, 1635–1650. [CrossRef]
42. Lin, Z.; Griffith, M.E. Origin of seed shattering in rice (Oryza sativa L.). *Planta* **2007**, *226*, 11–20. [CrossRef]
43. Xu, P.; Cai, W. RAN1 is involved in plant cold resistance and development in rice (Oryza sativa). *J. Exp. Bot.* **2014**, *65*, 3277–3287. [CrossRef] [PubMed]
44. Kanneganti, V.; Gupta, A.K. Overexpression of OsiSAP8, a member of stress associated protein (SAP) gene family of rice confers tolerance to salt, drought and cold stress in transgenic tobacco and rice. *Plant Mol. Biol.* **2008**, *66*, 445–462. [CrossRef] [PubMed]
45. Kim, H.; Hwang, H. A rice orthologous of the ABA receptor, OsPYL/RCAR5, is a positive regulator of the ABA signal transduction pathway in seed germination and early seedling growth. *J. Exp. Bot.* **2011**, *63*, 1013–1024. [CrossRef] [PubMed]

46. Chandraratna, M.F. A gene for photoperiod sensitivity in rice linked with apiculus colour. *Nature* **1953**, *171*, 1162–1163. [CrossRef] [PubMed]
47. Zhu, B.F.; Si, L. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Phys.* **2011**, *155*, 1301–1311. [CrossRef] [PubMed]
48. Furukawa, T.; Maekawa, M. The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **2007**, *49*, 91–102. [CrossRef]
49. Ma, X.D.; Han, B. Construction of chromosome segment substitution lines of Dongxiang common wild rice (Oryza rufipogon Griff.) in the background of the japonica rice cultivar Nipponbare (Oryza sativa L.). *Plant Physiol. Biochem.* **2019**, *144*, 274–282. [CrossRef]
50. Tan, L.B.; Zhang, P.J. Identification of quantitative trait loci controlling plant height and days to heading from Yuanjiang common wild rice (Oryza rufipogon Griff.). *Acta Genet. Sinica* **2004**, *31*, 1123–1128. [CrossRef]
51. Andaya, V.; Mackill, D. Qtls conferring cold tolerance at the booting stage of rice using recombinant inbred lines from a japonica × indica cross. *Theor. Appl. Genet.* **2003**, *106*, 1084–1090. [CrossRef]
52. Koji, S.; Yuriko, H.S. Map-based cloning of the rice cold tolerance gene ctb1. *Plant Sci.* **2010**, *179*, 97–102. [CrossRef]
53. Zheng, J.X.; Ma, Z.F. Identification and mapping of QTLs for cold tolerance at the seedling stage in common wild rice (Oryza rufipogon). *China J. Rice Sci.* **2011**, *25*, 52–58.

genes

MDPI

Article

Genome-Wide Association Study (GWAS) for Mesocotyl Elongation in Rice (*Oryza sativa* L.) under Multiple Culture Conditions

Hongyan Liu [1,†], Junhui Zhan [1,†], Jiaolong Li [1], Xiang Lu [1], Jindong Liu [1], Yamei Wang [1,*], Quanzhi Zhao [2,*] and Guoyou Ye [1,3]

1 CAAS-IRRI Joint Laboratory for Genomics-Assisted Germplasm Enhancement, Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China; hongyanliu@caas.cn (H.L.); jhzhanhau@126.com (J.Z.); lijiaolong@caas.cn (J.L.); luxiang@caas.cn (X.L.); Liujindong@caas.cn (J.L.); g.ye@irri.org (G.Y.)

2 Collaborative Innovation Center of Henan Grain Crops and Key Laboratory of Rice Biology in Henan Province, College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China

3 Strategic Innovation Platform, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines

* Correspondence: wangyamei@caas.cn (Y.W.); qzzhaoh@126.com (Q.Z.); Tel./Fax: 0755-2839-4429 (Y.W.)

† Both authors contributed equally to this work.

Received: 10 December 2019; Accepted: 27 December 2019; Published: 31 December 2019

Abstract: Mesocotyl is a crucial organ for pushing buds out of soil, which plays a vital role in seedling emergence and establishment in dry direct-seeded rice. However, the genetic mechanisms of mesocotyl elongation remains unclear. In our study, 208 rice accessions were used to identify the SNPs significantly associated with mesocotyl length under various culture conditions, including sand, water and soil. The mesocotyl length ranges from 0 to 4.88 cm, 0 to 3.99 cm and 0 to 4.51 cm in sand, water and soil covering, respectively. A total of 2,338,336 SNPs were discovered by re-sequencing of 208 rice accessions. Genome-wide association study (GWAS) based on mixed linear model (MLM) was conducted and 16 unique loci were identified on chromosomes 1, 2 (2), 3, 4, 5 (2), 6 (2), 7, 8, 9 (2) and 12 (3), respectively, explaining phenotypic variations ranging from 6.3 to 15.9%. Among these loci, 12 were stable across two or more environments. Ten out of the sixteen loci coincided with known genes or quantitative trait locus (QTL), whereas the other six were potentially novel loci. Furthermore, five high-confidence candidate genes related to mesocotyl elongation were identified on chromosomes 1, 3, 5, 9 and 12. Moreover, qRT-PCR analysis showed that all the five genes showed significant expression difference between short-mesocotyl accessions and long-mesocotyl accessions. This study provides new insights into the genetic architecture of rice mesocotyl, the associated SNPs and germplasms with long mesocotyl could be useful in the breeding of mechanized dry direct-seeded rice.

Keywords: dry direct-seeded rice; early vigor; QTL; candidate gene; phenotyping

1. Introduction

Rice (*Oryza sativa*) is one of the most important cereal crops grown worldwide. Dry direct-seeded rice refers to the process of establishing the crop from seeds sown on non-puddled and unsaturated soil; in contrast, seedlings from nursery are transplanted into puddled or submerged soil in transplanted rice [1]. Compared with traditional transplanted rice, dry direct-seeded rice has been proposed as a water-efficient and labor-saving approach, which can reduce the cost of water and labor at about 50% [2,3]. Also, dry direct-seeded rice could efficiently utilize early-season rainfall in drought-prone environments and complete its growth cycle within the wet season in rainfed lowlands [4]. To date,

direct seeding has been adopted by more than 25% of the worldwide rice cultivation area [1]. Dry direct-seeded rice is becoming a popular option in Philippines, India, Thailand, Cambodia, Laos and Indonesia in tropical Asia as well as the US, Australia and Latin America [5,6].

However, a few constraints in dry direct-seeded rice, e.g., poor seedling establishment, weed infestation and lodging susceptibility could lead to large reduction in grain yield and quality [7–11]. Seedling establishment is very important for high yielding and controlling weed infestation in direct seeding rice [10]. Meanwhile, deep sowing could be good for root and basal internodes elongation in deeper soil, which could gain the stability of rice plant and increase lodging resistance of direct seeding rice [12]. However, deep sowing is always in contradiction to seedling uniformity and leads to grain yield reduction [13]. Rapid and good seedling establishment is important for weed competitiveness and good harvesting. Mesocotyl, an organ between the coleoptilar node and the basal part of seminal root in young monocot seedlings, plays a key role in pushing buds out of the deep water or soil during germination for successful seedling establishment. Longer mesocotyl can facilitate seedling establishment under deep sown conditions in dry direct-seeded rice [14–18]. Besides, Mgonja et al. (1988) also reported the strong association between mesocotyl length and seedling vigor [19]. However, the molecular mechanisms of rice mesocotyl length variation are poorly understood.

Genome-wide association study (GWAS) based on linkage disequilibrium (LD) has been widely adopted to identify loci significantly associated with important and complex agronomic traits in rice [20–22]. Extremely high resolution can be achieved by dense SNPs identified in diverse germplasm panels based on the next-generation sequencing (NGS) or SNP chip approaches [20,21,23,24]. Thousands of rice landraces or cultivars have been screened for mesocotyl length by GWAS and dozens of QTL have been reported in previous studies [17,18,25,26]. Wu et al. (2015) screened 270 rice accessions and 16 loci were identified associated with mesocotyl elongation [17]. Furthermore, 469 Indica accessions were used to measure mesocotyl length, and 23 loci were significantly associated with mesocotyl length [18]. Zhao et al. (2018) evaluated the mesocotyl length of 621 rice accessions and detected 13 QTLs [26]. Also, Sun et al. (2018) have identified three QTL for mesocotyl length from 510 rice accessions [25].

Identifying the QTLs in multi-environments could provide accurate information for gene cloning and molecular breeding. In this study, GWAS based on resequencing was conducted in a set of 208 rice accessions for mesocotyl elongation under sand, water and soil (2.0 cm, 4.0 cm and 6.0 cm). The objectives of this study were: (1) To dissect the genetic architecture of mesocotyl elongation, (2) to identify SNPs significantly associated with mesocotyl length, (3) to search for candidate mesocotyl elongation genes for further study, and (4) to select accessions with longer mesocotyl for the breeding of mechanized dry direct-seeded rice.

2. Materials and Methods

2.1. Plant Materials

The panel for GWAS consisted of two parts of rice germplasms, one part (114 accessions) includes advanced lines from International Rice Research Institute (IRRI) and mega varieties released in southeast Asian countries, such as India, Philippines and Bangladesh. The others (94 accessions) are landraces and released varieties mainly from China. All the 208 accessions could be divided into two ecotypes, *Indica* (170) and *Japonica* (38) (Table S1).

2.2. Genotyping

Total genomic DNA for re-sequencing was extracted from young leaves according to the CTAB method. The 208 accessions were genotyped using the Illumina HiSeq 2000 (PE150) (50X) by Berry Genomics Corporation (https://www.berrygenomics.com/) (Beijing, China). The average sequencing depth of each accession genome was 50×. Reads were aligned to the Nipponbare RefSeq (IRGSP-1.0) using BWA-MEM (release 0.7.10). Then, the mapped reads were sorted and duplicated were removed

by Picard tools (http://broadinstitute.github.io/picard/). The variants for each accession were called by the GATK best practices (release 3.2-2).

2.3. Measurement of Mesocotyl Length

To evaluate the variation of mesocotyl length (the distance from the basal part of seminal root to the coleoptilar node) at varied sowing depths, two replications were set up, and in each replication 15 good-quality seeds from each accession were sown at a depth of 2, 4 and 6 cm in soil contained by plastic trays. The size of plastic tray contained 50 holes with the size of 9.5 cm depth, 4.5 cm top diameter and 2.1 cm bottom diameter. After sowing, the plastic tray was kept in a plastic pallet with 3-cm-deep soil, and the whole system were maintained in a 30 °C dark incubator and the soil in each pallet was kept water saturated for seed germination and seedling growth. Emergence rate was recorded every day, until the emergence rate of any accession reached 100%. Three days after that, seedlings from each hole were carefully excavated and washed for measuring mesocotyl length using Image J (https://imagej.en.softonic.com/).

To quantify the mesocotyl length under different culture media, sand and water culture were used, in addition to soil culture. Fifteen good-quality seeds were sown at a depth of 6 cm in sand, the culture method and sampling timing in sand was the same as that in soil. As for water culture, 15 good-quality seeds were put on a gauze in each hole of a plastic tray (50 holes with a depth of 4.8 cm, top diameter at 4.1 cm and bottom diameter at 2.1 cm). The plastic tray was put on a plastic pallet containing water in level of 2 cm from the bottom. Each plastic tray was wrapped with tin foil paper to avoid illumination and kept in phytotron with constant temperature of 30 °C for 10 days. After that, ten uniform seedlings out of 15 seeds of each accession were used for mesocotyl length measurement with Image J. Two replications were set up in both culture methods and mean of the two replications was used for GWAS analysis.

2.4. Population Structure and LD Decay Analysis

The population structure of the 208 accessions was analyzed using Admixture 1.3.0 [27] based on 2000 SNP markers with default cross-validation (K value ranged from 2 to 5). An adhoc quantity statistic ΔK was used to predict the real number of subpopulations and k = 2 was chosen. Besides, the principal component analysis (PCA) and neighbor-jointing trees were also used to validate population stratification with the software Tassel v5.0 (https://www.maizegenetics.net/tassel).

LD decay analysis was done for the whole population. LD decay was measured by correlation coefficients (r^2) for all pairs of SNPs within 500 Kb that were calculated using PopLDdecay v3.27 (https://github.com/BGI-shenzhen/PopLDdecay) with the following parameters: -MaxDist 500-MAF 0.05-Het 0.88-Miss 0.999.

2.5. Genome-Wide Association Analysis and Candidate Genes Identification

Associations between genotypic and phenotypic data were analyzed using the kinship matrix in an MLM (Mixed linear model) by GAPIT (http://www.zzlab.net/GAPIT/) based on R 3.6.1 to control background variation and eliminate the spurious MTAs. Due to Bonferroni-Holm correction for multiple testing ($\alpha = 0.05$) was too conserved and only a few significant MTAs were detected with this criterion, markers with an adjusted $-\log_{10}$ (p-value) ≥ 6.0 were regarded as the significant ones.

Candidate genes for the loci consistently identified in two or more environments were identified. The following steps were conducted to identify candidate genes for important QTL. Firstly, found all the genes located in LD block region around the peak SNP (± 150 Kb based on LD decay analysis) of each important QTL from the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/). Then, all available SNPs located inside of these genes were searched. The genes (except for expressed protein, hypothetical protein, transposon protein and retrotransposon protein) with SNP in coding region and which could further lead to sense mutation were considered as candidate genes. Besides, mesocotyl elongation is highly regulated by various phytohormones, including

strigolactones (SLs), cytokinin (CTK), brassinosteroid (BR), abscisic acid (ABA), jasmonates (JAs), gibberellins (GA) and auxin (IAA). Thus, these genes involved in the metabolism of the phytohormones talked above were selected as high-confidence candidate genes for mesocotyl elongation.

2.6. Gene Expression Analysis

Quantitative real-time PCR (qRT-PCR) was conducted to test the expression differences of candidate genes between accessions with extreme mesocotyl length (Table S4). The mesocotyl part was sampled at the 52 h after germination before the coleoptile excavated. Total RNA was extracted according to the Trizol method, cDNA was synthesized with the HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, China). Primers (Table S5) were designed with the primer premier 5.0 software (http://www.premierbiosoft.com/; last accessed Jan 2019). The PCR procedure was conducted in a volume of 20 μL, containing 2μL cDNA, 0.4 μL of each primer (μM), 10 μL ChamQ Universal SYBR qPCR Master Mix. *OsActin1* was used as the internal control to normalize the expression level of different samples. All assays were performed in two independent experiments, each with three repetitions.

3. Results

3.1. Phenotypic Variation of Mesocotyl Elongation

Mesocotyl length of 208 rice accessions were evaluated in water, sand and soil (2, 4 and 6 cm). Continuous variation was observed across all environments, presenting a wide range of mesocotyl length and indicating that this diversity panel was ideal for conducting GWAS (Figure 1). In sand culture, the mesocotyl lengths ranged from 0 to 5.03 cm, with an average of 1.67 cm; while in water, the mesocotyl lengths ranged from 0 to 3.57 cm, with an average of 1.21 cm. The seeds under soil were sown at a depth of 2, 4 and 6 cm, with the average mesocotyl lengths of 0.33 cm (0–0.71 cm), 0.51 cm (0–1.66 cm) and 2.47 cm (0–4.51 cm), respectively. Among different sowing depth, most mesocotyl lengths of 6 cm sowing depth were significantly longer than those of 2 cm or 4 cm sowing depth (Figure 1, Figure 2a). The maximum of mesocotyl lengths were 0.70 cm, 1.66 cm and 4.51 cm at 2 cm, 4 cm and 6 cm sowing depths in soil, respectively, indicating that mesocotyl elongation of these accessions had larger variation at 6 cm sowing depth than those at 2 cm or 4 cm. Furthermore, there was only a small proportion, about 32.4%, 7.2% and 70.0% accessions with longer mesocotyls (≥ 2.0 cm) in sand, water and soil (6 cm), respectively.

The correlation analysis showed that moderate correlations were observed among different culture media and high correlations were found among different sowing depth under soil. Mesocotyl length evaluated in sand showed active correlation with that under water (r = 0.70), soil (2 cm) (r = 0.74), soil (4 cm) (r = 0.75) and soil (6 cm) (r = 0.78). Also, mesocotyl length evaluated in water showed active correlation with that under soil (2 cm) (r = 0.76), soil (4 cm) (r = 0.78) and soil (6 cm) (r = 0.75). High correlations were observed among different sowing depth under soil. Mesocotyl length evaluated in soil (2 cm) showed significant active correlation with mesocotyl length at soil (4 cm) (r = 0.93) and soil (6 cm) (r = 0.85), whereas the correlation coefficient (r) between soil (4 cm) and soil (6 cm) was 0.86.

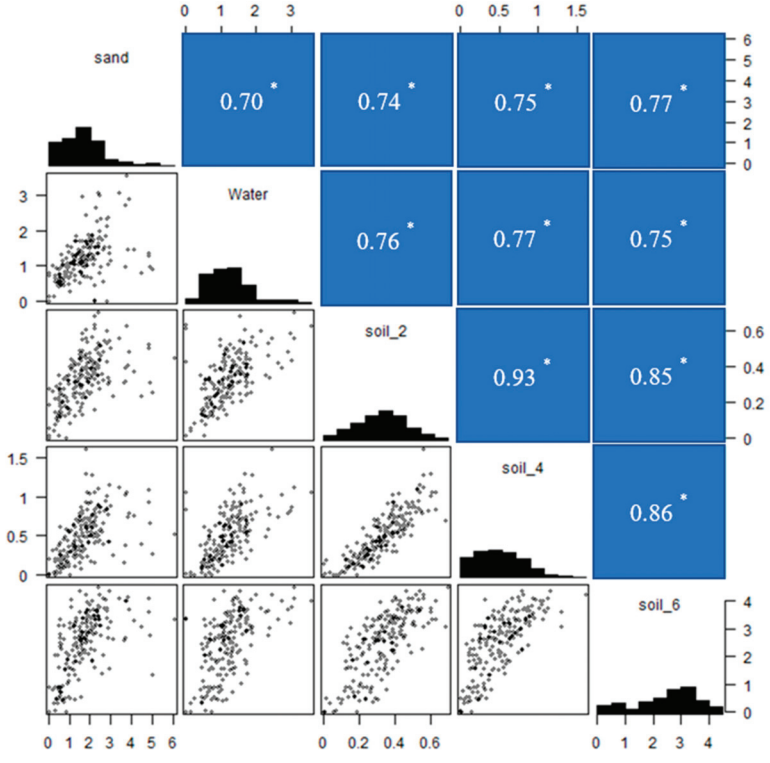


Figure 1. Summary of the phenotypic variations of mesocotyl length among different culture conditions, including correlations between mesocotyl lengths under different culture conditions and mesocotyl length distribution under varied culture conditions. Soil_2: soil culture with 2 cm sowing depth; Soil_4: soil culture with 4 cm sowing depth; Soil_6: soil culture with 6 cm sowing depth, numbers in blue boxes are correlation coefficients (r), * indicates the significance at $p \leq 0.05$.

Analysis of variance (ANOVA) for mesocotyl length in 208 rice accessions revealed significant differences ($P \leq 0.001$) among genotypes (G), environments (E), and genotype × environment interactions (GEI) (Table 1). The V_E is much larger than V_G and V_{GEI}, indicating that mesocotyl length were greatly affected by environments. Besides, the broad sense heritability (h_b^2) estimate for mesocotyl length across all environments was 0.49. Thus, identifying the loci significantly associated with mesocotyl length under multiple environments is crucial for MAS breeding.

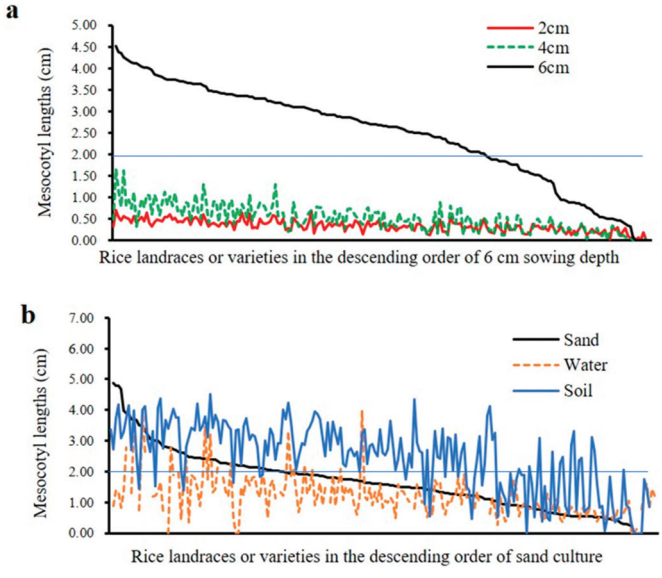


Figure 2. Phenotypic variations of mesocotyl elongation among three culture media. (**a**) Mesocotyl length of 208 rice accessions measured in soil culture with three sowing depths. (**b**) Mesocotyl length variation in sand, soil culture with 6 cm sowing depth and water.

Table 1. ANOVA for mesocotyl length under different environment.

Source	df	Sum of Square	Mean Square	*F*-Value
Block (Environment)	5	0.9063	0.1813	2.5698 ***
Genotype	207	583.1571	2.8172	39.9407 ***
Environment	4	1270.5662	317.6415	4503.3657 ***
Genotyped × Environment	813	533.9877	0.6568	9.3120 ***

*** indicates the significance at $p \leq 0.001$.

3.2. Marker Coverage

The sequencing data were mapped to the Nipponbare reference genome (IRGSP 1.0). After removing the SNPs with minor allele frequency (MAF) < 5% and missing data > 10%, 2,338,386 SNPs were left and employed for GWAS. The chromosome size varied from 22.8 Mb for chromosome 9 to 43.2 Mb for chromosome 1. These markers spanned a physical distance of 373 Mb, with an average density of 0.16 Kb per marker.

3.3. Population Structure and Linkage Disequilibrium

The diversity panel could be divided into two subgroups, the *Japonica* subpopulation and the other derived from the *Indica* accessions, whose characterization were largely consistent with geographic origins. Besides, admixture between *Indica* and *Japonica* were also observed in the present study (Figure 3c). Numerous studies have shown that the lack of appropriate correction for population structure can lead to spurious maker-trait associations (MTAs). Consequently, to eliminate spurious MTAs resulting from population structure, an MLM implemented in GAPIT were adopted for association analysis in the current study. PCA analysis indicated that the top three PCs could explain

28.5%, 8.2% and 3.2% of the total variation of population structure, respectively, and this panel consists of two subgroups (Figure 3a). The neighbor-joining (NJ) tree showing two clades also represented the two subpopulations (Figure 3b).

The decay of LD along physical distances was computed for the 208 rice accessions. A scatter r^2 against physical distance showed a clean pattern of LD decay in the 208 rice accessions. A critical value of the determination coefficients $r^2 > 0.2$ was determined to be the threshold for LD decay and the LD decay distance was about 150 Kb (Figure 3d).

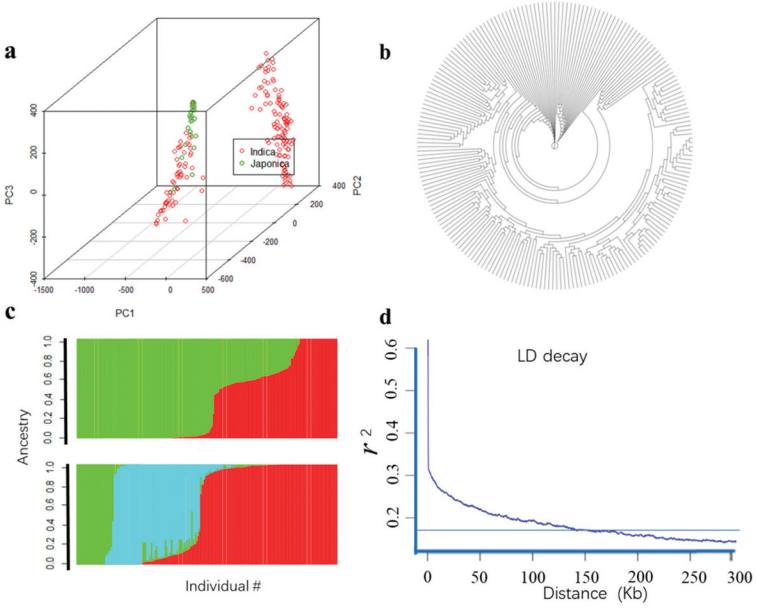


Figure 3. Population structure and LD decay of 208 rice accessions. (**a**) Principal component plot, different colors represent different ingredients; (**b**) neighbor-Jointing tree; (**c**) structure; (**d**) LD decay analysis.

3.4. GWAS of Mesocotyl Lengths

As Bonferroni correction was extremely conservative, a compromised threshold of $-\log_{10}(P) \geq 6.0$ was used as the threshold for significantly associated SNPs. The Manhattan plots for the markers significantly associated with mesocotyl length under different conditions were shown in Figure 4. A total of 16 unique loci were detected, which explained the phenotypic variations ranged from 6.3% to 15.9%. Totally, 12 loci for mesocotyl length were identified on chromosomes 1, 2, 3, 5, 6, 8, 9 and 12 in sand, and each explained phenotypic variation ranged from 6.6 to 15.7%. Besides, seven loci were detected on chromosomes 2, 3, 5, 6 and 9 under water culture, which explain phenotypic variations ranging from 7.8 to 15.5%. There were remarkable differences among three sowing depths under soil culture. As far as the 2 cm sowing depth, a total of seven notable loci were detected on chromosomes 1, 4, 5, 7, 9 and 12, each explaining 7.6 to 15.7% of the phenotypic variation. When coming to the sowing depth of 4 and 6 cm, only two (chromosomes 1 and 12) and three (chromosomes 2, 6 and 12) loci were detected respectively, with an explanation rate of 14.4–14.8% and 6.3–15.9%.

Among all the above QTLs, 12 on chromosomes 1, 2, 3, 5 and 9 were identified under two or more environments. The loci located on chromosomes 1 (14.1–17.3 Mb) and 12 (13.3–15.0 Mb) were identified in sand, soil (2 cm) and soil (6 cm); whereas the locus on chromosome 12 (4.4-4.5 Mb) was detected in

sand, soil (4 cm) and soil (6 cm). Besides, five loci on chromosomes 2 (5.6–8.7 Mb), 3 (15.2–15.3 Mb), 5 (5.8–6.3 Mb), 6 (7.3–9.7 Mb) and 9 (1.3–2.8 Mb) were identified in water and sand; two loci identified on chromosomes 5 (10.6–12.0 Mb) and 9 (7.1–9.1 Mb) were detected in sand and soil (6 cm); the locus on chromosome 6 (15.3–16.6 Mb) was identified on sand and soil (4 cm); the locus on chromosomes 2 (11.7–15.4 Mb) was identified on water and soil (4 cm) (Table 2, Table S2, Figure 4).

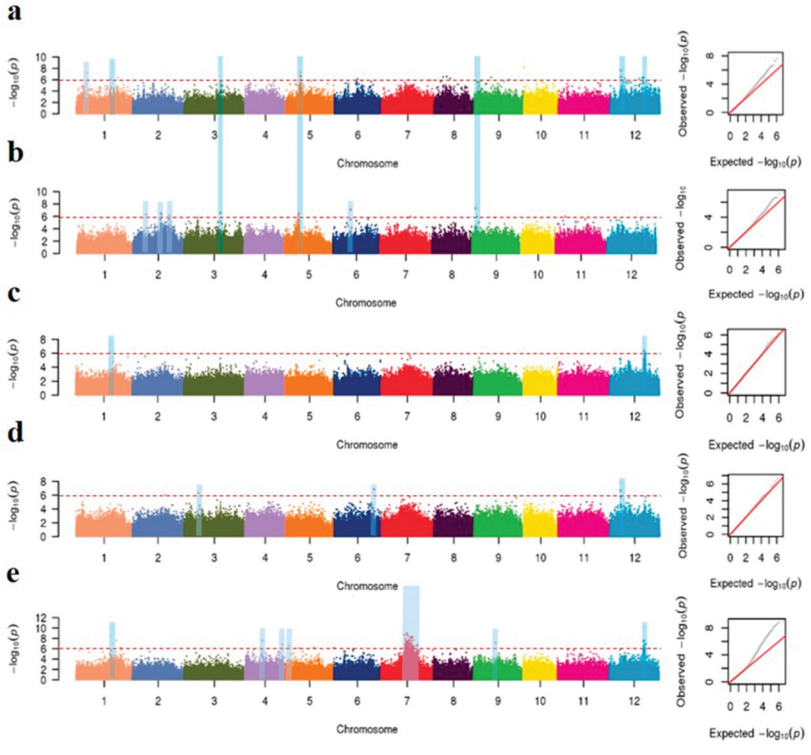


Figure 4. Genome-wide association studies for mesocotyl length among three culture media. Manhattan plots and Quantile-quantile plots of mesocotyl lengths in sand culture (**a**), water (**b**) and soil culture with a sowing depths of 2 cm (**c**), 4 cm (**d**) and 6 cm (**e**), respectively. Negative $\log_{10}$-transformed p values from a genome-wide scan are plotted against position on each of 12 chromosomes. The red horizontal dashed lines indicate the genome-wide significance threshold. Blue stripes show important signals in different GWAS.

Table 2. Loci significantly associated with mesocotyl length under different environment.

Peak Marker	Chr.	Genomic Regions (Mb)	*P* Value	R^2 (%)	Environment	Reported Locus	References
rs_1_16752644	1	14.1~17.3	1.9 E-08~7.7 E-07	7.2–15.2	Sand, Soil (2 cm), Soil (6 cm)	*qml1-1 and qML1*	Cao et al. 2002a, Huang et al., 2010
rs_2_5647116	2	5.6-8.7	4.8~9.4 E-07	7.6–8.1	Water, Sand	*qML2*	Huang et al., 2010
rs_2_11740641	2	11.7-15.4	2.8-7.3 E-07	8.1–15.5	Water, Soil (4 cm)		
rs_3_15245138	3	15.2-15.3	2.5-5.7 E-07	14.8–14.9	Water, Sand	*qml3-1*	Ouyang et al., 2005
rs_4_7234284	4	7.2	8.2 E-08	14.9	Soil (6 cm)	*Qml4-1*	Ouyang et al., 2005
rs_5_6206136	5	5.8-6.3	9.5E-09~9.7E-07	6.6–15.3	Water, Sand	*Qml5-5*	Ouyang et al. 2005
rs_5_11950679	5	10.6-12.0	6.6~7.6E-07	6.6–8.1	Sand, Soil (6 cm)		
rs_6_7313005	6	7.3-9.7	7.7E-08~8.4E-07	7.1–8.1	Water, Sand	*qML6*	Huang et al., 2010
rs_6_16643928	6	15.3-16.6	1.2~3.7 E-07	6.7–14.4	Sand, Soil (4 cm)		
rs_7_10769188	7	10.0-13.6	2.8E-09~9.9E-07	6.4–15.9	Soil (6 cm)	*qFML7-1*	Zhao et al., 2018
rs_8_5219449	8	2.8-5.2	6.8E-08~3.5E-07	14.6–15.3	Sand	*qLOE-8*	Zhang et al., 2006
rs_9_1446824	9	1.3-2.8	4.8E-08~4.8E-07	14.9–15.2	Water, Sand	*seq-rs4080*	Lu et al., 2016
rs_9_7059246	9	7.1-9.1	8.5 E-08~1.2 E-07	14.7–14.8	Sand, Soil (6 cm)		
rs_12_4444693	12	4.4-4.5	4.3E-09~5.9E-07	14.8–15.7	Sand, Soil (4 cm), Soil (6 cm)		
rs_12_6008455	12	6.0-7.8	1.2~4.6 E-07	14.7–15.4	Sand		
rs_12_14155486	12	13.3-15.0	2.3E-08~9.9E-07	7.5–15.3	Sand, Soil (2 cm), Soil (6 cm)	*qMel-12*	Lee et al. 2012b

3.5. Candidate Genes Related with Mesocotyl Elongation

According to the results of GWAS, 107 candidate genes were found to be related with mesocotyl elongation (Table S3). Also, considering that mesocotyl elongation is highly regulated by various phytohormones, such as SLs, CTK, BR, ABA, JAs, GA and IAA, five genes involved in the biological metabolism of the phytohormones, cell elongation and division were selected as the high-confidence candidate genes for mesocotyl elongation (Table 3). That is the zinc finger CCCH type family protein (*LOC_Os05g10670*) and transcription factor jumonji (*LOC_Os05g10770*) on chromosome 5 (5.6–6.1 Mb), a gene encoding cytokinin-O-glucosyl transferase 2 on chromosome 9 (1.2–1.5 Mb) and two genes (4.3–4.5 Mb and 14.1–14.5 Mb) encoding flavin monooxygenase and 9-cis-epoxycarotenoid dioxygenase 1, respectively.

The expression of five candidate genes were detected using qRT-PCR (Figure 5). The results showed that there was significant difference between accessions with extreme mesocotyl length, with the longer one exhibited higher expression abundance. In detail, five genes showed more than 4.4-fold (*LOC_Os05g10670*), 2.8-fold (*LOC_Os05g10770*), 3.0-fold (*LOC_Os09g03140*), 8.0-fold (*LOC_Os12g08780*) and 32.6-fold (*LOC_Os12g24800*) higher expression in long-mesocotyl accessions compared to short-mesocotyl accessions, respectively.

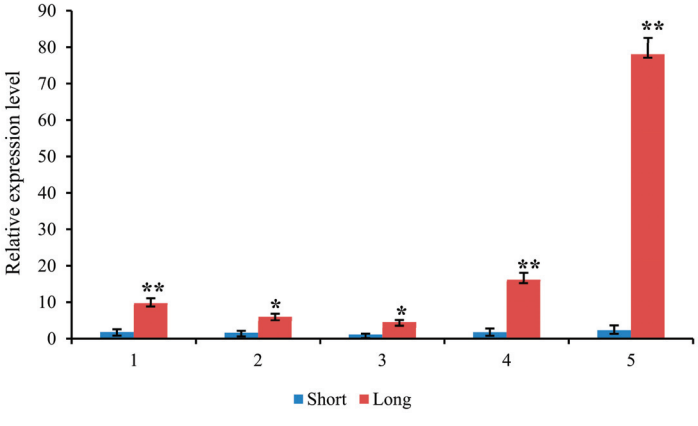


Figure 5. Expression analysis of five candidate genes. Relative expressions of the five genes in long- and short -mesocotyl length rice accessions were tested by qRT-PCR. Transcription levels were normalized to that of *Osactin1* and compared to the short control. 1 to 5 represent five candidate genes, that is 1, *LOC_Os05g10670*; 2, *LOC_Os05g10770*; 3, *LOC_Os09g03140*; 4, *LOC_Os12g08780*; 5, *LOC_Os12g24800*. Values of fold change are showed in mean ± SEM. Relative transcription levels were calculated by the $\log 2^{-\Delta\Delta CT}$ method. * and ** indicate the significance at $p \leq 0.05$ and $p \leq 0.01$, respectively.

Table 3. High confidence candidate genes for mesocotyl lengths in 208 rice landraces or varieties.

Chromosome	Genomic Regions	Candidate Gene	Functional Annotation
5	5688534-6067598	*LOC_Os05g10670*	Zinc finger CCCH type family protein, putative, expressed
5	5688534-6067598	*LOC_Os05g10770*	Transcription factor jumonji, putative, expressed
9	1245230-1521824	*LOC_Os09g03140*	Cytokinin-O-glucosyltransferase 2, putative, expressed
12	4369693-4551492	*LOC_Os12g08780*	Flavin monooxygenase, putative, expressed
12	14080486-14520450	*LOC_Os12g24800*	9-cis-epoxycarotenoid dioxygenase 1, chloroplast precursor, putative, expressed

4. Discussion

The GWAS panel, including released cultivars, advanced lines and landraces from different ecological regions, has a high genetic diversity with a wide range of mesocotyl length. In this study, *Indica* accessions had 32.4% and 70.0% accessions with longer mesocotyls (≥2.0 cm) in sand and 6 cm soil culture, respectively, but only 30.0% and 45.0% accessions in the *Japonica*, indicating that mesocotyl length had a great potential for improvement. Although mesocotyl is a crucial developmental trait of rice and is imminently required to improve crop adaptability to modern cultivation mode, the molecular mechanisms of its elongation and domestication remains unclear. Here, a GWAS based on re-sequencing were employed to identify associated SNP markers and understand the genetic basis of mesocotyl length.

4.1. Mesocotyl Elongation Varied among Culture Conditions

The elongation of mesocotyl was influenced by various environmental factors, such as light, temperature, moisture [28,29]. Thus, rice mesocotyl elongation were varied among different culture media, such as soil [26], sand [17], water [30]. In our study, mesocotyl length ranged from 0 to 3.99 cm under water culture, which was close to 0-5.05 cm in a previous study [17]. While, the mesocotyl length under 6 cm sand culture ranged from 0 to 4.88 cm, which differed from the previous result ranging from 0 to 2.05 cm under 5 cm sand culture [17]. Moreover, more than half of the 208 accessions had mesocotyl length longer than 1 cm under water culture. However, Wu et al. (2015) reported that only 10.5% accessions had mesocotyl length longer than 1 cm under water culture [17]. It is probably that the diverse results were mainly caused by different genotypes. Previous studies suggested that under varied sowing depths of soil cover, mesocotyl length was promoted by the depth in soil culture [11,18,26]. Lee et al. (1999) also reported that mesocotyl length under 3-cm sowing depth was shorter than that under 5-cm sowing depth [31]. Similarly, our results showed that mesocotyl length had a large variation among sowing depth, with an order of 6 cm > 4 cm >2 cm in soil. These results indicated that the variation of mesocotyl length existed in varied culture media and sowing depths.

4.2. LD Pattern

LD decay affects the precision of GWAS and was influenced by population structure, allele frequency, recombination rate and selection. Previous studies reported that LD decay in rice ranged between 120 Kb and 270 Kb [18,23,31,32]. In this panel, the LD decay was about 150 Kb for the whole genome (Figure 3e), consisting with previous reports [18,23,32,33].

4.3. GWAS Was Effective to Identify QTLs Controlled Complex Quantitative Traits

So far, several previous studies have reported the associated QTLs and candidate genes for rice mesocotyl elongation using linkage mapping or GWAS [20,25,34–36]. In this study, the loci identified on chromosome 7 (10.03–13.61 Mb) at the 6 cm sowing depth of soil was overlapped with QTLs *qFML7-1*, *qFML7-2*, *qFML7-3*, *qFML7-4* and *qFML7-5* [26]. The QTL on chromosome 1 (14.1–17.3 Mb) were overlapped with *qml1-1* [37,38] and *qML1* [23]. Furthermore, two QTLs on chromosome 2 (5.6–8.7 Mb) and 6 (7.3–9.7 Mb) identified under water and sand were overlapped with the QTL *qML2* and *qML6* detected from Shennong 265/Lijiangxintuanheigu RIL population [20]. Ouyang et al. (2015) have reported *qml3-1*, *qml4-1* and *qml5-1* by a RIL population [39], which nearly coincide with the QTL identified on chromosome 3 (15.2–15.3 Mb), 4 (8.3 Mb) and 5 (5.8–6.3 Mb) in our study. Zhang et al. (2006) detected four QTLs for submergence tolerance in seedling stage from a DH population, which explained phenotypic variations from 10.7% to 29.3% [35]. Among these, *qLOE-8* were overlapped with the QTL regions located on chromosome 8 (208–5.2 Mb) identified in sand. Lee et al. (2012) reported two QTLs associated with mesocotyl elongation, *qMel-1* (R^2 = 37.3%) and *qMel-3* (R^2 = 6.5%) in weedy rice on chromosomes 1 and 3, respectively [40]. Moreover, Lu et al. (2016) reported 17 loci and 289 candidate genes for rice mesocotyl by GWAS on chromosomes 1 (4), 4 (4), 5 (1), 6 (3), 7 (1), 9 (3) and

11 (1), respectively [18]. QTLs identified on chromosomes 9 (1.3–2.8 Mb) and 12 (13.3–15.0 Mb) were overlapped with those QTLs reported by previous studies [18,40].

Wu et al. (2015) found that when under water culture and at a significant level of $-\log_{10}(P) \geq 8.0$, a total of 13 loci were detected to be significantly associated with mesocotyl length, while only three significant SNPs were declared when the medium was merely changed to sand, and two of the three SNPs were co-localized under both conditions [17]. Lu et al. (2016) reported that six and seven loci were detected under two environments, respectively [18]. Only two of them on chromosome 6 were identified under both environments. In different studies, nearly 40 loci significantly associated with mesocotyl length have been detected on all twelve chromosomes by GWAS, and most of these loci in each study cannot be validated by others [17,18,25,26]. The interaction of G (genotype) × E (environment) can influence QTL and association mapping results, indicating that the effect of identified genomic regions was required to be estimated for each environment [41,42]. Only the major genomic regions could be co-localized under multi-environments.

4.4. Potential Candidate Genes for Mesocotyl Elongation were Identified

According to the result of GWAS, 107 candidate genes were found to be related to mesocotyl elongation. Among them, five involved in the biological metabolism of phytohormones, cell elongation and division were selected as the high-confidence candidate genes. Bioinformatics analysis indicated that loci on chromosome 5 (5.6–6.1 Mb) corresponding to zinc finger CCCH type family protein (*LOC_Os05g10670*) and transcription factor jumonji (*LOC_Os05g10770*). CCCH-type zinc finger proteins comprise a large family that is induced by drought, high temperature stress and hydrogen peroxide, and is also induced by abscisic acid, methyl jasmonate and salicylic acid [43,44]. Noh et al., (2004) reported that jumonji class transcription factor controls stem elongation, transposon activity and panicle development [45]. Besides, Feng et al. (2017) have identified several candidate genes for mesocotyl in rice including zinc-finger protein genes, which involved in the JA biosynthesis and signaling pathways [29]. Besides, a candidate gene on chromosome 9 (1.2–1.5 Mb) encoding cytokinin-O-glucosyltransferase 2 were identified. CK is a class of plant hormones that were first identified as cell division-promoting factors and were subsequently identified as factors that control various processes in plant growth and development, including mesocotyl elongation [46]. Besides, CK played an important role in the biosynthesis of BRs, a group of steroid plant hormones essential in plant growth and development [47]. Two candidate genes in loci 4.3–4.5 Mb and 14.1–14.5 Mb on chromosomes 12 were identified, which encodes flavin monooxygenase and 9-cis-epoxycarotenoid dioxygenase 1, respectively. Flavin monooxygenase catalyzes hydroxylation of the amino group of tryptamines, a rate-limiting step in tryptophan-dependent auxin biosynthesis, which regulate many processes in plant development [48]. The 9-cis-epoxycarotenoid dioxygenase is essential for the biosynthesis of ABA [49]. The 9-cis-epoxycarotenoid dioxygenase gene could lead to the over-production of abscisic acid, which plays an important role in mesocotyl elongation [50]. Expression of the five candidate genes in different accessions indicated that five genes were all functional in regulating mesocotyl elongation. The most significant difference in expression level was observed in gene, *LOC_Os12g24800*, which indicated that this gene could be the most likely functional gene for mesocotyl length. These results also proved that the loci identified by GWAS were reliable. Mesocotyl elongation is a consequence of complicated biological processes and its mechanism remains unclear; more detailed experimental analyses are needed to confirm the function of candidate genes in mesocotyl elongation.

5. Conclusions

In the present study, a GWAS for mesocotyl length in 208 rice accessions was conducted. Mesocotyl length varied in different culture conditions and sowing depths. GWAS have identified 16 loci significantly associated with mesocotyl elongation, which explained phenotypic variations ranged from 7.1% to 10.0%. Furthermore, five high-confidence candidate genes were identified on

chromosomes 1, 3, 5, 9 and 12. The accessions with longer mesocotyl and the markers significantly associated with mesocotyl length can be used to promote the process of rice breeding. This study will facilitate our uanderstanding on the genetic architecture of rice mesocotyl length.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/1/49/s1, Table S1: Detailed information of 208 rice accessions, Table S2: SNPs significantly associated with mesocotyl length under different environments, Table S3: Candidate genes for mesocotyl length in 208 rice landraces or varieties, Table S4: Accessions used in qRT-PCR analysis, Table S5: Primers used in this study. The raw Illumina sequencing data will be available upon the request.

Author Contributions: Conceived and designed the experiments: H.L., Y.W., G.Y. Performed the experiments: H.L., J.Z., X.L. Analyzed the data: H.L., J.L. (Jiaolong Li), G.Y. Wrote and revised the paper: H.L., J.Z., J.L. (Jindong Liu), Y.W., G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission (Grant No. CAAS-XTCX2016001) and Shenzhen Science and Technology Projects (No. JSGG20160608160725473).

Acknowledgments: We thank Guosheng Xiong for providing some of the germplasms used in this study. We are grateful to the field research assistant Bingmei Lin in phenotyping and other field work.

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

References

1. Kumar, V.; Ladha, J.K. Direct seeding of rice: Recent developments and future research needs. *Adv. Agron.* **2011**, *111*, 297–413.
2. Kato, Y.; Katsura, K. Rice adaptation to aerobic soils: Physiological considerations and implications for agronomy. *Plant Prod. Sci.* **2014**, *17*, 1–12. [CrossRef]
3. Liu, H.; Hussain, S.; Zheng, M.; Peng, S.; Huang, J.; Cui, K.; Nie, L. Dry direct-seeded rice as an alternative to transplanted-flooded rice in Central China. *Agron. Sustain. Dev.* **2015**, *35*, 285–294. [CrossRef]
4. Ohno, H.; Banayo, N.P.; Bueno, C.; Kashiwagi, J.I.; Nakashima, T.; Iwama, K.; Corales, A.M.; Garcia, R.; Kato, Y. On-farm assessment of a new early-maturing drought-tolerant rice cultivar for dry direct seeding in rainfed lowlands. *Field Crop. Res.* **2018**, *219*, 222–228. [CrossRef]
5. GRiSP. *Rice Almanac*, 4th ed.; International Rice Research Institute: Manila, Philippines, 2013; p. 283.
6. Haefele, S.M.; Kato, Y.; Singh, S. Climate ready rice: Augmenting drought tolerance with best management practices. *Field Crop. Res.* **2013**, *190*, 60–69. [CrossRef]
7. Kono, M. Physiological aspects of lodging. In *Science of the Rice Plant, Physiology, 2*; Matsuo, Y., Kumazawa, K., Ishii, R., Ishihara, K., Hirata, H., Eds.; Food and Agricultural Policy Research Center: Tokyo, Japan, 1995; pp. 971–982.
8. Ramzan, M. Evaluation of various planting methods in rice-wheat cropping system, Punjab, Pakistan. *Rice Crop Rep.* **2003**, *4*, 4–5.
9. Kashiwagi, T.; Sasaki, H.; Ishimaru, K. Factors responsible for decreasing sturdiness of the lower part in lodging of rice (*Oryza sativa* L.). *Plant Prod. Sci.* **2005**, *8*, 166–172. [CrossRef]
10. Mahender, A.; Anandan, A.; Pradhan, S.K. Early seedling vigour, an imperative trait for direct-seeded rice: An overview on physio-morphological parameters and molecular markers. *Planta* **2015**, *241*, 1027–1050. [CrossRef]
11. Lee, H.S.; Sasaki, K.; Kang, J.W.; Sato, T.; Song, W.Y.; Ahn, S.N. Mesocotyl elongation is essential for seedling emergence under deep-seeding condition in rice. *Rice* **2017**, *10*, 32. [CrossRef]
12. Ekanayake, I.J.; Garrity, D.P.; O'Toole, J.C. Influence of Deep Root Density on Root Pulling Resistance in Rice. *Crop Sci.* **1986**, *26*, 1181–1186. [CrossRef]
13. Kirby, E. Effect of sowing depth on seedling emergence, growth and development in barley and wheat. *Field Crop. Res.* **1993**, *35*, 101–111. [CrossRef]
14. Turner, F.T.; Chen, C.C.; Bollich, C.N. Coleoptile and mesocotyl lengths in semi-dwarf rice seedlings. *Crop Sci.* **1982**, *22*, 43–46. [CrossRef]
15. Chung, N.J. Elongation habit of mesocotyls and coleoptiles in weedy rice with high emergence ability in direct-seeding on dry paddy fields. *Crop Pasture Sci.* **2010**, *61*, 911–917. [CrossRef]

16. Trachsel, S.R.; Messmer, P.; Stamp, N.R.; Hund, A. QTLs for early vigor of tropical maize. *Mol. Breed.* **2010**, *25*, 91–103. [CrossRef]
17. Wu, J.; Feng, F.; Lian, X.; Teng, X.; Wei, H.; Yu, H.; Xie, W.; Yan, M.; Fan, P.; Li, Y.; et al. Genome-wide Association Study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. *BMC Plant Biol.* **2015**, *15*, 218. [CrossRef]
18. Lu, Q.; Zhang, M.; Niu, X.; Wang, C.; Xu, Q.; Feng, Y.; Wang, S.; Yuan, X.; Yu, H.; Wang, Y. Uncovering novel loci for mesocotyl elongation and shoot length in indica rice through genome-wide association mapping. *Planta* **2016**, *243*, 645–657. [CrossRef]
19. Mgonja, M.; Dilday, R.; Skinner, S.; Collins, F. Association of mesocotyl elongation with seedling vigor in rice. *J. Ark. Acad. Sci.* **1988**, *42*, 52–55.
20. Huang, C.; Jiang, S.K.; Feng, L.; Xu, Z.J.; Chen, W.F. QTL analysis for mesocotyl length in rice (*Oryza sativa* L.). *Acta Agron. Sin.* **2010**, *36*, 1108–1113.
21. Huang, X.; Zhao, Y.; Li, C.; Wang, A.; Zhao, Q.; Li, W.; Guo, Y.; Deng, L.; Zhu, C.; Fan, D.; et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **2012**, *44*, 32–39. [CrossRef]
22. Chen, W.; Gao, Y.; Xie, W.; Gong, L.; Lu, K.; Wang, W.; Li, Y.; Liu, X.; Zhang, H.; Dong, H.; et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **2014**, *46*, 714–721. [CrossRef]
23. Huang, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; Li, M.; et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef] [PubMed]
24. Zhao, K.; Tung, C.W.; Eizenga, G.C.; Wright, M.H.; Ali, M.L.; Price, A.H.; Norton, G.J.; Islam, M.R.; Reynolds, A.; Mezey, J.; et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat. Commun.* **2011**, *2*, 467. [CrossRef] [PubMed]
25. Sun, S.; Wang, T.; Wang, L.; Li, X.; Jia, Y.; Liu, C.; Huang, X.; Xie, W.; Wang, X. Natural selection of a GSK3 determines rice mesocotyl domestication by coordinating strigolactone and brassinosteroid signaling. *Nat. Commun.* **2018**, *9*, 2523. [CrossRef] [PubMed]
26. Zhao, Y.; Zhao, W.; Jiang, C.; Wang, X.; Xiong, H.; Todorovska, E.G.; Yin, Z.; Chen, Y.; Wang, X.; Xie, J.; et al. Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. *Front Plant Sci.* **2018**, *9*, 332. [CrossRef] [PubMed]
27. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **2009**, *19*, 1655–1664. [CrossRef] [PubMed]
28. Radford, B.J.; Henzell, R.G. Temperature affects the mesocotyl and coleoptile length of grain sorghum genotypes. *Aust. J. Agric. Res.* **1990**, *41*, 79–87. [CrossRef]
29. Feng, F.; Mei, H.; Fan, P.; Li, Y.; Xu, X.; Wei, H.; Yan, M.; Luo, L. Dynamic transcriptome and phytohormone profiling along the time of light exposure in the mesocotyl of rice seedling. *Sci. Rep.* **2017**, *7*, 11961. [CrossRef]
30. Cai, H.; Morishima, H. QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* **2002**, *104*, 1217–1228. [CrossRef]
31. Lee, S.; Kim, J.; Hong, S. Effects of priming and growth regulator treatment of seed on emergence and seedling growth of rice. *Korean J. Crop Sci.* **1999**, *44*, 134–137.
32. Qiu, X.; Pang, Y.; Yuan, Z.; Xing, D.; Xu, J.; Dingkuhn, M.; Li, Z.; Ye, G. Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS ONE* **2015**, *10*, e0145577. [CrossRef]
33. McCouch, S.R.; Wright, M.H.; Tung, C.W.; Maron, L.G.; McNally, K.L.; Fitzgerald, M.; Singh, N.; DeClerck, G.; Agosto-Perez, F.; Korniliev, P.; et al. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* **2016**, *7*, 10532. [CrossRef] [PubMed]
34. Redona, E.D.; Mackill, D.J. Mapping quantitative trait loci for seedling vigor in rice using RFLPs. *Theor. Appl. Genet.* **1996**, *92*, 395–402. [CrossRef] [PubMed]
35. Zhang, G.; Zeng, D.; Hu, S.; Su, Y.; Ajia, L.; Guo, L.; Qian, Q. QTL analysis of traits concerned submergence tolerance at seedling stage in rice (*Oryza sativa* L.). *Acta Agron. Sin.* **2006**, *32*, 1280–1286.
36. Lim, J.H.; Yang, H.J.; Jung, K.H.; Yoo, S.C.; Paek, N.C. Identification of chromosomal regions influencing mesocotyl elongation by bulked segregation analysis based on genome re-sequencing in rice. *Acta Agric. Shanghai* **2017**, *33*, 10–15.

37. Cao, L.; Zhu, J.; Ren, L.; Zhao, S.; Yan, Q. Mapping QTLs and epistasis for seedling vigor in rice (*Oryza sativa* L.). *Acta Agron. Sin.* **2002**, *28*, 809–815.
38. Cao, L.; Zhu, J.; Yan, Q.; He, L.; Wei, X.; Cheng, S. Mapping QTLs with epistasis for mesocotyl length in a DH population from indica-japonica cross of rice (*Oryza sativa*). *Chin. J. Rice Sci.* **2002**, *16*, 221–224.
39. Ouyang, Y.N.; Zhang, Q.Y.; Zhang, K.Q.; Yu, S.M.; Zhuang, J.Y.; Jin, Q.Y.; Cheng, S.H. QTL mapping and interaction analysis of genotype × environment (Fe^{2+}-concentrations) for mesocotyl length in rice (*Oryza sativa* L.). *Acta Genet. Sin.* **2005**, *32*, 712–718.
40. Lee, H.S.; Sasaki, K.; Higashitani, A.; Ahn, S.N.; Sato, T. Mapping and characterization of quantitative trait loci for mesocotyl elongation in rice (*Oryza sativa* L.). *Rice* **2012**, *5*, 1–10. [CrossRef]
41. Zhang, K.; Tian, J.; Zhao, L.; Wang, S. Mapping QTLs with epistatic effects and QTL x environment interactions for plant height using a doubled haploid population in cultivated wheat. *J Genet Genom.* **2008**, *35*, 119–127. [CrossRef]
42. Tetard-Jones, C.; Kertesz, M.A.; Preziosi, R.F. Identification of plant quantitative trait loci modulating a rhizobacteria-aphid indirect effect. *PLoS ONE* **2012**, *7*, e41524. [CrossRef]
43. Jan, A.; Maruyama, K.; Todaka, D.; Kidokoro, S.; Abo, M.; Yoshimura, E.; Shinozaki, K.; Nakashima, K.; Yamaguchi-Shinozaki, K. OsTZF1, a CCCH-tandem zinc finger protein, confers delayed senescence and stress tolerance in rice by regulating stress-related genes. *Plant Physiol.* **2013**, *161*, 1202–1216. [CrossRef]
44. Xiong, Q.; Ma, B.; Lu, X.; Huang, Y.H.; He, S.J.; Yang, C.; Yin, C.C.; Zhao, H.; Zhou, Y.; Zhang, W.K.; et al. Ethylene-Inhibited Jasmonic Acid Biosynthesis Promotes Mesocotyl/Coleoptile Elongation of Etiolated Rice Seedlings. *Plant Cell* **2017**, *29*, 1053–1072. [CrossRef]
45. Noh, B.; Lee, S.H.; Kim, H.J.; Yi, G.; Shin, E.A.; Lee, M.; Jung, K.J.; Doyle, M.R.; Amasino, R.M.; Noh, Y.S. Divergent roles of a pair of homologous Jumonji/Zinc-Finger–class transcription factor proteins in the regulation of Arabidopsis flowering time. *Plant Cell* **2004**, *16*, 2601–2613. [CrossRef] [PubMed]
46. Hu, Z.; Yamauchi, T.; Yang, J.; Jikumaru, Y.; Tsuchida-Mayama, T.; Ichikawa, H.; Takamure, I.; Nagamura, Y.; Tsutsumi, N.; Yamaguchi, S.; et al. Strigolactone and cytokinin act antagonistically in regulating rice mesocotyl elongation in darkness. *Plant Cell Physiol.* **2014**, *55*, 30–41. [CrossRef] [PubMed]
47. Yuldashev, R.; Avalbaev, A.; Bezrukova, M.; Vysotskaya, L.; Khripach, V.; Shakirova, F. Cytokinin oxidase is involved in the regulation of cytokinin content by 24-epibrassinolide in wheat seedlings. *Plant Physiol. Biochem.* **2012**, *55*, 1–6. [CrossRef] [PubMed]
48. Zhao, Y.; Christensen, S.K.; Fankhauser, C.; Cashman, J.R.; Cohen, J.D.; Weigel, D.; Chory, J. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science* **2001**, *291*, 306–309. [CrossRef] [PubMed]
49. Wan, X.R.; Li, L. Regulation of ABA level and water-stress tolerance of Arabidopsis by ectopic expression of a peanut 9-cis-epoxycarotenoid dioxygenase gene. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 1030–1038. [CrossRef]
50. Thompson, A.J.; Jackson, A.C.; Symonds, R.C.; Mulholland, B.J.; Dadswell, A.R.; Blake, P.S.; Burbidge, A.; Taylor, I.B. Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. *Plant J.* **2000**, *23*, 363–374. [CrossRef]

genes

MDPI

Article

Insights into the Genetic Architecture of Bran Friability and Water Retention Capacity, Two Important Traits for Whole Grain End-Use Quality in Winter Wheat

Sviatoslav Navrotskyi [1], Vikas Belamkar [2,*], P. Stephen Baenziger [2] and Devin J. Rose [1,2,*]

[1] Department of Food Science & Technology, University of Nebraska-Lincoln, Lincoln, NE 68588, USA; snavrotskyi@huskers.unl.edu

[2] Department of Agronomy & Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA; pbaenziger1@unl.edu

* Correspondence: vikas.belamkar@unl.edu (V.B.); drose3@unl.edu (D.J.R.); Tel.: +1-402-472-2811 (V.B.); +1-402-472-2802 (D.J.R.)

Received: 27 June 2020; Accepted: 21 July 2020; Published: 23 July 2020

Abstract: Bran friability (particle size distribution after milling) and water retention capacity (WRC) impact wheat bran functionality in whole grain milling and baking applications. The goal of this study was to identify genomic regions and underlying genes that may be responsible for these traits. The Hard Winter Wheat Association Mapping Panel, which comprised 299 lines from breeding programs in the Great Plains region of the US, was used in a genome-wide association study. Bran friability ranged from 34.5% to 65.9% (median, 51.1%) and WRC ranged from 159% to 458% (median, 331%). Two single-nucleotide polymorphisms (SNPs) on chromosome 5D were significantly associated with bran friability, accounting for 11–12% of the phenotypic variation. One of these SNPs was located within the *Puroindoline-b* gene, which is known for influencing endosperm texture. Two SNPs on chromosome 4A were tentatively associated with WRC, accounting for 4.6% and 4.4% of phenotypic variation. The favorable alleles at the SNP sites were present in only 15% (friability) and 34% (WRC) of lines, indicating a need to develop new germplasm for these whole-grain end-use quality traits. Validation of these findings in independent populations will be useful for breeding winter wheat cultivars with improved functionality for whole grain food applications.

Keywords: wheat quality; wheat milling; wheat hardness; puroindolines; water absorption capacity

1. Introduction

Numerous studies have associated the consumption of high-fiber foods with decreased risk of type 2 diabetes, obesity, and heart disease [1,2]. Consumer awareness of the health benefits of dietary fiber-rich foods has driven product development of high-fiber foods in the baking industry. As a good source of fiber, whole wheat flour has become increasingly popular globally with the number of whole grain products increasing from 164 in 2000 to 3378 in 2011 [3].

Unlike refined white flour, whole wheat flour contains the all the anatomical portions of the grain are together in the same proportions as they occur in the intact kernel [4]. The bran component (including the outer portions of the kernel: pericarp, germ, and aleurone) has a significant impact on whole wheat flour functionality [5–7]. Due to its high fiber concentration, wheat bran binds considerable amounts of water, which interferes with proper gluten hydration and leads to detrimental effects on bread quality [8]. Bran water retention capacity (WRC), which is a measure of how much water the bran will imbibe, varies among genotypes and growing conditions and is correlated with whole grain flour functionality [5,9]. Therefore, WRC is an important trait in the evaluation of bran quality.

As wheat is milled, the bran may fracture into many small particles or remain as larger pieces. The tendency for bran to fracture into smaller pieces is known as friability. Initial research on wheat milling was focused on a reduction in bran friability (ability to retain large particles) to facilitate efficient separation of white flour from the bran by sifting [10]. However, as whole wheat flour has become more popular, there have been numerous studies showing that bran particle size affects whole wheat flour functionality [6,11]. Reported results are somewhat contradictory, but they do show that very large particles are undesirable due to interference with gluten network formation [6]. Furthermore, bran WRC also increases with increased bran particle size [12]. Therefore, for whole wheat flour, it is desirable to have high bran friability.

In recent years, genome-wide association studies (GWAS) have been used extensively to determine the genetic basis of agronomic, physiological, and disease resistance traits in bread wheats [13–33]. Relatively fewer studies have used GWAS to identify loci that control end-use quality traits, including grain protein concentration, flour yield and color, gluten strength, and mixing properties, among others [16,34–41]. However, these studies have focused on end-use quality for refined wheat flour applications. No studies have focused on quality traits related to whole grain end-use quality.

Our previous study showed that among eight traits measured on wheat bran, including phenolic compounds, antioxidant activity, protein, ash, lipoxygenase activity, and free thiol groups, bran particle size and WRC showed the most significant impact on whole wheat bread volume and texture among the other chemical and physical parameters tested [9]. Therefore, the objectives of this study were to: (1) investigate the variation in bran friability and WRC in the Hard Winter Wheat Association Mapping Panel (HWWAMP), a diverse set of 299 winter wheat lines representing germplasm from the hard winter wheat growing regions of the US [42,43]; (2) conduct a genome-wide association study (GWAS) to identify genomic regions associated with bran friability and WRC; and (3) identify underlying genes associated with bran friability and WRC and investigate their annotations to determine potential genes that may be important for these two traits.

2. Materials and Methods

2.1. Plant Material

The study comprised 299 genotypes from the HWWAMP. These genotypes included varieties released since the 1940s and represent existing germplasm of the hard winter wheat growing regions of the US, including Colorado, Kansas, Michigan, Montana, Nebraska, North Dakota, Oklahoma, South Dakota, and Texas [42,43]. The HWWAMP represents diverse germplasm with unique haplotypes as compared to a breeding program nursery from a single breeding program. A list of the genotypes comprising the HWWAMP along with their pedigrees and geographical origin has been published previously in the supplementary material of Guttieri et al. [44]. Many studies in bread wheat rely on using this HWWAMP for mapping genomic regions important for a trait, and subsequently validate those marker-trait associations in their own breeding programs prior to using the markers for marker-assisted selection. The HWWAMP has been successfully used for mapping genomic regions for various biotic and abiotic traits including wheat spot blotch resistance [45], drought resistance [46], Cd accumulation [44,47], and bacterial leaf streak resistance [48]. All genotypes in HWWAMP were grown at Lincoln, NE in 2018 in 1 m^2 plots with fungicide application, and the harvested seeds from these genotypes were used in this study.

2.2. Bran Functionality Traits

Wheat kernels from each of the HWWAMP genotypes were cleaned and tempered to 15.3% moisture for 24 h. Milling was performed on a laboratory mill (Quadrumat, C.W., Jr.; Barabender Instruments Inc., South Hackensack, NJ, USA) as described previously [49]. The mill produced two fractions, bran and flour, that were collected and weighed. Bran yield was calculated as

mass of bran divided by the combined mass of the flour and bran, expressed as a percentage. Bran yield was used in downstream analysis as a covariate.

Friability of bran was measured by sieving the bran fraction through two testing sieves stacked on top of each other (No. 20 and No. 60 containing 850 µm and 250 µm openings, respectively). Sieves were selected according to published guidelines, where coarse bran is separated from "shorts", or fine bran, by sifting on a No. 20 sieve, and any remaining flour is separated from shorts by sifting through a No. 60 sieve [49]. Friability was calculated as the weight fraction of bran remaining on sieve No. 60 (fine bran particles) divided by the combined weights of bran on the No. 20 and No. 60 sieves (coarse and fine bran particles), expressed as a percentage.

The WRC of bran was analyzed according to the approved method for flour [49], with some modifications. In particular, 1 g of bran was weighed into a tared tube to which 5 mL of water was added. After vortex mixing for 5 s, samples were shaken on a horizontal shaking platform at room temperature and 100 rpm for 20 min. Then samples were centrifuged at 1000× *g* for 15 min. The supernatant was then carefully decanted and test tubes were drained upside down over paper towels for 10 min. Following decanting, the tubes were weighed and the weight of the bran plus absorbed water was calculated. WRC was calculated as the weight of bran plus absorbed water divided by the dry weight of bran, expressed as a percentage.

2.3. Statistical Analysis of Bran Functionality Traits

Data distributions of friability and WRC were visualized using histograms. The normality of the data was analyzed using the Shapiro–Wilk test using the "shapiro.test" package in R (version 1.2.1335, RStudio, Boston, MA, USA).

2.4. Genotyping and Population Structure

The HWWAMP panel was previously genotyped using the 90 K SNP wheat iSelect assay and the marker data is available at the T3 wheat database [50]. The majority of SNPs included in the 90 K SNP assay were SNPs discovered using RNA-seq transcripts from 19 hexaploid and 18 tetraploid wheat lines in addition to SNPs previously discovered in hexaploid wheat [51]. The source of majority of SNPs, being RNA-Seq transcripts, resulted in nonrandom distribution of SNPs in the genome [52,53]. Additional details on the wheat iSelect assay have been presented previously [54]. The dataset comprised a total of 16,054 SNPs across 299 genotypes. In order to avoid spurious marker–trait associations, SNP markers with minor allele frequency <0.05 and missing data of >20% were excluded from the analysis. Genotypes with >20% missing SNP sites (Dawn and Parker) were also excluded from the analyses (leaving 14,661 SNPs). The population structure of the HWWAMP was studied using principal component analysis (PCA). The PCA was performed using TASSEL (version 5.0, Buckler Lab, Ithaca, NY, USA) [55] and a biplot of PC1 versus PC2 was generated.

2.5. Estimating Heritability Using Genome-Wide SNP Markers

The two traits recorded in this study, bran friability and WRC, had one phenotypic value per line. This was similar to one phenotypic measurement generally available on other quality traits in a cultivar development program [56]. A single value per line did not allow the estimation of heritability using variance components derived from a linear mixed model used for phenotypic analysis. Hence, the SNP markers were used to estimate the heritability of friability and WRC traits. The quality-filtered markers used in GWAS were imputed using mean imputations and a genomic relationship matrix (GRM) was built [57]. The details of building a GRM and fitting a genomic best linear unbiased prediction (GBLUP) model to the phenotypic measurements is described previously in our earlier study [58]. The GBLUP model was fit using the Bayesian Generalized Linear Regression (BGLR) function in the BGLR package in R and variance components for GRM (genomic variance) and residual term were estimated [57]. The heritability was estimated as the ratio of genomic variance to the sum of genomic and error variances.

2.6. Genome-Wide Association Study (GWAS)

The GWAS was conducted in TASSEL. Several linear models were examined to achieve the best fit possible, including: general linear model (GLM) [59] using only the SNP markers as fixed effect; mixed linear model (MLM) [60] with kinship as a random effect [59]; MLM with kinship and adding principal components as covariates; and MLM with kinship and phenotypic traits (bran yield and friability) as covariates. The model fit was evaluated by visualizing QQ-plots plotted using the "qqman" package in R [61]. The model that provided the best fit was used to generate the significance values, additive effect scores, and the amount of phenotypic variation explained by the markers. Manhattan plots displaying the significance of the SNPs after Bonferroni correction for each of the traits were plotted using the "qqman" package in R.

2.7. Candidate Gene Analysis

The annotations of the candidate genes underlying or flanking (±5 kb) the SNPs associated with a trait were obtained from the International Wheat Genome Sequencing Consortium (IWGSC) [62]. If annotations for underlying genes were not available, then a default BLASTN search was performed using the wheat gene coding sequence (CDS) against the Ensembl Plants database of monocot and dicot sequences [63]. The BLAST search was performed to determine the likely annotation of the candidate genes. The BLAST search was limited to the top hit with an E-value cut off of less than 1×10^{-40} and sequence identity greater than 50%. The annotations for the underlying or flanking genes were compared against the trait of interest to determine the candidate genes that were likely important for bran functionality traits.

2.8. Linkage Disequilibrium between Trait-Associated SNP Markers on Chromosome 5D and Pin Locus

The Pina-D1, Pinb-D1b, and Pinb-Wild molecular marker information on 44 lines in the HWWAMP was obtained from the regional breeding programs, Northern Regional Performance Nursery (NRPN) and Southern Regional Performance Nursery (SRPN) from 2005–2009 dataset [64]. The "cor.test" procedure was conducted in R using "ggpubr" package to calculate correlation coefficients (as a measure of linkage disequilibrium (LD)) between trait associated SNP markers on chromosome 5D and Pin markers.

3. Results and Discussion

3.1. Bran Functionality Traits

With the emphasis on consumption of whole grain foods by governmental agencies [63], it is relevant to identify and develop wheat lines with specialized traits for whole grain products. Bran particle size and bran WRC have been identified as important determinants of whole grain breadmaking quality [5,8,9]. Therefore, we examined bran friability and WRC in a diverse set of 299 hard winter wheat genotypes to identify genomic regions that may be associated with whole grain end-use quality.

Bran friability was normally distributed from 34.5% to 65.9% with a median of 51.1% (Figure 1a; Shapiro–Wilk, W = 0.99, $p = 0.83$). The wide range in this trait suggested a significant variation in milling performance among wheat genotypes in the HWWAMP.

The WRC ranged from 159–458% with a median of 331% (Figure 1b). The WRC data distribution was close to normal, but had a significant Shapiro–Wilk test (W = 0.97; $p < 0.001$). The broad range in WRC data suggested a significant genetic effect on this trait. However, unlike friability, the distribution departed slightly from normal.

Correlations among bran yield, friability, and water retention were calculated in order to determine their application as a covariates in the GWAS models. Only bran yield and WRC were significantly correlated (r = 0.21; $p = 0.0002$).

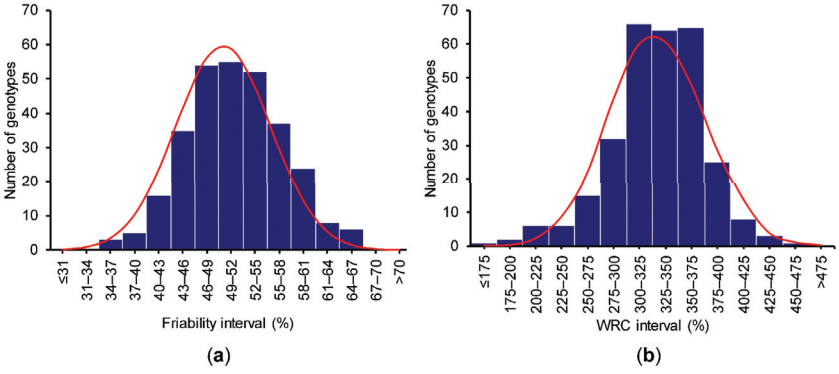


Figure 1. Data distribution of the selected traits: (**a**) bran friability and (**b**) water retention capacity; normal distribution curve shown in red.

3.2. Genotyping and Population Structure

Of the 14,731 SNPs identified across the 299 genotypes, two genotypes and 70 SNPs were dropped from analysis due to minor allele frequency <0.05 and missing data of >20%. The remaining high-quality 14,661 SNPs across 297 lines were used in the analysis. These SNP markers were distributed across all 21 chromosomes with an average of 688 SNPs per chromosome (Figure 2a). Population structure, as visualized by PCA, indicated the presence of a few outliers among the genotypes in the HWWAMP (Figure 2b). A model-based clustering performed using the program STRUCTURE has previously indicated presence of four subgroups in the HWWAMP, and the plots are available in the supplementary files of Ayana et al. [45] and Sidhu et al. [65]. The STRUCTURE plots indicated presence of just a few lines with high (>80–90%) membership to each subgroup and the rest of the lines as admixed. These results of PCA and STRUCTURE suggesting a few lines with distinct backgrounds and a large number of remaining lines as admixed is expected since many of the public wheat breeding programs routinely share germplasm among each other for use as parental lines in breeding programs.

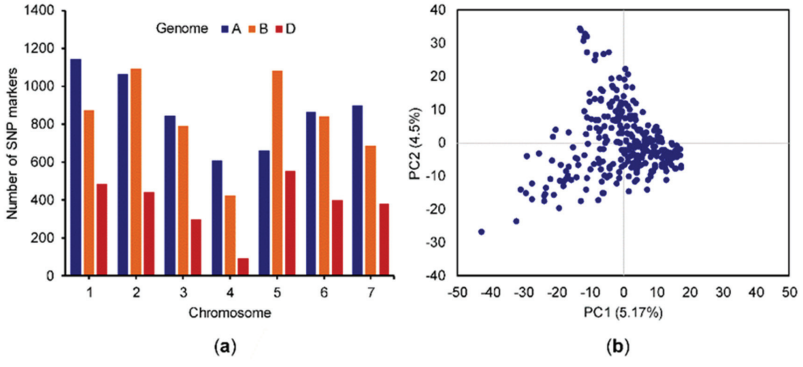


Figure 2. (**a**) Count of single-nucleotide polymorphisms (SNPs) on the wheat chromosomes (208 SNPs located on unanchored scaffold are not included in the plot); (**b**) principle component analysis of the 297 genotypes in the Hard Winter Wheat Association Mapping Panel (HWWAMP) using marker data.

3.3. Heritability Estimates for Bran Functionality Traits Using Genome-Wide SNP Markers

The genomic and error variance estimates for bran friability were 17.64 and 20.53, and for WRC they were 473.78 and 1768.09. The marker-based heritability of friability and WRC were 0.46 and 0.21. These estimates are conservative as compared to broad-sense heritability estimates obtained using the

phenotypic data, because only the additive variance is captured with the markers. Friability had a moderately high heritability value, and this is in the range of heritability values recorded for most quality traits [56]. The relatively low heritability value of WRC suggests a slightly larger influence of nongenetic factors influencing these phenotypic measurements, and additional experiments will be needed to confirm the heritability for this trait. Nevertheless, since the marker-based heritability value was above zero WRC does appear to be influenced at least somewhat by genetics (the phenotypic values are not just noise or nongenetic factors), and thus can be further considered for marker–trait analysis.

3.4. GWAS for Bran Friability

Among models tested, the MLM with kinship and bran yield as a covariate provided the best fit to the data and was used in further analysis (Figure 3). GWAS analysis identified two SNPs located on the 5D chromosome: BS00000020_51 (3,609,894 bp) and Excalibur_c49805_63 (1,614,602 bp), that were significantly associated with bran friability (Bonferroni corrected p-value < 0.05; Figure 4). Of the 297 genotypes, 46 carried the GG genotype (BS00000020_51 marker) and 43 carried the AA genotype (Excalibur_c49805_63 marker), which were associated with increased bran friability (Table 1; Supplementary Table S1). These genotypes were randomly distributed on a PCA biplot generated using the marker data, suggesting the absence of groupings among lines carrying the favorable alleles (Supplementary Figure S1).

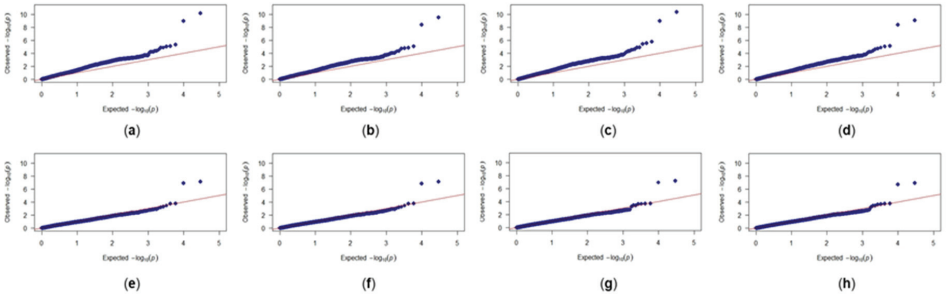


Figure 3. Quantile-quantile plots for bran friability trait: (**a**) General linear model (GLM) with no covariates; (**b**) GLM with bran yield (BY) as covariate; (**c**) GLM with water retention capacity (WRC) as covariate; (**d**) GLM with BY and WRC as covariates; (**e**) mixed linear model (MLM) with kinship and no covariates; (**f**) MLM with kinship and BY as covariates; (**g**) MLM with kinship and WRC as covariates; (**h**) MLM with kinship and BY and WRC as covariates. red line: X = Y; blue diamonds: actual values.

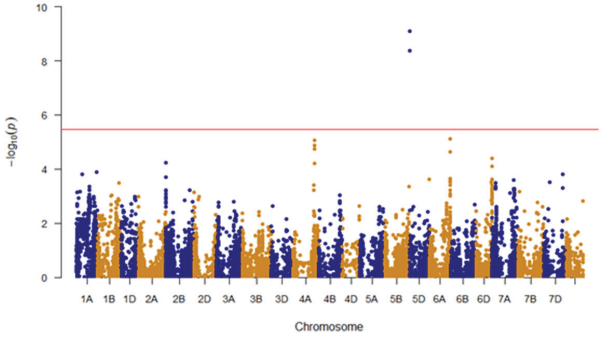


Figure 4. Manhattan plot showing trait associated SNPs for friability; Un, unanchored scaffold; red line, Bonferroni correction (adjusted p < 0.05) cutoff.

Table 1. Trait-associated SNPs and their annotations.[a]

Trait	Marker	SNP	Chr.	Position (bp)	*p*-Value	Marker R^2 [b]	Additive Effect Score (Associated Genotype)	Genotype at SNP Site Associated with an Increase in the Phenotypic Value of the Trait (Number of Lines with This Genotype) [a]	Genotype at SNP Site Associated with a Decrease in the Phenotypic Value of the Trait (Number of Lines with This Genotype)
Friability	BS00000020_51	A/G	5D	3,609,894	<0.001	0.12	−5.95 (AA)	GG (46)	AA (250)
	Excalibur_c49805_63	A/G	5D	1,614,602	<0.001	0.11	5.93 (AA)	AA (43)	GG (254)
WRC	IWA4867	C/T	4A	562,407,344	0.0002	0.05	−20.5 (CC)	TT (192)	CC (97)
	IWA4698	C/T	4A	562,453,542	0.0002	0.04	−19.9 (CC)	TT (194)	CC (98)

[a] Out of a total of 297 lines analyzed from the Hard Winter Wheat Association Mapping Panel; [b] Proportion of phenotypic variance explained by the SNP marker.

Low bran friability is desirable for refined flour production because bran particles remain large and intact and are thus easier to separate from the endosperm flour. As refined flour represents the majority of wheat flour production [3], this may be why so few lines carried the high bran friability alleles. However, for whole wheat flour milling, the high bran friability trait would be beneficial because smaller bran particle sizes have been shown to result in bread with higher volume [6]. As demand for whole wheat flour increases [3], wheat genotypes that carry the high bran friability trait may become more sought after. The markers associated with bran friability in this study, after validation in independent population, will be a useful resource for developing lines with increased bran friability. The 43 to 46 genotypes carrying increased bran friability favorable alleles can be used as parental lines while making crosses to develop lines with increased bran friability (Supplementary Table S1).

3.5. GWAS for Bran WRC

Among models tested, the best fit for the WRC trait was obtained using a GLM model with bran yield and friability as covariates (Figure 5). Unlike the friability trait, GWAS analysis did not reveal any makers that were significantly associated with WRC after Bonferroni correction (Figure 6). These results could be because the WRC data were not normally distributed or the phenotypic measurements recorded for WRC in this study are significantly influenced by nongenetic factors as noted by the low value of marker-based heritability. There were two markers located on the chromosome 4A that stood out from among the other SNPs: IWA4867 (562,407,344 bp) and IWA4698 (562,453,542 bp). Despite being not significant, these markers were examined further to determine if there may be any logical link to bran WRC. A majority of genotypes had these alleles for increased WRC: 192 genotypes for marker IWA4867 (TT genotype) and 194 for marker IWA4698 (TT genotype; Table 1; Supplementary Table S1). These genotypes were randomly distributed on a PCA biplot generated using the marker data, suggesting the absence of groupings among lines carrying the favorable alleles (Supplementary Figure S2).

As indicated, most genotypes carried the tentative alleles for increased bran WRC. Unfortunately, higher WRC of bran is unfavorable, because it is associated with production of bread with low loaf volume [9]. While the vast majority of breeding efforts have been focused on improving the performance of the refined flour, limited attention was dedicated to bran quality. Therefore, more efforts are needed to select lines for whole grain baking. Perhaps this can be done by selecting lines that do not have these alleles.

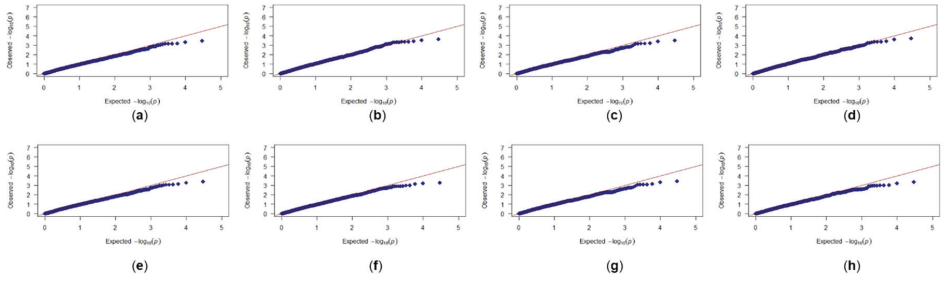


Figure 5. Quantile-quantile plots for water retention capacity trait: (**a**) GLM model with no covariates; (**b**) GLM model with bran yield (BY) as covariate; (**c**) GLM model with bran friability (BF) as covariate; (**d**) GLM model with BY and BF as covariates; (**e**) MLM model with kinship and no covariates; (**f**) MLM model with kinship and BY as covariate; (**g**) MLM model with kinship and BF as covariate; (**h**) MLM model with kinship and BY and BF as covariates.

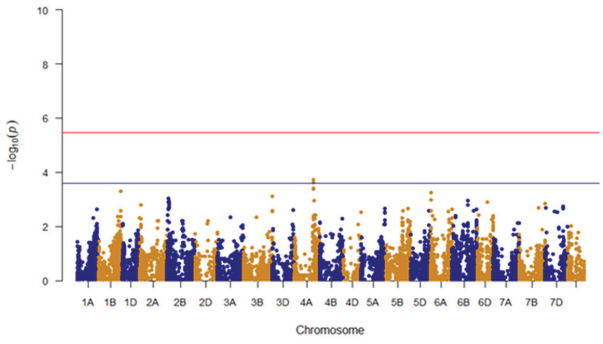


Figure 6. Manhattan plot showing trait associated SNPs for water retention capacity; Un, unanchored scaffold; red line, Bonferroni correction (adjusted $p < 0.05$) cutoff; blue line, notable SNPs above the background.

3.6. Candidate Genomic Regions for Bran Friability and WRC

The two markers that were significantly associated with bran friability, BS00000020_51 (5D: 3,609,894 bp) and Excalibur_c49805_63 (5D: 1,614,602 bp), were located in genes *TraesCS5D02G004300* (5D: 3,609,672–3,610,121 bp) and TraesCS5D02G001200 (5D: 1,609,486–1,614,664 bp). *TraesCS5D02G001200* encodes for sucrose membrane transfer proteins.As a part of cell membranes, these proteins could play an additional role in the friability of bran during milling. The other gene, *TraesCS5D02G004300*, was annotated as *Pinb-D1b* at Ensembl Plants [63] and as *Puroindoline-b* in the IWGSC functional annotations [62]. A BLASTN search with Pinb-D1 forward primer sequence (obtained from Graingenes; [66]) as query against the IWGSC RefSeq v1.0 wheat genome matched exactly to the first 20 bp of this gene. Further, the correlation coefficient between the SNP marker associated with bran friability (BS00000020_51) and *Pinb-D1b* marker (r = 0.75; $p < 0.001$) and *Pinb-Wild* (r = −0.70; $p < 0.001$) was relatively stronger compared to *Pina-D1* (r = 0.35; $p = 0.02$). This result is probably because the distance between *Pinb-D1* forward primer start position and the friability-associated SNP marker (BS00000020_51) is 222 bp. Overall, these observations indicate that the SNP marker associated with bran friability (BS00000020_51; 5D: 3 609 894 bp) is in linkage disequilibrium and potentially tagged to the Pinb-D1 marker, and is a functional variant for *Pinb* locus. This also suggests that *Pinb-D1* locus may be involved in regulating bran friability.

Pinb-D1 genes encode for puroindoline proteins, which make up a complex multicomponent complex called fraibilin [67]. Friabilin is involved in the association between starch and protein in the wheat kernel and is responsible for soft endosperm texture [68]. The results from this study suggest that, in addition to controlling endosperm texture, the puroindoline genes may also affect bran texture.

The two SNP markers that were tentatively associated with WRC, IWA4867 (4A: 562,407,344 bp) and IWA4698 (4A: 562,453,542 bp), were located close to *TraesCS4A02G251100* (562,408,612–562,414,155 bp) and *TraesCS4A02G251300* (562,457,963–562,459,338 bp), respectively. The *TraesCS4A02G251300* gene did not have a characterized function in wheat. However, it did have high similarity to the *Zm00001d033902_T001* gene present in maize (%ID—89.5; E-value: 1.6×10^{-63}), which was reported to encode for cysteine-rich secretory proteins. Interestingly, in wheat, nongluten cysteine-rich proteins are harmful for breadmaking because they interfere with gluten macropolymerization [69].

4. Conclusions

In conclusion, our study identified novel SNPs associated with bran friability and tentative SNPs associated with WRC. One of the SNP markers associated with bran friability was tagged to the well-characterized Pinb locus in wheat, which suggests that in addition to endosperm texture, Pinb may also have a role in bran texture (friability). The favorable alleles at the trait-associated SNP sites were

present in only a fraction of lines (mostly recent releases and very few heritage or older cultivars) in the HWWAMP, which comprises breeding lines from many of the breeding programs in the US. This result indicates that the traits beneficial for whole wheat flour are neither selected for or against, just randomly, and are relatively new in the breeding history. Additionally, the lines carrying favorable alleles for both traits were randomly distributed in a PCA biplot, suggesting the absence of groupings among lines carrying favorable alleles. A conscious effort is thus needed to breed for cultivars that would be more desirable for whole-grain milling. Upon validation in independent populations and multiple environments, the novel markers identified in this study may help breeders select lines that have high bran friability and low WRC, which are preferred in whole grain bread-making. Additionally, the lines carrying favorable alleles identified in this study can be used as parental lines while making new crosses in breeding programs for developing cultivars suitable for whole wheat grain flour. Further, if the genomic regions for bran friability and WRC are validated in multiple backgrounds and environments, they can then become potential targets for function characterization studies such as knockout or overexpression experiments and thus help narrow down genes important for these traits in bread wheat. These efforts overall could lead to release of wheat lines with high whole grain flour functionality and identification and characterization of genes important for whole grain bread-making.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/8/838/s1, Figure S1: Principle components analysis biplot of the 297 genotypes in the HWWAMP using marker data, with the 46 lines carrying the GG favorable genotype for the friability marker, BS00000020_51 (A), highlighted in color, and the 43 lines carrying the AA favorable genotype for the friability marker, Excalibur_c49805_63 (B), highlighted in color, Figure S2: Principle components analysis biplot of the 297 genotypes in the HWWAMP using marker data, with the 97 lines carrying CC favorable genotype for the WRC marker, IWA4867 (A), highlighted in color, and the 98 lines carrying CC favorable genotype for the WRC marker, IWA4698, highlighted in color, Table S1: Genotype information across HWWAMP lines for the markers associated with Friability and Water Retention Capacity traits.

Author Contributions: S.N., P.S.B. and D.J.R. contributed conception and design of the study; S.N. performed the experiments; P.S.B. organized, maintained, and produced the mapping population; S.N. and V.B. performed the quantitative genetics and statistical analyses; S.N. wrote the first draft of the manuscript; S.N., V.B., and D.J.R. wrote sections of the manuscript. All authors contributed to manuscript revision. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by a grant from the Agricultural Research Division of the University of Nebraska-Lincoln.

Acknowledgments: The authors acknowledge the helpful assistance of Eric Nkurikiye in completion of the bran friability and bran water retention capacity measurements. The authors would like to thank Triticeae Coordinated Agricultural Project (TCAP) for making the genotypic data publicly available. Also, we want to acknowledge Marry Guttieri and Nicholas Garst for their efforts in organization and growing of HWWAMP in Nebraska. Finally, we are grateful to the numerous wheat plant breeders and geneticists from the public and private sector who donated their lines to create the HWWAMP.

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. Aune, D.; Norat, T.; Romundstad, P.; Vatten, L.J. Whole grain and refined grain consumption and the risk of type 2 diabetes: A systematic review and dose-response meta-analysis of cohort studies. *Eur. J. Epidemiol.* **2013**, *28*, 845–858. [CrossRef]
2. Cho, S.S.; Qi, L.; Fahey, G.C.J.; Klurfeld, D.M. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. *Am. J. Clin. Nutr.* **2013**, *98*, 594–619. [CrossRef] [PubMed]
3. Whole Grains Council. Whole Grain Statistics [Internet]. 2020. Available online: https://wholegrainscouncil.org/newsroom/whole-grain-statistics (accessed on 13 March 2020).
4. Whole Grain Initiative. Definition of Whole Grain as Food Ingredient [Internet]. 2019. Available online: https://wgi.meetinghand.com/projectData/775/webData/Definition-of-Whole-Grain-as-Food-Ingredient-Version-20190501C.pdf (accessed on 8 July 2020).

5. Cai, L.; Choi, I.; Lee, C.K.; Park, K.K.; Baik, B.K. Bran characteristics and bread-baking quality of whole grain wheat flour. *Cereal Chem.* **2014**, *91*, 398–405. [CrossRef]
6. Noort, M.W.J.; van Haaster, D.; Hemery, Y.; Schols, H.A.; Hamer, R.A. The effect of particle size of wheat bran fractions on bread quality-Evidence for fibre-protein interactions. *J. Cereal Sci.* **2010**, *52*, 59–64. [CrossRef]
7. Jacobs, P.J.; Hemdane, S.; Claes, S.; Mulders, L.; Langenaeken, N.A.; Dewettinck, K.; Courtin, C.M. Wheat bran-associated subaleurone and endosperm proteins and their impact on bran-rich bread-making. *J. Cereal Sci.* **2018**, *81*, 99–107. [CrossRef]
8. Jacobs, P.J.; Hemdane, S.; Dornez, E.; Delcour, J.A.; Courtin, C.M. Study of hydration properties of wheat bran as a function of particle size. *Food Chem.* **2015**, *179*, 296–304. [CrossRef]
9. Navrotskyi, S.; Guo, G.; Baenziger, P.S.; Xu, L.; Rose, D.J. Impact of wheat bran physical properties and chemical composition on whole grain flour mixing and baking properties. *J. Cereal Sci.* **2019**, *89*, 102790. [CrossRef]
10. Khalid, K.H.; Manthey, F.; Simsek, S. Centrifugal milling of wheat bran. *Cereal Chem.* **2018**, *95*, 330–341. [CrossRef]
11. Zhang, D.; Moore, W.R. Effect of wheat bran particle size on dough rheological properties. *J. Sci. Food Agric.* **1997**, *74*, 490–496. [CrossRef]
12. Hemdane, S.; Jacobs, P.J.; Dornez, E.; Verspreet, J.; Delcour, J.A.; Courtin, C.M. Wheat (*Triticum aestivum* L.) bran in bread making: A critical review. *Compr. Rev. Food Sci. Food Saf.* **2016**, *15*, 28–42. [CrossRef]
13. Maccaferri, M.; Zhang, J.; Bulli, P.; Abate, Z.; Chao, S.; Cantu, D.; Bossolini, E.; Chen, X.; Pumphrey, M.; Dubcovsky, J. A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3 Genes Genomes Genet.* **2015**, *5*, 449–465.
14. Jaiswal, V.; Gahlaut, V.; Meher, P.K.; Mir, R.R.; Jaiswal, J.P.; Rao, A.; Balyan, H.S.; Gupta, P. Genome Wide Single Locus Single Trait, Multi-Locus and Multi-Trait Association Mapping for Some Important Agronomic Traits in Common Wheat (*T. aestivum* L.). *PLoS ONE* **2016**, *11*, e0159343. [CrossRef] [PubMed]
15. Lehnert, H.; Serfling, A.; Friedt, W.; Ordon, F. Genome-Wide Association Studies Reveal Genomic Regions Associated With the Response of Wheat (*Triticum aestivum* L.) to Mycorrhizae Under Drought Stress Conditions. *Front. Plant Sci.* **2018**, *9*, 1728. [CrossRef] [PubMed]
16. Yan, X.; Zhao, L.; Ren, Y.; Dong, Z.; Cui, D.; Chen, F. Genome-wide association study revealed that the TaGW8 gene was associated with kernel size in Chinese bread wheat. *Sci. Rep.* **2019**, *9*, 2702. [CrossRef] [PubMed]
17. Jamil, M.; Ali, A.; Gul, A.; Ghafoor, A.; Napar, A.A.; Ibrahim, A.M.H.; Naveed, N.H.; Yasin, N.A.; Mujeeb-Kazi, A. Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* **2019**, *19*, 149. [CrossRef] [PubMed]
18. Yang, X.; Pan, Y.; Singh, P.K.; He, X.; Ren, Y.; Zhao, L.; Zhang, N.; Cheng, S.; Chen, F. Investigation and genome-wide association study for Fusarium crown rot resistance in Chinese common wheat. *BMC Plant Biol.* **2019**, *19*, 153. [CrossRef]
19. Sheoran, S.; Jaiswal, S.; Kumar, D.; Raghav, N.; Sharma, R.; Pawar, S.; Paul, S.; Iquebal, M.A.; Jaiswar, A.; Sharma, P.; et al. Uncovering Genomic Regions Associated With 36 Agro-Morphological Traits in Indian Spring Wheat Using GWAS. *Front. Plant Sci.* **2019**, *10*, 527. [CrossRef]
20. Afzal, F.; Li, H.; Gul, A.; Subhani, A.; Ali, A.; Mujeeb-Kazi, A.; Ogbonnaya, F.; Trethowan, R.; Xia, X.; He, Z.; et al. Genome-Wide Analyses Reveal Footprints of Divergent Selection and Drought Adaptive Traits in Synthetic-Derived Wheats. *G3 Genes Genomes Genet.* **2019**, *9*, 1957–1973. [CrossRef]
21. Pradhan, S.; Babar, A.; Robbins, K.; Bai, G.; Mason, R.E.; Khan, J.; Shahi, D.; Avci, M.; Guo, J.; Hossain, M.M.; et al. Understanding the Genetic Basis of Spike Fertility to Improve Grain Number, Harvest Index, and Grain Yield in Wheat Under High Temperature Stress Environments. Front. *Plant Sci.* **2019**, *10*, 1481. [CrossRef]
22. Rahimi, Y.; Bihamta, M.R.; Taleei, A.; Alipour, H.; Ingvarsson, P.K. Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. *BMC Plant Biol.* **2019**, *19*, 1–19. [CrossRef]
23. Basile, S.M.L.; Ramírez, I.A.; Crescente, J.M.; Conde, M.B.; Demichelis, M.; Abbate, P.; Rogers, W.J.; Pontaroli, A.; Helguera, M.; Vanzetti, L.S. Haplotype block analysis of an Argentinean hexaploid wheat collection and GWAS for yield components and adaptation. *BMC Plant Biol.* **2019**, *19*, 1–16.

24. Bin Safdar, L.; Andleeb, T.; Latif, S.; Umer, M.J.; Tang, M.; Li, X.; Liu, S.; Quraishi, U.M. Genome-Wide Association Study and QTL Meta-Analysis Identified Novel Genomic Loci Controlling Potassium Use Efficiency and Agronomic Traits in Bread Wheat. *Front. Plant Sci.* **2020**, *11*, 70. [CrossRef] [PubMed]
25. Li, G.; Xu, X.; Bai, G.; Carver, B.F.; Hunger, R.; Bonman, J.M.; Kolmer, J.A.; Dong, H. Genome-Wide Association Mapping Reveals Novel QTL for Seedling Leaf Rust Resistance in a Worldwide Collection of Winter Wheat. *Plant Genome* **2016**, *9*, 1–12. [CrossRef] [PubMed]
26. Daba, S.D.; Tyagi, P.; Brown-Guedira, G.; Mohammadi, M. Genome-wide association study in historical and contemporary U.S. winter wheats identifies height-reducing loci. *Crop J.* **2020**, *8*, 243–251. [CrossRef]
27. Sun, C.; Zhang, F.; Yan, X.; Zhang, X.; Dong, Z.; Cui, D.; Chen, F. Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnol. J.* **2017**, *15*, 953–969. [CrossRef]
28. Ogbonnaya, F.C.; Rasheed, A.; Okechukwu, E.C.; Jighly, A.; Makdis, F.; Wuletaw, T.; Hagras, A.; Uguru, M.I.; Agbo, C.U. Genome-wide association study for agronomic and physiological traits in spring wheat evaluated in a range of heat prone environments. Theor. *Appl. Genet.* **2017**, *130*, 1819–1835. [CrossRef]
29. Liu, J.; He, Z.; Rasheed, A.; Wen, W.; Yan, J.; Zhang, P.; Wan, Y.; Zhang, Y.; Xie, C.; Xia, X. Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2017**, *17*, 220. [CrossRef]
30. Sukumaran, S.; Lopes, M.; Dreisigacker, S.; Reynolds, M. Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theor. Appl. Genet.* **2018**, *131*, 985–998. [CrossRef]
31. Peng, Y.; Liu, H.; Chen, J.; Shi, T.; Zhang, C.; Sun, D.; He, Z.; Hao, Y.; Chen, W. Genome-Wide Association Studies of Free Amino Acid Levels by Six Multi-Locus Models in Bread Wheat. Front. *Plant Sci.* **2018**, *9*, 1196.
32. Bhatta, M.; Morgounov, A.; Belamkar, V.; Baenziger, P.S. Genome-Wide Association Study Reveals Novel Genomic Regions for Grain Yield and Yield-Related Traits in Drought-Stressed Synthetic Hexaploid Wheat. *Int. J. Mol. Sci.* **2018**, *19*, 3011. [CrossRef]
33. Liu, J.; Xu, Z.; Fan, X.; Zhou, Q.; Cao, J.; Wang, F.; Ji, G.; Yang, L.; Feng, B.; Wang, T. A Genome-Wide Association Study of Wheat Spike Related Traits in China. *Front. Plant Sci.* **2018**, *9*, 1584. [CrossRef] [PubMed]
34. Bhatta, M.; Shamanin, V.; Shepelev, S.S.; Baenziger, P.S.; Pozherukova, V.; Pototskaya, I.; Morgounov, A. Marker-Trait Associations for Enhancing Agronomic Performance, Disease Resistance, and Grain Quality in Synthetic and Bread Wheat Accessions in Western Siberia. *G3 Genes Genomes Genet.* **2019**, *9*, 4209–4222. [CrossRef] [PubMed]
35. Zhu, Y.; Wang, S.; Wei, W.; Xie, H.; Liu, K.; Zhang, C.; Wu, Z.; Jiang, H.; Cao, J.; Zhao, L.; et al. Genome-wide association study of pre-harvest sprouting tolerance using a 90K SNP array in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2019**, *132*, 2947–2963. [CrossRef] [PubMed]
36. Chen, J.; Zhang, F.; Zhao, C.; Lv, G.; Sun, C.; Pan, Y.; Guo, X.; Chen, F. Genome-wide association study of six quality traits reveals the association of the TaRPP13L1 gene with flour colour in Chinese bread wheat. *Plant Biotechnol. J.* **2019**, *17*, 2106–2122. [CrossRef] [PubMed]
37. Kristensen, P.S.; Jensen, J.; Andersen, J.R.; Guzmán, C.; Orabi, J.; Jahoor, A. Genomic Prediction and Genome-Wide Association Studies of Flour Yield and Alveograph Quality Traits Using Advanced Winter Wheat Breeding Material. *Genes* **2019**, *10*, 669. [CrossRef] [PubMed]
38. Battenfield, S.D.; Sheridan, J.L.; Silva, L.D.C.E.; Miclaus, K.J.; Dreisigacker, S.; Wolfinger, R.D.; Peña, R.J.; Singh, R.P.; Jackson, E.W.; Fritz, A.K.; et al. Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS ONE* **2018**, *13*, e0204757. [CrossRef] [PubMed]
39. Zhai, S.; Liu, J.; Xu, D.; Wen, W.; Yan, J.; Zhang, P.; Wan, Y.; Cao, S.; Hao, Y.; Xia, X.; et al. A Genome-Wide Association Study Reveals a Rich Genetic Architecture of Flour Color-Related Traits in Bread Wheat. *Front. Plant Sci.* **2018**, *9*, 1136. [CrossRef]
40. Liu, J.; Feng, B.; Xu, Z.; Fan, X.; Jiang, F.; Jin, X.; Cao, J.; Wang, F.; Liu, Q.; Yang, L.; et al. A genome-wide association study of wheat yield and quality-related traits in southwest China. *Mol. Breed.* **2017**, *38*, 1–11. [CrossRef]
41. Li, J.; Rasheed, A.; Guo, Q.; Dong, Y.; Liu, J.; Xia, X.; Zhang, Y.; He, Z. Genome-wide association mapping of starch granule size distribution in common wheat. *J. Cereal Sci.* **2017**, *77*, 211–218. [CrossRef]

42. Grogan, S.M.; Anderson, J.; Baenziger, P.S.; Frels, K.; Guttieri, M.J.; Haley, S.D.; Kim, K.-S.; Liu, S.; McMaster, G.S.; Newell, M.; et al. Phenotypic Plasticity of Winter Wheat Heading Date and Grain Yield across the US Great Plains. *Crop Sci.* **2016**, *56*, 2223–2236. [CrossRef]
43. Guttieri, M.J.; Baenziger, P.S.; Frels, K.; Carver, B.; Arnall, B.; Waters, B. Variation for Grain Mineral Concentration in a Diversity Panel of Current and Historical Great Plains Hard Winter Wheat Germplasm. *Crop Sci.* **2015**, *55*, 1035–1052. [CrossRef]
44. Guttieri, M.J.; Baenziger, P.S.; Frels, K.; Carver, B.; Arnall, B.; Wang, S.; Akhunov, E.; Waters, B. Prospects for Selecting Wheat with Increased Zinc and Decreased Cadmium Concentration in Grain. *Crop Sci.* **2015**, *55*, 1712–1728. [CrossRef]
45. Ayana, G.T.; Ali, S.; Sidhu, J.S.; Hernandez, J.L.G.; Turnipseed, B.; Sehgal, S.K. Genome-Wide Association Study for Spot Blotch Resistance in Hard Winter Wheat. *Front. Plant Sci.* **2018**, *9*, 926.
46. Awad, W. Genome-Wide Association Study and Drought Tolerance Evaluation of a Winter Wheat Association Mapping Panel. Ph.D. Thesis, Colorado State University, Fort Collins, CO, USA, 2015.
47. Bhatta, M.; Baenziger, P.S.; Waters, B.; Poudel, R.; Belamkar, V.; Poland, J.; Morgounov, A. Genome-Wide Association Study Reveals Novel Genomic Regions Associated with 10 Grain Minerals in Synthetic Hexaploid Wheat. *Int. J. Mol. Sci.* **2018**, *19*, 3237. [CrossRef] [PubMed]
48. Ramakrishnan, S.M.; Sidhu, J.S.; Ali, S.; Kaur, N.; Wu, J.; Sehgal, S.K. Molecular characterization of bacterial leaf streak resistance in hard winter wheat. *PeerJ* **2019**, *7*, e7276. [CrossRef]
49. Cereals & Grains Association. Approved methods: 26-50.01–Brabender Quadrumat Jr. (Quadruplex) method; 56-11.02–Solvent retention capacity profile. In *Cereals & Grains Association, editor. Approv. Methods Anal*, 11th ed.; Cereals & Grains Association: St. Louis, MO, USA, 2020. Available online: http://methods.aaccnet.org/default.aspx (accessed on 22 July 2020).
50. T3/Wheat [Internet]. Available online: https://triticeaetoolbox.org/wheat/genotyping/display_genotype.php?trial_code=TCAP90K_HWWAMP (accessed on 22 July 2020).
51. El-Basyoni, I.S.; Lorenz, A.; Guttieri, M.; Frels, K.; Baenziger, P.; Poland, J.; Akhunov, E. A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci.* **2018**, *270*, 123–130. [CrossRef]
52. The International Wheat Genome Sequencing Consortium (IWGSC); Appels, R.; Eversole, K.A.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar719.
53. Belamkar, V.; Farmer, A.D.; Weeks, N.T.; Kalberer, S.R.; Blackmon, W.J.; Cannon, S.B. Genomics-assisted characterization of a breeding collection of Apios americana, an edible tuberous legume. *Sci. Rep.* **2016**, *6*, 34908. [CrossRef]
54. Wang, S.; Wong, D.; Forrest, K.; Allen, A.; Chao, S.; Huang, B.E.; Maccaferri, M.; Salvi, S.; Milner, S.G.; Cattivelli, L.; et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **2014**, *12*, 787–796. [CrossRef]
55. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]
56. Battenfield, S.D.; Guzman, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* **2016**, *9*, 1–12. [CrossRef] [PubMed]
57. Pérez, P.; Campos, G.D.L. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]
58. Belamkar, V.; Guttieri, M.J.; Hussain, W.; Jarquín, D.; El-Basyoni, I.; Poland, J.; Lorenz, A.J.; Baenziger, P.S. Genomic Selection in Preliminary Yield Trials in a Winter Wheat Breeding Program. *G3 Genes Genomes Genet.* **2018**, *8*, 2735–2747. [CrossRef] [PubMed]
59. Oraguzie, N.; Rikkerink, E.; Gardine, S.; De Silva, N. *Association Mapping in Plants*; Oraguzie, N., Rikkerink, E., Gardine, S., De Silva, N., Eds.; Springer: New York, NY, USA, 2007.
60. Collins, A. *Linkage Disequilibrium and Association Mapping: Analysis and Applications*; Collins, A., Ed.; Humana Press: Totowa, NJ, USA, 2007.
61. Turner, S.D. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*, 005165. [CrossRef]

62. U.S. Department of Health and Human Services, U.S. Department of Agriculture. Dietary Guidelines for Americans 2015–2020. Available online: https://health.gov/dietaryguidelines/2015/guidelines/ (accessed on 12 July 2019).
63. Ensembl Plants [Internet]. Available online: https://plants.ensembl.org/index.html (accessed on 22 July 2020).
64. Hard Winter Wheat Regional Nursery Program [Internet]. Available online: https://www.ars.usda.gov/plains-area/lincoln-ne/wheat-sorghum-and-forage-research/docs/hard-winter-wheat-regional-nursery-program/research/ (accessed on 22 July 2020).
65. Sidhu, J.S.; Singh, D.; Gill, H.S.; Brar, N.K.; Qiu, Y.; Halder, J.; Al Tameemi, R.; Turnipseed, B.; Sehgal, S.K. Genome-Wide Association Study Uncovers Novel Genomic Regions Associated With Coleoptile Length in Hard Winter Wheat. *Front. Genet.* **2020**, *10*, 1345. Available online: https://www.frontiersin.org/article/10.3389/fgene.2019.01345/full (accessed on 13 July 2020). [CrossRef]
66. GrainGenes [Internet]. Available online: https://wheat.pw.usda.gov/GG3/ (accessed on 22 July 2020).
67. Iftikhar, A.; Ali, I. Kernel softness in wheat is determined by starch granule bound Puroindoline proteins. *J. Plant Biochem. Biotechnol.* **2016**, *26*, 247–262. [CrossRef]
68. Pauly, A.; Pareyt, B.; Fierens, E.; Delcour, J.A. Wheat (*Triticum aestivum* L. and *T. turgidum* L. ssp. durum) Kernel Hardness: I. Current View on the Role of Puroindolines and Polar Lipids. *Compr. Rev. Food Sci. Food Saf.* **2013**, *12*, 413–426. Available online: http://doi.wiley.com/10.1111/1541-4337.12019 (accessed on 13 July 2020).
69. Noctor, G.; Mhamdi, A.; Chaouch, S.; Han, Y.; Neukermans, J.; García, B.M.; Queval, G.; Foyer, C.H. Glutathione in plants: An integrated overview. *Plant Cell Environ.* **2011**, *35*, 454–484. [CrossRef]

genes

MDPI

Article

A Genome-Wide Analysis of the'*Pentatricopeptide Repeat* (PPR) Gene Family and PPR-Derived Markers for Flesh Color in Watermelon (*Citrullus lanatus*)

Saminathan Subburaj [1], Luhua Tu [1], Kayoun Lee [1], Gwang-Soo Park [1,2], Hyunbae Lee [1,2], Jong-Pil Chun [1], Yong-Pyo Lim [1], Min-Woo Park [3], Cecilia McGregor [4] and Geung-Joo Lee [1,2,*]

1 Department of Horticulture, Chungnam National University, Daejeon 34134, Korea; sami_wheat@cnu.ac.kr (S.S.); luluguniang@gmail.com (L.T.); kayoun200@cnu.ac.kr (K.L.); kps21641001@hanmail.net (G.-S.P.); lhb7982@gmail.com (H.L.); jpchun@cnu.ac.kr (J.-P.C.); yplim@cnu.ac.kr (Y.-P.L.)
2 Department of Smart Agriculture Systems, Chungnam National University, Daejeon 34134, Korea
3 Breeding Institute, Hyundai Seed Co Ltd., Yeoju, Gyeonggi-do 12660, Korea; p.minwoo@gmail.com
4 Department of Horticulture, University of Georgia, Athens, GA 30602, USA; cmcgre1@uga.edu
* Correspondence: gjlee@cnu.ac.kr; Tel.: +82-42-821-5734; Fax: +82-42-823-1382

Received: 24 August 2020; Accepted: 23 September 2020; Published: 24 September 2020

Abstract: Watermelon (*Citrullus lanatus*) is an economically important fruit crop grown for consumption of its large edible fruit flesh. *Pentatricopeptide-repeat* (PPR) encoding genes, one of the large gene families in plants, are important RNA-binding proteins involved in the regulation of plant growth and development by influencing the expression of organellar mRNA transcripts. However, systematic information regarding the PPR gene family in watermelon remains largely unknown. In this comprehensive study, we identified and characterized a total of 422 *C. lanatus* PPR (*ClaPPR*) genes in the watermelon genome. Most *ClaPPRs* were intronless and were mapped across 12 chromosomes. Phylogenetic analysis showed that *ClaPPR* proteins could be divided into P and PLS subfamilies. Gene duplication analysis suggested that 11 pairs of segmentally duplicated genes existed. In-silico expression pattern analysis demonstrated that *ClaPPRs* may participate in the regulation of fruit development and ripening processes. Genotyping of 70 lines using 4 single nucleotide polymorphisms (SNPs) from 4 *ClaPPRs* resulted in match rates of over 0.87 for each validated SNPs in correlation with the unique phenotypes of flesh color, and could be used in differentiating red, yellow, or orange watermelons in breeding programs. Our results provide significant insights for a comprehensive understanding of *PPR* genes and recommend further studies on their roles in watermelon fruit growth and ripening, which could be utilized for cultivar development of watermelon.

Keywords: watermelon; pentatricopeptide-repeat (PPR) gene family; comprehensive analysis; expression profiling; flesh color

1. Introduction

Pentatricopeptide (PPR) proteins are one of the largest gene families in plants, and are usually characterized by an array of 2–30 tandem repeats of a degenerated unit consisting of 30–40 amino acid (aa) sequence motifs [1]. According to the domain architecture, PPR proteins are divided into subfamilies of P and PPR-like long and short (PLS), which are characterized by motifs without space and motifs with interspaced either of short (31 aa) or long (35–36 aa) PPR-like motifs, respectively. Based on the domain assembly in the C terminal of a PPR protein, the PLS subfamily is further

classified into five subgroups: PLS, E1, E2, E+, and DYW [1–3]. Since the discovery of PPR proteins in yeast (*Saccharomyces cerevisiae* L.) [4], these have been reported in various terrestrial plants. Being a large gene family, more than 400 PPR proteins have been reported in various plants, including *Arabidopsis* (441), foxtail millet (486), poplar (626), maize (491), and rice (477) [1,3,5–7]. PPR proteins have been found to exhibit RNA-binding properties, which facilitate in mediating gene expression through posttranscriptional processes associated with transcripts in the mitochondria, chloroplast, and nucleus. Thus, PPR genes are thought to have a major impact on organelle stability, including biogenesis, and function through their involvement in various posttranscriptional processes such as RNA-editing [1,8], RNA-splicing [9], and RNA-processing [10].

The functions of PPR proteins have been reported to be associated with plant growth and development and organelle formation. In Arabidopsis, a chloroplast-localized PPR protein, called EMBRYO-DEFECTIVE175 (EMB175), has been found to influence embryo morphogenesis [11]. Similarly, in maize, a mitochondria-localized P-type PPR protein, EMPTY PERICARP12 (EMP12), has been reported to be essential for embryogenesis and endosperm development through trans-splicing of mitochondrial *nad2* introns [12]. Hsieh et al. [13] showed that the *SLOW GROWTH3* (*SLO3*) gene, encoding a PPR protein, was involved in the splicing of *nad7* intron 2 in Arabidopsis, and its mutant, *slo3,* exhibited a dysfunctional mitochondrion, which resulted in growth retardation and delayed development. In rice, few DYW-type PPR proteins, such as OPAQUE AND GROWTH RETARDATION 1 (OGR1) and PPS1, play important roles in C→U RNA editing of mitochondrial transcripts; upon silencing of these genes using T-DNA insertion and RNAi, mutant plants exhibited various pleiotropic phenotypes, including late seed germination, retarded growth, delayed development, dwarfing, and partial pollen sterility at both vegetative and reproductive stages [14,15]. In addition to plant developmental process, PPR proteins have been reported to be involved in the responses to biotic and abiotic stresses. For instance, in Arabidopsis, several PPR proteins such as *SOAR1* [16], *PGN* [17], *SLG1* [18], and *PPR96* [19], have been shown to participate in responses to abiotic stresses.

PPR proteins have also been characterized to be involved in cytoplasmic male sterility (CMS); it is an important intriguing issue in plants [20], as only the male gametes are impaired, resulting in a failure to produce functional pollens. Some of the PPR proteins encoded as restorers of fertility (*Rf*) genes mask the mitochondrial transcripts that cause CMS and thus restores fertility. The *Rf-PPR* genes have been reported in various terrestrial plants, including petunia (*RF952*) [21], radish (orf687) [22], pepper (*CaPPR6*) [23], and Arabidopsis (*RFL2*) [24]. Increasing molecular evidence has clearly emphasized the roles of PPR genes in fruit development, ripening, and flesh color of plants [25–27]. *GUN1,* which encodes a plastid-located PPR protein, has been reported to be involved in the plastid-to-nucleus retrograde signaling pathway during fruit development and ripening in tomato [26]. In relation to this, ripening impaired tomato mutants such as *Cnr* dramatically repressed expression of genes associated with ripening, cell wall-degrading enzymes and PPR repeat-containing proteins, resulting in mature fruits with colorless pericarp tissue, thereby indicating that PPR proteins play a significant role in fruit development [25]. Recently, in melon (*Cucumis melo* L.), the *white-flesh* gene, named *CmPPR1* (MELO3C003069), encoding a plastid-targeted P-type PPR protein, has been reported to be a candidate gene in one of the two major QTL, which determine flesh color intensity [27]. Furthermore, the polymorphic SNP markers in P-class motifs of *CmPPR1* have been found to contribute to genetic variation in orange, green, and white fruit flesh colors within the species. It has also been established that this *CmPPR1* is possibly involved in plastid-to-nucleus retrograde signaling, thereby affecting the expression of plastid-targeted genes, indicating the involvement of PPR proteins in fruit flesh color variation [27].

Watermelon (*Citrullus lanatus*) is an important fruit crop with overall annual production of more than 103 million tons worldwide (http://www.fao.org/statistics/en/. Watermelon exhibits diverse variation in fruit-quality traits, including soluble sugars, firmness, fruit size, shape, skin, and flesh color, along with functional factors such as lycopene and β-carotene [28,29]. This extensive polymorphic variation motivates researchers to investigate the genetics of watermelon fruit-quality traits. PPR

genes related to fruit development and flesh color variation in watermelon have not been studied yet. Fortunately, the recently released watermelon (97103 v2) genome sequence (http://cucurbitgenomics.org) provides an excellent opportunity to perform a genome-wide analysis of important gene families, including the PPR gene family in watermelon. In the present study, we identified and characterized 464 putative *PPR* genes from 97103 watermelon genome. Furthermore, we investigated their intron-exon organization, chromosomal localization, types of PPR motifs, functional diversification, subcellular locations, and phylogenetic analysis. With a focus on the involvement of PPR genes in fruit development and flesh color variation in watermelon, we also examined their expression patterns through several RNA-seq analyses from the cucurbit expression atlas (http://cucurbitgenomics.org/rnaseq/home). Finally, diagnostic SNP-CAPS markers of PPR genes were developed to study their association with fruit flesh color variation. Thus, the findings of this study will contribute to the understanding of PPR gene distribution and functions in watermelon, and also improve our understanding of the relationship between PPR genes and flesh color variations.

2. Materials and Methods

2.1. Plant Materials

All the watermelon lines used in this study were obtained from domestic seed companies (Hyundai Seed Company, Gyeonggi-do, South Korea) in Republic of Korea. A total of 70 lines with red (33), yellow (17), and orange (20) flesh colors were used in this study (Table S1). Seeds of the lines were sown in 72-cell polyethylene flats and cultivated under greenhouse at 25 and 20 °C under 16 and 8 h light and dark conditions, respectively, until the appearance of second and third true leaves. Thereafter, the leaf samples were collected and the genomic DNA from the leaves were isolated using the WizPrep™ Plant DNA Mini Kit (Wizbiosolutions, Seongnam, South Korea).

2.2. Sequence Retrial and Identification of the PPR Family Members in Watermelon

To retrieve PPR genes in watermelon, the PPR motif "PF01535" from the Pfam (http://pfam.sanger.ac.uk/) database was used to BlastP searches against Watermelon *Citrullus lanatus* subsp. *vulgaris* cv. 97103 protein (version 2) sequences on the cucurbit genomics database (CuGenDB; http://cucurbitgenomics.org). Additionally, a BlastP search was also investigated with Arabidopsis and rice PPR proteins [1,7]. As queries (e-value set at1×10^{-5}) against watermelon (version 2) protein sequences. Apart from Blastp analysis, '*Pentatrichopeptide repeat*' was used as a keyword in a functional annotation search in the genome of Watermelon (97103) version 2 at the CuGenDB. After combining the results from above searches, redundant sequences were removed. The non-redundant PPR proteins in watermelon with the presence of PPR motif with confidence (E-value < 0.1) in SMART (http://smart.embl-heidelberg.de/) were taken into further analysis. To analyze the protein structure and PPR motif types in translated PPR protein sequences, the HMMsearch program from the HMMER package [30] was used to classify P or PLS (E1, E2, E+, and DYW). The PPR proteins with zero or less than 2 P motifs were excluded from further analysis following reports of previous studies on rice and cotton [7,31]. Finally, the identified candidate PPR genes were named as *Citrullus lanatus* PPR (*ClaPPR*).

2.3. Chromosomal Locations, Genomic Distribution, Exons/intron Organization, and Synteny Analysis

Information on accession number, chromosomal locations, CDS, and protein sequences of each non-redundant PPR gene of watermelon 97103 (version 2) were finally retrieved from the cucurbit genomics database (CuGenDB; http://cucurbitgenomics.org). The identified *ClaPPR* genes were mapped proportionally on watermelon chromosomes using CIRCOS image software [32]. The exon/intron organization of *ClaPPR* genes were predicted and investigated using the Gene Structure Display Server (GSDS2.0; http://gsds.cbi.pku.edu.cn/). To analyze the gene duplication and syntenic relationship of *ClaPPR* genes, the multiple Collinearity Scan toolkit (MCScanX) was employed with default parameters as previously reported [33]. Watermelon 97103 (version 2) genes were classified into

segmental or tandem duplication types using an all-against-all BLASTP comparison (e-value ≤1 × 10^{-10}). Putative segmentally duplicated gene pairs of *ClaPPR* in watermelon genome were visualized in CIRCOS. To evaluate the selection pressure on duplicated *ClaPPR* genes, the rates of non-synonymous (Ka) and synonymous substitution (Ks) were determined using the PAL2NAL web program (http://www.bork.embl.de/pal2nal/).

2.4. Gene Ontology (GO), Motif Identification, and Subcellular Location Prediction

The *ClaPPR* genes were Blast searched against the Arabidopsis genome (https://www.arabidopsis.org/Blast/cereon.jsp), and corresponding Arabidopsis homolog accessions (selected at an E-value of1 × 10^{-10}) for each *ClaPPR* were retrieved. GO annotations were performed for the *ClaPPR* using the Arabidopsis accessions. AgriGO web-based tool (v1.2)153 was used for gene ontology (GO) enrichment analysis ($p < 0.05$) of *ClaPPR* genes (http://systemsbiology.cau.edu.cn/agriGOv2/index.php) [34]. The conserved motifs among *ClaPPR* proteins were investigated with the following parameters: motif width between 13–50 residues, maximum number of 25 motifs, and remain parameters at default, using the MEME (Multiple Em for Motif Elicitation) software version 5.5.1 (http://meme-suite.org) [35]. The subcellular distribution of *ClaPPR* proteins were predicted using TargetP 2.0 (https://services.healthtech.dtu.dk/service.php?TargetP-2.0) and Predotarv.1.04 [36] with default parameters.

2.5. Phylogenetic Analysis of PPR Proteins

A total of 422 PPR protein sequences from watermelon and 44 typical PPR proteins from Arabidopsis, which utilized were in a previous study [3], were used to build a phylogenetic evolutionary tree. The translated sequences of the whole coding regions of PPR proteins were aligned using the MUSCLE method. The tree was constructed using the neighbor-joining (NJ) method with MEGA X software [37] and bootstrap analysis of 1000 replicates.

2.6. Expression Pattern of ClaPPR Genes in Watermelon

To study the expression patterns of *ClaPPR* genes in 97103 watermelon, the transcriptome RNA-seq data (BioProject: SRP012849) from the cucurbit expression atlas (http://cucurbitgenomics.org/rnaseq/home) were used [38]. These RNA-seq data contained data on fruit flesh (FF) and fruit rind (FR) at four stages (10, 18, 26, and 34 days after pollination (DAP)) during the development of the watermelon cultivar 97103. Furthermore, to examine the contribution and regulatory roles of *ClaPPR* genes, which are involved in fruit ripening, among different flesh-colored watermelons, the comparative watermelon transcript data (BioProject: PRJNA338036), containing five different fruit ripening stages (10, 18, 26, 34, and 42 DAP) of a pale yellow-flesh cultivated watermelon ('COS') and red-flesh cultivated watermelon ('LSW-177'), were used. DEG (differentially expressed gene) values for corresponding *ClaPPR* were obtained using the list of *ClaPPR* accession numbers. The obtained DEG values were log2 base transformed. The visualization of expression in heatmap and hierarchical clustering was investigated using Cluster 3.0 and TreeView software [39].

2.7. Identification of Single Nucleotide Polymorphisms (SNPs) for ClaPPR Genes and Match Rate Analysis with Flesh Color

Single-nucleotide polymorphisms (SNPs) for *ClaPPR* genes in different flesh-colored (red, yellow, and orange) watermelons were identified from the whole genome resequencing (WGRS) data (Bioproject: PRJNA516776) of our recent study [29]. The SNP variant calling procedure was carried out according to our recent study [29] in which the reads of different flesh-colored watermelons from WGRS were mapped on to the watermelon 97103 reference genome. From this SNP variant calling, we searched for the SNPs that were specific to either of the flesh types such as red, yellow, and orange in the SNP subset matrix by identifying protein-coding *ClaPPR* genes bearing SNPs that were 1) monomorphic among a chosen flesh color-type, 2) monomorphic among other unchosen flesh color-types, and 3) polymorphic

between a chosen and an unchosen flesh color-types. Then, the sequences of selected *ClaPPR* SNP variants were converted into CAPS markers using SNP2CAPS software [40] and simultaneously, primers were designed using Primer3Plus software (http://www.bioinformatics.nl/primer3plus). To validate CAPS markers, genomic DNA extraction, PCR assays, restriction digestion of PCR amplicons, and match rate analyses were carried out according to a previous study [29].

3. Results

3.1. Genome-Wide Identification, Classification, and Conserved Motif Analysis of PPR Genes in Watermelon

Genome-wide search analysis revealed that there were 464 putative PPR genes present in watermelon genome (97103 v2). After analyzing the domain and P motif patterns, a total of 422 PPR genes were predicted and identified in this study. These watermelon PPRs were designated as *Citrullus lanatus* PPR (*ClaPPR*) from *ClaPPR1* to *ClaPPR422* proteins in the order of their chromosomal position and accession number in the CuGenDB database (http://cucurbitgenomics.org). The chromosomal location, accession number, length of ORF and protein, and number of introns are listed in Table S2. The identified number of *ClaPPR* genes on each chromosome ranged from 20 to 56. (Figure 1A). The determined exon-intron organization of 422 *ClaPPR* genes showed that 71.8% of them were intronless (303/422), while the remaining were with 1 intron (14.7%), 2–5 introns (9.2), and ≥6 introns (4.3%) as shown in Figure 1B. Analysis of repeated motifs structures in *ClaPPRs* indicated that they could be classified into P and PLS subfamilies, containing approximately equal number of genes, representing 46.4% (197 of 422) and 53.6% (225 of 422) PPR proteins, respectively (Figure 1C). In the PLS subfamily, DYW and E2 subgroups both accounted for almost half of the genes (representing 100 of 197 and 97 of 197, respectively), followed by PLS (15), E+ (10), and E1 (3) with the least number of *ClaPPR* genes (Figure 1C). Based on the tandem array of PPR motifs, in watermelon, the estimated number of PPR motifs per protein characterized was 3–27 motifs. A basic motif organization of several typical *ClaPPR* proteins representing the subfamily and subgroups are shown in Figure 1D. A strong peak was noted in the distribution at around 7–12 and 13–17 PPR motifs in P- and PLS-class proteins of watermelon, respectively (Table S2).

Analysis of conserved motifs in PPR proteins have been suggested to rule out the common molecular functions of PPR genes in different subgroups [6]. Therefore, we investigated the conserved motifs in *ClaPPR* proteins using MEME Suite (Figure S1), and results indicated that there were 25 motifs in the 422 *ClaPPR* proteins (Figure S2). Almost all of the *ClaPPR* proteins contained 16 of the 25 motifs except these nine motifs (Motifs 3, 5, 7, 13, 17, 20, 21, 24, and 25), indicating that these *ClaPPR* proteins might have a conserved domain. In addition, the majority of these motifs were analyzed mostly in P, DYW, and E2 (197, 100, and 97 *ClaPPR* genes, respectively) subgroups because they were more dominant in total number than others: E+, E1, and PLS were least in number (Figure 1C). We also found out that the different subgroups possessed specific motifs; for example, motif 21 and 25 exist only in the P subfamily. In the PLS subfamily, the DYW subgroup mainly contained motif 5, 13, and 20. Similarly, motif 24 was mainly present in the E2 subgroup. Some motifs were found to be conserved in two subgroups, such as motif 3, 7, and 17, which mainly exist in DYW and E2 subgroups (Figure S2).

Duplication events of whole genome (tandem) or segmental had been portrayed as a major factor responsible for the expansion of a gene in gene families, including PPR and their subsequent evolution in plants [5,41]. Therefore, we investigated gene duplication events to determine the expansion mechanism of the *ClaPPR* members in watermelon. A total of 11 segmentally duplicated *ClaPPR* gene pairs (5 from the P subfamily and 6 from the PLS subfamily) were identified in the watermelon genome (Table S3). All the gene pairs were inter-chromosomal, involving two different chromosomes (Figure 2). Further analysis showed that the PLS subfamily consisted of three special duplicated gene pairs involving different subgroups: *ClaPPR205*, *ClaPPR264*, and *ClaPPR221*, which belong to DYW and *ClaPPR366* (E2), *ClaPPR307* (E+), and *ClaPPR286* (E2) of the E-subgroups (Table S3). In addition, the calculated ratio of non-synonymous (Ka) and synonymous (Ks) substitution ratios (Ka/Ks or ω) for

these 11 duplicated gene-pairs were found to be $\omega < 1$, suggesting that these duplicated *ClaPPR* gene pairs were under purifying selection.

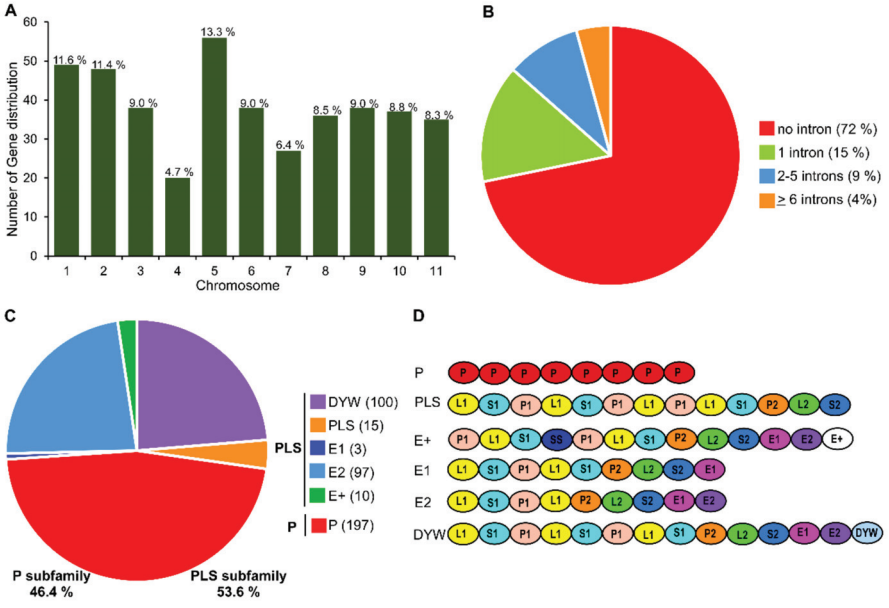


Figure 1. Number, distribution, and structures of *Pentatricopeptide-repeat* (*PPR*) genes in watermelon. (**A**) Number of *ClaPPR* genes in each chromosome. (**B**) Number of introns in *ClaPPR* genes. (**C**) Number of *ClaPPR* proteins belonging to the P subfamily and PPR-like long and short (PLS) subfamily with subgroups. (**D**) Typical motif structures of *ClaPPR* proteins from different subfamilies and subgroups.

3.2. Chromosomal Distribution and Duplication of PPR Members in Watermelon

To investigate the chromosomal distribution of *ClaPPR* genes, the detailed position of *ClaPPR1–422* genes on watermelon (97103 v2) chromosomes were obtained from the CuGenDB database. The results showed that the identified 422 *ClaPPR* genes were distributed unevenly and widely in all the 12 chromosomes; for example, chromosome 5 and 4 were found to have the largest and fewest *ClaPPR* genes at 13.3% and 4.7%, respectively (Figures 1A and 2). PPR genes usually appear in clusters or individually on chromosome [3,5]. In the present study, most of the *ClaPPR* genes clustered together either proximally or distally with very few *ClaPPR* genes positioned in the pericentromeric region of the chromosomes, indicating gene duplications during evolution.

3.3. Phylogenetic Analysis of PPR Members in Watermelon

In order to determine the evolutionary relationships among the *ClaPPR* family members, we constructed a phylogenetic tree based on the deduced full-length amino acid sequences of the 422 ClaPPR proteins along with the 48 PPR proteins from Arabidopsis. As expected, the tree was divided into two distinct clusters: one containing the P subfamily and the other containing the PLS subfamily (Figure 3). Interestingly, the PLS subfamily member *ClaPPR53* was clustered into the P subfamily members; similarly, several P subfamily members, including *ClaPPR338*, *ClaPPR368*, and *ClaPPR394*, were clustered into the PLS subfamily members regardless of the corresponding structure of their repeated motifs. This finding is consistent with the findings from the phylogenetic analysis of PPR proteins in rice and poplar, where P or PLS subfamily members of PPR proteins were found to be clustered into their opposite members [3,7]. These deviations in clustering could be explained by

the shared structural similarities of the C-terminal motifs between P and PLS members which might have arisen via duplication of PPR motif coding regions during evolution of the aforementioned plant species including watermelon.

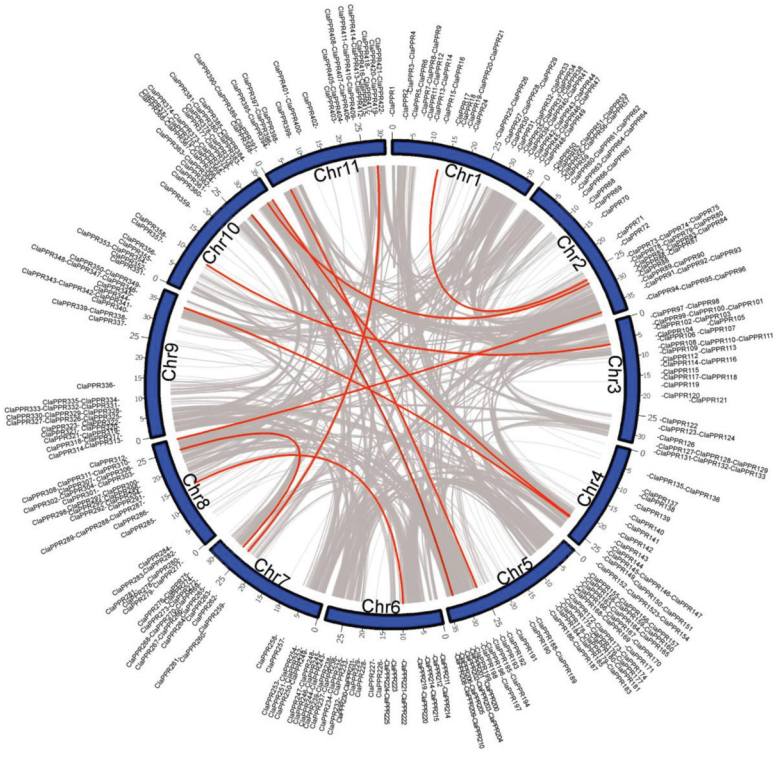


Figure 2. Putative chromosomal localization and gene duplication of *ClaPPRs* in watermelon. Collinear blocks in whole watermelon genome are indicated by grey lines, while the distributions of segmentally duplicated *ClaPPR* pairs are connected with red lines.

3.4. Predicted Subcellular Localization of PPR Proteins

Increasing molecular evidence suggested that PPR proteins play a pivotal role in RNA editing of transcripts in mitochondria and chloroplast organelles. Therefore, we determined the subcellular location of *ClaPPR* proteins in watermelon using TragetP2.0 and Predotar v.1.04 programs. The TragetP2.0 results showed that approximately 32% *ClaPPR* proteins were targeted to mitochondria and 6% to chloroplast, whereas Predotar v.1.04 results showed that approximately 44% and 16% were targeted to mitochondria and chloroplast, respectively (Table S4). Combining the results from both programs, we were able to predict that approximately 65% of *ClaPPR* proteins were targeted to the organelles of chloroplast (73 of 422) and mitochondria (204 of 422); however, few proteins were found to be targeted to ER (5%), and the remaining (30%) protein distributions were uncertain. In the P subfamily, almost half of the proteins (106 of 196) were predicted to be in the mitochondria and 18% (36 of 197) were in the chloroplast (Figure S3). Similarly, in the PLS subfamily, the DYW, E2, and E+ subgroup proteins had a similar localization with almost half in the mitochondria (43–50%) and some proportions in the chloroplasts (19%, 10%, and 30%, respectively). As in the case of PLS and E1, 47% and 67% were predicted in uncertain and mitochondrial localizations, respectively, and approximately 33% of both PLS and E1 proteins were targeted in the chloroplast and ER, respectively (Figure S3).

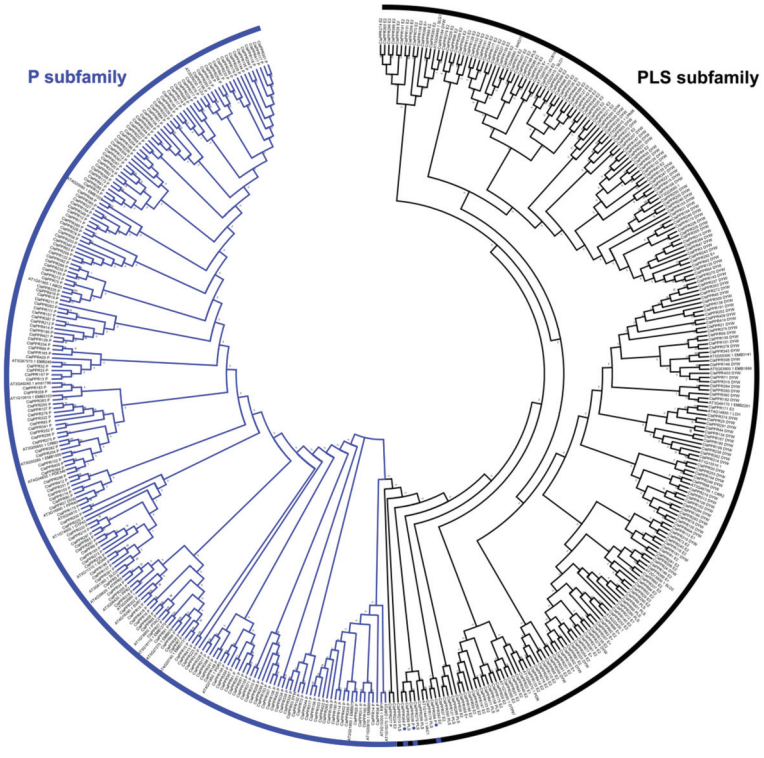


Figure 3. Phylogenetic relationships among the *ClaPPR* family genes. The full coding amino acid sequences of 422 *ClaPPR* proteins and 44 PPR proteins from Arabidopsis were aligned, and the NJ tree was built with 1000 bootstrap replicates using MEGA7.0. P; PLS subfamilies are represented using blue and black lines, respectively. P subfamily members which were clustered into the PLS subfamily members are indicated by a dot (blue) symbol.

3.5. Gene Ontology (GO) Annotation of ClaPPR Genes

To elucidate the role of PPR genes in watermelon, GO annotations were performed for *ClaPPR*s. The results suggested that 364 of the 422 *ClaPPR* transcripts participated in biological processes (82.69%), cellular components (92.85%), and molecular functions (55.22%) (Figure 4). Further insights into the functional categorization indicated that a large portion of *ClaPPR*s were likely related to metabolic processes (179), followed by nucleobase-containing compound metabolic (168), unknown biological (91), and other cellular (50) processes (Figure 4A; Table S5). A total of 191 and 136 *ClaPPR* genes were found to be targeted to mitochondria and chloroplast, respectively; 146 to other intracellular components, 37 to nucleus, and 19 to plastids (Figure 4B; Table S5). For molecular functions, a total of 130 *ClaPPR* genes showed putative participation in binding functions such as protein (28), RNA (15), and DNA binding (5) and other bindings (82). We also found out that several *ClaPPRs* were predicted to be involved in activities, including transferase (15), catalytic (11), transporter (5), kinase (5), hydrolase (7), and unknown molecular functions (69) (Figure 4C; Table S5). In addition, GO enrichment analysis using AgriGO [34] also provided similar results as GO annotation. In the biological process category, all the *ClaPPR* families enriched for RNA modification (GO: 0009451) (Figure S4A). Among the molecular function, the binding functions such as zinc ion (GO: 0008270), translational metal ion (GO: 0043169), cation (GO: 0043169), and protein (GO: 0005515) binding were the enriched category (Figure S4B).

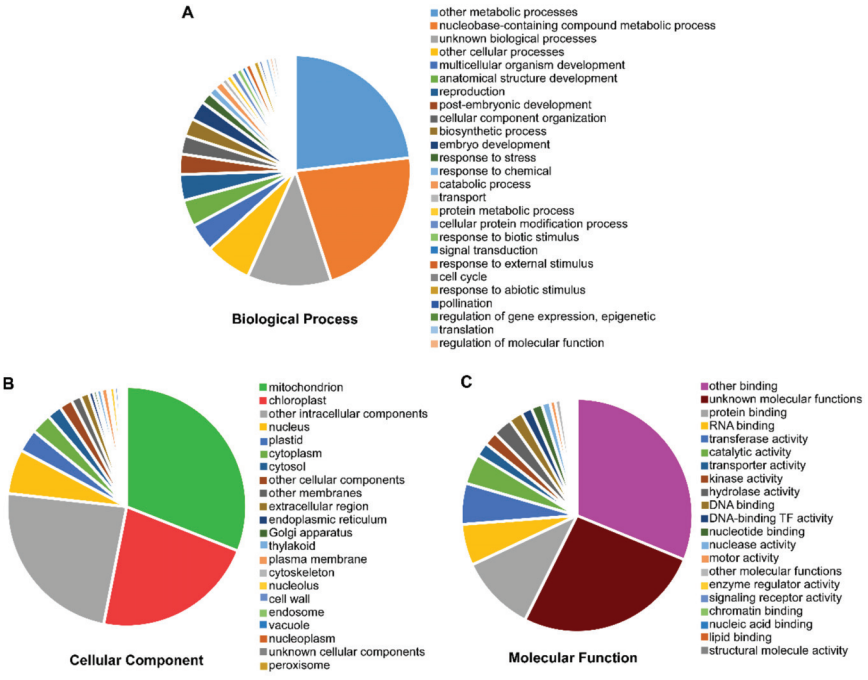


Figure 4. Functional annotation of ClaPPR proteins by Gene Ontology (GO) analysis. According to the GO annotation, the *ClaPPR* proteins were annotated into functional categories of (**A**) biological process, (**B**), cellular component, and (**C**) molecular function.

3.6. Expression Profiles of ClaPPR Genes in Different Stages of Fruit Development in 97103 Watermelon

Different PPR members may exhibit variations in the levels of mRNA accumulation among different tissues during the physiological processes of plants. To explore the putative biological functions of PPR members in watermelon, expression profiles of P- and PLS-subfamily members, including their subgroups (PLS, DYW, E1, E2, and E+), were investigated in fruit rind and flesh on different DAP during fruit development of watermelon 97103 using the RNA-seq data from the cucurbit expression atlas. Based on hierarchical clustering and expression heat map (Figure 5), *ClaPPR* genes from each subgroup were differentially expressed in rind and flesh parts during fruit development. However, the majority of the members from each subgroup exhibited preferential accumulation in rind compared to flesh and could therefore be clustered into different expression groups/clades. Based on the expression pattern, the P-subfamily members of 197 *ClaPPR* genes were distributed into seven distinct clades. Expression Clade-II of the P-subfamily includes a total of 27 genes displaying abundant expression at earlier DPAs (10 and 16 DPA) in both rind and flesh; however, expression levels declined at later stages (Figure 5A). The results suggest that these *ClaPPR* genes might be involved in the early stages of each tissue development (Figure 5A). Clade-III is comprised of 13 genes (*ClaPPR334*, *ClaPPR304*, *ClaPPR341*, *ClaPPR331*, *ClaPPR294*, *ClaPPR377*, *ClaPPR277*, *ClaPPR102*, *ClaPPR222*, *ClaPPR13*, *ClaPPR391*, *ClaPPR228*, and *ClaPPR32*), and was strongly upregulated in the flesh at almost all stages. Clade IV comprising 22 genes showed preferential accumulation in flesh; however, in the rind, the genes showed significant expression at only 10 DAP. Clade V, which is the largest one with 81 genes, displayed expression abundance only in the rind at almost all stages, suggesting that these genes could have a functional role in rind development. Clade VI and VII contain 31 and 11 genes (*ClaPPR421*, *322*, *333*, *364*, *27*, *183*, *285*, *263*, *160*, *298*, *119*, and *16*) with transcript accumulation

majorly in the rind at all stages, however, their expression in the flesh was higher in early (10 DPA) and later (26 and 34 DPA) stages, respectively (Figure 5A).

Analysis of the expression patterns of other subgroups in the PLS subfamily indicated that clade I of DYW (47 genes), clade II (13 genes) and III (31 genes) of E2 and clade I of E+ (6 genes) of *ClaPPR* members showed significantly higher expression levels in the rind tissues of both stages compare to those in the flesh, where their expressions were only on early 10 DPA (Figure 5B). Similarly, in the DYW subgroup, some of the genes in clades II and III displayed upregulated expression patterns in the rind (10–34 DPA) and flesh (26–34 DPA). Flesh-specific expression of *ClaPPR* genes were also noted in clade IV of DYW (*ClaPPR330*, *143*, *240*, *31*, *185*, *98*, *55*, *168*, *226*, *189*, *313*, and *191*), where these genes were highly expressed at 26–34 DPA of flesh. Clade-I from the E2 subgroup, which comprises 13 genes, showed an up-regulated expression pattern in flesh tissues at 26 DPA (*ClaPPR281*, *216*, *171*, *139*, *320*, *230*, *201*, *37*, and *271*) and 34 DPA (*ClaPPR201*, *37*, *271*, and *131*) and displayed higher expressions in rind at 10 DPA (Figure 5C). Furthermore, clade IV of E2 (containing 29 genes) also possessed few genes (*ClaPPR321*, *178*, *81 50*, *242*, *395*, and *85*) that responded highly in flesh at 34 DPA. However, in the PLS-subgroup, a total of 7 genes (*ClaPPR401*, *90*, *280*, *350*, *121*, *179*, and *411*) were found to be more expressed in the rind than in the flesh at all DPA stages in clade II (Figure 5D). E1 and E+, which are smaller subgroups with only 3 and 10 *ClaPPR* members, respectively, showed a relatively higher expression in the rind (clade I of each) than in the flesh; however, only a few genes were expressed at a higher level in the flesh (*ClaPPR312* in clade of E1; *ClaPPR52*, *307*, and *395* in clade two E+) (Figure 5E,F). The result of the analysis indicates that *ClaPPR* genes in each subgroup showed high expression levels in both rind and flesh at all stages or particular stages of DPA; this facilitates the preliminary understanding of their possible participation in watermelon fruit development.

3.7. Sequence Variation in ClaPPR Genes and Development of CAPS Markers for Flesh Color

For the utilization of *ClaPPR* genes in watermelon breeding, we investigated the relationship between the watermelon flesh color and sequence variations in *ClaPPR* genes. We identified the SNPs in the sequences of *ClaPPR* encoding genes from 24 different flesh colored watermelons (red, yellow, and orange) using our WGRS data (Bioproject: PRJNA516776) [29]. A total of 368 SNPs were observed from 139 *ClaPPR* genes in the WGRS data. After detailed analysis of all SNPs, we selected 4 SNP-carrying genes, including *ClaPPR11*, *ClaPPR25*, *ClaPPR95*, and *ClaPPR140* from 9 red, 9 yellow, and 6 orange flesh-colored watermelons, and compared them with the reference 97103 genome (Table S7). The SNPs in *ClaPPR* genes were found to be monomorphic among a chosen flesh color-type and polymorphic between a chosen and an unchosen flesh color-types; each SNP in the corresponding *ClaPPR* gene almost showed a co-segregation with a particular flesh color phenotype (Orange: *ClaPPR11*, Yellow: *ClaPPR25* and *ClaPPR95*, and Red: *ClaPPR140*). Among the four SNP-carrying genes, 3 of them were classified as non-synonymous substitutions (*ClaPPR11*, *ClaPPR25*, and *ClaPPR140*) with altered amino acid residue in the PPR motifs, which could cause functional variation, in those corresponding genes, between a desired and non-desired flesh types. Therefore, to detect the association among the four SNP-carrying *ClaPPR* genes and flesh colors, CAPS marker primer sets were designed and analyzed by restriction digestion. Using these four CAPS markers, genotyping was carried out on 70 different commercial cultivars comprising red (33), yellow (17), and orange (20) flesh color for their reliability and applicability on watermelon breeding (Figure 6; Table S7). The genotyping results for flesh color determination, based on the SNPs of the four *ClaPPR* genes, are described in Table S8. *ClaPPR11* showed a higher match rate of 0.87 (87%) among genotypes of the markers and phenotypes of the orange-flesh color in all surveyed lines. However, *ClaPPR140* had a perfect co-segregation with red-flesh color with a match rate of 1 (100%). With regard to genotyping for *ClaPPR25* and *ClaPPR95*, they co-segregated well with yellow-flesh, exhibiting a match rate of 0.79 and 0.76, respectively (Table S8). Furthermore, a joint, *ClaPPR25* + *ClaPPR95* genotyping provided an average match rate of 0.94, indicating a high reliability and applicability of flesh type specific *ClaPPR* gene-based SNPs identified in this study.

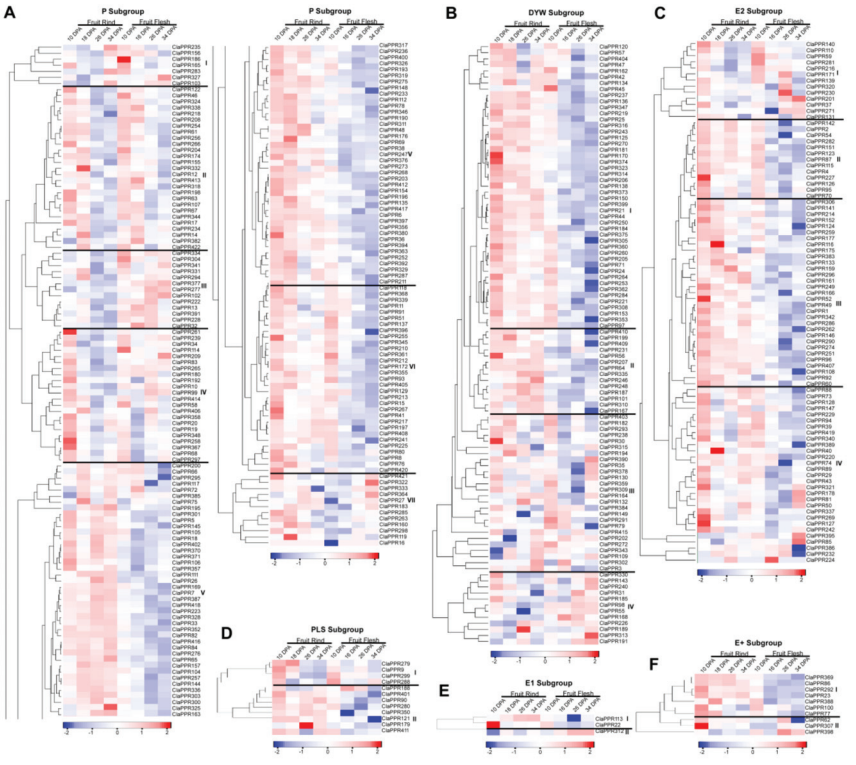


Figure 5. Expression profiles of *ClaPPR* genes during watermelon fruit development (Cultivar 97103; BioProject: SRP012849). All heat maps showing various expression levels of *ClaPPR* genes and subdivided into various clusters (labeled as roman numerals) were built using log2- transformed FPKM values of fruit rind and flesh at the developmental stages of 10, 18, 26, and 34 days after pollination (DAP). (**A**) Expression profiles of *ClaPPR* genes in the P subfamily, (**B**–**F**) Expression profiles of *ClaPPR* genes in the PLS subfamily and subgroups (**B**) DYW, (**C**) E2, (**D**) PLS, (**E**) E1, and (**F**) E+. Differences in transcript abundances such as high (red) and low (blue) levels are shown in color as the scale bar of *Z*-score.

3.8. Comparative Expression Patterns of ClaPPR Genes under Different Fruit Ripening Stages of Two Cultivated Watermelon Varieties with Red- and Pale-Yellow-Fleshed

To explore the role of *PPR*s in fruit ripening of different flesh-colored watermelons, we investigated the comparative expression profiles of *ClaPPR* genes in fruit growth and ripening stages, including immature white (10 DPA), white-pink (18 DPA), red (26 DPA), and full-ripe (34 and 42 DPA), of two cultivated watermelons differing in flesh color, 'COS' (pale yellow-flesh) and 'LSW-177' (red-flesh) (Table S6) (BioProject: SRP012849). Almost all of the 422 *ClaPPR* genes showed expression in at least one of the DPA stages of the two watermelon varieties. It was noted that *ClaPPR* genes exhibited preferential and stage-specific expression between varieties. In the P-subfamily, most of the *ClaPPR* members showed a uniform upregulated expression in both COS and LSW177 at 10 DPA. However, at 26 DPA, these genes apparently exhibited abundant expression in LSW177 than in COS (Figure S5A). At 10 DPA, almost all of the *ClaPPR*s genes in both DYW and E2 subgroups showed significantly upregulated expression in the red flesh of LSW177 compared to that in COS. At 18 DPA, these genes exhibited robust, uniform upregulated expression patterns in both COS and LSW177, but relative to the remaining DPA stages (Figure S5B–C). The PLS subgroup members showed relatively similar

expression levels in both COS and LSW177 at 10 DPA; and a higher level in LSW177 than in COS at 26 DPA as observed in the P-subgroup (Figure S5D). Both E1 and E+ subgroup members exhibited significant expression in LSW117 than inCOS at 10 DPA, while their expression at 18 DPA were almost uniform between LSW117 and COS as observed in DYW and E2 subgroups (Figure S5E–F). At the full-ripe stages (34 and 42 DPA), only the P-subfamily exhibited significant expression; some of the *ClaPPR* members had high transcript accumulation in COS than in LSW117, particularly at 34 DPA (Figure S5A and D). These results suggest that *ClaPPR* genes can be considered as candidate genes that are associated with growth and ripening of watermelon fruits.

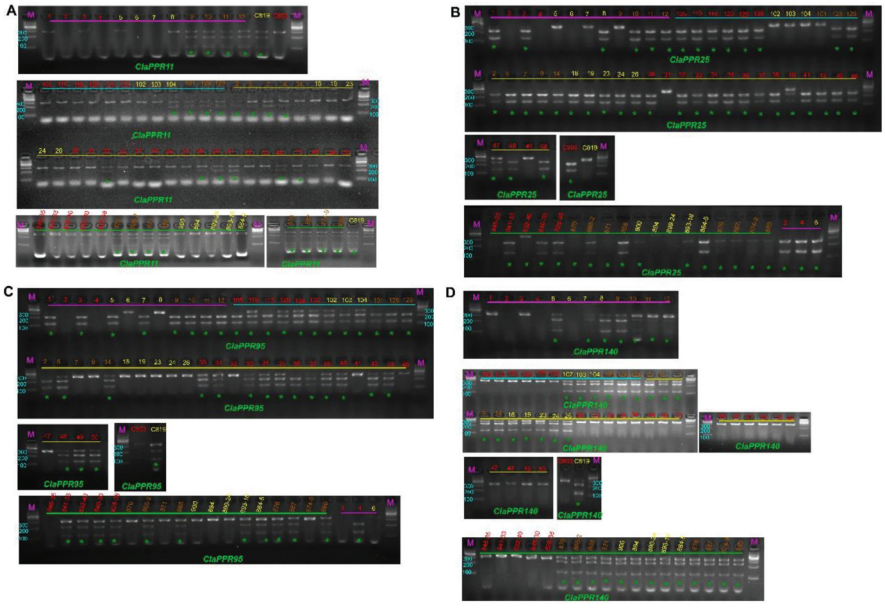


Figure 6. Gel pictures for the validated cleaved amplified polymorphic sequence (CAPS) markers based on single-nucleotide polymorphisms (SNPs) in *PPR* genes. The name of validated CAPS markers such as *ClaPPR11* (**A**), *ClaPPR25* (**B**), *ClaPPR95* (**C**), and *ClaPPR140* (**D**) are shown below their corresponding gel pictures. Numbers in lanes underlined with different colors indicates the DNA sample names of the 70 lines, same as those listed in Table S7. M represents a 100 bp ladder. A "*green asterisk*" represents the enzyme-cleavage of the DNA samples. DNA samples of C803 (red-flesh) and C819 (yellow-flesh) were used as control (Table S7) during the restriction digestion of PCR amplicons.

4. Discussion

Watermelon (*Citrullus lanatus*) is one of the important economic crops in the *Cucurbitaceae* family. Fruits of watermelon contain sugars, carotenoids (lycopene, beta-carotene, and phytoene), and various health-promoting nutritional compounds (glutathione, citrulline, and arginine) which significantly contributes to the human diet [28]. With a relatively compact genetic complement (~425 Mbp), the gene families of watermelon are being investigated owing to their sequenced or re-sequenced draft genomes [29,42]. PPR proteins are one of the largest gene families in terrestrial plants. In the present study, 422 PPR proteins, of which 197 and 225 members belonged to the P subfamily and PLS subfamily, respectively, were identified in the 97103 watermelon genome (Figure 1; Table S2). This number of proteins is in accordance with previous studies that reported the presence of >400 PPR genes per plant genome, including *Arabidopsis*, foxtail millet, maize, and rice [1,5–7]. Analysis of gene structure revealed that more than 70% of the *ClaPPR* genes were intronless, indicating that a majority of the plant

PPR genes (for example, 80% of *PPR* genes in Arabidopsis) lack introns [3,5,7], as it could be the result of amplification by retrotransposition events, in which intron-poor genes might have originated from intron-rich PPRs [1,43,44].

Although analysis of conserved protein motif analyses did not reveal the same results as that for the motifs used to distinguish different types of PPR proteins, it can be used to determine the conserved molecular functions in P and PLS subfamilies of PPR genes [6]. Conserved motif analysis showed that 25 motifs are present in *ClaPPR* proteins; motif 21 and 25 are present only in the P subfamily; the PLS subfamily of the DYW subgroup only exhibits motif 5, 13, and 20; and both DYW and E2 share motifs 3, 7, and 17 (Figure S2). This type of motif distribution has also been observed for PPR proteins in maize where some motifs were even conserved between their genomes [6]. These identified subgroup-specific motifs could be a significant component of the corresponding *ClaPPR* genes in different subgroups that may determine the conserved molecular functions. However, extensive future studies on the characterization of these *ClaPPRs* are required to elucidate their conserved functions. The *ClPPR*s were unevenly distributed on each chromosome, and often clustered together in short regions of the chromosomes (Figures 1A and 2). These results indicated that the size of chromosomes was not relatively associated with the number of genes [31] and that duplication events could have resulted in the expansion of these genes as suggested in a previous study [5]. In duplicated *ClPPR* gene analysis, we noted a total of 11 segmentally duplicated genes located on each of all the chromosomes (Figure 2; Table S3), suggesting that segmental duplication is the most prevalent, having a higher frequency than tandem duplication events in watermelon, and correspond to that reported in previous genome-wide studies [45,46]. In our phylogenetic analysis, *ClaPPR* proteins, based on the pattern of PPR motifs, can be classified into two groups of proteins: P subfamily and PLS subfamily (Figure 3). It has been observed that several P subfamily proteins clustered together with PLS subfamily proteins, showing similarity in the evolutionary relationship of PPR proteins in poplar and rice [3,7].

PPR proteins have been predominantly predicted to be located in the mitochondria and plastids [7]. In the present study, most of the identified *ClaPPR* proteins (approximately 65%) were predicted to be commonly localized in subcellular regions of the chloroplast (73 of 422) and mitochondria (204 of 422) (Figure S3). GO analysis also indicated that many *ClaPPR* proteins seem to be located in the mitochondria (191) and chloroplast (136) (Figure 4). PPR proteins modulate gene expression via post-transcriptional or translational regulations in organelles at the RNA level by acting as RNA-binding proteins [47]. It has also been observed that a large number of *ClPPRs* had binding functions, including DNA, RNA, and protein binding and other-binding functions (Figure 4 and Figure S4), corroborating the results of PPR protein functions. Therefore, any defects or mutations in organelle-targeted PPR proteins often result in organelle dysfunction, which ultimately leads to altered phenotypes, including cytoplasmic male sterility [24], defective embryo development [12], abnormal photosynthetic ability and aberrant pigmentation in seeds [48], and flesh color variation [27]. Future studies on the functional characterization of *ClaPPR* genes will clarify their significant implications in watermelon breeding.

There is increasing evidence that PPR genes play a significant role in plant growth and developmental process, and their mRNA expression patterns have been explored in cotton floral buds [31], maize kernels [6], and rice panicles [7]. In the present study, we investigated the expression patterns of *ClaPPR* genes in watermelon fruit development (BioProject: SRP012849). The results of in-silico expression analyses indicated that *ClaPPRs* were differentially expressed in the rind and flesh tissues (Figure 5). Watermelon fruits have rapid cell division and expansion in their early fruit development stages, resulting in changes in cell wall structure and accumulation of compounds (carbohydrates and organic acids) in vacuoles; followed by fruit ripening stages, which involves changes in carotenoid profiles and conversion of carbohydrates to sugars [49]. The expression levels of some of the *ClaPPR* members in the P-subfamily were relatively high in fruit flesh (Clade III), or fruit rind (Clade V), and or both flesh and rind tissues (Clade II, IV, VI, and VII), suggesting that they could be important in correlating the development of rind and flesh in watermelon (Figure 5A). In the PLS-subfamily, most of the subgroups were preferentially expressed in rind with a higher level

at all DPA as well as flesh at 10 DPA (clade I of DYW; clade II, III of E2; and clade I of E+), suggesting that these genes might be important in fruit rind and early fruit flesh development (Figure 5B–F). DYW (clade II, III, and IV) and E2 (clade IV) members presented high expression levels in the fruit flesh at the late ripening stages (26 and 34 DAP), indicating that these genes might be essential for carotenoid accumulation and fruit ripening of watermelon (Figure 5B–C). These results imply that *ClaPPR* genes might be involved in watermelon fruit development and chloroplast-to-chromoplast transition.

Fruit flesh color is an important trait of watermelon; variations in carotenoid profiles often result in colors of red (lycopene), yellow (phytoene), and orange (β-carotene) fleshes [29]. Fruit ripening has been reported to be influenced by environmental factors, hormones, and developmental gene regulation [49,50]. Therefore, identification and characterization of genes, which govern fruit growth and ripening, would be helpful in watermelon breeding. The present study explored the expression profiles of *ClaPPR* genes to evaluate the possible roles of *ClaPPR* in fruit growth and ripening stages (10–42 DPA), between 'COS' (pale yellow-flesh) and 'LSW-177' (red-flesh) watermelons (BioProject: PRJNA338036). At 10 and 18 DPA, *ClaPPR* genes from various subgroups were found to be generally upregulated in both COS and LSW177 (Figure S5). However, DYW, E2, and E+ subgroup genes displayed robust upregulated expression of their transcripts in LSW177 than in COS. Similarly, both P- and PLS- subgroups were also upregulated only in LSW177 at 26 DPA. In contrast, these two subgroups showed high transcript accumulation in COS than in LSW117 at full-ripening stages of 34 DPA (Figure S5A,D). From the expression profiles, it was also speculated that the expression of 242 and 226 *ClaPPR* genes were upregulated at all stages in LSW177 and COS, respectively, among which 60 and 52 genes showed high expression levels (a log2 value between 3–5) (Figure S5; Table S6). This indicated that LSW177 had 1.07 (242/226 = 1.07) fold higher number of upregulated *ClaPPR* genes than that of COS. Therefore, it appears that the mechanism of fruit growth and ripening in LSW177 is more complex than that in COS, and that the PPR family also has functional involvement in the growth and ripening of fruits. However, further studies are required to elucidate the complete role of these upregulated *ClaPPR*s in watermelon fruit.

mRNA expression of PPR genes have been reported to be regulated by microRNAs (miRNA) through cleavage or translational repression in plants [3,7,51]. In watermelon, a previous study showed that eight PPR genes (Cla008388, Cla012681, Cla015802, Cla018752, Cla011015, Cla005585, Cla019381, and Cla006187) have complementary sites of miRNAs [52]. These miRNAs have been reported to be involved in melatonin-mediated cold tolerance in watermelon by suppressing the expression of the abovementioned PPR genes through either cleavage (gene names as in the present study; *ClaPPR29*:*miR399-5p*, *ClaPPR268*:*miR8029-39*, and *ClaPPR59*:*novel-m0058-5p*), or translational inhibition (*ClaPPR234*:*miR159-5p*, *ClaPPR21*:*miR6284-3p*, *ClaPPR348*:*novel-m0030-5p*, *ClaPPR104*:*novel-m0051-5p*, and *ClaPPR179*:*novel-m0051-5p*) [52]. In addition, a recent study showed that a total of 218 PPR genes have complementary sites of 160 miRNAs [53]. Hence, further research is required on the dynamic expression of miRNAs and their corresponding *ClaPPR* targets, and a crosstalk between miRNAs and PPRs will contribute to the regulation of plant growth and fruit development in watermelon.

Based on the SNPs in *ClaPPR* genes (Table S7), the developed four CAPS were found to perfectly co-segregated with their corresponding flesh colors, and match rate ranged from 0.87 to 1 (Figure 6; Table S8). Notably, *ClaPPR11* co-localized with β-carotene-related QTL on chromosome Chr1 [54], whereas *ClaPPR140* co-localized with lycopene-related QTL on chromosome Chr4 [55]. With regard to *ClaPPR25* and *ClaPPR95*, they were not co-localized with any QTL for flesh color. Therefore, the identified SNPs in these *ClaPPR* genes might be used for fine mapping of flesh color locus in watermelon genome. Few amino acid positions (4th and 34th) in a PPR motif have been found to act as attachment points, which help PPR proteins to binds to target mRNAs, thus inhibiting translation [56]. In Arabidopsis, a single nonsynonymous mutation at the 4th amino acid in the 12th PPR motif inhibited the complete function of a PPR gene called *Proton Gradient Regulation3* [57]. Among the selected SNP-carrying candidate genes, *ClaPPR11* and *ClaPPR140* had nonsynonymous mutation at the 2nd amino acid location in the 13th and 11th motif, respectively, while *ClaPPR25* had nonsynonymous

mutation at the 23rd amino acid location in the 18th motif (Table S7), suggesting that these SNPs could influence the binding action of corresponding *ClaPPR* and therefore, play a role in their functions. Furthermore, the aforementioned *ClaPPR* protein-encoding genes could be considered as important candidates for watermelon fruit related traits and the developed CAPS markers will be helpful for breeders to economically distinguish fruit flesh colors at watermelon seedling stage.

5. Conclusions

In this study, a total of 422 PPR protein genes were identified in the watermelon genome. Based on the PPR motif type, *ClaPPR* genes were divided into five subgroups. Most of the genes were intronless and distributed widely across all watermelon chromosomes, and encoding proteins were targeted to organelles of chloroplast or mitochondria which gives valuable information for future studies on characterization of *ClaPPR* genes. Duplication analyses suggested that 11 segmentally duplicated *ClaPPR* pairs exist in the genome. We conducted an in-silico expression pattern analysis in watermelon fruit rind and flesh tissues. In addition, a comparative expression study was performed in the fruit ripening stages of red- and pale yellow-fleshed watermelons, which provides preliminary understanding about *ClaPPR* participation in fruit development. Based on sequence variation analyses of *ClaPPR* genes, four CAPS markers were developed and found to have co-segregation with distinct flesh types, which could be used to distinguish different flesh colors, including red, yellow, and orange. Taken together, the findings of this study provide comprehensive understanding of the *ClaPPR* gene family and clarify candidate *ClaPPR* genes for functional validation in the future.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/10/1125/s1, Figure S1. Conserved motifs in *ClaPPR* proteins using MEME Suite, Figure S2. Conserved motifs of *ClaPPR* proteins. The consensus sequence of motif logos shown in the left panel. The number of motifs in the different subfamilies and/or subgroups of the *ClaPPR* proteins are shown in the right panel, Figure S3. Subcellular localizations of *ClaPPR* genes, using on-line tools of TargetP 2.0 and Predotarv.1.04. The locations of each *ClaPPR* subfamily and its subgroups in the chloroplast (blue), mitochondria (orange), endoplasmic reticulum (ER) (grey), and other locations (yellow) were determined, Figure S4. GO enrichment analysis of *ClaPPR* genes. (A) Graphical results of biological process. (B) Graphical results of molecular function, Figure S5. A comparative expression profiles of *ClaPPR* genes during watermelon fruit development between red-fleshed (LSW177) and pale-yellow-fleshed cultivated watermelons (BioProject: PRJNA338036). The heat map with clustering (labeled as roman numerals) was generated based on the log2- transformed FPKM value of *ClaPPR* genes at developmental stages of 10, 18, 26, 34, and 42 days after pollination (DAP). (A) Expression profiles of *ClaPPR* genes in the P subfamily, (B–F) Expression profiles of *ClaPPR* genes in the PLS subfamily of subgroups (B) DYW, (C) E2, (D) PLS, (E) E1, and (F) E+. Differences in transcript abundances such as high (red) and low (blue) levels are shown in color as the scale, Table S1. List of watermelon commercial cultivars or inbred lines used in this study for CAPS marker validation and their representative fruit characteristics, Table S2. Detailed information of PPR genes in watermelon genome, Table S3. Segment duplication of PPR genes in watermelon genome, Table S4. Subcellular localization of predicted PPR genes in watermelon, Table S5. Details of GO annotation of *ClaPPR* genes, Table S6. Log2 FPKM values for *ClaPPR* genes in different fruit ripening stages of two cultivated watermelon varieties with red-fleshed (LSW-177) and pale-yellow-fleshed (COS), Table S7. Identification of putative sequence variation (SNPs) between lines with flesh-colored (red, yellow, and orange) watermelons from the WGRS data, Table S8. Summary of result analysis of match rate between validated *ClaPPR* CAPS markers and SNPs in 70 watermelons.

Author Contributions: Conceptualization, S.S. and G.-J.L.; data curation, S.S.; formal analysis, S.S., and G.-J.L.; funding acquisition, Y.-P.L. and G.-J.L.; investigation, L.T., K.L., G.-S.P., H.L., and J.-P.C.; methodology, S.S., C.M., and G.-J.L.; project administration, G.-J.L.; resources, M.-W.P.; software, S.S.; visualization, S.S.; writing—original draft, S.S.; writing—review and editing, C.M. and G.-J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Golden Seed Project (Center for Vegetable Seed Development, No. 213003-05-1-SBW30), Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Basic Science Research Program (NRF-2020R1A2C1015119) through the National Research Foundation (NRF) of Korea funded by the Ministry of Science and ICT.

Acknowledgments: We thank Ian Small, The University of Western Australia, for the technical assistance in classification of PPR genes in this study.

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

References

1. Lurin, C.; Andrés, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyère, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell.* **2004**, *16*, 2089–2103. [CrossRef] [PubMed]
2. Cheng, S.; Gutmann, B.; Zhong, X.; Ye, Y.; Fisher, M.F.; Bai, F.; Castleden, I.; Song, Y.; Song, B.; Huang, J.; et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* **2016**, *85*, 532–547. [CrossRef]
3. 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]
4. Manthey, G.M.; McEwen, J.E. The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *EMBO J.* **1995**, *14*, 16–4031. [CrossRef]
5. Liu, J.M.; Xu, Z.S.; Lu, P.P.; Li, W.W.; Chen, M.; Guo, C.H.; Ma, Y.Z. Genome-wide investigation and expression analyses of the pentatricopeptide repeat protein gene family in foxtail millet. *BMC Genom.* **2016**, *17*, 840. [CrossRef]
6. Chen, L.; Li, Y.X.; Li, C.; Shi, Y.; Song, Y.; Zhang, D.; Li, Y.; Wang, T. Genome-wide analysis of the pentatricopeptide repeat gene family in different maize genomes and its important role in kernel development. *BMC Plant Biol.* **2018**, *18*, 366. [CrossRef]
7. 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]
8. Hayes, M.L.; Dang, K.N.; Diaz, M.F.; Mulligan, R.M. A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. *J. Biol. Chem.* **2015**, *290*, 10136–10142. [CrossRef]
9. Ichinose, M.; Tasaki, E.; Sugita, C.; Sugita, M. A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in *Physcomitrella patens*. *Plant J.* **2012**, *70*, 271–278. [CrossRef]
10. Hao, Y.; Wang, Y.; Wu, M.; Zhu, X.; Teng, X.; Sun, Y.; Zhu, J.; Zhang, Y.; Jing, R.; Lei, J.; et al. The nucleus-localized PPR protein OsNPPR1 is important for mitochondrial function and endosperm development in rice. *J. Exp. Bot.* **2019**, *70*, 4705–4720. [CrossRef]
11. Cushing, D.A.; Forsthoefel, N.R.; Gestaut, D.R.; Vernon, D.M. Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **2005**, *221*, 424–436. [CrossRef] [PubMed]
12. Sun, F.; Xiu, Z.; Jiang, R.; Liu, Y.; Zhang, X.; Yang, Y.Z.; Tan, B.C. The mitochondrial pentatricopeptide repeat protein EMP12 is involved in the splicing of three nad2 introns and seed development in maize. *J. Exp. Bot.* **2018**, *70*, 963–972. [CrossRef] [PubMed]
13. Hsieh, W.Y.; Liao, J.C.; Chang, C.Y.; Harrison, T.; Boucher, C.; Hsieh, M.H. The SLOW GROWTH 3 pentatricopeptide repeat protein is required for the splicing of mitochondrial nad7 intron 2 in Arabidopsis. *Plant Physiol.* **2015**, *168*, 490–501. [CrossRef] [PubMed]
14. Kim, S.R.; Yang, J.I.; Moon, S.; Ryu, C.H.; An, K.; Kim, K.M.; Yim, J.; An, G. Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J.* **2009**, *59*, 738–749. [CrossRef]
15. Xiao, H.; Zhang, Q.; Qin, X.; Xu, Y.; Ni, C.; Huang, J.; Zhu, Y. Rice PPS1 encodes a DYW motif-containing pentatricopeptide repeat protein required for five consecutive RNA-editing sites of nad3 in mitochondria. *New Phytol.* **2018**, *220*, 878–892. [CrossRef]
16. Oren, R.; Ellsworth, D.S.; Johnsen, K.H.; Phillips, N.; Ewers, B.E.; Maier, C.; Schäfer, K.V.; McCarthy, H.; Hendrey, G.; McNulty, S.G.; et al. Soil fertility limits carbon sequestration by forest ecosystems in a CO2-enriched atmosphere. *Nature* **2001**, *411*, 469–472. [CrossRef]
17. Laluk, K.; Abuqamar, S.; Mengiste, T. The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiol.* **2011**, *156*, 2053–2068. [CrossRef]

18. Yuan, H.; Liu, D. Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. *Plant J.* **2012**, *70*, 432–444. [CrossRef]
19. 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]
20. Dahan, J.; Mireau, H. The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. *RNA Biol.* **2013**, *10*, 1469–1476. [CrossRef]
21. Bentolila, S.; Alfonso, A.A.; Hanson, M.R. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic malesterile plants. *Proc. Natl Acad. Sci. USA* **2002**, *99*, 10887–10892. [CrossRef] [PubMed]
22. Koizuka, N.; Imai, R.; Fujimoto, H.; Hayakawa, T.; Kimura, Y.; Kohno-Murase, J.; Sakai, T.; Kawasaki, S.; Imamura, J. Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile kosena radish. *Plant J.* **2003**, *34*, 407–415. [CrossRef] [PubMed]
23. Jo, Y.D.; Ha, Y.; Lee, J.H.; Park, M.; Bergsma, A.C.; Choi, H.I.; Goritschnig, S.; Kloosterman, B.; Van Dijk, P.J.; Choi, D.; et al. Fine mapping of restorer-of-fertility in pepper (*Capsicum annuum* L.) identified a candidate gene encoding a pentatricopeptide repeat (PPR)-containing protein. *Theor. Appl. Genet.* **2016**, *129*, 2003–2017. [CrossRef]
24. Fujii, S.; Suzuki, T.; Giegé, P.; Higashiyama, T.; Koizuka, N.; Shikanai, T. The restorer of- fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. *Plant J.* **2016**, *86*, 504–513. [CrossRef] [PubMed]
25. Eriksson, E.M.; Bovy, A.; Manning, K.; Harrison, L.; Andrews, J.; De Silva, J.; Tucker, G.A.; Seymour, G.B. Effect of the Colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. *Plant Physiol.* **2004**, *136*, 4184–4197. [CrossRef] [PubMed]
26. Pesaresi, P.; Mizzotti, C.; Colombo, M.; Masiero, S. Genetic regulation and structural changes during tomato fruit development and ripening. *Front. Plant Sci.* **2014**, *5*, 124. [CrossRef]
27. Galpaz, N.; Gonda, I.; Shem-Tov, D.; Barad, O.; Tzuri, G.; Lev, S.; Fei, Z.; Xu, Y.; Mao, L.; Jiao, C.; et al. Deciphering genetic factors that determine melon fruit-quality traits using RNA-Seq-based high-resolution QTL and eQTL mapping. *Plant J.* **2018**, *94*, 169–191. [CrossRef]
28. Gusmini, G.; Wehner, T.C. Qualitative inheritance of rind pattern and flesh color in watermelon. *J. Hered.* **2006**, *97*, 177–185. [CrossRef]
29. Subburaj, S.; Lee, K.; Jeon, Y.; Tu, L.; Son, G.; Choi, S.; Lim, Y.P.; McGregor, C.; Lee, G.J. Whole genome resequencing of watermelons to identify single nucleotide polymorphisms related to flesh color and lycopene content. *PLoS ONE* **2019**, *14*, e0223441. [CrossRef]
30. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [CrossRef] [PubMed]
31. Han, Z.; Qin, Y.; Li, X.; Yu, J.; Li, R.; Xing, C.; Song, M.; Wu, J.; Zhang, J. A genome-wide analysis of pentatricopeptide repeat (PPR) protein-encoding genes in four Gossypium species with an emphasis on their expression in floral buds, ovules, and fibers in upland cotton. *Mol. Genet. Genom.* **2020**, *295*, 55–66. [CrossRef] [PubMed]
32. 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]
33. Zhou, Y.; Cheng, Y.; Wan, C.; Li, J.; Yang, Y. Genome-wide characterization and expression analysis of the Dof gene family related to abiotic stress in watermelon. *PeerJ* **2020**, *8*, e8358. [CrossRef] [PubMed]
34. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z. agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W122–W129. [CrossRef] [PubMed]
35. Bailey, T.L.; Bodén, 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]
36. Small, I.; Peeters, N.; Legeai, F.; Lurin, C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **2004**, *4*, 1581–1590. [CrossRef]
37. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
38. Guo, S.; Sun, H.; Zhang, H.; Liu, J.; Ren, Y.; Gong, G.; Jiao, C.; Zheng, Y.; Yang, W.; Fei, Z.; et al. Comparative Transcriptome Analysis of Cultivated and Wild Watermelon during Fruit Development. *PLoS ONE* **2015**, *10*, e0130267. [CrossRef]

39. 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]
40. Thiel, T.; Kota, R.; Grosse, I.; Stein, N.; Graner, A. SNP2CAPS: A SNP and INDEL analysis tool for CAPS marker development. *Nucleic Acids Res.* **2004**, *32*, e5. [CrossRef]
41. Gutmann, B.; Royan, S.; Schallenberg-Rüdinger, M.; Lenz, H.; Castleden, I.R.; McDowell, R.; Vacher, M.A.; Tonti-Filippini, J.; Bond, C.S.; Knoop, V.; et al. The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Mol. Plant* **2020**, *13*, 215–230. [CrossRef] [PubMed]
42. 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] [PubMed]
43. O'Toole, N.; Hattori, M.; Andres, C.; Iida, K.; Lurin, C.; Schmitz-Linneweber, C.; Sugita, M.; Small, I. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* **2008**, *25*, 1120–1128. [CrossRef]
44. Wang, W.; Wu, Y.; Messing, J. Genome-wide analysis of pentatricopeptide-repeat proteins of an aquatic plant. *Planta* **2016**, *244*, 893–899. [CrossRef] [PubMed]
45. Celik Altunoglu, Y.; Baloglu, M.C.; Baloglu, P.; Yer, E.N.; Kara, S. Genome-wide identification and comparative expression analysis of LEA genes in watermelon and melon genomes. *Physiol. Mol. Biol. Plants* **2017**, *23*, 5–21. [CrossRef]
46. Yang, Y.; Ahammed, G.J.; Wan, C.; Liu, H.; Chen, R.; Zhou, Y. Comprehensive Analysis of TIFY Transcription Factors and Their Expression Profiles under Jasmonic Acid and Abiotic Stresses in Watermelon. *Int. J. Genom.* **2019**, *2019*, 6813086. [CrossRef] [PubMed]
47. Barkan, A.; Small, I. Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 415–442. [CrossRef]
48. Bryant, N.; Lloyd, J.; Sweeney, C.; Myouga, F.; Meinke, D. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol.* **2011**, *155*, 1678–1689. [CrossRef]
49. Wechter, W.P.; Levi, A.; Harris, K.R.; Davis, A.R.; Fei, Z.; Katzir, N.; Giovannoni, J.J.; Salman-Minkov, A.; Hernandez, A.; Thimmapuram, J.; et al. Gene expression in developing watermelon fruit. *BMC Genom.* **2008**, *9*, 275. [CrossRef]
50. Adams-Phillips, L.; Barry, C.; Giovannoni, J. Signal transduction systems regulating fruit ripening. *Trends Plant Sci.* **2004**, *9*, 331–338. [CrossRef]
51. Park, Y.J.; Lee, H.J.; Kwak, K.J.; Lee, K.; Hong, S.W.; Kang, H. MicroRNA400-guided cleavage of Pentatricopeptide repeat protein mRNAs Renders *Arabidopsis thaliana* more susceptible to pathogenic bacteria and fungi. *Plant Cell Physiol.* **2014**, *55*, 1660–1668. [CrossRef]
52. Li, H.; Dong, Y.; Chang, J.; He, J.; Chen, H.; Liu, Q.; Wei, C.; Ma, J.; Zhang, Y.; Yang, J.; et al. High-Throughput MicroRNA and mRNA Sequencing Reveals That MicroRNAs May Be Involved in Melatonin-Mediated Cold Tolerance in *Citrullus lanatus* L. *Front. Plant Sci.* **2016**, *7*, 1231. [CrossRef] [PubMed]
53. Liu, L.; Ren, S.; Guo, J.; Wang, Q.; Zhang, X.; Liao, P.; Li, S.; Sunkar, R.; Zheng, Y. Genome-wide identification and comprehensive analysis of microRNAs and phased small interfering RNAs in watermelon. *BMC Genom.* **2018**, *19*, 111. [CrossRef] [PubMed]
54. Branham, S.; Vexler, L.; Meir, A.; Tzuri, G.; Frieman, Z.; Levi, A.; Wechter, W.P.; Tadmor, Y.; Gur, A. Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. *Mol. Breed.* **2017**, *37*, 146. [CrossRef]
55. Liu, S.; Gao, P.; Wang, X.; Davis, A.R.; Baloch, A.M.; Luan, F. Mapping of quantitative trait loci for lycopene content and fruit traits in *Citrullus lanatus*. *Euphytica* **2015**, *202*, 411–426. [CrossRef]
56. Manna, S. An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* **2015**, *113*, 93–99. [CrossRef] [PubMed]
57. Fujii, S.; Sato, N.; Shikanai, T. Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON GRADIENT REGULATION3. *Plant Cell* **2013**, *25*, 3079–3088. [CrossRef]

genes

MDPI

Article

Genetic Diversity and Genome-Wide Association Study of Seed Aspect Ratio Using a High-Density SNP Array in Peanut (*Arachis hypogaea* L.)

Kunyan Zou [1,†], Ki-Seung Kim [2,†], Kipoong Kim [3], Dongwoo Kang [1], Yu-Hyeon Park [1], Hokeun Sun [3], Bo-Keun Ha [4], Jungmin Ha [5] and Tae-Hwan Jun [1,6,*]

1 Department of Plant Bioscience, Pusan National University, Miryang 50463, Korea; 601588zky@pusan.ac.kr (K.Z.); kk7ing@pusan.ac.kr (D.K.); eksvnd951@pusan.ac.kr (Y.-H.P.)
2 FarmHannong, Ltd., Nonsan 33010, Korea; leehan26@snu.ac.kr
3 Department of Statistics, Pusan National University, Busan 46241, Korea; kkp7700@gmail.com (K.K.); hsun@pusan.ac.kr (H.S.)
4 Department of Applied Plant Science, Chonnam National University, Gwangju 61186, Korea; bkha@jnu.ac.kr
5 Department of Plant Science, Gangneung-Wonju National University, Gangneung 25457, Korea; j.ha@gwnu.ac.kr
6 Life and Industry Convergence Research Institute, Pusan National University, Miryang 50463, Korea
* Correspondence: thjun76@pusan.ac.kr; Tel.: +82-55-350-5507
† These authors contributed equally to this work.

Abstract: Peanut (*Arachis hypogaea* L.) is one of the important oil crops of the world. In this study, we aimed to evaluate the genetic diversity of 384 peanut germplasms including 100 Korean germplasms and 284 core collections from the United States Department of Agriculture (USDA) using an Axiom_Arachis array with 58K single-nucleotide polymorphisms (SNPs). We evaluated the evolutionary relationships among 384 peanut germplasms using a genome-wide association study (GWAS) of seed aspect ratio data processed by ImageJ software. In total, 14,030 filtered polymorphic SNPs were identified from the peanut 58K SNP array. We identified five SNPs with significant associations to seed aspect ratio on chromosomes Aradu.A09, Aradu.A10, Araip.B08, and Araip.B09. AX-177640219 on chromosome Araip.B08 was the most significantly associated marker in GAPIT and Regularization method. Phosphoenolpyruvate carboxylase (PEPC) was found among the eleven genes within a linkage disequilibrium (LD) of the significant SNPs on Araip.B08 and could have a strong causal effect in determining seed aspect ratio. The results of the present study provide information and methods that are useful for further genetic and genomic studies as well as molecular breeding programs in peanuts.

Keywords: peanut; core collection; genetic diversity; population structure; genome-wide association study; linkage disequilibrium

Citation: Zou, K.; Kim, K.; Kim, K.; Kang, D.; Park, Y.; Sun, H.; Ha, B.; Ha, J.; Jun, T. Genetic Diversity and Genome-Wide Association Study of Seed Aspect Ratio Using a High-Density SNP Array in Peanut (*Arachis hypogaea* L.). *Genes* **2021**, *12*, 2. https://dx.doi.org/10.3390/genes12010002

Received: 22 August 2020
Accepted: 17 December 2020
Published: 22 December 2020

1. Introduction

1.1. Peanut Information

Peanut or groundnut (*Arachis hypogaea* L.) is an important oil and cash crop of the world [1]. Peanut seeds are rich in oil (48–50%) and protein (25–28%) and they contain certain vitamins and minerals which allows them to be used as an energy source for humans [2,3]. In addition, peanuts contain rich functional elements, such as oleic acid, linoleic acid, resveratrol, fiber, and vitamins [4–6].

Since the beginning of agriculture, food grains have been subjected to selection and breeding for size and most of the grains have seeds far larger than their wild relatives [7]. In the United States, peanut seed size is one of the standards used to determine the grade of shelled peanuts and to evaluate the commercial potential of advanced peanut breeding lines prior to the release of varieties [8].

There have been some studies on seed size in peanut. Quantitative trait loci (QTL) study were conducted to identify loci controlling seed size using a 142 backcross population (87 BC3F1 and 55 BC2F2) with two parents under two water regimes in peanut, while several QTLs associated with increased seed width were detected under water-limited treatment [9]. Simple sequence repeat (SSR) marker PM375 associated with seed length was identified in a total of 88 F2:6 recombinant inbred lines (RILs), representing that increase in seed length may influence in an increase in the weight of a hundred seeds, or in the length of the pod [10]. Florida-07 by GP-NC WS 16. A major seed size QTL on chromosome A05 was identified in the US peanut mini core collection using RILs from a cross between Florida-07 and GP-NC WS 16 [8]. However, there are few studies on seed shape in peanuts so far.

1.2. Peanut Germplasms and Core Collection

Various germplasms with large genetic diversity are excellent resources for peanut breeders to broaden the genetic basis of breeding materials and integrate important alleles related to valuable traits [11]. Diverse germplasms in peanuts have been used to enrich genetic resources, introduce resistance to diseases and pests and, finally, to improve the yield potential through continuous breeding programs.

Recently, effective methods to evaluate and introduce a genetic diversity of germplasm resources have been performed in various studies. Core collections were first defined as a limited set of accessions "representing, with a minimum of repetitiveness, the genetic diversity of a crop species and its wild relatives" [12,13]. The use of core collections has many advantages and they also represent a good starting material for association mapping. Recently, core collections have been established in various crops, including rice [14], wheat [15], maize [16], and *Brassica napus* [17]. The peanut core collections were developed from the US germplasm collection [18], and information on the accessions of the core collection are available at the Germplasm Resource Information Network (GRIN) (https://www.ars-grin.gov).

To promote and improve the utilization of germplasm resources in peanut breeding programs, the peanut mini core collection was established by utilizing the stratification strategy of the United States (US) peanut germplasm resource center [19]. The majority of the accessions in the mini core collection were unrelated individuals, which may be a good starting material for initiating the peanut association study. The purpose of establishing a core or mini core collection for any crop is to promote the efficient and economical use of plant materials by end-users and to identify germplasms with desirable characteristics.

The Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences in China established a core collection with 576 *A. hypogaea* genotypes and a mini core collection with 298 accessions representing the majority of the genetic diversity of cultivated peanut in China. They conducted an association study using the mini core collection, and a total of 89 simple sequence repeat (SSR) alleles were identified as associated with 15 agronomic traits. The results showed that there was a great possibility to combine association analysis and marker-assisted breeding using the peanut mini core collection [20,21]. The US mini core collection was evaluated and mapped using quantitative trait loci (QTL) for several traits, such as resistance to Tomato spotted wilt virus (TSWV) [21]. In the ICRISAT mini core collection, several candidate regions associated with non-redundant leaf proteins were identified as being related to tolerance to water deficit stress; however, little has been reported regarding these traits in the US germplasms [22].

1.3. Characteristic of Peanut Genome

Cultivated peanut is allotetraploid (2n = 4× = 40, AABB) with a genome size of 2800 Mb/1C and the genome composition of cultivated peanut was shown to have derived from a recent hybridization of *A. duranensis* (A subgenome) and *A. ipaensis* (B subgenome) [23–26]. As the polyploidization event occurred recently, the genetic diversity of cultivated peanut is extremely low [27]. Peanut subgenomes are very closely related [28,29] and

have an estimated repetition rate of 64% [1], which makes the assembly of peanut genome sequences extremely difficult [1,26,30]. The genome sequences of the diploid ancestors (*A. duranensis* and *A. ipaensis*) of cultivated peanut were reported in 2016, which became the basis for understanding the genome of cultivated peanut [26]. The sequencing results of A. *duranensis* (A genomic progenitor) and A. *ipaensis* (B genome progenitor) provided new insights into the biology, evolution, and genome changes of cultivated peanut and accelerated the molecular breeding of peanut varieties [31].

Recently, the cultivated peanut allotetraploid A. *hypogaea* genome was sequenced in 2019 and compared with the related diploid *A. duranensis* and *A. ipaensis* genomes. A total of 39,888 A subgenome genes and 41,526 B subgenome genes were annotated in the allotetraploid subgenome [32].

1.4. Development of Molecular Markers Using Next Generation Sequencing (NGS) Technology

In 2005, pyrosequencing technology was implemented using large-scale parallel sequencing or deep sequencing, revolutionizing next generation sequencing (NGS) technology and biological genomic research [33]. In the past decade, NGS technology made significant progress, and the cost of sequencing dropped sharply [27]. In addition, there have been innovative improvements in the productivity and accuracy of sequencing data. In particular, genome-wide studies using de novo assembly, resequencing, and a variety of bioinformatic methods have enabled the production of large numbers of single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSR) in complex genomes [26,34–36]. In recent work, high-throughput genotyping was conducted using NGS technology through double-digest restriction-site-associated DNA sequencing (ddRADseq), a total of 14,663 SNPs were developed, and a genetic linkage map based on SNPs was constructed using 1765 SNP markers in 166 F9 RIL population from a cross between Zhonghua 5 and ICGV86699 [37]. Numerous SNP and cleaved amplified polymorphic sequence (CAPS) markers were developed from the re-sequencing of two Korean peanut germplasms of K-Ol and Pungan, which indicates that the molecular marker information can provide valuable guidance and information for peanut breeding programs [27].

Due to the relatively large genome size and the low genetic diversity in cultivated peanut, developing SNP array chips for high-throughput genotyping is necessary [38]. By DNA resequencing and the RNA sequencing of 41 peanut genetic materials and wild diploid ancestors, a total of 163,782 SNPS were obtained. A total of 58,233 unique SNP sequences with large amounts of information were selected to construct the high-density SNP array Axiom_Arachis with 58K SNPs [39]. The high-density SNP Axiom_Arachis array with 58K SNPs could be used to accelerate the process of high-resolution mapping and molecular breeding in peanuts.

1.5. Applications of High-Density SNP Arrays in Crops

As the most abundant type of DNA sequence variation in the genome, SNPs could be successfully used to associate the genotypic variations with target phenotypes. High-density SNP arrays have been developed for high-resolution mapping of crops and are widely used in many applications that require a large number of molecular markers, such as high-density genetic profiling, genome-wide association study (GWAS), and genomic selection [38,40,41]. One hundred and seven U.S. peanut mini core collections were genotyped using a 58K Affymetrix SNP array and a total of 13,527 highly polymorphic SNP markers were selected for marker-trait associations in arachidic and behenic fatty acid compositions [42]. A total of 2882 polymorphic SNPs retained from the second edition of the Axiom_Arachis array (Axiom_Arachis2) were used to identify loci controlling pod construction trait using 195 F7 recombinant inbred lines (RILs) [43]. The 48K Axiom Arachis2 SNP array was applied to identify single nucleotide polymorphisms (SNP) among the two sets of RILs and the two original Nod+ parental lines to explore the genetic factors and genetic regions controlling nodulation in peanut [44].

Genomic-assisted breeding (GAB) using large amounts of genomic data related to important agronomic traits could be used to develop new varieties faster than when using traditional breeding methods. Detailed genetic maps consisting of thousands of array-based SNPs have been used for the identification of genes controlling target traits [41,45]. GWAS, also known as whole-genome association study, is an observational study of a genome-wide set of genetic variants in different individuals to investigate whether any variant is associated with the target traits [46]. Any phenotypic differences could then be connected back to the underlying causative loci via various mapping approaches, including quantitative trait loci (QTL) mapping. Many research groups have used GWAS to identify associations between genotypes and phenotypes as well as to discover novel biological mechanisms [47]. Currently, most GWAS have been performed using high-throughput SNP data obtained by SNP arrays with a greater density of variants and a wide range of allele frequencies [48–51]. The GWAS format is easy to share and generate, and GWAS can be conducted using various applications and software [46].

1.6. Purpose

In this study, we aimed to (1) evaluate the population structure and genetic diversity of 384 peanut germplasms including 100 Korean germplasms and 284 United States Department of Agriculture (USDA) core collections using Axiom_Arachis array with 58K SNPs, and (2) to conduct GWAS for seed shape and identify candidate genes associated with this trait. Our results could provide useful tools for improving various agronomic traits in molecular breeding programs for peanuts.

2. Materials and Methods

2.1. Plant Materials, DNA Extraction, and Genotyping

A total of 384 peanut accessions were used for the present study (Supplementary Table S1). Among those, 284 peanut accessions were obtained from the core collections of the US Department of Agriculture (USDA) according to the proportion of the number of germplasms, which were widely distributed in East Asia, South Asia, West Asia, East Africa, South Africa, West Africa, North America, South America, Europe, and the Australian continent. In addition, 100 peanut germplasms were obtained from the National Agrobiodiversity Center Korean, RURAL DEVELOPMENT ADMINISTRATION (RDA)-GenBank Information Center, South Korea, including landraces, breeding lines, and cultivars. A young leaf from each individual accession was collected to extract the genomic DNA. A total of 384 peanut genomic DNA were extracted for each accession using the cetyltrimethylammonium bromide (CTAB) protocol with slight modifications [52]. The quality and quantity of the extracted DNA were determined using a NanoDrop ND-1000 (Thermos Fisher Scientific Inc., Wilmington, DE, USA) and 1% agarose gel electrophoresis.

A high-density SNP array Axiom_Arachis with 58K SNPs was used to obtain the genotyping data [39]. Reference genome builds were acquired from arahy.Tifrunner.gnm1.KYV3 (https://www.ncbi.nlm.nih.gov/assembly) to serve as controls in the array design.

2.2. Screening of Seed Aspect Ratio

The seed aspect ratio data (Supplementary Table S2) were obtained by scanning seed images. The scanning images were processed by ImageJ 1.52a software (https://imagej.nih.gov/ij/notes.html) to generate phenotype data for the genome-wide association study. The seed aspect ratio was calculated as the seed major axis divided by the seed minor axis. Ten seeds per accession were scanned at the same time, and the seed aspect ratios of the ten seeds were averaged (Supplementary Figure S1). The phenotype data were analyzed using the R program to conduct a t-test and normal distribution in the accessions.

2.3. Population Structure Analysis

A principle coordinate analysis (PCoA) was conducted using the software GenAlEx V6.503 [53,54]. The population structure of 384 peanut accessions was evaluated by Struc-

ture v2.3.4 software (https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html) under the admixture model. We compared the structures following the same parameters with K-values ranging from 1 to 10, and 20,000 Markov chain Monte Carlo iterations after a burn-in period of 10,000 iterations were carried out for three independent runs per K value. To make a decision for the optimum number for K, the delta K (ΔK) method used the software online "harvester structure".

2.4. Genome-Wide Association Analysis

We analyzed the SNPs in Axiom_Arachis with a 58K array of the cultivated peanut using R software analysis tools. In the present study, the GAPIT package of R software—was used to conduct GWAS, and the enriched compressed mixed linear model (ECMLM) was selected for the analysis of association between SNPs and the phenotype data of interest [55]. The cutoff for significant association was a false discovery rate (FDR) adjusted *p*-value of less than 0.05. Candidate genes covering significantly associated SNPs were selected from the PEANUTBASE website tool (https://www.peanutbase.org) within a 150 kb region upstream or downstream of peak SNPs according to the linkage disequilibrium (LD) decay results.

2.5. Linkage Disequilibrium (LD) Analysis

We performed linkage disequilibrium analysis for all possible pairs of SNPs with a minor allele frequency (MAF) greater than 0.01 in a dataset. To determine the degree of resolution achieved in the association analysis, both the genome and chromosome-wide LDs were estimated [56].

LD blocks were viewed using Haploview4.2, which uses permutation tests to determine the *p*-values for each pairwise correlation. The LD decay was calculated with PopLDdecay [57,58]. The physical distance of the LD decay plot was determined based on the *D′* values and distances between each pair of SNPs on each chromosome using a nonlinear model [59].

The standard descriptive LD parameter D′ was estimated as previously described by [60,61]. The average D′ value was calculated for each chromosome using Haploview software [60].

2.6. Regularization Method

In human genome-wide association studies, regularization methods based on penalized likelihood are popular regarding their application to identify disease-related genes or genetic regions as they are computationally efficient when used in analysis of high-dimensional genomic data [62–68]. The penalized likelihood function using an elastic-net penalty is defined as

$$Q(\beta) = -l(\beta) + \lambda\alpha \sum_{j=1}^{p} |\beta_j| + \lambda(1-\alpha) \sum_{j=1}^{p} \beta_j^2 \quad (1)$$

where $l(\beta)$ is a log-likelihood function, β is the p-dimensional coefficient vector, $\lambda \geq 0$ is a tuning parameter for sparsity, and $\alpha \in [0,1]$ is a tuning parameter for smoothness. When $\alpha = 1$, the coefficient vector β becomes the solution of the least absolute and shrinkage selection operator (LASSO) [69]. The estimated coefficient β consists mostly of zero values and only a few nonzero values. Based on 100 bootstrap samples, the selection probability of individual SNPs was computed where only SNPs with nonzero coefficients were selected for each bootstrap sample. Finally, we were able to identify the top ranked SNPs by their selection probability.

In order to select significant SNPs, we used two types of threshold of selection probability which can control the number of falsely selected SNPs. The first one is the theoretical threshold proposed by [70]. The second one is the empirical threshold [71] which basically computes the quantile value of an empirical distribution of selection probability based on permutation. In their extensive simulation studies, it was demonstrated that the number of falsely selected SNPs can be controlled when the empirical threshold is applied to high-

dimensional genomic data. The theoretical threshold (π_θ) and the empirical threshold (π^*_θ) can be written as:

$$\pi_\theta = \frac{q^2_\Lambda}{2\theta p} + \frac{1}{2} \text{ and } \pi^*_\theta = \frac{1}{\mathrm{B}} \sum_{b=1}^{\mathrm{B}} SP^{[\theta]}_{(b)}(I_b)\,, \tag{2}$$

where θ is the upper bound of the expected number of false discoveries, q_Λ is the average number of selected SNPs, B is the number of permutations and I_b is the b-th random permuted sample. We denote $SP^{[\theta]}_{(b)}$ by the top θ-th ranked selection probability when they were sorted in descending order for the b-th permuted sample such as $SP^{[1]}_{(b)} > \cdots > SP^{[p]}_{(b)}$. We chose the expected number of false discoveries $\theta = 1$, and thereby the number of falsely selected SNPs by each threshold can be guaranteed to be less than $\theta = 1$.

3. Results

3.1. SNP Genotyping

Of the 58K informative SNPs, a total of 47,837 polymorphic SNPs were selected (Supplementary Table S3). Of the 47,837 SNPs, 19,554 and 21,876 SNPs were derived from the subgenomes A and B, respectively, and 6407 SNPs were derived from scaffolds (Supplementary Table S3 and Figure 1a). A total of 14,030 SNPs were selected for association analysis after eliminating SNPs with high levels of missing data (>20%), heterozygosity (>20%), or low a minor allele frequency (MAF) (<0.01). Of the 14030 SNPs, 6623 and 7407 SNPs were derived from A and B subgenomes, respectively. The majority of SNPs were evenly distributed across the chromosomes; however, there were some large gaps between SNPs on the chromosomes Aradu.A09, Aradu.A10, Araip.B05, Araip.B06, Araip.B07, Araip.B09, and Araip.B10 (Figure 1b). The peanut genome had an overall SNP density of 5.91 SNPs/Mb, with the Aradu.A09 (3.45 SNPs/Mb) and Aradu.A08 (9.35 SNPs/Mb) chromosomal densities being the lowest and highest, respectively (Supplementary Table S4).

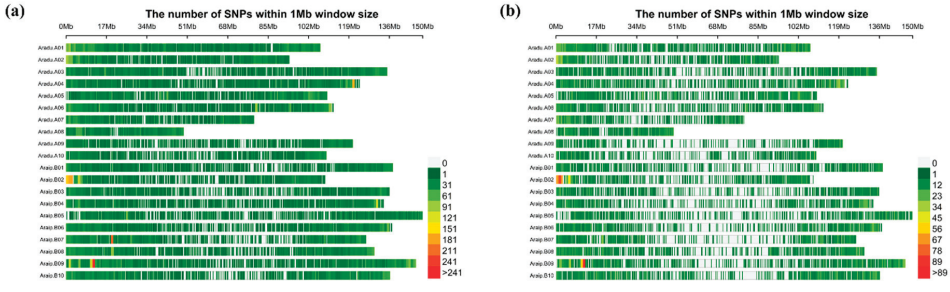


Figure 1. Single nucleotide polymorphisms (SNP) distribution in the 20 chromosomes of the cultivated peanut. The horizontal axis shows chromosome length (Mb), the shades of red represent SNP density. The vertical axis shows the 20 chromosomes. (**a**) Polymorphic SNPs except for scaffold markers; (**b**) Polymorphic SNPs (except for scaffold markers) after filter by GAPIT coding.

3.2. Phenotype Data Analysis

The mean value for the seed aspect ratio was 1.6325 (Figure 2a). The normal distribution test showed that the scatter points of the quantile–quantile (QQ plot) graph (Figure 2b) were clustered around the fixed line; therefore, we assumed that the data were normally distributed (p = 0.05).

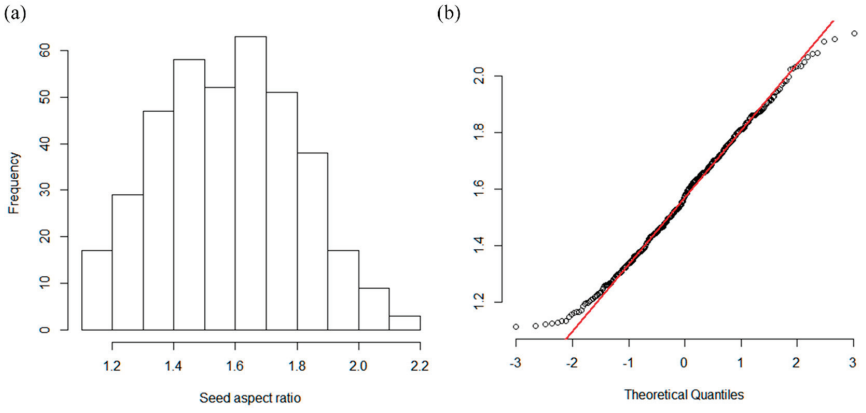


Figure 2. (**a**) Peanut seed aspect ratio data histogram; (**b**) The normal distribution test by the quantile-quantile (QQ plot) graph.

3.3. Genetic Diversity

The pattern of PCoA (Figure 3) showed that the first two axes accounted for 30.19% and 6.91%, respectively, of the total variation and the 384 peanut accessions were divided into three broad groups across the first two axes. The first axes separated the South Korean (clustered filled diamonds) and South American (green filled squares) peanut accessions into two very different parts, and, at the same time, assigned East Asian, South Asian, and West Asian peanut accessions (brown filled triangles, pink filled diamonds, and green filled circles, respectively) to another part. Additionally, the peanut accessions that originated from East Africa, South Africa, West Africa, North America, and Europe formed two concentrated groups by the first and second axes. Interestingly, the accessions from South Korea were genetically very different from those from South America, which is the origin of the cultivated peanut.

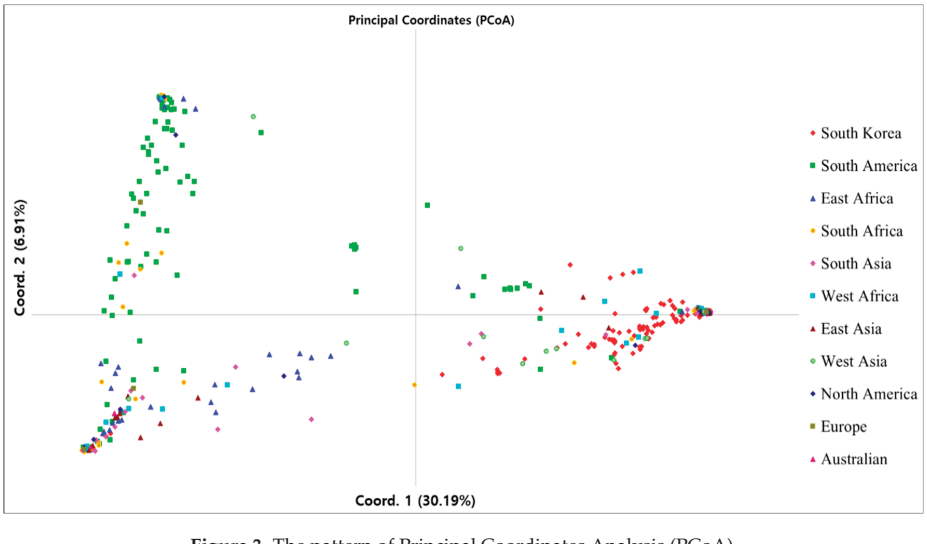


Figure 3. The pattern of Principal Coordinates Analysis (PCoA).

3.4. Genetic Structure

At K = 2, we found maximum Δk values that were plotted against the K to confirm the number of populations, while another lower peak was shown at K = 7 (Supplementary Figure S2). When most individuals were divided into the two subpopulations (K = 2, Figure 4), the peanut accessions, including 64.9% from Asia (of which approximately 74% individuals were from South Korea and 26% from other origins in Asia), 24.4% from Africa, 10.2% from South America, and 0.5% from Europe, belonged to one subgroup (red), while another subgroup (green) revealed features of accessions, including 16.8% from Asia (comprising about 6.7% from South Korea and 93.3% from other Asia origins), 35.2% from Africa, 42.5% from South America, 2.8% from North America, 1.7% from Europe, and 1% from Australia.

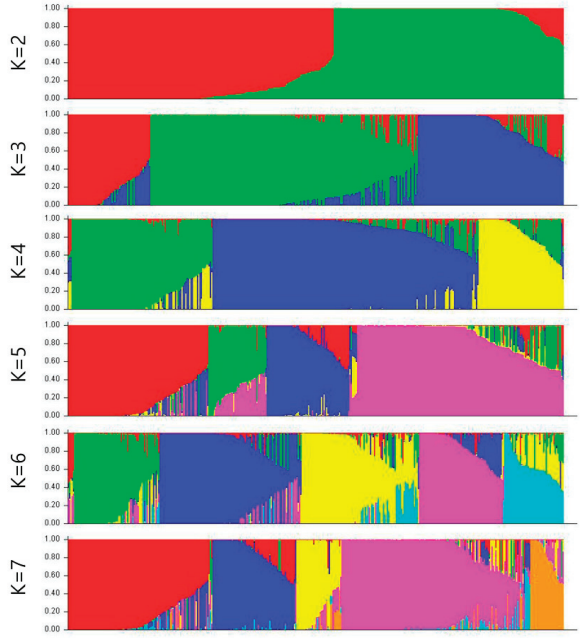


Figure 4. Structure clustering results obtained at K = 2 to K = 7 of the 384 peanut accessions. Each individual is represented by a bar corresponding to the sum of assignment probabilities to the K cluster.

As we continued to divide the subgroups carefully, there were new divisions into the subgroups. The most divergent subgroups were formed at K = 7. Of the peanut accessions, 26.4% originating from Asia (of which approximately 10.3% were from South Korea and 89.7% from other Asia origins), 45.5% from Africa, 20.9% from South America, 3.6% from North America, 1.8% from Europe, and 1.8% from Australia belonged to the red subgroup. The green subgroup revealed features of 50% accessions from South Korea and 50% from Africa. The dark blue subgroup showed features of 1.5% accessions from Asia, 20% from Africa, 75.5% from South America, 1.5% from North America, and 1.5% from Europe. The yellow subgroup showed features of 91.4% accessions from Asia (including about 87.5% from South Korea and 12.5% from other Asia origins), 2.9% from Africa, and 5.7% from South America. The pink subgroup consisted of 56.3% accessions from Asia (of which approximately 63.8% were from South Korea, and 36.2% from other Asia origins), 31% from Africa, 12% from South America, and 0.7% from Europe. The light blue subgroup showed features of accessions from only South America. The orange subgroup showed features of individuals with 76.9% of accessions from Asia (of which approximately 85%

accessions were from South Korea and 15% from other Asia origins), 15.4% from Africa, and 7.7% from South America.

3.5. Genome-Wide Association Study (GWAS)

The genotype data of 14,030 filtered polymorphic SNPs and the phenotypic data of the seed aspect ratios were analyzed for GWAS by GAPIT. A total of five candidate SNPs showing significant associations ($p < 0.0001$) with the seed aspect ratio were identified on chromosomes Aradu.A09, Aradu.A10, Araip.B08, and Araip.B09 (Table 1 and Figure 5a). The distribution of the observed −log10(p) for each SNP was compared with the expected distribution in the QQ plot representing that the population structure and kinship relationship were well controlled in the GWAS (Figure 5b). The significance of the marker–trait associations were determined using the FDR with adjusted *p*-value ($p = 0.05$). AX-177640219 on chromosome Araip.B08 was significantly associated with the seed trait at the significant threshold (Table 1).

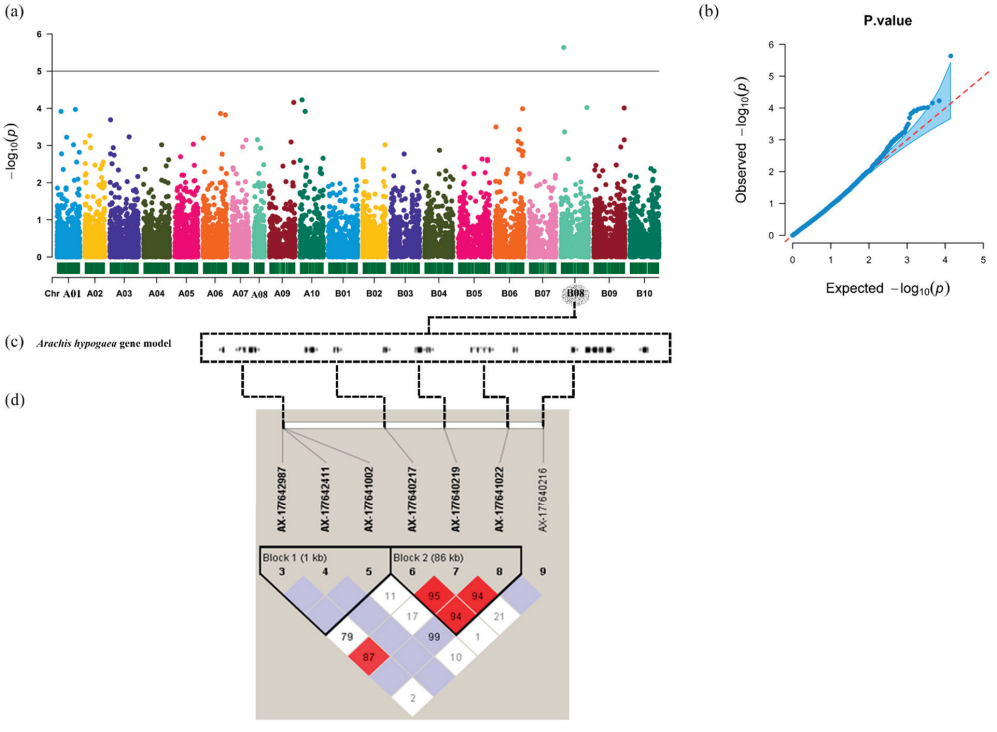


Figure 5. (**a**) Manhattan plot of a genome-wide association analysis by GAPIT; (**b**) Q-Q (quantile-quantile) plot; (**c**) Genes located at the association regions based on the *Arachis hypogaea* Tifrunner 1.0 reference genome; (**d**) Linkage disequilibrium (LD) plot generated using Haploview, D′ values that correspond to SNP pairs are shown within the respective squares. Higher D′ values are indicated with a brighter red color.

Table 1. Significant markers associated with seed aspect ratio of peanut identified using GAPIT analysis.

SNP	Chromosome	Position (bp)	*p*-Value (*p*)	FDR_Adjusted_*p*-Values
AX-177640219	Araip.B08	12829161	2.31×10^{-6}	0.032
AX-147235444	Aradu.A10	8911644	5.91×10^{-5}	NS [a]
AX-176807953	Aradu.A09	113907685	6.95×10^{-5}	NS
AX-176822392	Araip.B08	121783058	9.55×10^{-5}	NS
AX-147262340	Araip.B09	143554366	9.80×10^{-5}	NS

[a] FDR_adjusted_*p*-value is not significant at the level of 0.05.

3.6. LD and Candidate Genes Analysis

Pairwise comparisons were performed between all SNPs for the estimation of LD decay. At a cutoff value of $r^2 = 0.1$, the averaged LD decay distance of the 384 peanuts was approximately 150 kb (Supplementary Figure S3). The pattern of LD across the entire genome presented a number of haplotype blocks containing SNPs that can be used to determine the range of the candidate gene. The genomic locations harboring significant SNPs from the GWAS were investigated to identify putative candidate genes based on the peanut reference genome (*A. hypogaea* Tifrunner 1.0). Strong and extensive pairwise LD was observed among highly significant SNPs around AX-177640219 (*p*-value = 0.000015) on chromosome Araip.B08 from the 12,629,161 to 13,029,161 bp region ($D' > 0.80$) in which D' varied from 0.036 to 1 (Figure 5d and Supplementary Table S5).

Fifteen annotated genes at the association regions flanked by SNP AX-177640219 on chromosome Araip.B08 were identified within the estimated ±150 kb window based on the reference genome (Figure 5c and Supplementary Table S6).

3.7. Regularization Method

Alternatively, we also conducted regularization methods, such as LASSO, to identify candidate regions associated with the seed aspect ratio (Figure 6) [71]. The regularization method was performed using an entire dataset at a time and could select several putative markers most likely related to the trait based on the value of selection probability, whereas the ECMLM analysis only tested one marker at a time. As a result, one SNP locus (AX-177640219 on Araip.B08) was identified as being most likely related to the seed aspect ratio based on the selection probability at the permuted threshold 0.894, and was also found to be highly significantly associated in the GAPIT analysis (Figure 6). When loosening the strict threshold to 0.506, a total of six SNPs were additionally identified, AX-177640938 on chromosome Araip.B08, AX-147218661 on Aradu.A03, AX-147251864 on Araip.B06, AX-176802342 on Araip.B04, AX-176791478 on Aradu.A02, and AX-176800768 on Aradu.A01, which presented significant associations with ECMLM results indicating that the regions flanked with these markers might be candidate regions for possible determination of seed shape in peanuts. Therefore, the use of both methods to conduct association studies is beneficial in (1) boosting confidence in the case where common markers are identified and (2) to maximize the possibility of finding new significant markers associated with a trait of interest.

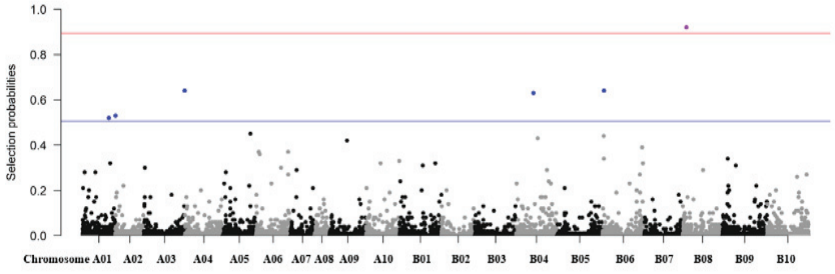


Figure 6. Manhattan plot of a genome-wide association analysis by the least absolute and shrinkage selection operator (LASSO).

3.8. Evaluation of Heterozygous Rate

The same filtering conditions with maximum missing data of 20% and MAF of 0.01%, different heterozygosity rates (starting from 5% and 10% and every 10% until 100% maximum heterozygous SNPs (Figure 7 and Supplementary Table S7) were used to filter the genotype data in our study, and different significance cutoff thresholds were used to assess the effect of the SNPs on the seed aspect ratio.

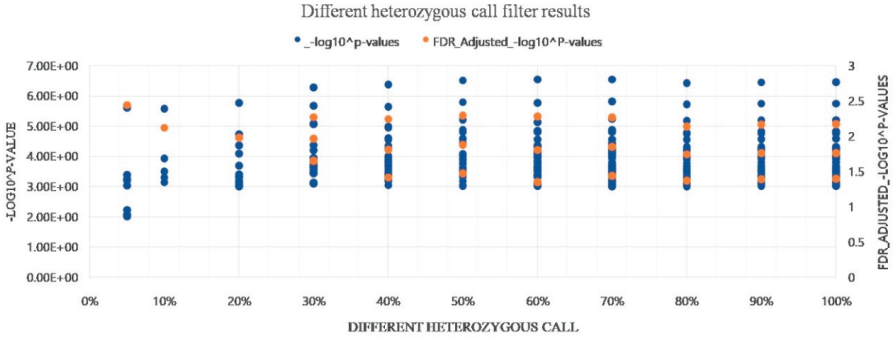


Figure 7. Different heterozygosity rates to filter genotype data and different significance cut-off thresholds used to assess the effect of the SNPs on seed aspect ratio.

When the genotype data filtered by a 5% to 20% maximum heterozygous rate were used for GWAS analysis, a higher specificity of the results was obtained; however, only one significance marker was evaluated at the 0.05 critical threshold for the false discovery rate (FDR) adjusted p-value. On the contrary, when a high heterozygosity rate of 30% to 100% was used for data filtering, additional significant markers were detected; however, those markers require validation.

4. Discussion

The trait of seed ratio (length-width ratio) screened in this study has been reported to have very high broad-sense of heritability in recently published peanut research. Zhang et al. [72] reported that it has a high broad-sense of heritability (0.81) in peanuts. For other legume crops, Hu et al. [73] reported a very high broad-sense of heritability ranged from 92.46 to96.25 in three traits related to seed shape in soybeans. If a phenotypic trait has a high level of heritability, the influence of the environmental factors might be relatively small, and in this case, it could be possible that genes (or QTL) with relatively large effects on the trait could be identified even if the trait were not measured in the same conditions. Of course, even in this case, the influence of the environmental conditions cannot be overlooked.

The genotyping data from the 58K SNP array chip could play an important role in understanding the evolutionary history of peanuts and the domestication of cultivated peanut [74]. The application of the array chip also demonstrated that it is a powerful and reliable tool for peanut germplasm background selection and evolutionary studies [75]. In the present study, it is the first to conduct GWAS analysis using a large number of Korean peanut germplasms as well as the USDA peanut core collection with a high-density SNP chip data that can be used toward increasing the genetic diversity of the US peanut germplasm collection.

The cultivated peanut species (A. *hypogaea*) is known to originate from southern Bolivia to northwestern Argentina based on the occurrence of the two progenitor species, A. *duranensis* and A. *ipaensis*, and archaeological evidence gathered in those regions [76–78]. Researchers also suggested that the eastern slopes of Cordillera may be a possible area for the origin of A. *hypogaea* due to the favorable environment for peanut growth [78,79]. However, the present study showed an interesting result in that South American peanuts, generally regarded as the origin of peanuts, were revealed as having significant genetic differences from peanuts of other regions, including South Korea.

The evaluation results for the evolutionary relationships among the entirety of the 384 peanut germplasms indicated that most of the peanut individuals from South Korea and South America separated into two distinct groups and were also independent from the peanuts from the other origins. This might indicate that there was a great genetic difference between the peanut germplasms from South Korea and South America. Likely, due to the lack of interactions between South Korean peanut germplasms and others, it might be possible that an independent breeding history by human selection and/or environmental influences for a long period have caused these genetic differences.

In human genetic association studies with high-dimensional genomic data, regularization methods, such as LASSO and elastic-net, have been widely applied to identify outcome-related genetic sites and genes as they have certain advantages over univariate analysis. First, regularization methods can easily handle highly correlated genomic measurements and covariate effects as they are based on a regression model. Secondly, the majority of regularization methods have been implemented into very efficient computational algorithms such R package 'glmnet' and 'gglasso'. These packages can detect outcome-related genetic-sites and genes in less than a minute for more than 100K dimensional genomic data. Lastly, there are various types of regularization methods that can be applied to different types of genomic data. For example, we applied LASSO and elastic-net to SNP data in the GWAS or QTL analysis; however, sparse group LASSO [80] and network-based regularization [81] are ideal for group structured genomic data, such as gene expression data and DNA methylation data. Despite these advantages of regularization methods, they have rarely been applied to detect QTLs or genes of interest in crops. In this study, LASSO was able to identify potentially outcome-related SNPs that were not identified in general GWAS methods although further validation studies are required for these SNPs.

Data filtering is the primary process of genome-wide association analysis, which includes huge amounts of data and requires strict quality control standards. Data filtering is divided into two sections, one for marker variables and another for individuals. The former considers the minor allele frequency (MAF) and the degree of missing data and heterozygosity, etc., whereas the latter mostly considers missing levels, population stratification, and independency among individuals [82]. The entire set of heterozygous SNPs are typically used in human GWAS analysis [83]. In peanuts, a high level of heterozygosity may not be expected as peanut is a self-pollinating crop revealed to have a low outcrossing rate ranging from 1.9% to 8% [84]. However, our array chip data showed a large number of heterozygous SNPs, which can affect the GWAS results. According to Figure 7, the significant SNPs identified from using 5% to 20% maximum heterozygous rate showed the same GWAS results, with one significant marker at FDR 0.05, while the results from using 30% to 100% maximum heterozygous rate showed similar GWAS results with three

significant markers. Therefore, we filtered the genotype data with maximum heterozygous SNPs of 20%, and we used less heterozygous SNPs for analysis.

Carbon assimilated by photosynthesis is transported into seeds with multiple purposes, such as the biosynthesis of starch, oil, amino acids, and cellulose. The most important aspect of oil accumulation in developing seeds lies in the activation of metabolic pathways driving incoming carbon into fatty acid biosynthesis at the expense of competitive pathways. Within the genomic region of ~300 kb associated with seed development, phosphoenolpyruvate (PEP) carboxylase (PEPC; Arahy.HT9EWH) was among eleven genes located within the LD of significant SNPs on the chromosome Araip.B08 (Supplementary Table S6). PEP is catalyzed into oxaloacetate (OAA), a protein precursor, by PEPC [85]. OAA can be converted to malate and then to pyruvate (a precursor for oil). PEPC had been reported to regulate the metabolic network of glycolytic carbon into precursors for both oil and protein in soybean seed development [86]. The activation status of PEPC has been reported to play a key role in the partitioning of assimilates into the different storage products in barley (*Hordeum vulgare*), alfalfa (*Medicago sativa*), and fava bean (*Vicia faba*) [87–89]. In peanuts, researchers reported that the expression levels of PEPC genes were significantly associated with lipid accumulation [90]. In the present study, only fifteen annotated genes were identified within the genomic region as being highly associated with seed development through high-throughput GWAS analysis. Among them, the PEPC gene could have a strong causal effect within this region associated with diverse metabolic pathways that includes including protein and oil biosynthesis.

5. Conclusions

Peanut is one of the most important food/oil crops and improving the quality and yield potential of crops is an important challenge in most breeding programs. Our study demonstrated the feasibility of GWAS analysis using the core germplasm from diverse origins and high-density array chips. Five candidate markers with a significant correlation with the aspect ratio of peanut seeds were identified and lay a foundation for further research. The Arahy.HT9EWH, phosphoenolpyruvate carboxylase (PEPC) gene corresponding to the most significantly associated marker was a promising candidate gene that is involved in many metabolic pathways, including those involved in seed development processes. Therefore, the results of the present study provide valuable information and methods for the genetic and genomic study as well as molecular breeding programs in peanuts.

Supplementary Materials: The following are available online at https://www.mdpi.com/2073-4425/12/1/2/s1, Figure S1: ImageJ protocol, Figure S2: Δk values to confirm the number of populations, Figure S3: The estimation of LD decay, Table S1: Information on the 384 peanut accessions, Table S2: Information on the 384 peanut accessions seed aspect ratio, Table S3: A high-density SNP array 'Axiom_Arachis' with 58K SNPs, Table S4: Information of the peanut genome from 58K SNP array chip, Table S5: Pairwise LD on chromosome Araip.B08, Table S6: List of genes in the significant region (chromosome Araip.B08) with annotations identified by *Arachis hypogaea* Tifrunner 1.0 reference, Table S7: Genome-wide association study (GWAS) (p-value < 0.001) of seed aspect ratio were constructed by different heterozygous filter methods.

Author Contributions: Conceptualization, T.-H.J.; methodology, T.-H.J. and H.S.; software, K.Z. and K.K.; formal analysis, K.Z. and K.K.; investigation, K.Z., D.K. and Y.-H.P.; resources, T.-H.J.; data curation, J.H. and K.-S.K.; writing—original draft preparation, K.Z. and K.-S.K.; writing—review and editing, B.-K.H., H.S., K.-S.K. and T.-H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was carried out with the support of the "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ013125022020)" Rural Development Administration, Republic of Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Please refer to suggested Data Availability Statements in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

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

References

1. Dhillon, S.S.; Rake, A.V.; Miksche, J.P. Reassociation kinetics and cytophotometric characterization of peanut (*Arachis hypogaea* L.) DNA. *Plant Physiol.* **1980**, *65*, 1121–1127. [CrossRef] [PubMed]
2. Win, M.M.; Abdul-Hamid, A.; Baharin, B.S.; Anwar, F.; Sabu, M.C.; Pak-Dek, M.S. Phenolic compounds and antioxidant activity of peanut's skin, hull, raw kernel and roasted kernel flour. *Pak. J. Bot.* **2011**, *43*, 1635–1642.
3. Pasupuleti, J.; Nigam, S.N.; Pandey, M.K.; Nagesh, P.; Varshney, R.K. Groundnut improvement: Use of genetic and genomic tools. *Front. Plant Sci.* **2013**, *4*, 23. [CrossRef]
4. Radhakrishnan, R.; Pae, S.B.; Lee, B.K.; Baek, I.Y. Evaluation of luteolin from shells of Korean peanut cultivars for industrial utilization. *Afr. J. Biotechnol.* **2013**, *12*, 4477–4480. [CrossRef]
5. Musa, Ö.M. Some nutritional characteristics of kernel and oil of peanut (*Arachis hypogaea* L.). *J. Oleo Sci.* **2010**, *59*, 1–5. [CrossRef]
6. Sales, J.M.; Resurreccion, A.V. Resveratrol in peanuts. *Crit. Rev. Food Sci. Nutr.* **2014**, *54*, 734–770. [CrossRef]
7. Sundaresan, V. Control of seed size in plants. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17887–17888. [CrossRef]
8. Chu, Y.; Chee, P.; Isleib, T.G.; Holbrook, C.C.; Ozias-Akins, P. Major seed size QTL on chromosome A05 of peanut (Arachis hypogaea) is conserved in the US mini core germplasm collection. *Mol. Breed.* **2020**, *40*, 6. [CrossRef]
9. Fonceka, D.; Tossim, H.A.; Rivallan, R.; Vignes, H.; Faye, I.; Ndoye, O.; Rami, J.F. Fostered and left behind alleles in peanut: Interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. *BMC Plant Biol.* **2012**, *12*, 26. [CrossRef]
10. Gomez Selvaraj, M.; Narayana, M.; Schubert, A.M.; Ayers, J.L.; Baring, M.R.; Burow, M.D. Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. *Electron. J. Biotechnol.* **2009**, *12*, 3–4. [CrossRef]
11. Zhao, Z.; Tseng, Y.C.; Peng, Z.; Lopez, Y.; Chen, C.Y.; Tillman, B.L.; Dang, P.; Wang, J. Refining a major QTL controlling spotted wilt disease resistance in cultivated peanut (*Arachis hypogaea* L.) and evaluating its contribution to the resistance variations in peanut germplasm. *BMC Genet.* **2018**, *19*, 17. [CrossRef] [PubMed]
12. Arber, W.; Illmensee, K.; Peacock, W.J.; Starlinger, P. (Eds.) *Genetic Manipulation: Impact on Man and Society (No. 1)*; Cambridge University Press: Cambridge, UK, 1984.
13. Frankel, O.H.; Brown, A.H.D. *Current Plant Genetic Resources—A Critical Appraisal*; Oxford and IBH Publishing Co.: New Delhi, India, 1984.
14. Zhang, H.; Zhang, D.; Wang, M.; Sun, J.; Qi, Y.; Li, J.; Han, L.; Qiu, Z.; Tang, S.; Li, Z. A core collection and mini core collection of *Oryza sativa* L. in China. *Theor. Appl. Genet.* **2011**, *122*, 49–61. [CrossRef] [PubMed]
15. Hao, C.Y.; Zhang, X.Y.; Wang, L.F.; Dong, Y.S.; Shang, X.W.; Jia, J.Z. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Mol. Breed.* **2006**, *17*, 69–77. [CrossRef]
16. Yang, X.; Gao, S.; Xu, S.; Zhang, Z.; Prasanna, B.M.; Li, L.; Li, J.; Yan, J. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* **2011**, *28*, 511–526. [CrossRef]
17. Xiao, Y.; Cai, D.; Yang, W.; Ye, W.; Younas, M.; Wu, J.; Liu, K. Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. *Theor. Appl. Genet.* **2012**, *125*, 437–447. [CrossRef]
18. Holbrook, C.C.; Anderson, W.F.; Pittman, R.N. Selection of a core collection from the US germplasm collection of peanut. *Crop Sci.* **1993**, *33*, 859–861. [CrossRef]
19. Jiang, H.F.; Ren, X.P.; Huang, J.Q.; Liao, B.S.; Lei, Y. Establishment of peanut mini core collection in China and exploration of new resource with high oleat. *Chin. J. Oil Crop Sci.* **2008**, *30*, 294–299.
20. Hui-Fang, J.; Xiao-Ping, R.; Bo-Shou, L.; Jia-Quan, H.; Yong, L.; Ben-Yin, C.; Guo, B.Z.; Holbrook, C.C.; Upadhyaya, H.D. Peanut core collection established in China and compared with ICRISAT mini core collection. *Acta Agron. Sin.* **2008**, *34*, 25–30. [CrossRef]
21. Holbrook, C.C.; Dong, W. Development and evaluation of a mini core collection for the US peanut germplasm collection. *Crop Sci.* **2005**, *45*, 1540–1544. [CrossRef]
22. Kottapalli, K.R.; Rakwal, R.; Shibato, J.; Burow, G.; Tissue, D.; Burke, J.; Puppala, N.; Burow, M.; Payton, P. Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. *Plant Cell Environ.* **2009**, *32*, 380–407. [CrossRef]
23. Smartt, J.; Gregory, W.C.; Gregory, M.P. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. *Euphytica* **1978**, *27*, 665–675. [CrossRef]
24. Seijo, G.; Lavia, G.I.; Fernández, A.; Krapovickas, A.; Ducasse, D.A.; Bertioli, D.J.; Moscone, E.A. Genomic relationships between the cultivated peanut (*Arachis hypogaea*, L.) and its close relatives revealed by double GISH. *Am. J. Bot.* **2007**, *94*, 1963–1971. [CrossRef] [PubMed]
25. Robledo, G.; Lavia, G.I.; Seijo, G. Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. *Theor. Appl. Genet.* **2009**, *118*, 1295–1307. [CrossRef] [PubMed]
26. Bertioli, D.J.; Cannon, S.B.; Froenicke, L.; Huang, G.; Farmer, A.D.; Cannon, E.K.; Liu, X.; Gao, D.; Clevenger, J.; Dash, S.; et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* **2016**, *48*, 438–446. [CrossRef]

27. Kim, K.S.; Lee, D.; Bae, S.B.; Kim, Y.C.; Choi, I.S.; Kim, S.T.; Lee, T.H.; Jun, T.H. Development of SNP-Based Molecular Markers by Re-Sequencing Strategy in Peanut. *Plant Breed. Biotechnol.* **2017**, *5*, 325–333. [CrossRef]
28. Moretzsohn, M.C.; Gouvea, E.G.; Inglis, P.W.; Leal-Bertioli, S.C.; Valls, J.F.; Bertioli, D.J. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* **2012**, *111*, 113–126. [CrossRef] [PubMed]
29. Nielen, S.; Vidigal, B.S.; Leal-Bertioli, S.C.; Ratnaparkhe, M.; Paterson, A.H.; Garsmeur, O.; D'Hont, A.; Guimaraes, P.M.; Bertioli, D.J. Matita, a new retroelement from peanut: Characterization and evolutionary context in the light of the *Arachis* A–B genome divergence. *Mol. Genet. Genom.* **2012**, *287*, 21–38. [CrossRef] [PubMed]
30. Temsch, E.M.; Greilhuber, J. Genome size variation in *Arachis hypogaea* and A. *monticola* re-evaluated. *Genome* **2000**, *43*, 449–451. [CrossRef] [PubMed]
31. Chen, X.; Li, H.; Pandey, M.K.; Yang, Q.; Wang, X.; Garg, V.; Li, H.; Chi, X.; Doddamani, D.; Hong, Y.; et al. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6785–6790. [CrossRef]
32. Chen, X.; Lu, Q.; Liu, H.; Zhang, J.; Hong, Y.; Lan, H.; Li, H.; Wang, J.; Liu, H.; Li, S.; et al. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant* **2019**, *12*, 920–934. [CrossRef]
33. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; et al. Genome sequencing in micro fabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376–380. [CrossRef] [PubMed]
34. Yang, H.; Tao, Y.; Zheng, Z.; Li, C.; Sweetingham, M.W.; Howieson, J.G. Application of next-generation sequencing for rapid marker development in molecular plant breeding: A case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genom.* **2012**, *13*, 318. [CrossRef] [PubMed]
35. Lee, J.; Izzah, N.K.; Jayakodi, M.; Perumal, S.; Joh, H.J.; Lee, H.J.; Lee, S.C.; Park, J.Y.; Yang, K.W.; Nou, I.S.; et al. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC Plant Biol.* **2015**, *15*, 32. [CrossRef]
36. Kang, Y.J.; Ahn, Y.K.; Kim, K.T.; Jun, T.H. Resequencing of Capsicum *annuum* parental lines (YCM334 and Taean) for the genetic analysis of bacterial wilt resistance. *BMC Plant Biol.* **2016**, *16*, 235. [CrossRef] [PubMed]
37. Zhou, X.; Xia, Y.; Ren, X.; Chen, Y.; Huang, L.; Huang, S.; Liao, B.; Lei, Y.; Yan, L.; Jiang, H. Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genom.* **2014**, *15*, 351. [CrossRef]
38. Pandey, M.K.; Monyo, E.; Ozias-Akins, P.; Liang, X.; Guimarães, P.; Nigam, S.N.; Upadhyaya, H.D.; Janila, P.; Zhang, X.; Guo, B.; et al. Advances in *Arachis* genomics for peanut improvement. *Biotechnol. Adv.* **2012**, *30*, 639–651. [CrossRef]
39. Pandey, M.K.; Agarwal, G.; Kale, S.M.; Clevenger, J.; Nayak, S.N.; Sriswathi, M.; Chitikineni, A.; Chavarro, C.; Chen, X.; Upadhyaya, H.D.; et al. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* **2017**, *7*, 40577. [CrossRef]
40. Varshney, R.K.; Mohan, S.M.; Gaur, P.M.; Gangarao, N.V.P.R.; Pandey, M.K.; Bohra, A.; Sawargaonkar, S.L.; Chitikineni, A.; Kimurto, P.K.; Janila, P.; et al. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnol. Adv.* **2013**, *31*, 1120–1134. [CrossRef]
41. Pandey, M.K.; Roorkiwal, M.; Singh, V.K.; Ramalingam, A.; Kudapa, H.; Thudi, M.; Chitikineni, A.; Rathore, A.; Varshney, R.K. Emerging genomic tools for legume breeding: Current status and future prospects. *Front. Plant Sci.* **2016**, *7*, 455. [CrossRef]
42. Otyama, P.I.; Wilkey, A.; Kulkarni, R.; Assefa, T.; Chu, Y.; Clevenger, J.; Anglin, N.L. Evaluation of linkage disequilibrium, population structure, and genetic diversity in the US peanut mini core collection. *BMC Genom.* **2019**, *20*, 481. [CrossRef]
43. Patil, A.S.; Popovsky, S.; Levy, Y.; Chu, Y.; Clevenger, J.; Ozias-Akins, P.; Hovav, R. Genetic insight and mapping of the pod constriction trait in Virginia-type peanut. *BMC Genet.* **2018**, *19*, 93. [CrossRef] [PubMed]
44. Peng, Z.; Zhao, Z.; Clevenger, J.P.; Chu, Y.; Paudel, D.; Ozias-Akins, P.; Wang, J. Comparison of SNP Calling Pipelines and NGS Platforms to Predict the Genomic Regions Harboring Candidate Genes for Nodulation in Cultivated Peanut. *Front. Genet.* **2020**, *11*, 222. [CrossRef] [PubMed]
45. Li, X.; Singh, J.; Qin, M.; Li, S.; Zhang, X.; Zhang, M.; Khan, A.; Zhang, S.; Wu, J. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol. J.* **2019**, *17*, 1582–1594. [CrossRef] [PubMed]
46. Tam, V.; Patel, N.; Turcotte, M.; Bossé, Y.; Paré, G.; Meyre, D. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **2019**, *20*, 467–484. [CrossRef]
47. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. [CrossRef]
48. Salem, M.; Al-Tobasei, R.; Ali, A.; Lourenco, D.; Gao, G.; Palti, Y.; Leeds, T.D. Genome-wide association analysis with a 50K transcribed gene SNP-chip identifies QTL affecting muscle yield in rainbow trout. *Front. Genet.* **2018**, *9*, 387. [CrossRef]
49. Bayer, M.M.; Rapazote-Flores, P.; Ganal, M.; Hedley, P.E.; Macaulay, M.; Plieske, J.; Waugh, R. Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* **2017**, *8*, 1792. [CrossRef]
50. Comadran, J.; Kilian, B.; Russell, J.; Ramsay, L.; Stein, N.; Ganal, M.; Hedley, P. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **2012**, *44*, 1388–1392. [CrossRef]

51. Allen, A.M.; Winfield, M.O.; Burridge, A.J.; Downie, R.C.; Benbow, H.R.; Barker, G.L.; Scopes, G. Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol. J.* **2017**, *15*, 390–401. [CrossRef]
52. Saghai-Maroof, M.A.; Soliman, K.M.; Jorgensen, R.A.; Allard, R.W.L. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 8014–8018. [CrossRef]
53. Ya, N.; Raveendar, S.; Bayarsukh, N.; Ya, M.; Lee, J.R.; Lee, K.J.; Shin, M.J.; Cho, G.T.; Ma, K.H.; Lee, G.A. Genetic diversity and population structure of Mongolian wheat based on SSR markers: Implications for conservation and management. *Plant Breed. Biotechnol.* **2017**, *5*, 213–220. [CrossRef]
54. Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.K.; Singh, N.K.; Singh, R. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE* **2013**, *8*, e84136. [CrossRef] [PubMed]
55. Tang, Y.; Liu, X.; Wang, J.; Li, M.; Wang, Q.; Tian, F.; Buckler, E.S. GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *Plant Genome* **2016**, *9*. [CrossRef] [PubMed]
56. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006**, *17*, 155–160. [CrossRef] [PubMed]
57. Bradbury, P.J.; Zhang, Z.; 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]
58. Zhang, C.; Dong, S.S.; Xu, J.Y.; He, W.M.; Yang, T.L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [CrossRef]
59. Chu, C.W.; Zhang, G.P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *Int. J. Prod. Econ.* **2003**, *86*, 217–231. [CrossRef]
60. Barrett, J.C.; Fry, B.; Maller, J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [CrossRef]
61. Lewontin, R.C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **1964**, *49*, 49–67.
62. Wu, T.T.; Chen, Y.F.; Hastie, T.; Sobel, E.; Lange, K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* **2009**, *25*, 714–721. [CrossRef]
63. Zhou, H.; Sehl, M.E.; Sinsheimer, J.S.; Lange, K. Association screening of common and rare genetic variants by penalized regression. *Bioinformatics* **2010**, *26*, 2375. [CrossRef] [PubMed]
64. Alexander, D.H.; Lange, K. Stability selection for genome-wide association. *Genet. Epidemiol.* **2011**, *35*, 722–728. [CrossRef] [PubMed]
65. Sun, H.; Wang, S. Penalized logistic regression for high-dimensional DNA methylation data with case-control studies. *Bioinformatics* **2012**, *28*, 1368–1375. [CrossRef] [PubMed]
66. Sun, H.; Wang, S. Network-based regularization for matched case-control analysis of high-dimensional DNA methylation data. *Stat. Med.* **2013**, *32*, 2127–2139. [CrossRef]
67. Okser, S.; Pahikkala, T.; Airola, A.; Salakoski, T.; Ripatti, S.; Aittokallio, T. Regularized machine learning in the genetic prediction of complex traits. *PLoS Genet.* **2014**, *10*, e1004754. [CrossRef]
68. Sun, H.; Wang, Y.; Chen, Y.; Li, Y.; Wang, S. pETM: A penalized Exponential Tilt Model for analysis of correlated high-dimensional DNA methylation data. *Bioinformatics* **2017**, *33*, 1765–1772. [CrossRef]
69. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]
70. Nicolai, M.; Bühlmann, P. Stability Selection: Stability Selection. *J. R. Stat. Soc. Ser. B* **2010**, *72*, 417–473. [CrossRef]
71. Kim, K.; Koo, J.; Sun, H. An empirical threshold of selection probability for analysis of high-dimensional correlated data. *J. Stat. Comput. Simul.* **2020**, *90*, 1606–1617. [CrossRef]
72. Zhang, S.; Hu, X.; Miao, H.; Chu, Y.; Cui, F.; Yang, W.; Wang, C.; Shen, Y.; Xu, T.; Zhao, L.; et al. QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* **2019**, *19*, 537. [CrossRef]
73. Hu, Z.; Zhang, H.; Kan, G.; Ma, D.; Zhang, D.; Shi, G.; Hong, D.; Zhang, G.; Yu, D. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* **2013**, *141*, 247–254. [CrossRef] [PubMed]
74. Bertioli, D.J.; Jenkins, J.; Clevenger, J.; Dudchenko, O.; Gao, D.; Samoluk, S.S. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature genetics* **2019**, *51*, 877–884. [CrossRef] [PubMed]
75. Otyama, P.I.; Kulkarni, R.; Chamberlin, K.; Ozias-Akins, P.K.; Chu, J.; Fernández-Baca, D.F. Genotypic characterization of the US peanut core collection. *BioRxiv* **2020**. [CrossRef]
76. Hammons, R.O.; Herman, D.; Stalker, H.T. Origin and early history of the peanut. In *Peanuts*; AOCS Press: Urbana, IL, USA, 2016; pp. 1–26. [CrossRef]
77. Simpson, C.E.; Krapovickas, A.; Valls, J.F.M. History of Arachis including evidence of A. *hypogaea L. progenitors. Peanut Sci.* **2001**, *28*, 78–80. [CrossRef]
78. Stalker, H.T.; Dhesi, J.S.; Kochert, G. Genetic diversity within the species *Arachis duranensis* Krapov. & W. C. Gregory, a possible progenitor of cultivated peanut. *Genome* **1995**, *38*, 1201–1212. [CrossRef]

79. Stalker, H.T.; Tallury, S.P.; Seijo, G.R.; Leal-Bertioli, S.C. Biology, speciation, and utilization of peanut species. In *Peanuts*; AOCS Press: Urbana, IL, USA, 2016; pp. 27–66. [CrossRef]
80. Simon, N.; Friedman, J.; Hastie, T.; Tibshirani, R. A sparse-group lasso. *J. Comput. Graph. Stat.* **2013**, *22*, 231–245. [CrossRef]
81. Li, C.; Li, H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics* **2008**, *24*, 1175–1182. [CrossRef]
82. Anderson, C.A.; Pettersson, F.H.; Clarke, G.M.; Cardon, L.R.; Morris, A.P.; Zondervan, K.T. Data quality control in genetic case-control association studies. *Nat. Protoc.* **2010**, *5*, 1564–1573. [CrossRef]
83. Rao, X.; Thapa, K.S.; Chen, A.B.; Lin, H.; Gao, H.; Reiter, J.L.; Hargreaves, K.A.; Ipe, J.; Lai, D.; Xuei, X.; et al. Allele-specific expression and high-throughput reporter assay reveal functional genetic variants associated with alcohol use disorders. *Mol. Psychiatry* **2019**, 1–10. [CrossRef]
84. Oliveira, J.C.D.; Rufino, P.B.; Azêvedo, H.S.F.D.S.; Sousa, A.C.B.D.; Assis, G.M.L.D.; Silva, L.M.D.; Campos, T.D. Inferring mating system parameters in forage peanut, *Arachis pintoi*, for Brazilian Amazon conditions. *Acta Amazonica* **2019**, *49*, 277–282. [CrossRef]
85. Baud, S. Seeds as oil factories. *Plant Reprod.* **2018**, *31*, 213–235. [CrossRef]
86. Smith, A.J.; Rinne, R.W.; Seif, R.D. Phosphoenolpyruvate carboxylase and pyruvate kinase involvement in protein and oil biosynthesis during soybean seed development. *Crop Sci.* **1989**, *29*, 349–353. [CrossRef]
87. Feria, A.B.; Alvarez, R.; Cochereau, L.; Vidal, J.; García-Mauriño, S.; Echevarría, C. Regulation of phosphoenolpyruvate carboxylase phosphorylation by metabolites and abscisic acid during the development and germination of barley seeds. *Plant Physiol.* **2008**, *148*, 761–774. [CrossRef] [PubMed]
88. Aivalakis, G.; Dimou, M.; Flemetakis, E.; Plati, F.; Katinakis, P.; Drossopoulos, J.B. Immunolocalization of carbonic anhydrase and phosphoenolpyruvate carboxylase in developing seeds of Medicago sativa. *Plant Physiol. Biochem.* **2004**, *42*, 181–186. [CrossRef] [PubMed]
89. Golombek, S.; Rolletschek, H.; Wobus, U.; Weber, H. Control of storage protein accumulation during legume seed development. *J. Plant Physiol.* **2001**, *158*, 457–464. [CrossRef]
90. Pan, L.J.; Yang, Q.L.; Chi, X.Y.; Chen, M.N.; Zhen, Y.A.N.G.; Na, C.H.E.N.; Yu, S.L. Functional analysis of the phosphoenolpyruvate carboxylase on the lipid accumulation of peanut (*Arachis hypogaea* L.) seeds. *J. Integr. Agric.* **2013**, *12*, 36–44. [CrossRef]

genes

MDPI

Article

Mapping-by-Sequencing via MutMap Identifies a Mutation in *ZmCLE7* Underlying Fasciation in a Newly Developed EMS Mutant Population in an Elite Tropical Maize Inbred

Quan Hong Tran [1,†], Ngoc Hong Bui [1,†], Christian Kappel [2], Nga Thi Ngoc Dau [1], Loan Thi Nguyen [1], Thuy Thi Tran [1], Tran Dang Khanh [1,3], Khuat Huu Trung [1], Michael Lenhard [2] and Son Lang Vi [1,*]

1 Department of Genetic Engineering, The Agricultural Genetics Institute, Km2 Pham Van Dong Street, Hanoi, Vietnam
2 Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Straße 24–25, House 26, 14476 Potsdam-Golm, Germany
3 Center for Expert, Vietnam National University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam
* Correspondence: vilangson@gmail.com
† These authors contributed equally.

Received: 14 February 2020; Accepted: 3 March 2020; Published: 6 March 2020

Abstract: Induced point mutations are important genetic resources for their ability to create hypo- and hypermorphic alleles that are useful for understanding gene functions and breeding. However, such mutant populations have only been developed for a few temperate maize varieties, mainly B73 and W22, yet no tropical maize inbred lines have been mutagenized and made available to the public to date. We developed a novel Ethyl Methanesulfonate (EMS) induced mutation resource in maize comprising 2050 independent M2 mutant families in the elite tropical maize inbred ML10. By phenotypic screening, we showed that this population is of comparable quality with other mutagenized populations in maize. To illustrate the usefulness of this population for gene discovery, we performed rapid mapping-by-sequencing to clone a fasciated-ear mutant and identify a causal promoter deletion in *ZmCLE7 (CLE7)*. Our mapping procedure does not require crossing to an unrelated parent, thus is suitable for mapping subtle traits and ones affected by heterosis. This first EMS population in tropical maize is expected to be very useful for the maize research community. Also, the EMS mutagenesis and rapid mapping-by-sequencing pipeline described here illustrate the power of performing forward genetics in diverse maize germplasms of choice, which can lead to novel gene discovery due to divergent genetic backgrounds.

Keywords: EMS; MutMap; mutagenesis; CLE7; tropical maize; fasciation; mapping

1. Introduction

Maize is an important staple crop worldwide used for food, feed and industrial products including biofuels [1]. The demand for maize has steadily risen in past decades and is expected to continue to increase, with maize demand predicted to double by 2050 [2,3]. Maize is not only of agronomic importance; it has also been a very important model organism for genetic studies.

Forward genetics in maize has been very fruitful and there are several mutant populations in maize; however, most of these are in standard inbred backgrounds. Most induced mutants in maize were created by two methods: pollen mutagenesis with Ethyl Methanesulfonate (EMS) or crossing with maize lines that carry highly active transposons [4]. As a mutagen, EMS has the unique advantage that it can potentially create allelic series ranging from null, hypomorphic, hypermorphic to gain-of-function

alleles, which can be very useful for studying gene function and applied breeding. For introducing point mutations, treating pollen with EMS has been the preferred method, because it is much more efficient than treating maize seeds or the use of radiation [4,5]. Using pollen mutagenesis, Prof. Gerald Neuffer has created a large number of mutants that have been used by researchers globally till now [6]. Several additional mutagenesis populations have been created: a TILLING population in B73 comprising 750 M1 plants [7], an indexed population in B73 that has been exome-sequenced comprising 1086 M1 plants [8], and other populations in W22 [7] and in the Iodent inbred PH207 [9]. By transposon mutagenesis, several large mutation populations have been created for reverse genetics (reviewed in [10]): the rescueMu created in A188 × B73 (HiII hybrid) [11]; the MTM Mu in B73 [12]; TUSC (Trait Utility System for Corn) in mixed background developed by the Pioneer Hi-Breed Company, the Mu population developed by Biogemma, UniformMu in W22 [13,14], the Photosynthetic Mutant Hunt population [15], and the ChinaMu in B73 [16]. A γ-irradiation mutant population in B73 was also created [17]. Despite the tremendous usefulness of these populations, in the South Asia tropical area, for example in Vietnam, the temperate standard maize inbreds (e.g., B73, W22, Mo17) grow very poorly, are highly susceptible to leaf disease, and have large anthesis-silking intervals making them very difficult to conduct genetics with.

Diversifying the genetic resources can help identify novel genes, whose function is masked by the genetic background of the standard inbreds, and is essential for studying adaptive, abiotic and biotic-stress related traits. Tropical maize inbreds have proved to be very valuable germplasm resources for several important traits. For examples, two drought-tolerant genes were cloned from the tropical inbred CIMBL55 [18,19], a head smut resistance and other disease resistance genes were identified from a Thailand Suwan and other tropical maize varieties [20–23]. Many of the studies on exotic germplasm in maize exploit natural variation, for example using GWAS [24]. So far, no artificial mutant population has been made in a non-standard maize inbred to allow forward genetic studies directly on the exotic germplasms.

The lack of materials is not the only issue; to successfully conduct forward genetic studies, one must consider standing challenges in mutant mapping/map-based cloning. First, maize has a large and complex genome (~2300 Mb) with lots of repetitive and hypervariable regions [25,26]. As such, the maize genome is several times larger than that of the two model plant species *Arabidopsis thaliana* (~135 Mb) and rice (~430 Mb). Second, heterosis is so strong and prevalent in maize that many mutant phenotypes cannot be scored accurately in F2 mapping populations of the original mutant outcrossed with an unrelated inbred. Subtle, quantitative, and physiology-based phenotypes such as kernel row number, plant height, root length, flowering time, nitrogen use efficiency, abiotic and biotic-stress related phenotypes, etc. require an isogenic background to correctly classify phenotypes. Third, background modifiers from the second parent, which are common in maize, can interfere with phenotype scoring. Therefore, in many cases several rounds of backcrossing, each of which can show different modifying effects, and careful pedigree/segregation analysis are required to create good mapping populations. This process is time-consuming and sometimes impossible for complex traits and genetics.

To solve these difficulties, several approaches combining next generation sequencing and mapping populations generated in the same genetic background have been developed. These include crossing to the unmutagenized parent in the Mutmap method in rice [27] and in maize [28], or crossing to a normal looking plant of a different M2 family (the so called "evil twin" in Sorghum [29]) or even not crossing at all, but only sequencing M3 plants as in Mutmap + (rice) [30], or sequencing individual M2 plants combined with zygosity analysis via progeny in maize [9]. These approaches utilized only the segregating mutagen-induced variants between the two parents or siblings to map, and hence they all shared the limitation of relying on the low number of mutagen-induced variants. It is unclear in maize whether these so called modified versions of the Mutmap approach can robustly identify causal mutations given the predicted low number of mutagen-induced variants for such a big, complex genome, especially when using non-reference germplasm.

In this study, we sought to create a new mutant population resource and establish a simple and rapid mutant mapping approach to facilitate the forward genetic study in a non-standard maize variety. Using the EMS pollen-mutagenesis method, we have created a high-quality mutant population in an elite tropical inbred ML10, one of the inbred parents for a very popular maize hybrid in tropical South Asia. We have used a modified Mutmap pipeline in maize that is simple, cost-effective and does not require crossing to an unrelated inbred. Instead the mutant in question was crossed to a ML10 line that was mutagenized for two consecutive generations with the aim to introduce more segregating variants for mapping. Despite this, the mutation was mapped based solely on the EMS-induced mutations in the focal mutant relative to the unmutagenized background, suggesting that even simple backcrossing to the unmutagenized background should be sufficient. We have illustrated the usefulness of this population and our mapping-by-sequencing strategy by cloning a fasciation mutant and identifying a promoter deletion in *ZmCLE7*.

2. Materials and Methods

2.1. Mutagenesis, Phenotypic Screening and Plant Growth Conditions

ML10 is the maternal parent of the popular elite hybrid LVN10 in Vietnam [31]. ML10 was used to create the mutagenized population. The inbred was developed by selfing followed by sib-mating and was released in 1994. The pedigree is unclear, but historical records indicate that it was developed from a CP hybrid (Charoen Pokphand Group, C.P. Tower, 313 Silom Road, Bangrak, Bangkok 10500, Thailand). BL10 is the paternal parent of hybrid LVN10, also developed by selfing and maintained by sib-mating. B73 is the reference inbred strain and was obtained from Professor David Jackson, Cold Spring Harbor Laboratory. Maize was planted in two seasons in Ha Noi, Vietnam (map coordinate: 21°06'23.8" N 105°49'34.7" E): for the spring season, sowing was done in January and for the autumn season, sowing was done in August. These sowing times were optimized as shown in Table 1. The EMS pollen mutagenesis was carried out according to [5]. In brief, maize pollen was collected in the morning from tassel bags that had been put on the tassels overnight. To prepare the EMS stock solution, EMS (Sigma M0880-1G) was diluted to 1% (v/v) in mineral oil (Sigma 69794-500ML); the bottle was covered in aluminum foil, and mixed thoroughly by a magnetic stirrer overnight before the mutagenesis experiment. This stock solution was used on the next day for making the final EMS solutions in mineral oil. About 12−20 mL pollen was incubated with 70 mL of mineral oil containing EMS at the concentration stated in the results for 30 min. The pollen mixtures were then pipetted onto the silks. The pollinated ears were covered with shoot bags. The concentration of EMS of 0.1% (v/v) was used to generate a large population of over 6000 M1 seeds. The resulting M1 plants were then grown on the field and selfed to create M2 families. 40 seeds per M2 ear were sown and scored for visible phenotypes. Researchers can obtain seeds from this mutant population and/or arrange a screen for the specific phenotype in this population for research/breeding purpose via a Material Transfer Agreement by directly contacting the corresponding author S.L.V.

Table 1. Optimized sowing date for EMS mutagenesis in ML10.

Season	Sowing Date	Pollen Shedding Date	Estimated Time from Sowing to Shedding
Spring	Jan 11 (±10 d)	Apr 5 (±10 d)	85 days
Autumn	Aug 8 to Sep 8	Oct 12 to Nov 12	65 days

2.2. Generating E-ML10

E-ML10 plants (the "Evil-twin" ML10 or "Evil" ML10) were ML10 mutagenized for two successive generations with 0.1% EMS. The first round of EMS was the same as in the generation of the large mutant population in Section 2.1. In the second round, we selected healthy wild-type looking M1 plants, which did not exhibit any visible dominant/deleterious mutant phenotypes, pooled pollen from these plants, treated the pollen with 0.1% EMS and used it to pollinate the sibling M1 plants from

which the pollen was harvested. The E-ML10 used for the cross with E1-9 to generate the mapping population was a wild-type looking individual grown from a mixture of seeds after the second round of EMS.

2.3. Mapping of E1-9

E1-9 was identified as an ear fasciation mutant in an M2 family. After first identification, this M2 family was re-grown; eight randomly selected M2 plants were selfed. The resulting M3 families were analyzed to select homozygous mutants. Three F2 mapping populations were made by crossing the M4 homozygous mutant E1-9 to a second parent (E-ML10, BL10 or B73) to make F1s; multiple F1 plants were selfed to generate F2 families.

Four samples used for next generation sequencing were one individual unmutagenized ML10 wild-type plant (named ML10), a pool of eleven mutant plants from a M5 homozygous mutant family (named E1-9), a pool of 72 plants with a mutant phenotype (named F2-mutant-pool) and a pool of 69 plants with wild-type phenotype (named F2-WT-pool) from the F2 mapping population E1-9 × E-ML10. The F2 plants in each pool were derived from four selfed F1 plants.

2.4. DNA Extraction

For PCR, DNA samples were extracted from maize leaves using CTAB protocols [32], then kept at −20 °C for long term storage. For next generation sequencing, DNA was first extracted with CTAB, treated with RNAse A, cleaned up with Phenol:Chloroform, then cleaned up one more time with Plant DNA mini extraction kit (Thermo). The DNA quality was checked by Nanodrop and gel electrophoresis.

2.5. Sequencing and Bioinformatic Analysis

Extracted DNA samples were sent to Macrogen for library preparation with TruSeq DNA PCR-Free kit and sequenced using 151 bp paired-end Illumina platform. We obtained 185,173,048 fragments for the ML10 wild-type pool, 173,761,432 for the M5 mutant, 207,190,254 for the pool of F2 wild-type individuals, 235,111,914 for the pool of F2 mutant individuals. Reads were mapped to the B73 reference genome (*Zea mays*, Ensemble release 40, http://plants.ensembl.org) using bwa mem [33] and further processed using samtools [34]. Variant calling was performed using bcftools [35,36]. Data analyses were done using R (R Core Team, 2017. https://www.R-project.org/), plots were generated using R/lattice [37] (http://lmdvr.r-forge.r-project.org). Variant callings with quality scores below 500 were discarded. Allele frequencies (ratio of sample over reference B73) for F2 pools were plotted for variants fixed between the E1-9 mutant sample (allele frequency equal to 1) and the ML10 wild-type sample (allele frequency equal to 0) and having an allele frequency between 0.25 and 0.75 in the F2 wild-type pool sample. Loess smoothing for average allele frequencies was plotted to visualize more general tendencies. Coding sequences (CDS) for known fasciation genes were extracted manually using IGV [38]. Sequences were translated to peptides using transeq and then multiple sequences aligned using muscle [39].

Sequencing data are available at NCBI SRA under accession number PRJNA602200.

2.6. PCR Analysis

All primers used were listed in Table S1. PCR was performed according to standard procedures. The PCR mix for a 20 μL reaction was: Taq 2X Mastermix (NEB): 10 μL; template DNA: 4 μL; primers (10 μM): 0.8 μL each, DMSO: 1 μL, deionised water to 20 μL total volume. For genotyping the *cle7* promoter deletion, the PCR thermal condition was 95 °C: 5 min; then 39 cycles of: 94 °C: 30 s, 59 °C: 30 s, 72 °C: 1 min; ending with 72 °C 10 min. For SSR analysis, the PCR thermal condition was 95 °C: 5 min; then 39 cycles of: 95 °C: 30 s, 55 °C: 40 s, 72 °C: 40 s; ending with 72 °C 10 min.

3. Results

3.1. Optimization of EMS Protocol for the ML10 Inbred and Development of the Mutant Population

ML10 is the maternal parent of a very popular hybrid, LVN10, in the South East Asian region, e.g., Vietnam, Laos and Cambodia [31]. The inbred was developed in the 1990s by phenotypic selection and compatibility testing. Due to its popularity, high adaptability, high combining ability, excellent seed germination and the ability to robustly shed non-clumpy pollen even under harsh environmental conditions, we chose ML10 as the inbred background for pollen mutagenesis.

We tested several EMS concentrations: 0.04%–0.06%–0.1%–0.185%. EMS treatment at 0.185% gave almost no seeds, while EMS treatment at 0.04, 0.06 and 0.1% gave seeds. M1 seeds from the 0.1% EMS experiment had a germination rate of approximately 50% compared to ML10 wild-type. As we wanted to make a population with as high a mutation density as possible, an EMS concentration of 0.1% was chosen to generate a large mutagenized population. Ears pollinated with 0.1% EMS-treated pollen contained on average 30 seeds/ear compared to approximately 220 seeds per ear in untreated plants (Figure 1A). This is also the concentration recommended for B73 in [5] and higher than the concentration used for B73 in [8] of 0.067%. We noticed that the weather condition during pollination is very important; in particular, partially sunny weather with temperatures from 28 to 32 °C and humidity from 50% to 70% seemed ideal, as otherwise the pollen got clumpy very quickly after harvest and had reduced viability. For the tropical conditions in Ha Noi, Vietnam, the best time for pollination to get this optimum weather is the first two weeks of April for the Spring season and from October the 12th to November the 12th for the Autumn season (see Methods).

Using the optimized sowing time and EMS concentration of 0.1%, we mutagenized ML10 and obtained over 6000 M1 seeds. We sowed out M1 seeds, selfed the resulting M1 plants, and got 2050 M2 families. 1185 M2 families were phenotyped to find mutants, and results are summarized in Table 2 and Figure 1B–I. Numerous mutants were found, many of which have phenotypes similar to classical mutants in maize like *sh2*, *su1*, *ramosa*, *fea*, *liguleless*, *dwarf*, etc., [6] demonstrating that our EMS mutagenesis generated a large number of induced mutations.

Table 2. Frequency of typical mutants observed among 1185 M2 families screened.

Phenotype	Number of M2 Families	Frequency (%)
Defective kernel	33	2.78
Small kernel	7	0.59
Dwarf plants	4	0.34
Kernel color	4	0.34
Narrow leaf	4	0.34
Albino	3	0.25
Purple stalk/leaf	3	0.25
Fasciated ear	2	0.17
Chlorotic lesion leaf	2	0.17
Liguleless/upright leaf	2	0.17
Wrinkled kernel	3	0.25
Tassel branch angle	2	0.17
Tassel branch number	2	0.17
Tassel seed	2	0.17
Branched ear (*ramosa*)	1	0.08
Anther color	1	0.08
Broad leaf	1	0.08
Early senescence	1	0.08

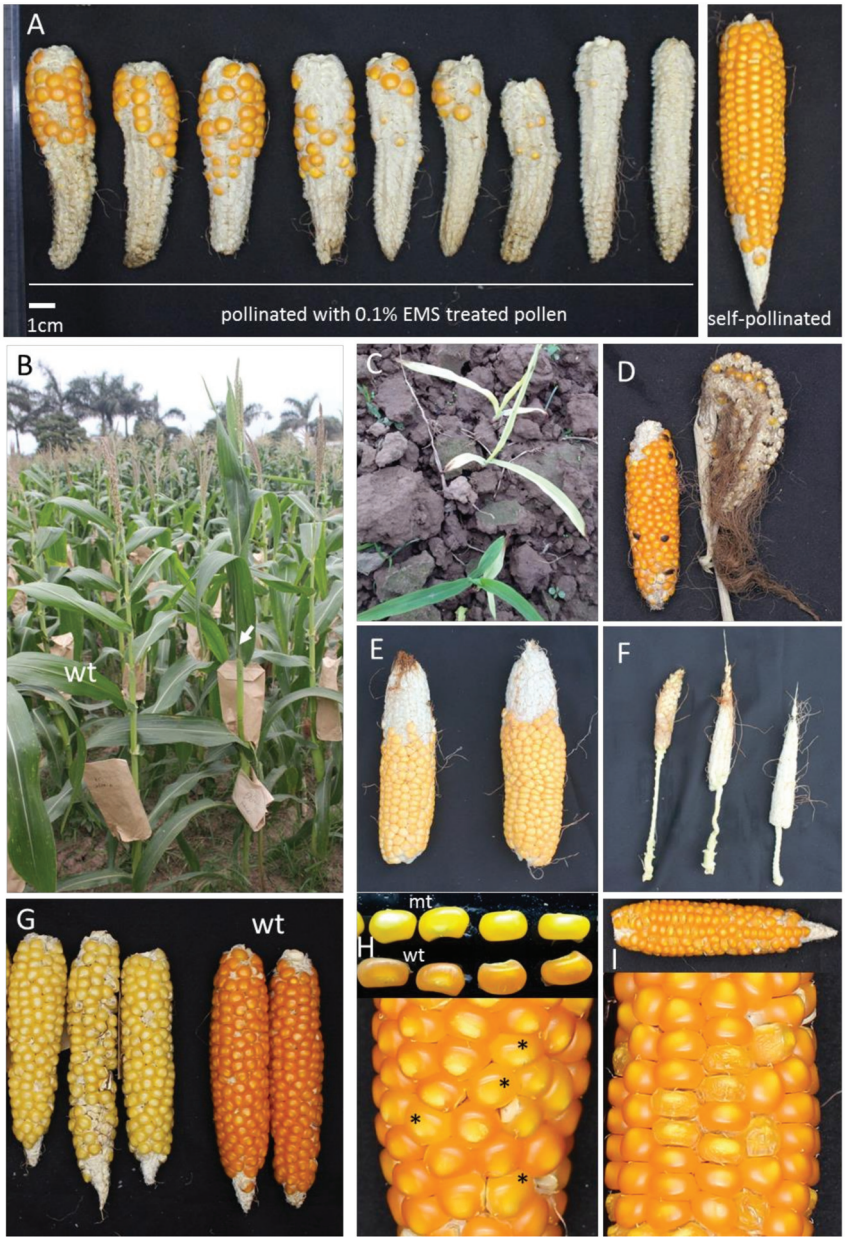


Figure 1. Typical M_0 ears and mutant phenotypes observed in the M_1, and later-generation mutagenized plants. (**A**) Typical M_0 ears of ML10 that were pollinated with 0.1% EMS-treated pollen and a control self-pollinated ear. (**B–I**) Examples of mutant phenotypes observed. (**B**) A dominant M1 *liguleless* mutant with upright leaf angle (arrow) next to wild-type (wt) control. (**C**) Albino. (**D**) Tassel-seed (ear and tassel). (**E**) Fasciated ear. (**F**) Small ear with long shank. (**G**) White kernels; wt ears have orange kernels. (**H**) Yellow kernel mutant (mt) (kernels with asterisk *). (**I**) Wrinkled kernels *sugary*.

3.2. Development of a Mapping-by-Sequencing Method via Mutmap to Map the E1-9 Mutant

3.2.1. F2 Mapping Population Generated with a Heavily Mutagenized ML10 (E-ML10)

A challenge for mapping mutations in maize is that when crossed to an unrelated parent to generate an F2 mapping population, mutant phenotypes may vary due to residual heterosis and segregating genetic modifiers. To circumvent this problem, we developed a heavily mutagenized ML10, named the E-ML10 for "Evil-twin" or "Evil" ML10, which underwent two generations of EMS pollen mutagenesis, and used it as an alternative parent to generate F2 mapping populations (Figure 2B). We reasoned that this E-ML10 will contain many EMS-induced SNPs compared to the original ML10 that can be used as molecular markers, abundant enough for fine mapping. Once an F2 population was made between E-ML10, and the mutant and phenotype scoring were successful, we used a similar method for mapping-by-sequencing as MutMap in rice [27] or BSA-Seq/MutMap in maize [28].

As the first pilot experiment to test the possibility of mapping without the need to cross to an unrelated inbred, we mapped the mild fasciation mutant E1-9 identified in our screen. The E1-9 mutant had its ear tip mildly fasciated and the ear had a higher number of kernel rows than wild-type (Figure 2A). Fasciation phenotypes can be subtle and difficult to score due to phenotypic suppression after crossing to an unrelated inbred; therefore, this phenotype was suitable to test the above mapping method. We crossed the homozygous M4 mutant to the E-ML10, and in the F2 derived from four selfed F1 plants we obtained 72 fasciated mutants from a total of 280 plants. This segregation ratio suggested that the fasciation phenotype in E1-9 was caused by a single recessive mutation. We sequenced pooled DNA from 72 fasciated mutants (F2-mutant pool) and 69 wild-type looking individuals (F2-WT pool) from this F2 mapping population. As the ML10 whole-genome sequence had not yet been available, we also sequenced one original unmutagenized ML10 individual (ML10), and a pool of eleven M5 homozygous E1-9 mutants (E1-9) to aid variant calling (Figure 2B).

3.2.2. Sequencing and Bioinformatic Analysis

Using the Illumina Hi-Seq paired-end sequencing platform, we obtained 52−71 Gb data per sample with > 20× coverage of the maize genome (2300 Mb) (Table 3). Reads were mapped to the B73 reference genome and variants were called.

Table 3. Sequencing data summary.

Sample	F2 Mutant Pool	F2 Wild-Type Pool	E1-9	ML10
Number of reads	470,223,828	414,380,508	347,522,864	370,346,096
Number of mapped reads	462,796,855	407,731,780	344,923,846	361,795,129
% of mapped reads	98.42%	98.40%	99.25%	97.69%
Properly paired reads	405,367,374	359,813,078	305,113,228	320,163,016
% of properly paired reads	86.21%	86.83%	87.80%	86.45%
Coverage (x)	26	23	19	21

To identify EMS-induced segregating variants in the F2 for mapping, we used only variants that were fixed between the M5 mutant (having an alternative allele frequency of 1) and the ML10 wild-type samples (having an alternative allele frequency of 0), and had an allele frequency between 0.25 and 0.75 in the F2 wild-type pool samples. This allele-frequency filter was used to exclude EMS-induced SNPs from E-ML10 that were only present in individual F1 plants from which part of the F2 was derived. Of note, because the F2 was derived from more than one selfed F1 plant, this filter means that we only based the mapping on EMS-induced variants present in the E1-9 line versus the unmutagenized ML10 and discarded other induced SNPs from E-ML10, as their expected frequencies could not be determined due to the pedigrees used (Figure 2B). We then plotted the allele frequencies of these filtered variants along the 10 chromosomes of maize to identify variants enriched in the F2-mutant pool, but not in the F2-wt pool (Figure 2C). This identified a broad region of distorted segregation

in the mutant pool on chromosome 4, with the strongest signal at the very top of this chromosome. In addition, we observed clusters containing a very high number of variants on chromosomes 1, 2, 3, 5 and 6; these likely reflect residual heterozygosity in the starting ML10 material.

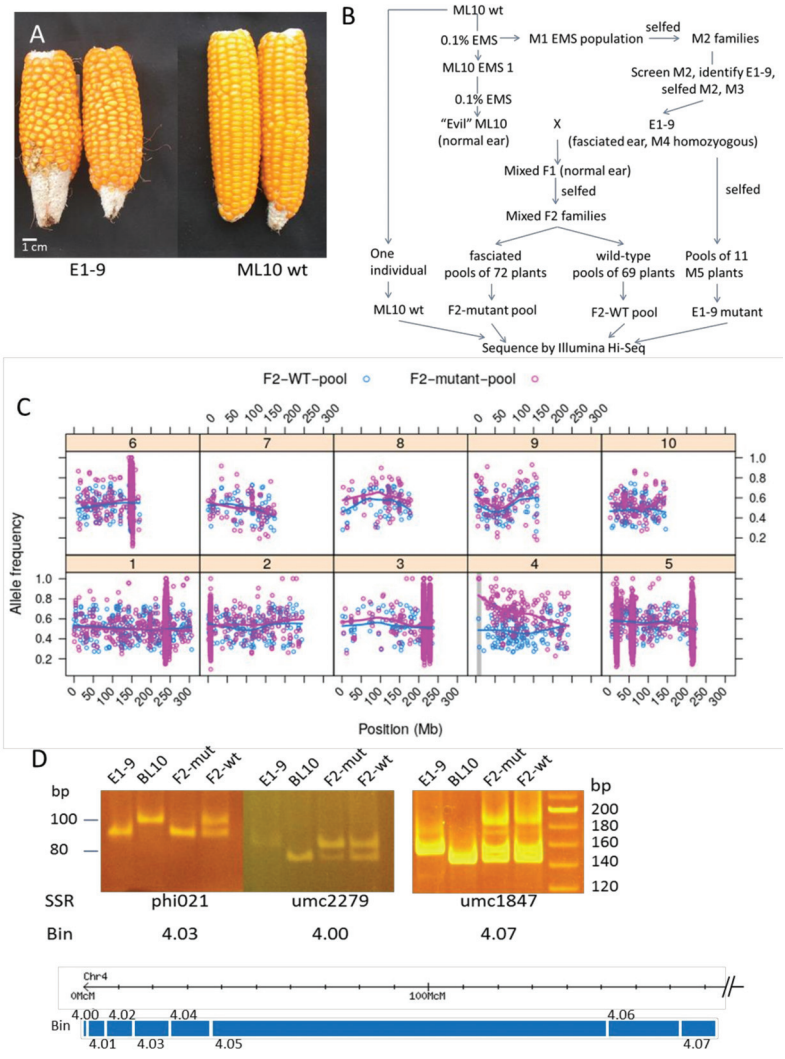


Figure 2. A modified Mutmap to map the E1-9 fasciation mutant by crossing with the "Evil" ML10. (**A**) Ears of E1-9 fasciation mutant and the wild-type ML10. (**B**) Crossing scheme and material preparation for bulk-sequencing. (**C**) Allele frequencies of filtered mutant vs wild-type markers for F2 mutant (purple) and F2 wild-type (blue) pools shown as dots. Lines represent Loess smoothing through allele frequency averages over 1 Mb windows. Vertical gray bar shows *CLE7* position. In the clusters with high variant densities on chromosomes 1, 2, 3, 5 and 6 the dots for the wild-type pools are obscured by the dots for the mutant pools. (**D**) SSR marker analysis in E1-9/BL10 F2 families confirms that the E1-9 mutation is linked to the top of chromosome 4. DNA samples are E1-9 and BL10 parents and their F2 pools of plants with mutant (F2-mut) and with wild-type phenotypes (F2-wt). Below is the genetic map of Bins on chromosome 4 (adapted from [6]). *CLE7* is in bin 4.02.

Similar to these observations, Klein et al. [28] also found that the B73 line used for their mutagenized population contained segments that differed from the B73 reference genome. Liang et al. [40] sequenced several B73 stocks from different laboratories and also reported the presences of clearly defined genomic blocks containing haplotypes that differ from the published B73 reference genome. Similarly, such residual heterozygosity in our ML10 material would explain why the E1-9 mutant contained genomic segments that differed from the E-ML10 used for crossing and the sequenced unmutagenized ML10 individuals. Thus, using our pipeline the E1-9 mutation was mapped to the top of chromosome 4.

3.2.3. Linkage Confirmation by PCR Analysis and Identification of Causal Mutation

To confirm the location of the mutation, bulk linkage analysis by SSR markers was performed in an independent F2 mapping population (E1-9 crossed with BL10). Results showed that the E1-9 mutation was linked to marker *phi021* in bin 4.03 on top of chromosome 4 (Figure 2D). These results showed that our map-by-sequencing approach allowed the rapid identification of the putative region harboring the causal mutation.

To identify the causal mutation in this region, we further looked at the above-filtered SNPs that were homozygous in the F2-mutant-pool sample and found only one such SNP at position Chr4: 8,764,229 (C in ML10 and T in E1-9). However, this SNP was in the intergenic region, and was 3 kb downstream from the closest gene LOC103655072, which encodes for a *Pectinesterase Inhibitor 38*. Hence, this SNP was unlikely to be the causal mutation in E1-9.

Therefore, we also checked the nucleotide sequences of candidate genes whose mutations are known to cause ear fasciation including *CT2* [41], *FEA2* [42], *TD1* [43], Zm*GB1* [44], *FEA3* [45], *FEA4* [46], and *CLE7* [47,48]. No differences were found in the coding region of these genes (Supplementary data) between ML10 and E1-9. However, we found a fixed 376 bp deletion (Chr4: 8,337,361−8,337,738) in the promoter of *CLE7* in E1-9 and F2-mutant-pool, but not in ML10, and segregating in the F2-wt pool (Figure 3A,B, Figure S1).

CLE7 lies in bin 4.02 on top of chromosome 4, very close to the region identified in our whole genome sequencing analysis. As CRISPR-Cas9 derived *cle7* maize mutants have recently been shown to cause ear fasciation [48], this makes the promoter deletion in *cle7* in our E1-9 mutant the prime candidate for the causal mutation. The *cle7-a1*, *cle7-a2* CRISPR-Cas9 derived alleles cause a deletion of one nucleotide at + 10 and + 11 after the start of translation, respectively, hence resulting in a frameshift after the third amino acid in the protein and causing a premature stop codon in these likely null-alleles. Our mutation is a 376bp deletion at position −2 relative to the ATG start codon, which is likely to be a strong, close-to-null mutant. The large promoter deletion explains why we did not find this variant in our bioinformatic analysis of the Mapping-by-Sequencing data, but only identified the closely linked marker SNP. Although the mechanism is unclear, EMS has been shown before to cause insertions/deletions [8,49].

To test for co-segregation of mutant phenotype and the promoter deletion in *CLE7*, we designed a PCR marker for the deletion (Figure 3B) and tested it in two different F2 populations. In one F2 (E1-9 crossed with B73), where a 1:3 segregation ratio of the mutant phenotype was observed (33 mutants: 95 wt plants), there was perfect co-segregation of the mutant phenotype with the homozygous deletion in all 33 mutant plants (Figure 3C). This result supports the hypothesis that the deletion in *CLE7* was the causal mutation in E1-9. In a different F2 (E1-9 crossed with BL10), we observed a much lower frequency of the mutant phenotype (18 mutants: 96 wt plants), significantly different from a 1:3 ratio (Chi-square test, $p = 0.02$). Although all of the 18 plants with the mutant phenotype were homozygous for the promoter deletion, confirming co-segregation, three plants, whose ears looked normal, were homozygous for the promoter deletion (Figure 3D). These three plants were genotypically mutant, but phenotypically wild-type, suggesting that the mutant phenotype was suppressed in these plants.

Taken together, these results strongly suggest that E1-9 was a *cle7* mutant; interestingly, they also suggest that there was a suppressor of the *cle7* mutant phenotype in the BL10 background.

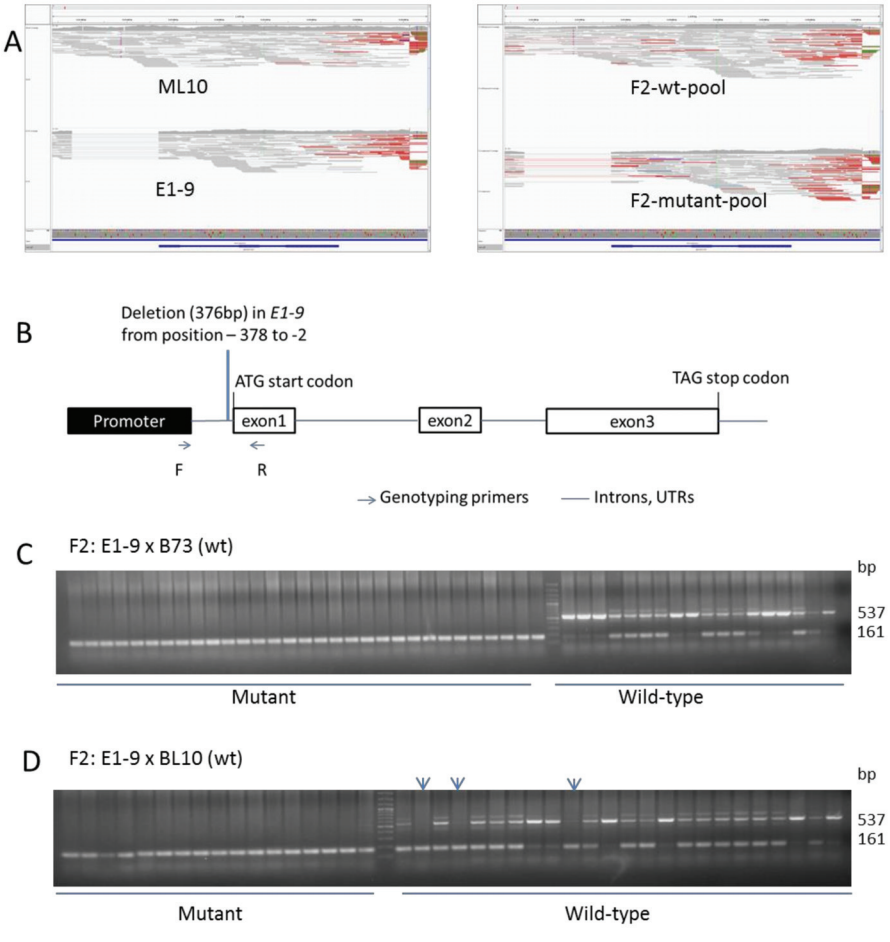


Figure 3. Identification of a promoter deletion in *CLE7* in E1-9. (**A**) IGV screenshot showing the alignments of reads to the B73 reference genome for ML10, E1-9, F2-wt pool and F2-mutant pool at the *CLE7* locus. (**B**) Illustration of the *CLE7* gene and the position of the promoter deletion in the E1-9 mutant. Genotyping primers were designed such that the wild-type allele gives a product of 537 bp, while the mutant allele gives a product of 161 bp. (**C**) and (**D**) Genotyping results for the phenotypically mutant and wild-type plants from two F2 populations made from crosses between the homozygous E1-9 mutant to B73 (**C**) and to BL10 (**D**), respectively. (**D**) In the F2 with BL10, arrows point at samples from plants with wild-type looking ears that had mutant genotypes, suggesting that BL10 contains a suppressor(s) of the *cle7* mutant phenotype.

4. Discussion and Conclusions

In this study, we successfully mutagenized an exotic elite tropical maize inbred by EMS treatment on pollen. We generated a 2050-family strong M2 population and identified numerous mutants with visible phenotypes. Many mutants identified in our population suggest the usefulness of this population as a complement to the B73 mutant collection. For example, B73 has yellow seeds while ML10 has orange seeds, and we identified several mutants that change the orange color of ML10 seeds to yellow and to white (Figure 1G–H). These mutants will be useful in identifying the natural variation that boosts provitamin A content in maize seeds [50,51]. Despite extensive mutagenesis screens,

mutants in *CLE7* had not been found in any EMS populations of standard inbreds, but were identified in our population. Since our allele was created by EMS mutagenesis in an elite parent, it can directly be used in a breeding program to improve kernel row number (KRN) without transgenic regulation as compared to the CRISPR-Cas9 alleles. Our results suggest that there was a genetic suppressor(s) of the *cle7* phenotype in BL10, the paternal parent in the elite hybrid paired with maternal ML10. Combining this suppressor and *cle7* in ML10/BL10 hybrids could lead to an improved hybrid with a higher KRN and yield without being fasciated. Mapping of this suppressor could also identify a new factor controlling meristem proliferation. The presence of a suppressor of the *cle7* mutant phenotype in the BL10 background, but not in B73 emphasizes the value of using diversified non-reference genotypes in genetic studies, allowing the study of natural suppressors and enhancers.

Our mapping-by-sequencing method is similar to MutMap in rice [27]. Like Mutmap, our mapping procedure does not require crossing to an unrelated inbred, which makes scoring of phenotypes easier and more accurate. First introduced in rice with a genome size of 370 Mb, MutMap was successfully used to map several mutants using fewer than 2300 segregating EMS-induced variants. For maize with an 8× bigger genome and many more repetitive sequences, it was questionable whether this method would work, in particular when using exotic germplasm for which no reference genome is available. To our knowledge, our study is the second example of MutMap in maize where EMS-derived variants were used for mapping without the need to cross to an unrelated inbred. The first example was from [28], where pools of nine F2 mutants and of nine wild-type siblings were sequenced, as well as the unmutagenized parent of unknown genetic background. Mutants with subtle phenotypes or ones whose phenotype is affected by residual heterosis, which is very common in maize, will benefit most from this approach. Our bioinformatic pipeline quickly identified the chromosomal region carrying the mutation based on only 1118 segregating variants spread largely evenly across the 10 chromosomes and pools of 69 wild-type and 72 mutant F2 plants. The number of variants used in our mapping was two-fold less than in MutMap in rice, which is due to our more stringent variant filtering to eliminate the potential false variant callings caused by the complex structure of the maize genome and the low-frequency EMS-induced marker mutations introduced only via individual F1 plants. A low number of segregating variants can be a disadvantage because it will make peak calling difficult and limit the resolution of mapping; at the same time, it has the potential advantage of having only a few putative SNPs in the identified region to pinpoint the causal mutation. Our success in mapping by Mutmap in a non-standard maize inbred, where no prior genetic studies had been performed, will facilitate the adoption of this method to allow the rapid cloning of maize genes in diversified germplasm.

As mentioned above, our analysis used only variants between ML10 and the M5 mutant. In principle, we could have used more variants from the more heavily mutagenized E-ML10 line; however, genomic complexity in the non-reference ML10 genome, the heterozygosity of E-ML10 for many of the induced mutations and the need to collect F2 samples from multiple F1 ears prevented us from doing so. In the future, a modification to the protocol will be to self E-ML10 several times to reduce heterozygosity of the induced marker mutations before crossing to the mutant of interest and to use plants from only one single selfed F1 plant for mapping. In addition, sequencing of the very parents from which the F1 plant(s) were derived (rather than sister plants from the parental generation) would be advisable.

Many mutations (for example weak *fea* mutants affecting kernel row number) in maize in B73 proved to be very difficult to map due to scoring F2 phenotypes in mixed genetic backgrounds (unpublished data from David Jackson). Based on our experience with mapping E1-9, we also generated a heavily mutagenized B73 line (the E-B73), and are now in the process of selfing it for several generations to fix the EMS-induced marker SNPs; this should result in an essentially isogenic second parent carrying a large number of SNP markers for mapping, which should be very useful to the maize research community.

In conclusion, we have developed an EMS population in an elite tropical inbred and found numerous interesting mutant phenotypes for forward genetic studies. We successfully identified a mutation in *CLE7* underlying ear fasciation by mapping-by-sequencing (MutMap). The EMS mutagenesis and rapid mapping-by-sequencing pipeline described here may encourage maize researchers to perform forward genetics in their maize germplasm of choice, which can lead to novel gene discovery due to diversified genetic backgrounds.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/3/281/s1, Table S1: Primers used in this study, Figure S1: Nucleotide alignment of the CLE7 locus among reference B73, unmutagenized ML10 and E1-9 mutant, Supplementary protein alignment files for FEA2, FEA3, FEA4, TD1, CT2, CLE7 and GB1.

Author Contributions: Conceptualization: S.L.V., M.L., C.K., T.D.K., K.H.T., and; Data analysis: C.K. and S.L.V.; Investigation and Methodology: Q.H.T., N.H.B., N.T.N.D., L.T.N., T.T.T. and S.L.V.; Writing – original draft: C.K. and S.L.V.; Writing – review & editing: S.L.V., M.L., T.D.K. and K.H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by The National Foundation for Science and Technology Development (NAFOSTED Vietnam) under grant number 106-NN.01-2014.48.

Acknowledgments: We thank Tran Thi Thu Nga, Kieu Thi Dung for their support on fieldwork. We thank David Jackson and Liu Lei for their advice on bioinformatics, mapping and comments on the manuscript. All authors have read and agreed to the published version of the manuscript.

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

References

1. Shiferaw, B.; Prasanna, B.M.; Hellin, J.; Bänziger, M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* **2011**, *3*, 307. [CrossRef]
2. OECD; FAO. *Agricultural Outlook 2018-2027*; OECD-FAO Agricultural Outlook 2018-2027; OECD Publishing: Paris, France; Food and Agriculture, Organization of the United Nations: Rome, Italy, 2018; pp. 109–112.
3. Rosegrant, M.W.; Msangi, S.; Ringler, C.; Sulser, T.B.; Zhu, T.; Cline, S.A. *International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT): Model Description*; International Food Policy Research Institute (IFPRI): Washington, DC, USA, 2012.
4. Candela, H.; Hake, S. The art and design of genetic screens: Maize. *Nat. Rev. Genet.* **2008**, *9*, 192–203. [CrossRef] [PubMed]
5. Neuffer, M. Mutagenesis. In *The Maize Handbook*; Springer: New York, NY, USA, 1994; pp. 212–219.
6. maizegdb.org. Available online: https://www.maizegdb.org/data_center/phenotype (accessed on 9 February 2020).
7. Till, B.J.; Reynolds, S.H.; Weil, C.; Springer, N.; Burtner, C.; Young, K.; Bowers, E.; Codomo, C.A.; Enns, L.C.; Odden, A.R.; et al. Discovery of induced point mutations in maize genes by TILLING. *BMC Plant. Biol.* **2004**, *4*, 12. [CrossRef] [PubMed]
8. Lu, X.; Liu, J.; Ren, W.; Yang, Q.; Chai, Z.; Chen, R.; Wang, L.; Zhao, J.; Lang, Z.; Wang, H.; et al. Gene-Indexed mutations in maize. *Mol. Plant.* **2018**, *11*, 496–504. [CrossRef] [PubMed]
9. Heuermann, M.C.; Rosso, M.G.; Mascher, M.; Brandt, R.; Tschiersch, H.; Altschmied, L.; Altmann, T. Combining next-generation sequencing and progeny testing for rapid identification of induced recessive and dominant mutations in maize M2 individuals. *Plant J.* **2019**, *100*, 851–862. [CrossRef] [PubMed]
10. Settles, A.M. Transposon tagging and reverse genetics. In *Molecular Genetic Approaches to Maize Improvement*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 143–159.
11. Raizada, M.N.; Nan, G.L.; Walbot, V. Somatic and germinal mobility of the RescueMu transposon in transgenic maize. *Plant Cell* **2001**, *13*, 1587–1608. [CrossRef] [PubMed]
12. May, B.P.; Liu, H.; Vollbrecht, E.; Senior, L.; Rabinowicz, P.D.; Roh, D.; Pan, X.; Stein, L.; Freeling, M.; Alexander, D.; et al. Maize-targeted mutagenesis: A knockout resource for maize. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11541–11546. [CrossRef]
13. McCarty, D.R.; Settles, A.M.; Suzuki, M.; Tan, B.C.; Latshaw, S.; Porch, T.; Robin, K.; Baier, J.; Avigne, W.; Lai, J.; et al. Steady-state transposon mutagenesis in inbred maize. *Plant J.* **2005**, *44*, 52–61. [CrossRef]

14. Settles, A.M.; Holding, D.R.; Tan, B.C.; Latshaw, S.P.; Liu, J.; Suzuki, M.; Li, L.; O'Brien, B.A.; Fajardo, D.S.; Wroclawska, E. Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genom.* **2007**, *8*, 116. [CrossRef]
15. Williams-Carrier, R.; Stiffler, N.; Belcher, S.; Kroeger, T.; Stern, D.B.; Monde, R.A.; Coalter, R.; Barkan, A. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *Plant J.* **2010**, *63*, 167–177. [CrossRef]
16. Liang, L.; Zhou, L.; Tang, Y.; Li, N.; Song, T.; Shao, W.; Zhang, Z.; Cai, P.; Feng, F.; Ma, Y.; et al. A Sequence-Indexed mutator insertional library for maize functional genomics study. *Plant Physiol.* **2019**, *181*, 1404–1414. [CrossRef] [PubMed]
17. Jia, S.; Li, A.; Morton, K.; Avoles-Kianian, P.; Kianian, S.F.; Zhang, C.; Holding, D. A Population of deletion mutants and an integrated mapping and exome-seq pipeline for gene discovery in maize. *G3 (Bethesda)* **2016**, *6*, 2385–2395. [CrossRef] [PubMed]
18. Mao, H.; Wang, H.; Liu, S.; Li, Z.; Yang, X.; Yan, J.; Li, J.; Tran, L.S.; Qin, F. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* **2015**, *6*, 8326. [CrossRef] [PubMed]
19. 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]
20. Nyaga, C.; Gowda, M.; Beyene, Y.; Muriithi, W.T.; Makumbi, D.; Olsen, M.S.; Suresh, L.M.; Bright, J.M.; Das, B.; Prasanna, B.M. Genome-wide analyses and prediction of resistance to MLN in large tropical maize germplasm. *Genes (Basel)* **2019**, *11*. [CrossRef] [PubMed]
21. Sitonik, C.; Suresh, L.M.; Beyene, Y.; Olsen, M.S.; Makumbi, D.; Oliver, K.; Das, B.; Bright, J.M.; Mugo, S.; Crossa, J.; et al. Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theor. Appl. Genet.* **2019**, *132*, 2381–2399. [CrossRef]
22. Zheng, H.; Chen, J.; Mu, C.; Makumbi, D.; Xu, Y.; Mahuku, G. Combined linkage and association mapping reveal QTL for host plant resistance to common rust (*Puccinia sorghi*) in tropical maize. *BMC Plant Biol.* **2018**, *18*, 310. [CrossRef]
23. Zuo, W.; Chao, Q.; Zhang, N.; Ye, J.; Tan, G.; Li, B.; Xing, Y.; Zhang, B.; Liu, H.; Fengler, K.A.; et al. A maize wall-associated kinase confers quantitative resistance to head smut. *Nat. Genet.* **2015**, *47*, 151–157. [CrossRef]
24. Xiao, Y.; Liu, H.; Wu, L.; Warburton, M.; Yan, J. Genome-wide association studies in maize: Praise and stargaze. *Mol. Plant* **2017**, *10*, 359–374. [CrossRef]
25. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [CrossRef]
26. Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.S.; et al. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524–527. [CrossRef] [PubMed]
27. Abe, A.; Kosugi, S.; Yoshida, K.; Natsume, S.; Takagi, H.; Kanzaki, H.; Matsumura, H.; Yoshida, K.; Mitsuoka, C.; Tamiru, M.; et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **2012**, *30*, 174–178. [CrossRef] [PubMed]
28. Klein, H.; Xiao, Y.; Conklin, P.A.; Govindarajulu, R.; Kelly, J.A.; Scanlon, M.J.; Whipple, C.J.; Bartlett, M. Bulked-segregant analysis coupled to whole genome sequencing (BSA-Seq) for rapid gene cloning in maize. *G3: Genes Genomes Genet.* **2018**, *8*, 3583–3592. [CrossRef]
29. Addo-Quaye, C.; Buescher, E.; Best, N.; Chaikam, V.; Baxter, I.; Dilkes, B.P. Forward genetics by sequencing EMS variation-Induced inbred lines. *G3 (Bethesda)* **2017**, *7*, 413–425. [CrossRef] [PubMed]
30. Fekih, R.; Takagi, H.; Tamiru, M.; Abe, A.; Natsume, S.; Yaegashi, H.; Sharma, S.; Sharma, S.; Kanzaki, H.; Matsumura, H.; et al. MutMap+: Genetic mapping and mutant identification without crossing in rice. *PLoS ONE* **2013**, *8*, e68529. [CrossRef] [PubMed]
31. Gerpacio, R.V. *Impact of Public-and Private-Sector Maize Breeding Research in Asia, 1966–1997/98*; CIMMYT: El Batán, Mexico, 2001; pp. 105–120.
32. Clarke, J.D. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harbor Protocols* **2009**, *2009*, pdb. prot5177. [CrossRef]

33. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
34. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
35. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
36. Danecek, P.; McCarthy, S.A. BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics* **2017**, *33*, 2037–2039. [CrossRef]
37. Sarkar, D. *Lattice: Multivariate Data Visualization with R*; Springer: New York, NY, USA, 2008.
38. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [CrossRef] [PubMed]
39. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
40. Liang, Z.; Schnable, J.C. RNA-Seq based analysis of population structure within the maize inbred B73. *PLoS ONE* **2016**, *11*, e0157942. [CrossRef] [PubMed]
41. Bommert, P.; Je, B.I.; Goldshmidt, A.; Jackson, D. The maize Gα gene COMPACT PLANT2 functions in CLAVATA signalling to control shoot meristem size. *Nature* **2013**, *502*, 555–558. [CrossRef]
42. Taguchi-Shiobara, F.; Yuan, Z.; Hake, S.; Jackson, D. The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. *Genes Dev.* **2001**, *15*, 2755–2766. [CrossRef]
43. Bommert, P.; Nardmann, J.; Vollbrecht, E.; Running, M.; Jackson, D.; Hake, S.; Werr, W. Thick tassel dwarf1 encodes a putative maize ortholog of the Arabidopsis CLAVATA1 leucine-rich repeat receptor-like kinase. *Development* **2005**, *132*, 1235–1245. [CrossRef]
44. Wu, Q.; Xu, F.; Liu, L.; Char, S.N.; Ding, Y.; Je, B.I.; Schmelz, E.; Yang, B.; Jackson, D. The maize heterotrimeric G protein β subunit controls shoot meristem development and immune responses. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1799–1805. [CrossRef]
45. Je, B.I.; Gruel, J.; Lee, Y.K.; Bommert, P.; Arevalo, E.D.; Eveland, A.L.; Wu, Q.; Goldshmidt, A.; Meeley, R.; Bartlett, M. Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nat. Genet.* **2016**, *48*, 785. [CrossRef]
46. Pautler, M.; Eveland, A.L.; LaRue, T.; Yang, F.; Weeks, R.; Je, B.I.; Meeley, R.; Komatsu, M.; Vollbrecht, E.; Sakai, H. FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize. *Plant Cell* **2015**, *27*, 104–120. [CrossRef]
47. Je, B.I.; Xu, F.; Wu, Q.; Liu, L.; Meeley, R.; Gallagher, J.P.; Corcilius, L.; Payne, R.J.; Bartlett, M.E.; Jackson, D. The CLAVATA receptor FASCIATED EAR2 responds to distinct CLE peptides by signaling through two downstream effectors. *Elife* **2018**, *7*, e35673. [CrossRef]
48. Rodriguez-Leal, D.; Xu, C.; Kwon, C.-T.; Soyars, C.; Demesa-Arevalo, E.; Man, J.; Liu, L.; Lemmon, Z.H.; Jones, D.S.; Van Eck, J. Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nat. Genet.* **2019**, *51*, 786–792. [CrossRef] [PubMed]
49. Henry, I.M.; Nagalakshmi, U.; Lieberman, M.C.; Ngo, K.J.; Krasileva, K.V.; Vasquez-Gross, H.; Akhunova, A.; Akhunov, E.; Dubcovsky, J.; Tai, T.H.; et al. Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* **2014**, *26*, 1382–1397. [CrossRef] [PubMed]
50. Harjes, C.E.; Rocheford, T.R.; Bai, L.; Brutnell, T.P.; Kandianis, C.B.; Sowinski, S.G.; Stapleton, A.E.; Vallabhaneni, R.; Williams, M.; Wurtzel, E.T.; et al. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **2008**, *319*, 330–333. [CrossRef] [PubMed]
51. Yan, J.; Kandianis, C.B.; Harjes, C.E.; Bai, L.; Kim, E.-H.; Yang, X.; Skinner, D.J.; Fu, Z.; Mitchell, S.; Li, Q. Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nat. Genet.* **2010**, *42*, 322. [CrossRef] [PubMed]

genes

MDPI

Article

A New SNP in Rice Gene Encoding Pyruvate Phosphate Dikinase (PPDK) Associated with Floury Endosperm

Heng Wang [1,2,†], Tae-Ho Ham [3,†], Da-Eun Im [1], San Mar Lar [1], Seong-Gyu Jang [1], Joohyun Lee [3], Youngjun Mo [4], Ji-Ung Jeung [4], Sun Tae Kim [1] and Soon-Wook Kwon [1,*]

1 Department of Plant Bioscience, College of Natural Resources and Life Science, Pusan National University, Miryang 50463, Korea; wang_heng126@126.com (H.W.); ekdms0309@gmail.com (D.-E.I.); sanmarlar2010@gmail.com (S.M.L.); sgjang0136@gmail.com (S.-G.J.); stkim5505@gmail.com (S.T.K.)

2 State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

3 Department of Applied Bioscience, Konkuk University, Seoul 05029, Korea; lion78@daum.net (T.-H.H.); joohyun00@gmail.com (J.L.)

4 National Institute of Crop Science, Rural Development Administration, Jeonju 54874, Korea; moyj82@korea.kr (Y.M.); jrnj@korea.kr (J.-U.J.)

* Correspondence: swkwon@pusan.ac.kr; Tel.: +82-55-350-5506

† These authors contributed equally.

Received: 20 March 2020; Accepted: 22 April 2020; Published: 24 April 2020

Abstract: Rice varieties with suitable flour-making qualities are required to promote the rice processed-food industry and to boost rice consumption. A rice mutation, Namil(SA)-flo1, produces grains with floury endosperm. Overall, grains with low grain hardness, low starch damage, and fine particle size are more suitable for use in flour processing grains with waxy, dull endosperm with normal grain hardness and a high amylose content. In this study, fine mapping found a C to T single nucleotide polymorphism (SNP) in exon 2 of the gene encoding cytosolic pyruvate phosphate dikinase (cy*OsPPDK*). The SNP resulted in a change of serine to phenylalanine acid at amino acid position 101. The gene was named *FLOURY ENDOSPERM 4-5* (*FLO4-5*). Co-segregation analysis with the developed cleaved amplified polymorphic sequence (CAPS) markers revealed co-segregation between the floury phenotype and the *flo4-5*. This CAPS marker could be applied directly for marker-assisted selection. Real-time RT-PCR experiments revealed that PPDK was expressed at considerably higher levels in the *flo4-5* mutant than in the wild type during the grain filling stage. Plastid ADP-glucose pyrophosphorylase small subunit *(AGPS2a* and *AGPS2b)* and soluble starch synthase (*SSIIb* and *SSIIc)* also exhibited enhanced expression in the *flo4-5* mutant.

Keywords: *Oryza sativa* L.; PPDK; *flo4-5*; floury endosperm

1. Introduction

Rice (*Oryza sativa* L.) is a staple food for more than half the world's population; improving yield and grain quality is therefore of great importance. Grain yield can be increased considerably through the use of hybrid rice strains, but the grain quality of these varieties does not fully satisfy customer demands [1]. As such, grain quality improvement is the current focus of many rice geneticists worldwide. Furthermore, cultivars with a range of different grain qualities are desirable for specific production or medicinal purposes.

Starch, which is the major storage carbohydrate in rice grains, constitutes approximately 90% of a rice grain, of which approximately 18% is amylose and 82% is amylopectin. The ratio of amylose to amylopectin plays an important role in rice grain structure, appearance, and eating quality. Amylose

content is the most important determining factor for almost all of the physicochemical properties of rice starch, including its gelling, pasting, firmness, turbidity, freeze–thaw stability, syneresis, and retrogradation properties [2,3]. Amylopectin content governs the formation of crystalline granules and paste thickening [4]. Much research in recent decades has focused on understanding the genetics and biochemistry of starch biosynthesis, which is a complex process through which glucose generated by photosynthesis is converted to starch via several intermediaries. AGPase catalyzes the first committed step of starch biosynthesis and, as such, plays an important role in regulating starch synthesis in rice endosperms. Rice grains from plants carrying mutations in the AGPase gene exhibit shrunken endosperms and reduced starch content as a result of significant reductions in starch synthesis [5,6]. By contrast, the overexpression of AGPase genes increases seed weight and starch content [7]. Both amylose and amylopectin utilize ADPG as the activated glucosyl donor, and subsequent synthesis steps use different enzymes for amylose and amylopectin production. Amylose is synthesized by *GBSSI* (granule-bound starch synthase), which is encoded by *Wx* (waxy). Loss-of-function mutations at the *GBSSI*/*Wx* locus result in the elimination of amylose synthesis, leading to a lack of long-chain amylopectin, as well as a complete lack of amylose [8,9].

Starch synthesis is also transcriptionally regulated by transcription factors that presumably function within biosynthetic networks [1]. These include alkaline leucine zipper transcription factor OsbZIP58, which regulates *OsAGPL3*, *Wx*, *OsSSIIa*, *SBE1*, *OsBEIIb*, and *ISA2* transcription through binding to their promoters [10]. A lack of *ISA* in the rice sugary mutant impacts the expression of several other genes related to starch synthesis [11]. The high-resistant starch (RS) mutant, which is defective in SSIIIa, affects the activity of PPDK (pyruvate phosphate dikinase) and AGPase and leads to increases in lipid and amylose synthesis [12]. The white-core endosperm mutant, which is defective in *OsPK2*, affects the expression of genes involved in glycolysis/gluconeogenesis, pyruvate/phosphoenolpyruvate metabolism, fatty acid metabolism, and starch synthesis [13].

Among endosperm mutants, only those with a floury and white core contain round and loosely packed starch granules that grind easily [14]. To date, seven floury endosperm genes (*FLO1–7*) have been identified in rice [15–21], six of which have been cloned and their functions characterized. Three of these genes are not directly involved in starch synthesis and are instead involved in protein–protein interactions that regulate starch-synthesis-related genes. *Flo2* encodes a tetratricopeptide repeat motif that affects *Wx* expression, Flo6 encodes a C-terminal carbohydrate-binding module 48 domain that regulates isoamylase1, and Flo7 harbors an N-terminal transit peptide that does not directly interact with starch synthesis genes [19–21]. The three remaining floury endosperm genes participate in starch synthesis pathways; Flo4 encodes a PPDK, Flo5 encodes a rice-soluble starch synthase gene (*OsSSIIIa*), and *osagpl2–3* (also named *Flo6*) encodes AGPase [15,18,22]. In rice starch metabolism pathways, amylopectin-synthesizing enzymes, including starch synthases (*SSIIIa*, *SSIVb*, and *SSIIa*), starch-branching enzymes (*SBEI* and *SBEIIb*), and *PUL*, physically interact with each other to form multienzyme complexes [23,24]. The multienzyme complexes also contain other enzymes, such as PPDK and AGPase, that control the partitioning of ADP-Glc into starch and lipid [12,24,25]. This suggests that the proteins encoded by the three floury endosperm starch synthesis genes could interact with one another. The PPDK encoded by *Flo4* could promote AGPase activity in the complex by directly supplying PPi (pyrophosphate) for the conversion of ADP-Glc to Glc-1-P, which would enhance lipid biosynthesis. *SSIIIa* could inhibit the activity of PPDK/AGPase [25], and this interaction could be disrupted in the *SSIIIa* mutant (G to A in an exon), releasing PPDK and AGPase and increasing lipid synthesis. Simultaneously, loss of *SSIIIa* activity could divert the carbon flow needed for amylose synthesis by *GBSSI* (encoded by *Wx*), and as a result, the balance of lipid and amylose, which are components of RS, could be disrupted [12,25]. For AGPase, an SNP in the open reading frame (ORF) region of *osagpl2–3* may impair the activity of the functional domain that interacts with ADP-Glc. The effect of this mutation is particularly apparent at the endosperm development stage and results in the mutant floury endosperm phenotype [22]. Nevertheless, direct evidence of PPDK's function in the multienzyme complex remains unknown.

Rice has three PPDK-encoding genes organized at two loci: PPDKA, which encodes a cytosolic isoform (*OsPPDKA*), and PPDKB, which has two promoter sites and produces cytosolic and chloroplastic enzymes (cyPPDKB and chPPDKB, respectively) [8,15,26]. Of these, cyPPDKB is critical during the filling stage of rice grain development. During this stage, cyPPDKB is produced in abundance and acts to provide carbon skeletons for amino acid and lipid synthesis through the reversible interconversion of pyruvate and Pi to phosphoenolpyruvate and PPi. As seed development progresses from the early stage to the final storage product accumulation stage, the level and activity of cyPPDKB decrease rapidly in response to the combined posttranslational mechanisms of threonyl phosphorylation and protein degradation [27]. Cytosolic PPDKB is also involved in the glycolytic and gluconeogenic pathways and is impacted by anoxia [28–31]. To determine PPDK function in developing seeds, experiments to transgenically eliminate the PPDK gene from rice endosperms, to release PPDK from the phosphate-sufficient condition, or to upregulate PPDK expression during the middle and late rice grain-filling stages would be logical. However, the first report of a T-DNA insertional knockout mutant of the rice PPDK gene (*flo4-1*, *flo4-2* and *flo4-3*) indicated that rice with inactivated PPDK produce unexpected opaque seeds with a high lipid content [15].

The Namil(SA)-flo1 rice mutant was developed via sodium azide mutagenesis of *Oryza sativa* ssp. japonica cv. Namil [32]. By screening the mutant stock of Namil, two allelic mutants exhibiting floury endosperms were isolated and named "Namil(SA)-flo1" and "Namil(SA)-flo2" (Suweon 542). Suweon 542 exhibited a milky-white opaque endosperm, except for a thin peripheral area. Physicochemical analysis of the Suweon 542 endosperm revealed a loosely packed structure of irregular and globular-shaped starch granules, a low protein content and grain weight, and high amylose content compared with its wild-type progenitor, Namil. During dry milling, Suweon 542 had significantly lower grain hardness, finer particle size, more loosely packed starch granules, and lower starch damage than Namil and other rice cultivars [33]. Genetic analysis of the floury endosperm characteristics of Suweon 542 revealed that the location of the target gene was in the 19.33–19.73 Mbp region on chromosome 5 between markers RM18624 and RM18639, and map-based cloning revealed a G→A SNP in exon 8 of cy*OsPPDK* (*flo4-4*), which was responsible for the floury endosperm of Suweon 542 [33,34]. Moreover, the floury endosperm of Namil(SA)-flo1 controlled by one recessive gene and the locus of Namil(SA)-flo1 was localized to the 17.7–20.7 Mbp region of chromosome 5 [35]. However, the grain characteristics and molecular mechanisms regulating the floury endosperm in Namil(SA)-flo1 remain unknown. Fine mapping and molecular cloning of floury endosperm-related genes are needed to verify and elucidate the genetic framework of floury endosperm development in rice.

In this study, the agronomic traits and grain physicochemical properties, including suitability for dry milling of Namil(SA)-flo1, were investigated. Map-based cloning revealed that a C→T SNP in exon 2 of cy*OsPPDK* results in a missense mutation from Ser to Phe at amino acid position 101. Co-segregation analysis and qRT-polymerase chain reaction (PCR) confirmed that *flo4-5* was responsible for the development of the floury endosperm during the grain filling stage and indicated the involvement of cy*OsPPDK* in grain quality and seed number control.

2. Materials and Methods

2.1. Plant Materials and DNA Extraction

The Namil(SA)-flo1 floury endosperm rice mutant was produced via sodium azide mutagenesis of Namil (*O. sativa* L. ssp. japonica), an early maturing, high-yield, non-waxy Korean elite rice cultivar [12]. To evaluate the major agronomic traits and the grain/flour physicochemical properties, Namil(SA)-flo1 was cultivated with Namil, Seolgaeng (non-waxy opaque endosperm japonica cultivar), and Hwaseong (non-waxy japonica cultivar) in the experimental plot of the National Institute of Crop Science (NICS), Rural Development Administration (RDA), Suwon, Korea.

2.2. Evaluation of Agronomic Traits and Grain Physicochemical Properties

Replicated yield trials were conducted to evaluate the major agronomic traits, as well as the grain/flour physicochemical properties, in the field at the NICS, RDA, Suwon, Korea [35]. The seeds of each rice line were sown on April 25, and were then transplanted on May 25 under a randomized complete block design (RCBD) with three replication plots. Each plot, consisting of eight rows with 30 hills per row and three plants per hill, was planted with 30 × 15 cm spacing. The amount of fertilizer application was 90–45–57 Kg/ha for N–P_2O_5–K_2O, and the 10 hills in the middle rows were used to determine days-to-heading (HD), culm length (CL), panicle length (PL), tiller number (TN), spikelet number per panicle (SN), and ripened grains percentage (RGP). The 1000-grain weight (TGW) was measured in grams as the average weight of 1000 fully filled brown rice grains from each plot.

The grain hardness of the brown rice was assessed by determining the pressure at the grain breakage point using a 5 mm probe attachment of a TA.XT Plus instrument (Stable Micro Systems, Godalming, Surrey, UK), using parameters of 0.4 mm/s and 40.0 g for test speed and trigger force, respectively.

The damaged starch content was evaluated using a starch damage assay kit (Megazyme International Ireland, Wicklow, Ireland) following the manufacturer's instructions. The lightness of rice flour was measured with a JS-555 instrument (Color Techno System, Tokyo, Japan). The moisture, protein, lipid, and ash contents of the rice flour were determined using methods 44-15A, 46-30, and 08-01 of the American Association of Cereal Chemists (AACC) 2000. The amylose content of the rice flour was estimated as described previously [36].

Grain samples for scanning electron microscopy (SEM) analyses were prepared as described previously [15]. Cleaved endosperm surfaces were observed under an S-550 scanning electron microscope (Hitachi Hi-Tech, Tokyo, Japan) at an accelerating voltage of 20 kV.

2.3. Fine Mapping of FLO4-5

In a previous study, the F2 population of Namil(SA)-flo1/Milyang 23 (non-waxy Tongil cultivar) was used for genetic analysis [33]. To investigate the candidate genomic region for *FLO4-5*, F_2 lines heterozygous for the genomic region between RM18624 and RM18639 were grown to the F_3 generation. The F_3 plants were cultivated to generate $F_{3:4}$ seeds at the Pusan National University (PNU) experimental farm. Total genomic DNA was extracted from fresh leaves using a NucleoSpin®Plant II kit (MACHEREY-NAGEL GmbH & Co.KG, North Rhine-Westphalia, Germany), according to the manufacturer's instructions.

Parental lines and 96 randomly selected $F_{3:4}$ recombinants were dehulled for visual inspection of the endosperm, and then genotyped with six cleaved amplified polymorphic sequence (CAPS) markers developed based on the whole-genome resequencing data [34] of Namil(SA)-flo1 and Milyang 23. The whole genomes of Namil(SA)-flo1 and Milyang 23 were re-sequenced with a 75-fold average coverage using the Illumina HiSeq 2500 Sequencing Systems Platform (Illumina Inc., San Diego, CA, USA). Raw sequence reads were aligned against the rice reference genome (IRGSP 1.0) [37], and the predetermined CAPS marker orders were judged by e-landing of each marker on the reference rice genome [38]. ORFs and their functional products were annotated according to the MSU Rice Genome Annotation Project Database [39] based on the defined physical locations. PCRs were performed in a total volume of 20 µL containing 10 ng of DNA template, 10 pmol of each primer, 1× PCR buffer (50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, and 1.5 mM $MgCl_2$), 0.2 mM dNTPs, and 1 unit of Taq DNA polymerase (Nurotics, Daejeon, Korea). PCRs were performed with an MJ Research PTC-100 thermocycler (Waltham, MA, USA) using the following conditions: initial denaturation at 94 °C for 5 min, followed by 36 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 1 min, with a final extension at 72 °C for 10 min. The PCR products were digested using restriction enzymes (New England Biolabs, Ipswich, MA, USA), according to the manufacturer's instructions. The digestion of PCR products encompassing CAPS markers was detected using the WatCut program [40]. Digestion products were separated on 3% polyacrylamide gels using 6 M urea

and 1X TAE at 80 volts and visualized using the Molecular Imager®Gel Doc™ XR System (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

2.4. Cloning of FLO4-5 and Identification of the Mutation Site

The coding sequence of cy*OsPPDK* in Namil(SA)-flo1 and Namil was compared using the CLC Sequence Viewer 7.0 (QIAGEN, Hilden, Germany). To verify the mutation site, a CAPS marker containing two *MboII* restriction sites in the floury endosperm mutant Namil(SA)-flo1 and one *MboII* restriction site in Namil was developed using the WatCut program. A 177 bp fragment containing the mutation site was PCR amplified from Namil(SA)-flo1 and Namil using the primers flo4-5_F: CTCCAGTGGGTGGAGGAGTA and flo4-5_R: GATCGATCAGCAACGGAGAT. The PCR products were digested with *MboII* (New England Biolabs) in a volume of 15 µL containing 5 µL of PCR product, 1.5 µL of 10× NEBuffer, 0.5 µL of *MboII*, and 8 µL of ultrapure water, and then incubated at 37 °C for 2 h. The digestion products were separated using the Fragment Analyzer™ (Agilent, Santa Clara, CA, USA). The CAPS marker was also used for co-segregation analysis of $F_{3:4}$ families and 44 Korean rice cultivars with the endosperm phenotype. The complete genomic DNA of the cy*OsPPDK* gene was cloned in three overlapping segments using primers designed on the basis of the cy*OsPPDK* gene sequence of Nipponbare (Table S1).

2.5. Predicting the Functional Effect of Amino Acid Substitutions and Real-Time qRT-PCR Analysis

The PROVEAN (Protein Variation Effect Analyzer, v1.0) tool [41] was used to predict amino acid changes that affected protein function. Total RNA was isolated from the rice grains of Namil and Namil(SA)-flo1 at 12 days after flowering (DAF) using an RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturers' instructions. Genomic DNA was removed with DNase I (QIAGEN), and reverse transcription was performed using an RNA to cDNA EcoDry Premix Kit (Clontech, Mountain View, CA, USA). A QuantiNova SYBR Green RT-PCR kit (QIAGEN) and the Rotor-Gene Q instrument (QIAGEN) were used for qRT-PCR with the following conditions: 95 °C for 10 min followed by 40 cycles of 95 °C for 10 s and 60 °C for 20 s. Fold-change was calculated relative to Namil. Rice Actin 1 (OsACT1; Os03g0718150) was used as an internal control. The primer sequences used for qRT-PCR are listed in Table S1.

3. Results

3.1. Agronomic Traits and Seed Characteristics of Namil(SA)-flo1

The major agronomic and yield-related traits of the Namil(SA)-flo1 mutant and the corresponding Namil wild type were analyzed in plants grown under paddy field conditions. Compared with the wild type, heading was delayed by 3 days and CL was extended by 4 cm in the Namil(SA)-flo1 mutant. Panicle size, tiller number, and spikelets per panicle did not differ significantly between the wild type and Namil(SA)-flo1 mutant, but RGP was lower in Namil(SA)-flo1 than in the wild type (Table 1).

Table 1. Major agronomic traits of Namil(SA)-flo1 in comparison with its wild type, Namil.

Line	HD (Days)	CL (cm)	PL (cm)	TN (No.)	SN (No.)	RGP (%)
Namil	101 [b]	78 [b]	25 [a]	11 [a]	117 [a]	87 [a]
Namil(SA)-flo1	104 [a]	82 [a]	25 [a]	10 [a]	120 [a]	74 [b]

[a] Means with the same letter are not significantly different at $p < 0.05$ in the least significant difference test (LSD0.05).
[b] The means of each line were obtained from replicated yield trials with three replication plots. HD: days-to-heading after sowing, CL: culm length, PL: panicle length, TN: tiller number, SN: spikelet number per panicle, RGP: ripened grains percentage.

Cross-sectional observation of dehulled kernels revealed that most of the Namil(SA)-flo1 endosperm was white–opaque, except for a thin peripheral area (Figure 1a). SEM showed that the Namil(SA)-flo1 endosperm contained numerous small starch grains of irregular and rounded

shapes that were more loosely packed than in the wild type. The wild-type endosperm exhibited densely packed starch granules of polyhedral angular shapes (Figure 1b). In contrast with the appearance of the grain and the endosperm in Namil, Namil(SA)-flo1 exhibited a milky-white opaque endosperm. No differences in grain width, length, or thickness were observed between the mutant and the wild type (Figure 1c).

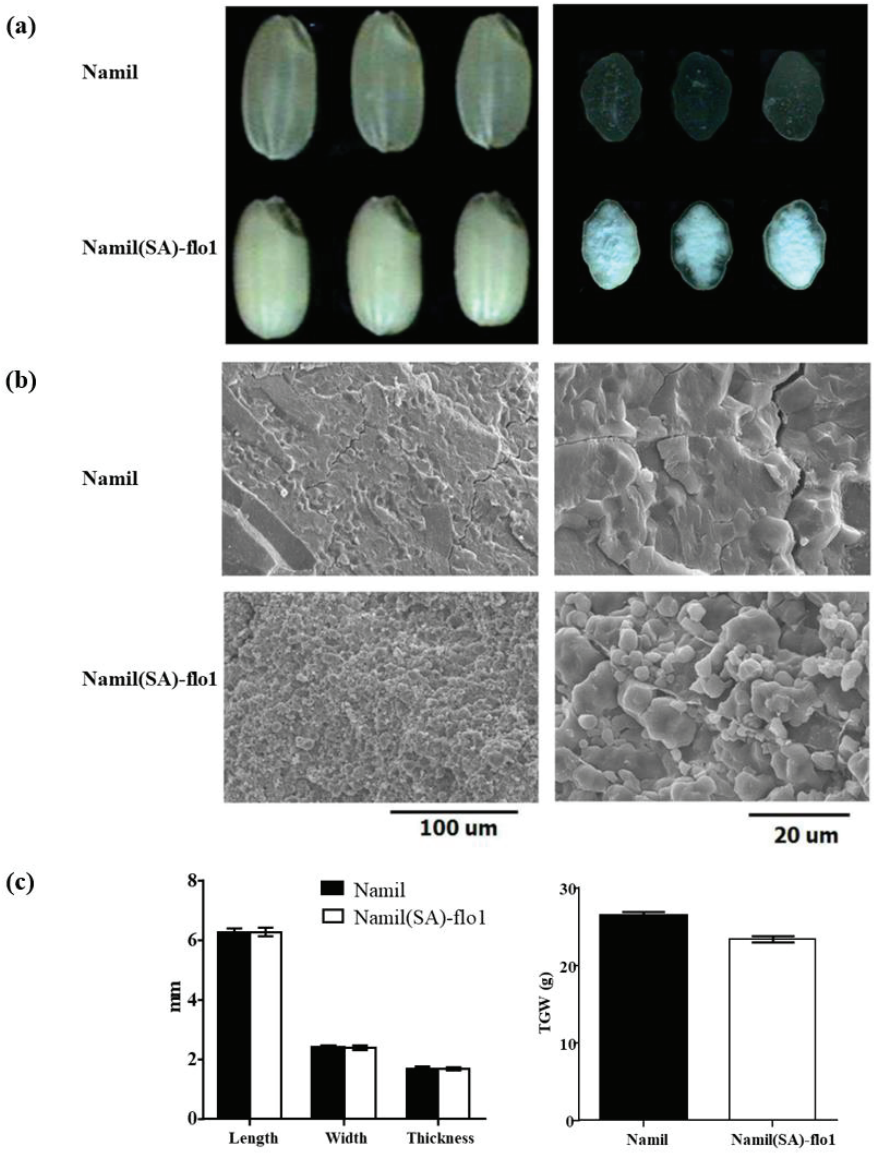


Figure 1. Phenotypic analyses of the mutant Namil(SA)-flo1 and wide type Namil. (**a**) Brown rice and transverse sections of Namil(SA)-flo1 and Namil; (**b**) Electron microscope visualization of mature endosperm. The Namil(SA)-flo1 is packed loosely with compound starch granules; (**c**) The grains' shape and weight (TGW: 1000-grain weight of brown rice).

The flour physicochemical properties were also examined. The results show that the flour derived from the mutant grains was significantly lighter than that from the wild-type grains. The ash content was lower in Namil(SA)-flo1 than in the wild type. The amylose content was comparable between the wild type and Namil(SA)-flo1, but the lipid content was higher and the protein content was significantly lower in Namil(SA)-flo1 than in the wild type (Table 2). These findings are consistent with the results of previous reports showing that increasing the lipid content could cause a floury endosperm [12].

3.2. Dry Milling Properties of Namil(SA)-flo1

The suitability of Namil(SA)-flo1 as a raw material for high-quality dry milled rice flour was assessed. The grain hardness of Namil(SA)-flo1 was 0.45 times that of the wild type, and was softer than that of Seolgaeng, a Korean rice cultivar with an opaque endosperm. The mean particle size of Namil(SA)-flo1 was 86.1 μm, smaller than that of the wild type (109.1 μm), Seolgaeng (97.6 μm), and Hwaseong (112.2 μm). The damaged starch content of dry milled flour in Namil(SA)-flo1 (4.9%) was significantly lower than that in Namil (9.2%), Seolgaeng (7.1%), and Hwaseong (10.3%) (Table 2). Overall, the low grain hardness, low starch damage, and fine particle size indicate that Namil(SA)-flo1 is suitable for dry milling to produce fine flour.

3.3. Fine Mapping of the Floury Endosperm Locus

The Namil(SA)-flo1 locus was previously refined to a region on chromosome 5 and, alongside *FLO4-4* (identified from Namil(SA)-flo2, also known as Suweon 542), was linked to chromosome 5 markers RM18624 and RM18639 [33,35]. To map the Namil(SA)-flo1 target gene more precisely, the whole genomes of Namil(SA)-flo1 and Milyang 23 were re-sequenced and aligned against one another to select homologous SNPs for developing CAPS markers. Six CAPS markers were developed (Table S2) and used to genotype 60 floury and 36 normal individuals derived from $F_{3:4}$ recombinant families. Multiple comparisons were conducted between the genotypes of these recombinants and the phenotypes of their offspring. Finally, the target region was mapped to BAC clones OJ1174_H11 and BAC OSJNBb0006J12 in a 33 kb region flanked by the markers CAPS6 and RM18639. Based on the Rice Annotation Project Database [42], four predicted genes were found in the critical 33 kb region (Figure 2a). Of these, *Os05g0404500* was annotated as encoding a hypothetical protein, Os05g0404700 was a functional gene similar to the methyl-binding protein gene MBD1, Os05g0404901 encoded a conserved hypothetical protein, and Os05g0405000 was annotated as a PPDK gene. Sequencing of all four candidates in Namil(SA)-flo1 and the Namil wild type revealed a C to T SNP in exon 2 of PPDK (Figure 2b and Table S3). In rice, PPDK, which has two promoter sites, encodes both a cytosolic- and a chloroplast-targeted PPDK protein (cy*OsPPDK* and ch*OsPPDK*, respectively). To analyze the transcript type of PPDK in developing grains of rice, the PPDK of Namil(SA)-flo1 and Namil was sequenced using the primer combinations PF2/PR3 (cy*OsPPDK*) and PF3/PR3 (ch*OsPPDK*) [15] and three additional primer pairs (Table S1). The full-length sequence of cy*OsPPDK* was identified as 7556 bp with a cDNA sequence of 2649 bp containing 19 exons and encoding an 882 amino acid protein. The SNP in exon 2 was a missense mutation that resulted in a serine to phenylalanine change at amino acid position 101 in cy*OsPPDK* (Figure 2b). Comparison of the cy*OsPPDK* sequences of Namil(SA)-flo1 and Namil with that of Nipponbare revealed nine SNPs and six InDels (insertion/deletion mutations) in the intron region and two SNPs in the coding region that were synonymous mutations (Table S3). Sequence analysis of Suweon 542 (*flo4-4*) did not reveal the same mutation site (the Suweon 542 phenotype was caused by a G→A SNP in exon 8 of cy*OsPPDK*) [34]. The novel recessive floury gene was named *floury4-5* (*flo4-5*). The full-length coding sequences of *FLO4-5* from Namil(SA)-flo1 and Namil were deposited in GenBank under accession numbers MG267057 and MG267056, respectively.

Table 2. Physicochemical properties of grains and rice flours.

Line	Hardness Index	Grain Hardness (Kg)	Mean Particle Size (μm)	Damaged Starch (%)	Lightness (CIE Value)	Ash (%)	Protein (%)	Amylose (%)	Lipid (%)
Hwaseong	1.04 a	7825 a	112.2 ± 0.40 a	10.3 ± 0.19 a	88.6 ± 0.01 b	0.84 ± 0.02 a	7.5 ± 0.16 c	18.5 ± 0.24	-
Seolgaeng	0.79 b	5962 b	97.6 ± 1.63 c	7.1 ± 0.10 c	90.3 ± 0.06 a	0.72 ± 0.01 c	6.6 ± 0.11 d	17.5 ± 0.60	-
Namil	1.00 a	7526 a	109.1 ± 0.62 b	9.2 ± 0.17 b	88.7 ± 0.12 b	0.82 ± 0.01 a	9.2 ± 0.25 a	17.7 ± 1.34	1.5 ± 0.56 b
Namil(SA)-flo1	0.45 c	3417 c	86.1 ± 0.81 d	5.1 ± 0.06 d	90.4 ± 0.09 a	0.77 ± 0.02 b	7.8 ± 0.04 b	17.8 ± 0.27	3.1 ± 0.54 a

Note: Different letter means indicate significant differences according to the Duncan multiple range test ($p < 0.05$).

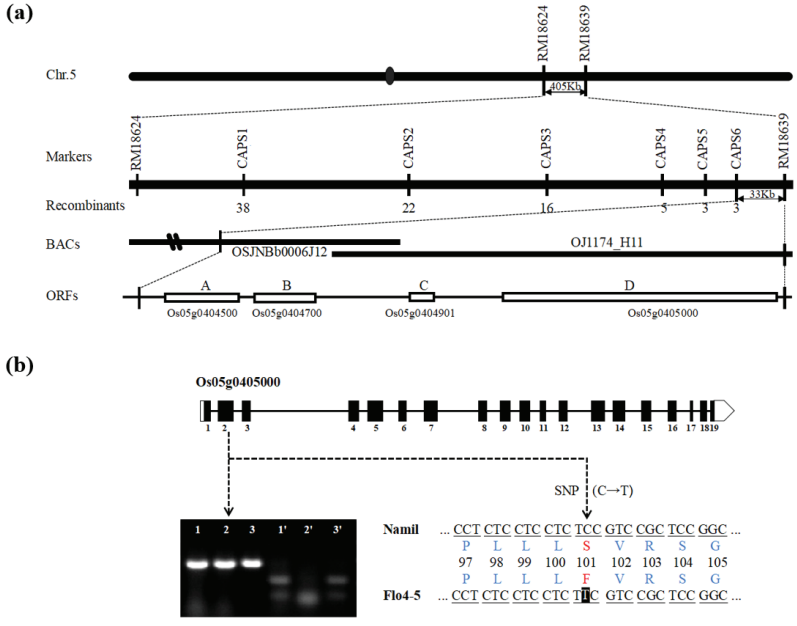


Figure 2. Map-based cloning of the *flo4-5* mutant. (**a**) Fine mapping of the *flo4-5* locus. The molecular markers and number of recombinants are shown. The 33kb virtual contig, composed of overlapping 2 BAC clones, was delimited by e-Landings of two significant markers on the reference rice genome, 'Os-Nipponbare-Reference-IRGSP-1.0'; (**b**) *flo4-5* gene structure and cDNA sequence comparison showing a nucleotide mutant (C to T) within exon 2 where Ser-101 of the wild was induced to Phe-101 of the *flo4-5*. White boxes represent untranslated regions, black boxes represent coding regions, and solid lines represent introns. 1,2,3 are the PCR results of wild type Namil, *flo4-5*(Namil (SA)–flo1) and *flo4-4*(Namil (SA)–flo2), respectively. 1′, 2′, 3′ are the digested results.

3.4. Co-Segregation and Expression Analyses

The mutation site in *flo4-5* was confirmed using a CAPS marker with a *MboII* restriction site. *MboII* digested the mutant Namil(SA)-flo1 allele at two restriction sites (Figure 3a,b, lane 3) and the wild-type Namil allele at a single site (Figure 3a,b, lane 4). For screening $F_{3:4}$ family populations using the CAPS marker, 23 plants with the Namil(SA)-flo1 phenotype and 22 with the Namil phenotype were selected. As shown in Figure 3a, sequences from all plants with the Namil(SA)-flo1 phenotype were digested at two restriction sites with *MboII*, whereas those with the Namil phenotype were digested at a single site. Fifteen $F_{3:4}$ plants were heterozygous and exhibited the Namil phenotype, suggesting that the floury endosperm phenotype was controlled by a recessive gene. The genotypes of the currently cultivated Korean japonica rice were evaluated with this marker. Of the 44 tested varieties, none of them were the mutant genotype, whereas two varieties were the Namil genotype and the others were the Nipponbare genotype (Figure 3b). Overall, the co-segregation analysis indicates that *FLO4-5* is responsible for the floury endosperm in rice.

The PROVEAN tool was used to predict the effect of a single nucleotide mutation (C to T) on the *FLO4-5* gene, resulting in a score of –5.802 and a rating of "deleterious". Next, qRT-PCR was used to investigate the transcript type of *FLO4-5* in rice during the grain filling stage. The result that relative expression of *FLO4-5* was much higher in Namil(SA)-flo1 than in Namil (Figure 4) was unexpected, because the mutation was in exon 2, not in the promotor region. Even though the higher expression level in the mutant was not clearly explained based on our experimental design, it is highly possible that the increased level of *FLO4-5* was caused by the mutation.

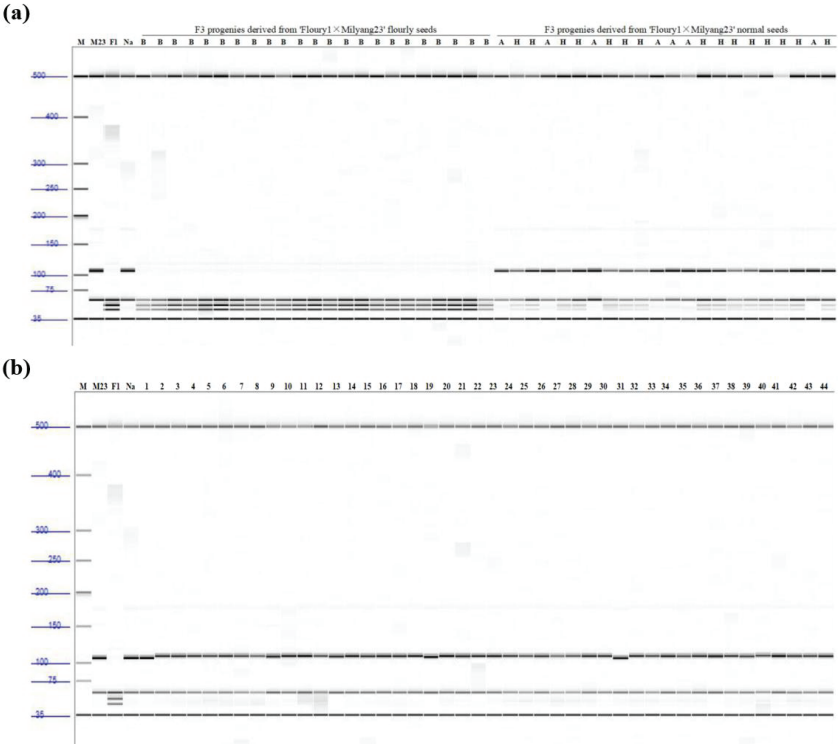


Figure 3. Co-segregation analysis of the *flo4-5* genotype with floury endosperm phenotype. (**a**) Verification of the CAPS marker (digestion by *MboII*) and tagging the *flo4-5* locus using a part of F3 individuals; (**b**) Verification of the CAPS marker (digestion by *MboII*) and tagging the *flo4-5* locus using Korean rice cultivars. 'A' and 'B' are homogeneous of Namil(SA)-flo1 and Milyang 23, respectively. 'H' is heterozygote. M (100bp ladder), F1 (Namil(SA)-flo1), M23 (Milyang 23), Na (Namil). 1. Namil(SA)-flo2(Suweon542), 2. Aranghyangchal, 3. Baegjinju1, 4. Baekogchal, 5. Boramchal, 6. Boramchan, 7. Borami, 8. Boseog, 9. Boseogchal, 10. Boseogheugchal, 11. Cheongnam, 12. Chindeul, 13. Chucheong, 14. Dabo, 15. Danmi, 16. Danpyeong, 17. Deuraechan, 18. Dodamssal, 19. Dongjin, 20. Dongjin1, 21. Dongjinchal, 22. Geonganghongmi, 23. Geonyang 2, 24. Goami, 25. Goami 2, 26. Goami 4, 27. Haepum, 28. Haiami, 29. Hanam, 30. Hanmaeum, 31. Heugjinmi, 32. Heughyangchal, 33. Heugjinju, 34. Heugnam, 35. Heugseol, 36. Hongjinju, 37. Hopum, 38. Hwanggeumnuri, 39. Hwaseong, 40. Hwawang, 41. Hwayeong, 42. Hyangnam, 43. Hyeonpum, 44. Ilmi.

3.5. Mutation of FLO4-5 Changed the Expression Levels of Major Starch Synthesis Enzymes

Because the starch content and the starch granules were altered in the *flo4-5* mutant, the expression levels of the starch synthesis-related genes in the developing grains were then analyzed using qRT-PCR. Compared with the wild-type Namil, the transcription of several genes decreased in the mutant, including *GBSSI*, *GBSSII*, *AGPL1*, *AGPL2*, *AGPL3*, *AGPL4*, *AGPLS1*, *SSI*, *SSIIa*, *SSIIIa*, *SSIIIb*, *SSIVa*, *SSIVb*, *BEI*, *BEIIa*, *BEIIb*, and *PUL*. The expression of *AGPS2a*, *AGPS2b*, *SSIIb*, and *SSIIc* increased in the mutant compared to the wild type, with substantial increases seen for *AGPS2a* and *AGPS2b* (Figure 5a). The expression analysis of Suweon 542 (*flo4-4*) revealed similar expression patterns (Figure 5b).

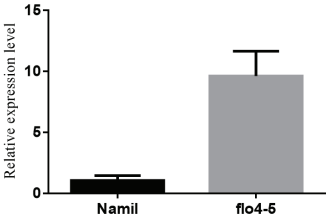


Figure 4. Expression analysis of PPDK between Namil and *flo4-5* at 12 DAF. Values shown are mean ± SD (n = 3).

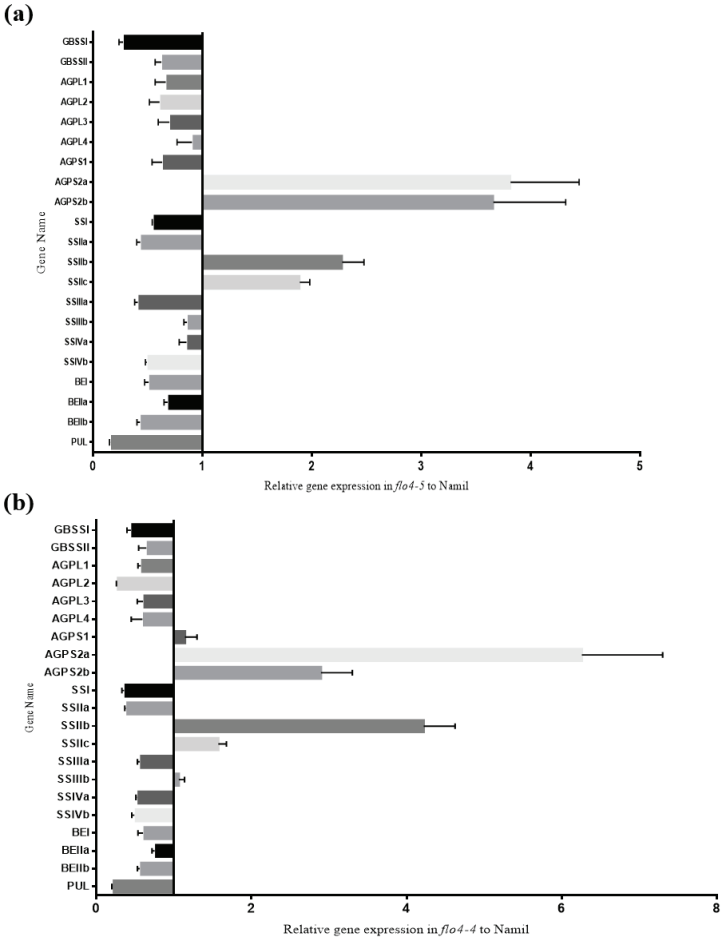


Figure 5. Expression profiles of the genes involved in production of storage starch in endosperm from the flo4-5 mutant (**a**) and Suweon542 (**b**) with the wild type Namil (bar begin with 1). Reference gene was Actin, and the expression levels of these genes are shown relative to the wild type Namil, which is set as 1. Each gene name is indicated by a simplified representation. *GBSSI* and *GBSSII*, granule bound starch synthase I and II; AGPL (*AGPL1*, *AGPL2*, *AGPL3* and *AGPL4*) and AGPS (*AGPS1*, *AGPS2a* and *AGPS2b*), ADP-glucose pyrophosphorylase large subunit and small subunite, respectively; *SSIIa* and *SSIIIa*, soluble starch synthase *IIa* and *IIIa*, respectively; *BEI* and *BEIIb*, branching enzyme *I* and *IIb*, respectively; *PUL*, pullulanase.

4. Discussion

Cereal crops accumulate high levels of starch in the seed endosperm as an energy reserve. Endosperm-defective mutants provide valuable genetic material for elucidating the gene networks that regulate starch synthesis and amyloplast development during grain filling. In this study, we isolated and characterized the Namsil(SA)-flo1 mutant, which exhibited an almost entirely milky-white opaque kernel, except for a thin peripheral area of the grain (Figure 1a). Compared to the wild type, the Namsil(SA)-flo1 mutant exhibited slightly higher amylose levels, lower protein levels, reduced kernel weight, and increased total lipid content (Table 2). No difference in seed size was observed (Figure 1c). Electron microscope visualization of the transverse sections of the Namsil(SA)-flo1 and wild-type mature endosperms showed that the Namsil(SA)-flo1 mutant was loosely packed with irregular and round-shaped compound starch granules, while the wild type contained densely packed polyhedral starch granules (Figure 1b). However, the physicochemical characteristics of Namsil(SA)-flo1 were different from those reported for *flo4-1*, *flo4-2*, and *flo4-3*, which were generated via T-DNA insertion into the cy*OsPPDK* gene [15,33,34]. We speculate that the cy*OsPPDK* gene may have diverse roles in the rice seed filling stage. Fine mapping, sequencing, and co-segregation analysis revealed that a C to T SNP in exon 2 of cy*OsPPDK* (the novel recessive floury gene was named *flo4-5*) was responsible for the Namsil(SA)-flo1 mutant phenotype (Figures 2 and 3). Based on our limited expression analysis method, this is not clearly explained; however, we hypothesize that the transcript from the mutated sequence compared to that of the normal sequence may not produce normal PPDK to increase the transcription level in the mutant, in order to compensate the normal transcript concentration in the cell. This hypothesis will be proven in a further study. Along with an increase in the transcripts of PPDK/AGPase, the expression of changes for starch synthesis genes (Figure 5) suggests that PPDK has a role in compound starch granule formation and starch synthesis during rice endosperm development.

PPDK catalyzes the freely reversible conversion of pyruvate, ATP, and Pi into phosphoenolpyruvate, AMP, and PPi, and is involved in diverse functions in various plant tissues. PPDK is ubiquitous in all rice tissues, but is generally of low abundance [43]. In developing rice kernels, PPDK is abundantly expressed during the syncytial/cellularization stage, with markedly reduced expression seen during the middle and late rice grain filling stages [15,27]. The presence of PPDK in high abundance in developing seeds suggests the importance of the enzyme; however, little is known regarding the molecular properties of rice PPDK.

The interconversion of phosphoenolpyruvate and pyruvate by PPDK leads to difficulties in evaluating PPDK function in seed development [43]. Previous studies have suggested several hypotheses for endosperm PPDK function, including (1) provision of pyruvate for lipid synthesis, (2) gluconeogenesis and provision of hexose for starch biosynthesis, (3) control of metabolic fluxes through its contribution to PPi homoeostasis, and (4) regulating glycolytic flux and energy charge [24,27,28,30,44,45]. In *Zea mays*, the PPi generated by PPDK can be channeled directly to AGPase within the protein complex, driving the plastidial AGPase reaction in the direction of ADPGlc breakdown to Glc-1-P, which can in turn support amino acid and lipid biosynthesis [24]. However, a loss-of-function mutation in the rice cy*OsPPDK* gene resulted in a lower amylose content, reduced kernel weight, increased total lipid content, and no change in storage protein or amino acid contents [15]. This suggests that PPDK may have several diverse roles in seed development, depending, in part, on the direction of the cycle [43].

In amylopectin biosynthesis, *SS* connects glucose in an α-1,4 glycosidic bond formation, and then *BE* (branching enzyme) connects a short chain to the C-6 hydroxyl group via hydrolysis of internal α-1,4 glycosidic bonds and formation of α-1,6 branches. *DBE* (starch-debranching enzyme: Isoamylase (*ISA*) and pullulanase (*PUL*)) cleaves aberrant branches through catalytic hydrolysis of α-1,6 glycosidic bonds. A lack of *SS* during rice endosperm amylopectin formation results in a decrease in polymerized chain formation and an increase in amylose characteristics [12,46–49]. Mutant *ae*, which is defective in the gene encoding *BE*, exhibits high proportional levels of amylose as a result of reductions in amylopectin production [50,51]. A loss-of-function mutation of the gene encoding *DBE* results in

a sugary phenotype in which endosperm starch is completely replaced by phytoglycogen [52,53]. Mutant *sug-h*, which is defective in *OsISA1* and *OsBEIIa*, exhibits a severely disrupted amylopectin structure [54].

A recent study revealed that rice amylopectin-synthesizing enzymes such as *SSI, SSIIIa, BEI, BEIIa, BEIIb, ISA1,* and *PUL* are physically associated with each other and form active protein complexes with various partnership patterns [12,23]. Further research in maize and rice indicates that PPDK and plastid AGPase may also participate in these enzyme complexes [12,24]. In these complexes, PPDK catalyzes the freely reversible conversion of pyruvate, ATP, and Pi into phosphoenolpyruvate, AMP, and PPi, but in photosynthesis, PPDK generates phosphoenolpyruvate by the hydrolysis of PPi used for CO_2 fixation in C4 plants [24]. Plastid AGPase also catalyzes a reversible reaction, but the direction of the reaction only depends on the relative concentrations of PPi and ATP [30]. Therefore, high PPi concentrations should orient the reaction to the direction of ADP-Glc degradation. However, the PPi levels in amyloplasts are very low as a result of high pyrophosphatase (PPase) activity, and the reaction therefore proceeds toward ADP-Glc synthesis. Zhou et al. [12] hypothesized that the starch biosynthetic enzymes in the complex might exert a constraining effect on PPDK and AGPase to control the partitioning of ADP-Glc into lipid and starch.

Based on previous research and our findings, we hypothesize that a single amino acid change (S to F) in the PPDK protein in the flo4-5 mutant may disrupt the amylopectin-synthesizing multienzyme complexes and release PPDK to allow an abundance of PPi in amyloplasts. The high concentration of PPi would then promote plastid AGPS2a and AGPS2b activity and direct the reaction toward ADP-Glc degradation, producing more Glc-1-P for lipid biosynthesis. At the same time, PPDK would actively increase transcription of the *SSIIb* and *SSIIc* genes without increasing starch content. This supports the view that the catalytic activity of one or more of the starch synthase isoforms is insufficient to accommodate the excess ADP-glc present in the amyloplast, but that other constraints within the amyloplast stroma control carbon flux into starch [55]. Further research is required to test this hypothesis.

The number of spikelets per panicle substantially increased (by approximately 32.5%) in the Suweon 542 mutant (*flo4-4*) compared with the wild type [33]. In the flo4-5 mutant, the number of spikelets per panicle (Table 1) increased only slightly. This may have been due to environmental effects on yield, whereby the average high temperature during the first 4 days after pollination must reach ~33 °C for yield increase to occur [56]. AGPase, the rate-limiting starch biosynthetic enzyme, is a heterotetramer composed of two identical small and two identical large subunits. The enzyme is allosterically controlled, and the isoform in the cereal endosperm is heat-labile [57]. In maize, the *Shrunken*-2 (*Sh*2) gene encodes the large subunit of endosperm AGPase, and overexpression of the gene enhances heat stability and reduces phosphate inhibition, leading to an increase in maize yield (up to 64%) via an increase in seed number [56]. The small subunit of the maize endosperm AGPase is encoded by the gene *Brittle*-2 (*Bt*2), and overexpression of this enzyme also increases yield by about 35%, albeit by an increase in seed number rather than seed weight. This increase is dependent on the average daily high temperature reaching ~33 °C during the first 4 days after pollination [57]. Furthermore, expression of an altered large subunit of endosperm AGPase increases potato tuber yield by 35% [58], wheat yield by 38% [59], and rice yield by 23% [60]. Notably, the yield increases in maize, rice, and wheat are due to increases in seed number rather than increases in individual seed weight. In rice, increases in seed number are a consequence of an enhanced probability of seed development rather than an increase in the number of ovaries. In wild-type rice, only about half of the ovaries develop into fully mature kernels. The remaining kernels cease development, and their contents disintegrate [56,57]. Overexpression of the small subunit of the maize endosperm AGPase (ZmBt1) in rice enhances ADP-Glc synthesis and imports into amyloplasts, but does not lead to further enhancement in seed weight, even under elevated CO_2 [55]. These observations suggest that the PPDK mutant might enhance the expression of *AGPS2a* and *AGPS2b*, as with AGPase transgene expression, leading to increases in the expression of *AGPS2a*, *AGPS2b*, *SSIIb*, and *SSIIc,* and an increase in seed

number rather than seed weight. Our results suggest that enhanced synthesis of ADP-Glc in the rice seed endosperm may increase seed number. Future research regarding the metabolic mechanisms underlying the increase in seed number will be useful for breeding programs aimed at improving rice yield.

5. Conclusions

Our data revealed that a novel SNP (C to T) in the coding region (exon 2) of the gene encoding cytosolic PPDK (cy*OsPPDK*) was responsible for the floury endosperm characteristics of low grain hardness, low starch damage, and fine particle size, and this mutation may be valuable in rice breeding programs. Co-segregation analysis with the developed CAPS marker (flo4-5_F/flo4-5_R) revealed co-segregation between the floury phenotype and the *flo4-5* using the segregation population and 44 japonica varieties. This CAPS marker could be applied directly to MAS. Real-time RT-PCR experiments revealed that PPDK was expressed at considerably higher levels in the *flo4-5* mutant than in the wild type during the grain filling stage. Plastids *AGPS2a*, *AGPS2b*, *SSIIb*, and *SSIIc* also exhibited enhanced expression in the *flo4-5* mutant. Although more studies are needed to fully understand the functions of PPDK, our data indicate that PPDK is involved in the endosperm development function through directly or indirectly regulating *AGPS2a*, *AGPS2b*, *SSIIb*, and *SSIIc*.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/4/465/s1, Table S1: Primers used for sequencing the whole genome sequence of *flo4-5* and qRT-PCR in this study, Table S2: Primer sequences of CAPS markers that developed for fine mapping, Table S3: Compare the whole genome sequences of *cyOsPPDK* gene among Nipponbare, *flo4-5* and its wild type Namil. InDel (Insertion or Deletion).

Author Contributions: H.W., T.-H.H. and S.-W.K.; developed ideas, designed and performed all experiments, and wrote the manuscript. Y.M. and J.-U.J.; developed the mutant lines and mapping population. D.-E.I. and J.L.; analyzed the whole genome resequencing dates. S.M.L. and S.-G.J.; produced rice plants for experiments. S.T.K.; support expression analysis. All authors read and approved the final manuscript.

Funding: This work was supported by the grant from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology and the research grant from Rural Development Administration, Republic of Korea.

Acknowledgments: This work was supported by the grant from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2018R1A4A1025158) and the research grant from the Rural Development Administration, Republic of Korea (PJ012890032020).

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

References

1. Tian, Z.; Liu, X.; Liu, G.; Wang, Y.; Li, J.; Qian, Q.; Yan, M.; Gao, Z.; Zeng, D.; Liu, Q.; et al. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21760–21765. [CrossRef] [PubMed]
2. Wani, A.A.; Singh, P.; Shah, M.A.; Schweiggert-Weisz, U.; Gul, K.; Wani, I.A. Rice Starch Diversity: Effects on Structural, Morphological, Thermal, and Physicochemical Properties-A Review. *Compr. Rev. Food Sci. Saf.* **2012**, *11*, 417–436. [CrossRef]
3. Wickramasinghe, H.A.M.; Noda, T. Physicochemical properties of starches from Sri Lankan rice varieties. *Food Sci. Technol. Res.* **2008**, *14*, 49–54. [CrossRef]
4. Whitt, S.R.; Buckler Iv, E.S.; Wilson, L.M.; Tenaillon, M.I.; Gaut, B.S. Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12959–12962. [CrossRef] [PubMed]
5. Lee, S.-K.; Hwang, S.-K.; Han, M.; Eom, J.-S.; Kang, H.-G.; Han, Y.; Choi, S.-B.; Cho, M.-H.; Bhoo, S.H.; An, G.; et al. Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa* L.). *Plant Mol. Biol. Rep.* **2007**, *65*, 531–546. [CrossRef]
6. Tang, X.-J.; Peng, C.; Zhang, J.; Cai, Y.; You, X.-M.; Kong, F.; Yan, H.-G.; Wang, G.-X.; Wang, L.; Jin, J.; et al. ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. *Plant Sci.* **2016**, *249*, 70–83. [CrossRef]

7. Li, N.; Zhang, S.; Zhao, Y.; Li, B.; Zhang, J. Over-expression of AGPase genes enhances seed weight and starch content in transgenic maize. *Planta* **2011**, *233*, 241–250. [CrossRef]
8. Hanashiro, I.; Kuratomi, Y.; Matsugasako, J.I.; Takeda, Y.; Itoh, K.; Yamazaki, M.; Igarashi, T. Granule-bound starch synthase I is responsible for biosynthesis of extra-long unit chains of amylopectin in rice. *Plant Cell Physiol.* **2008**, *49*, 925–933. [CrossRef]
9. Sano, Y.; Maekawa, M.; Kikuchl, H. Temperature effects on the Wx protein level and amylose content in the endosperm of rice. *J. Hered.* **1985**, *76*, 221–222. [CrossRef]
10. Wang, J.-C.; Xu, H.; Zhu, Y.; Liu, Q.-Q.; Cai, X.-L. OsbZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *J. Exp. Bot.* **2013**, *64*, 3453–3466. [CrossRef]
11. Dinges, J.R.; Colleoni, C.; Myers, A.M.; James, M.G. Molecular structure of three mutations at the maizesugary1 locus and their allele-specific phenotypic effects. *Plant Physiol.* **2001**, *125*, 1406–1418. [CrossRef] [PubMed]
12. Zhou, H.; Wang, L.; Liu, G.; Meng, X.; Jing, Y.; Shu, X.; Kong, X.; Sun, J.; Yu, H.; Smith, S.M. Critical roles of soluble starch synthase SSIIIa and granule-bound starch synthase Waxy in synthesizing resistant starch in rice. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12844–12849. [CrossRef] [PubMed]
13. Cai, Y.; Li, S.; Jiao, G.; Sheng, Z.; Wu, Y.; Shao, G.; Xie, L.; Peng, C.; Xu, J.; Tang, S. Os PK 2 encodes a plastidic pyruvate kinase involved in rice endosperm starch synthesis, compound granule formation and grain filling. *Plant Biotechnol. J.* **2018**, *16*, 1878–1891. [CrossRef] [PubMed]
14. Satoh, H.; Omura, T. New Endosperm Mutations Induced by Chemical Mutagens in Rice *Oryza sativa* L. *Jpn. J. Breed* **1981**, *31*, 316–326. [CrossRef]
15. Kang, H.G.; Park, S.; Matsuoka, M.; An, G. White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C4-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J.* **2005**, *42*, 901–911. [CrossRef]
16. Kaushik, R.; Khush, G. Genetic analysis of endosperm mutants in rice *Oryza sativa* L. *Theor. Appl. Genet.* **1991**, *83*, 146–152. [CrossRef]
17. Nishio, T.; Iida, S. Mutants having a low content of 16-kDa allergenic protein in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **1993**, *86*, 317–321. [CrossRef]
18. Ryoo, N.; Yu, C.; Park, C.-S.; Baik, M.-Y.; Park, I.M.; Cho, M.-H.; Bhoo, S.H.; An, G.; Hahn, T.-R.; Jeon, J.-S. Knockout of a starch synthase gene OsSSIIIa/Flo5 causes white-core floury endosperm in rice (*Oryza sativa* L.). *Plant Cell Rep.* **2007**, *26*, 1083–1095. [CrossRef]
19. She, K.-C.; Kusano, H.; Koizumi, K.; Yamakawa, H.; Hakata, M.; Imamura, T.; Fukuda, M.; Naito, N.; Tsurumaki, Y.; Yaeshima, M. A novel factor FLOURY ENDOSPERM2 is involved in regulation of rice grain size and starch quality. *Plant Cell* **2010**, *22*, 3280–3294. [CrossRef]
20. Peng, C.; Wang, Y.; Liu, F.; Ren, Y.; Zhou, K.; Lv, J.; Zheng, M.; Zhao, S.; Zhang, L.; Wang, C. FLOURY ENDOSPERM 6 encodes a CBM 48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. *Plant J.* **2014**, *77*, 917–930. [CrossRef]
21. Zhang, L.; Ren, Y.; Lu, B.; Yang, C.; Feng, Z.; Liu, Z.; Chen, J.; Ma, W.; Wang, Y.; Yu, X. FLOURY ENDOSPERM7 encodes a regulator of starch synthesis and amyloplast development essential for peripheral endosperm development in rice. *J. Exp. Bot.* **2016**, *67*, 633–647. [CrossRef] [PubMed]
22. Zhang, D.; Wu, J.; Zhang, Y.; Shi, C. Phenotypic and candidate gene analysis of a new floury endosperm mutant (osagpl2-3) in rice. *Mol Genet Genomics.* **2012**, *30*, 1303–1312. [CrossRef]
23. Crofts, N.; Abe, N.; Oitome, N.F.; Matsushima, R.; Hayashi, M.; Tetlow, I.J.; Emes, M.J.; Nakamura, Y.; Fujita, N. Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *J. Exp. Bot.* **2015**, *66*, 4469–4482. [CrossRef] [PubMed]
24. Hennen-Bierwagen, T.A.; Lin, Q.; Grimaud, F.; Planchot, V.; Keeling, P.L.; James, M.G.; Myers, A.M. Proteins from multiple metabolic pathways associate with starch biosynthetic enzymes in high molecular weight complexes: A model for regulation of carbon allocation in maize amyloplasts. *Plant Physiol.* **2009**, *149*, 1541–1559. [CrossRef]
25. Ordonio, R.L.; Matsuoka, M. Increasing resistant starch content in rice for better consumer health. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12616–12618. [CrossRef]
26. Imaizumi, N.; Ku, M.S.; Ishihara, K.; Samejima, M.; Kaneko, S.; Matsuoka, M. Characterization of the gene for pyruvate, orthophosphate dikinase from rice, a C3 plant, and a comparison of structure and expression between C3 and C4 genes for this protein. *Plant Mol. Biol.* **1997**, *34*, 701–716. [CrossRef]

27. Chastain, C.J.; Heck, J.W.; Colquhoun, T.A.; Voge, D.G.; Gu, X.-Y. Posttranslational regulation of pyruvate, orthophosphate dikinase in developing rice (*Oryza sativa*) seeds. *Planta* **2006**, *224*, 924–934. [CrossRef]
28. Lappe, R.R.; Baier, J.W.; Boehlein, S.K.; Huffman, R.; Lin, Q.; Wattebled, F.; Settles, A.M.; Hannah, L.C.; Borisjuk, L.; Rolletschek, H. Functions of maize genes encoding pyruvate phosphate dikinase in developing endosperm. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E24–E33. [CrossRef]
29. Lin, Q.; Huang, B.; Zhang, M.; Zhang, X.; Rivenbark, J.; Lappe, R.L.; James, M.G.; Myers, A.M.; Hennen-Bierwagen, T.A. Functional interactions between starch synthase III and isoamylase-type starch-debranching enzyme in maize endosperm. *Plant physiol.* **2012**, *158*, 679–692. [CrossRef]
30. Méchin, V.; Thévenot, C.; Le Guilloux, M.; Prioul, J.-L.; Damerval, C. Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant physiol.* **2007**, *143*, 1203–1219. [CrossRef]
31. Moons, A.; Valcke, R.; Van, M. Low-oxygen stress and water deficit induce cytosolic pyruvate orthophosphate dikinase (PPDK) expression in roots of rice, a C3 plant. *Plant J.* **1998**, *15*, 89–98. [CrossRef] [PubMed]
32. Shin, Y.S.; Park, C.S.; Seo, Y.W.; Jeung, J.U. Characteristics of endosperm starch of the rice mutant lines induced by sodium azide. *Korean J. Bredding Sci.* **2009**, *41*, 84–91.
33. Mo, Y.-J.; Jeung, J.-U.; Kang, K.-H.; Lee, J.-S.; Kim, B.-K. Genetic Analysis on Floury Endosperm Characteristics of'Namil (SA)-flo1', a Japonica Rice Mutant Line. *Korean J. Crop Sci.* **2013**, *58*, 283–291. [CrossRef]
34. Wang, H.; Mo, Y.-J.; Im, D.-E.; Jang, S.-G.; Ham, T.-H.; Lee, J.; Jeung, J.-U.; Kwon, S.-W. A new SNP in cyOsPPDK gene is associated with floury endosperm in Suweon 542. *Mol. Genet. Genom.* **2018**, *293*, 1151–1158. [CrossRef] [PubMed]
35. Mo, Y.-J.; Jeung, J.-U.; Shin, Y.-S.; Park, C.S.; Kang, K.-H.; Kim, B.-K. Agronomic and genetic analysis of Suweon 542, a rice floury mutant line suitable for dry milling. *Rice* **2013**, *6*, 37. [CrossRef] [PubMed]
36. Juliano, B. A simplified assay for milled rice amylose. *Cereal Sci. Today* **1971**, *16*, 334–360.
37. 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*, 4. [CrossRef]
38. Jeung, J.U.; Kim, B.R.; Cho, Y.C.; Han, S.S.; Moon, H.P.; Lee, Y.T.; Jena, K.K. A novel gene, Pi40(t), linked to the DNA markers derived from NBS-LRR motifs confers broad spectrum of blast resistance in rice. *Theor. Appl. Genet* **2007**, *115*, 1163–1177. [CrossRef]
39. Rice Genome Annotation Project. Available online: http://rice.plantbiology.msu.edu/index.shtml (accessed on 23 April 2020).
40. WatCut. Available online: http://watcut.uwaterloo.ca/template.php?act=snp_new (accessed on 23 April 2020).
41. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [CrossRef]
42. Rap-db. Available online: https://rapdb.dna.affrc.go.jp/index.html (accessed on 23 April 2020).
43. Chastain, C.J.; Chollet, R. Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C3 and C4 plants. *Plant Physiol. Biochem.* **2003**, *41*, 523–532. [CrossRef]
44. Huang, S.; Colmer, T.D.; Millar, A.H. Does anoxia tolerance involve altering the energy currency towards PPi? *Trends Plant Sci.* **2008**, *13*, 221–227. [CrossRef] [PubMed]
45. Walley, J.W.; Shen, Z.; Sartor, R.; Wu, K.J.; Osborn, J.; Smith, L.G.; Briggs, S.P. Reconstruction of protein networks from an atlas of maize seed proteotypes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4808–E4817. [CrossRef] [PubMed]
46. Fujita, N.; Yoshida, M.; Asakura, N.; Ohdan, T.; Miyao, A.; Hirochika, H.; Nakamura, Y. Function and characterization of starch synthase I using mutants in rice. *Plant Physiol.* **2006**, *140*, 1070–1084. [CrossRef] [PubMed]
47. Fujita, N.; Yoshida, M.; Kondo, T.; Saito, K.; Utsumi, Y.; Tokunaga, T.; Nishi, A.; Satoh, H.; Park, J.-H.; Jane, J.-L. Characterization of SSIIIa-deficient mutants of rice: The function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol.* **2007**, *144*, 2009–2023. [CrossRef]
48. Nakamura, Y.; Francisco, P.B.; Hosaka, Y.; Sato, A.; Sawada, T.; Kubo, A.; Fujita, N. Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. *Plant Mol. Biol. Rep.* **2005**, *58*, 213–227. [CrossRef]

49. Umemoto, T.; Yano, M.; Satoh, H.; Shomura, A.; Nakamura, Y. Mapping of a gene responsible for the difference in amylopectin structure between japonica-type and indica-type rice varieties. *Theor. Appl. Genet.* **2002**, *104*, 1–8. [CrossRef]
50. Nishi, A.; Nakamura, Y.; Tanaka, N.; Satoh, H. Biochemical and genetic analysis of the effects ofamylose-extender mutation in rice endosperm. *Plant Physiol.* **2001**, *127*, 459–472. [CrossRef]
51. Yang, R.; Sun, C.; Bai, J.; Luo, Z.; Shi, B.; Zhang, J.; Yan, W.; Piao, Z. A putative gene sbe3-rs for resistant starch mutated from SBE3 for starch branching enzyme in rice (*Oryza sativa* L.). *PLoS ONE* **2012**, *7*, e43026. [CrossRef]
52. Kubo, A.; Fujita, N.; Harada, K.; Matsuda, T.; Satoh, H.; Nakamura, Y. The starch-debranching enzymes isoamylase and pullulanase are both involved in amylopectin biosynthesis in rice endosperm. *Plant Physiol.* **1999**, *121*, 399–410. [CrossRef]
53. Kubo, A.; Rahman, S.; Utsumi, Y.; Li, Z.; Mukai, Y.; Yamamoto, M.; Ugaki, M.; Harada, K.; Satoh, H.; Konik-Rose, C. Complementation of sugary-1 phenotype in rice endosperm with the wheat isoamylase1 gene supports a direct role for isoamylase1 in amylopectin biosynthesis. *Plant Physiol.* **2005**, *137*, 43–56. [CrossRef]
54. Lee, Y.; Choi, M.-S.; Lee, G.; Jang, S.; Yoon, M.-R.; Kim, B.; Piao, R.; Woo, M.-O.; Chin, J.H.; Koh, H.-J. Sugary endosperm is modulated by starch branching enzyme IIa in rice (*Oryza sativa* L.). *Rice* **2017**, *10*, 33. [CrossRef] [PubMed]
55. Cakir, B.; Shiraishi, S.; Tuncel, A.; Matsusaka, H.; Satoh, R.; Singh, S.; Crofts, N.; Hosaka, Y.; Fujita, N.; Hwang, S.-K. Analysis of the rice ADP-glucose transporter (OsBT1) indicates the presence of regulatory processes in the amyloplast stroma that control ADP-glucose flux into starch. *Plant Physiol.* **2016**, *170*, 1271–1283. [PubMed]
56. Hannah, L.C.; Futch, B.; Bing, J.; Shaw, J.R.; Boehlein, S.; Stewart, J.D.; Beiriger, R.; Georgelis, N.; Greene, T. A shrunken-2 transgene increases maize yield by acting in maternal tissues to increase the frequency of seed development. *Plant Cell.* **2012**, *24*, 2352–2363. [CrossRef] [PubMed]
57. Hannah, L.C.; Shaw, J.R.; Clancy, M.A.; Georgelis, N.; Boehlein, S.K. A brittle-2 transgene increases maize yield by acting in maternal tissues to increase seed number. *Plant Direct.* **2017**, *1*, e00029. [CrossRef] [PubMed]
58. Stark, D.M.; Timmerman, K.P.; Barry, G.F.; Preiss, J.; Kishore, G.M. Regulation of the amount of starch in plant tissues by ADP glucose pyrophosphorylase. *Science* **1992**, *258*, 287–292. [CrossRef] [PubMed]
59. Smidansky, E.D.; Clancy, M.; Meyer, F.D.; Lanning, S.P.; Blake, N.K.; Talbert, L.E.; Giroux, M.J. Enhanced ADP-glucose pyrophosphorylase activity in wheat endosperm increases seed yield. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1724–1729. [CrossRef] [PubMed]
60. Smidansky, E.D.; Martin, J.M.; Hannah, C.L.; Fischer, A.M.; Giroux, M.J. Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. *Planta* **2003**, *216*, 656–664. [CrossRef]

genes

MDPI

Article

Drought-Induced Regulatory Cascades and Their Effects on the Nutritional Quality of Developing Potato Tubers

Letitia Da Ros [1,2], Raed Elferjani [3], Raju Soolanayakanahally [3,*], Sateesh Kagale [2], Shankar Pahari [3], Manoj Kulkarni [2], Jazeem Wahab [3] and Benoit Bizimungu [4]

1 Faculty of Forestry, University of British Columbia, Vancouver, BC V6T1Z4, Canada; lmdaros@mail.ubc.ca
2 National Research Council Canada, Saskatoon, SK S7N0W9, Canada; sateesh.kagale@nrc-cnrc.gc.ca (S.K.); ultimatekmanoj@gmail.com (M.K.)
3 Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK S7N0X2, Canada; cosmos.gob@gmail.com (R.E.); shankar.pahari@canada.ca (S.P.); jazeem.wahab@canada.ca (J.W.);
4 Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, NB E3B4Z7, Canada; benoit.bizimungu@canada.ca
* Correspondence: raju.soolanayakanahally@canada.ca

Received: 17 June 2020; Accepted: 28 July 2020; Published: 30 July 2020

Abstract: Competition for scarce water resources and the continued effects of global warming exacerbate current constraints on potato crop production. While plants' response to drought in above-ground tissues has been well documented, the regulatory cascades and subsequent nutritive changes in developing tubers have been largely unexplored. Using the commercial Canadian cultivar "Vigor", plants were subjected to a gradual drought treatment under high tunnels causing a 4 °C increase in the canopy temperature. Tubers were sampled for RNAseq and metabolite analysis. Approximately 2600 genes and 3898 transcripts were differentially expressed by at least 4-fold in drought-stressed potato tubers, with 75% and 69% being down-regulated, respectively. A further 229 small RNAs were implicated in gene regulation during drought. Expression of several small RNA clusters negatively correlated with expression of their six target patatin genes, suggesting involvement in the regulation of storage proteins during drought. The comparison of protein homologues between *Solanum tuberosum* L. and *Arabidopsis thaliana* L. indicated that down-regulated genes were associated with phenylpropanoid and carotenoid biosynthesis. As is indicative of reduced flow through the phenylpropanoid pathway, phenylalanine accumulated in drought-stressed tubers. This suggests that there may be nutritive implications to drought stress occurring during the potato tuber bulking phase in sensitive cultivars.

Keywords: crop genetics; *Solanum tuberosum*; abiotic stress; phenylpropanoids; essential amino acid; transcriptome; small RNA; comparative genomics; nutrition

1. Introduction

Potatoes are the fourth most consumed food crop worldwide and are an efficient source of energy, vitamins and minerals in the human diet [1]. High consumption rates and moderate concentrations of dietary antioxidants have led potatoes to be the third-largest source of total phenolics in the American diet [2]. Diets rich in phenolics have been implicated in the prevention of an array of degenerative diseases and concentrations of these compounds vary greatly based on the cultivar, highlighting the potential for the targeted breeding of potato (*S. tuberosum* L.) to enhance global human health [3]. In potato tubers, the primary polyphenol is chlorogenic acid with the remaining components comprised of carotenoids, anthocyanins, and flavonoids. While the phenolic content of tubers is largely

genotype-dependent, the phenolic profiles are driven by the environmental conditions present during growth, tuber bulking, and throughout storage. In general, the production and accumulation of these compounds are favored at lower temperatures, with potatoes grown in warm, dry regions producing lower amounts of phenolics [4]. Environmental parameters could be manipulated to manage the concentrations of desired phytonutrients [4–6].

Substrate entry into the general phenylpropanoid pathway in eudicots is driven by phenylalanine ammonia-lyase (PAL), an enzyme that regulates the deamination of phenylalanine to yield the cinnamic acid from which monolignols, flavonoids, and anthocyanins are produced [7]. PAL activity responds to a variety of developmental and environmental cues, with transcriptional regulation occurring by way of MYB, LIM, and KNOX transcription factors [8]. Furthermore, independent MYB transcription factors play a prominent role in the regulation of anthocyanin and flavonoid biosynthesis genes such as flavonol synthase (FLS), flavanone 3-hydroxylase (F3H) and flavonoid 3′–hydroxylase (F3′H), while expression of genes such as dihydroflavonol reductase (DFR) require MYB transcriptional complexes. MicroRNAs (miRNAs) and small interfering RNAs (siRNAs) have notable functions in the regulation of phenylpropanoid biosynthesis in eudicots through the targeting of MYBs, the most notable being miR858, miR828 and TAS4 [7,9]. Causal miRNA and the target MYB transcription factors have been previously identified in potato leaves under drought [10], however, the regulatory cascades present in potato tubers are still unknown.

In addition to phenolics, concentrations of available essential amino acids affect the nutritional value of potato tubers. Up to 50% of the amino acids in tubers are aspartic acid and glutamic acid, with the remaining portion made up of leucine, valine, alanine, lysine, and arginine with the total protein nutritional value of a potato being comparable to an egg white [11]. Genotypes for improving protein quality have been identified among non-traditional potato cultivars [12]. Essential amino acids function both as substrates for secondary metabolism and as a source of energy [13]. As a result, concentrations fluctuate in response to environmental stressors due to concurrent protein degradation and *de novo* synthesis. The transcriptional regulation of amino acid biosynthesis is highly complex and their function during stress response is still unclear [13].

Drought stress is one of the primary concerns for potato production given the projected increases in aridity. Potato is adapted to temperate climates with optimal tuber growth occurring at temperatures between 15 and 20 °C. Temperatures above this range, coupled with periodic drought, have resulted in reduced yields and increased incidences of tuber physiological defects [14]. Symptoms of drought in potato include reduced leaf size, increased chlorophyll content, reduced stomatal conductance, and wilting. However, rooting depth and plant recovery have been shown as the best indicators of plant susceptibility to drought [15,16]. Through comparisons between genotypes with differing tolerance to drought, novel potato drought-responsive genes and transcript markers for drought tolerance in potato leaves have been identified [17,18]. Gene responses in developing potato tubers to drought conditions are not well-documented outside of targeted metabolic pathways [3].

This study aimed at identifying drought-associated changes in developing potato tubers (i.e., tuber bulking phase) and their impacts on nutritional quality. The assessment of transcriptional changes in genes with metabolic functions and quantification of amino acid concentrations aims to guide production and harvest practices when optimizing the nutritional value of the crop. Analysis of small RNAs seeks to identify components of the drought regulatory cascade in potato tubers which, to our knowledge, has yet to be explored.

2. Materials and Methods

2.1. Experimental Design and Plant Growth Conditions

This study was conducted during the summer of 2017 at the Agriculture and Agri-Food Canada's Canada-Saskatchewan Irrigation Diversification Centre in Outlook (51°29′ N, 107°03′ W, 541 m), Saskatchewan. The cultivar "Vigor", a cross between "Agria" and "Wischip" from the Agriculture and

Agri-Food Canada Lethbridge Research and Development Centre, was evaluated for its performance under soil moisture stress during the tuber bulking phase. Prominent characteristics of the cultivar are its yellow-fleshed tubers and pigmented (red-violet) flowers. Plants were grown under optimum soil moisture conditions at 70% field capacity (FC) and restricted soil moisture conditions at 35% FC under two high tunnels using drip irrigation. Treatments were imposed at the start of the tuber bulking phase for gradual exposure to drought stress, thereby mimicking natural field conditions (Figure 1). High tunnels were opened from all sides but covered with plastic film on top to mimic the open field condition while preventing rainfall (Supplementary Figure S1A). Plots were laid out in a randomized complete block design containing four replicates with guard rows on either side. Each plot consisted of 12 hills. The two end hills were considered as guard hills for yield estimation purposes. Seed pieces were spaced 1 m between-rows and 20 cm within-rows and were planted on 30 May 2017. The crop was raised using standard management practices (i.e., fertility, irrigation, pest control, etc.). Pre-plant basal fertilizer included urea (46-0-0), mono-ammonium phosphate (11-52-0), and potash (0-0-60). Two applications of ammonium sulphate (21.5-0-0-4) were given at 4 and 7 weeks after planting. Soil moisture was monitored using Watermark sensors (Supplementary Figure S1B). Plots were harvested on 2 October 2017 and graded according to commercial-grade standards.

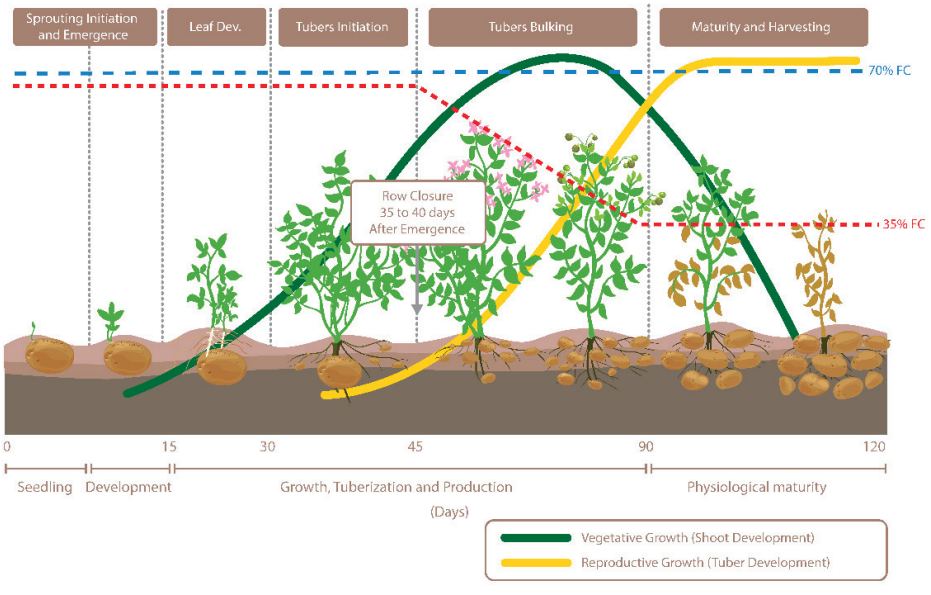


Figure 1. Visualization of potato growth stages and in-season soil moisture trends when soil moisture was maintained at 35% and 70% field capacity (FC).

2.2. Physiological Measurements

Physiological measurements were taken at the end of the tuber bulking phase (90th day after planting) with readings recorded consistently between 11:30 a.m. and 12:30 p.m. Canopy temperature was assessed with an infra-red thermal imaging camera (FLIR T530, FLIR Systems. Wilsonville, OR, USA), and leaflet chlorophyll content (CCI) was recorded using a chlorophyll content meter (CCM-200-Apogee Instruments, Logan, UT, USA). Quantum yield of dark-adapted leaflets (*Fv/Fm*) was measured using the portable fluorometer FluorPen FP 100 (PSI, DRASO Czech Republic). Detachable clips were used to dark-adapt the leaflets for 20 min, and *Fv/Fm* was measured on the adaxial surfaces of the top 3rd and 4th leaflet of each sampled plant (three plants per replication).

2.3. Amino Acid Profiling and Abscisic Acid Content

In both high tunnels from each replication, a tuber was collected from the middle of the plot on 13 September 2017 (106 DAP). The tubers were washed in running water, followed by distilled water, cut into cubes while evading the skin and immediately frozen in liquid nitrogen. Samples were stored at −80 °C until further use.

Amino acids were extracted from 10 mg of ground freeze-dried tissue following Inaba et al. (1994) [19] with some modifications. Briefly, 1 mL of 80% (*v/v*) ethanol solution (40 °C) was added to each sample, shaken for 30 min at 40 °C and the supernatant was recovered by centrifugation (4000 rpm for 10 min) at 4 °C. The pellets were re-extracted under the same conditions with an additional 500 µL of 80% (*v/v*) ethanol (40 °C). The supernatants were combined and stored at −20 °C until further use. Amino acids were derivatized following Waters AccQTag Reagent Kit (Waters, Milford, MA, USA; [20]). Briefly, a 10 µL aliquot of sample was mixed with 70 µL borate buffer and 20 µL AccQFluor reagent which was reconstituted in acetonitrile. AccQFluor reagent was reconstituted as follows: 1 mL of AccQFluor reagent diluent was transferred to a vial containing AccQFluor reagent powder and vortexed for 10 s before heating at 55 °C for a maximum of 10 min or until dissolved. The derivatized mixture was transferred to an autosampler vial and incubated at 55 °C for 10 min. High-performance liquid chromatography (HPLC) was conducted, as described in Waters AccQTag's chemistry package instruction manual, with samples separated on a Waters amino acid column −3.9 × 150 mm and quantified at an excitation wavelength of 285 nm and an emission wavelength of 320 nm using a 2475 scanning fluorescence detector (Waters, Milford, MA, USA). The column was set at 37 °C with a 5 µL injection volume. Waters AccQTag buffer (100 mL AccQTag Buffer concentrate +1000 mL deionized water), acetonitrile, and deionized water were used as mobile phases A, B, and C, respectively.

Abscisic acid content was determined following Yan et al., (2016) [21]. Samples were centrifuged to remove debris, and the pellet was washed twice. The supernatant was evaporated in a SpeedVac, and reconstituted in 1 mL of 1% (*v/v*) acetic acid. Abscisic acid (ABA) was purified by solid-phase extraction using Oasis HLB, MCX, and WAX cartridge columns (Waters, Milford, MA, USA). The solvent was removed under vacuum and subjected to LC-ESI-MS/MS analysis (Agilent 6410 TripleQuad LC/MS system). An LC (Agilent 1200 series) equipped with a 50 × 2.1 mm, 1.8-µm Zorbax SB-Phenyl column (Agilent Technologies, Santa Clara, CA, USA) was used with a binary solvent system comprised of 0.01% (*v/v*) acetic acid in water (solvent A) and 0.05% (*v/v*) acetic acid in acetonitrile (solvent B). Separations were performed using a gradient of increasing acetonitrile content at a flow rate of 0.2 mL min^{-1}. The gradient was increased linearly from 3% B to 50% B over 15 min. The retention time of ABA was 14 min.

2.4. Transcriptome and Small RNA Sequencing

Total RNA was extracted from 100 mg of tuber tissue partitioned from the sample taken for metabolite analysis using the RNeasy plant mini kit (Qiagen, Hilden, Germany). RNA quality and concentration were verified using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). TruSeq RNA and small RNA sequencing libraries were constructed following the standard preparation guide (Illumina, San Diego, CA, USA). All eight RNA samples (four replicates of each treatment) were multiplexed in a lane of a flow cell and paired-end sequencing (125 cycles) was performed using an Illumina HiSeq 2500. Similarly, for small RNA sequencing, all 8 samples were multiplexed in a lane of a flow cell and single-end sequencing was carried out on Illumina HiSeq 2500.

2.5. RNA and Small RNA Read Mapping and Analysis

Before read mapping and expression quantification, all RNA reads were filtered using Trimmomatic (version 0.36; [22]) by (i) removing adapter sequences, (ii) trimming leading and trailing low-quality sequences, (iii) removing sequences when the average quality per base dropped below 15 within a 4-base wide sliding window and (iv) keeping only those pairs where both reads were longer than 75 bp. Clean reads were aligned to the potato reference genome (SolTub_3.0, EnsemblPlants) with

STAR (v2.5.2b) and isoform expression was quantified with the RSEM (v1.3.3) algorithm [23]. The expected read counts generated by the RSEM algorithm were rounded off and fed into DESeq2.

The quality of small RNA sequencing reads was assessed using the FASTQC program (v0.11.8; [24]). Reads were quality-filtered and adapter-trimmed using cutadapt (v2.8; [25]). The alignment of filtered reads to the potato reference genome (SolTub_3.0, EnsemblPlants) and annotation and quantification of small RNAs was carried out using ShortStack (v3.8.5; [26]). psRNATarget [27] was used to predict the miRNA and small RNA target genes.

2.6. Differential RNA Expression Analysis

Raw read counts obtained from RNAseq were normalized and assessed for differential expression using the Statistical Software "R" version 3.6.0 and the package DESeq2 [28,29]. Log2 fold change threshold of 2 and a 5% false discovery rate (FDR) were used as cut-off values for continuing to the annotation step. R scripts are available on Bioconductor (https://bioconductor.org/) from the package developers and were adapted for the data presented in this paper. The same technique was repeated for the discovery of differentially expressed small RNA, with target gene identification done using psRNATarget [27]. Gene annotations for *S. tuberosum* and *A. thaliana* were obtained from the Ensembl Plants database (http://plants.ensembl.org). *Arabidopsis* homologs with >50% identity to the original potato gene were input into the online DAVID Bioinformatics Resources version 6.8 (https://david.ncifcrf.gov/) for functional annotation clustering and KEGG pathway mapping analyses.

3. Results

3.1. Physiological Response

A high degree of variability existed within the four control and four treatment plots for the agronomic and physiological traits measured during this study (Table 1). Both crop yield and tuber number per plot were not found to be significantly different from one another, although differences in plot averages were observed. Calculated values from spectral measurements, such as CCI and *Fv/Fm*, showed no significant differences. However, canopy temperatures measured in plots maintained at 35% FC were 3.9 °C higher than control plots with soil moisture maintained at 70% FC (Table 1). ABA concentrations in the well-watered control and water-deficit conditions were 43.9 and 55.4 ng gDW^{-1}, respectively. Variation was high among the control plants, resulting in no significant differences between treatments.

Table 1. Agronomic and physiological traits averaged for drought (35% field capacity) and control (70% field capacity) plots with standard error of the mean in parentheses. Bold numbers indicate significance ($p < 0.05$) between treatments.

Treatment	Yield (g)	No. of Tubers	CCI	Fv/Fm	Canopy Temp (°C)
35% FC	1381.7 (248.1)	12.8 (0.8)	19.5 (0.2)	0.24 (0.05)	**28.0 (0.3)**
70% FC	1737.3 (198.1)	18.1 (3.5)	15.4 (1.6)	0.38 (0.03)	24.1 (0.9)

3.2. Tuber Amino Acid Fluctuations in Response to Soil Moisture Deficit

Of the eight essential amino acids, lysine, phenylalanine, isoleucine, and leucine were found to be more abundant in drought-stressed tubers. The largest differences were observed in the concentrations of leucine, phenylalanine, and isoleucine which increased by 3×, 2×, and 1.9×, respectively (Figure 2A). Quantities of branched-chain amino acids, a group that includes leucine, isoleucine, and valine, were therefore significantly higher under drought treatment. Histidine and valine were the most abundant essential amino acids in developing potato tubers (Figure 2A). The majority of non-essential amino acids had similar concentrations in developing tubers regardless of the treatment. Only glutamic acid showed a marked increase of 5.75 μmol g^{-1} under reduced soil moisture conditions. Concentrations of cysteine, proline, and serine were highest among all amino acids measured (Figure 2B).

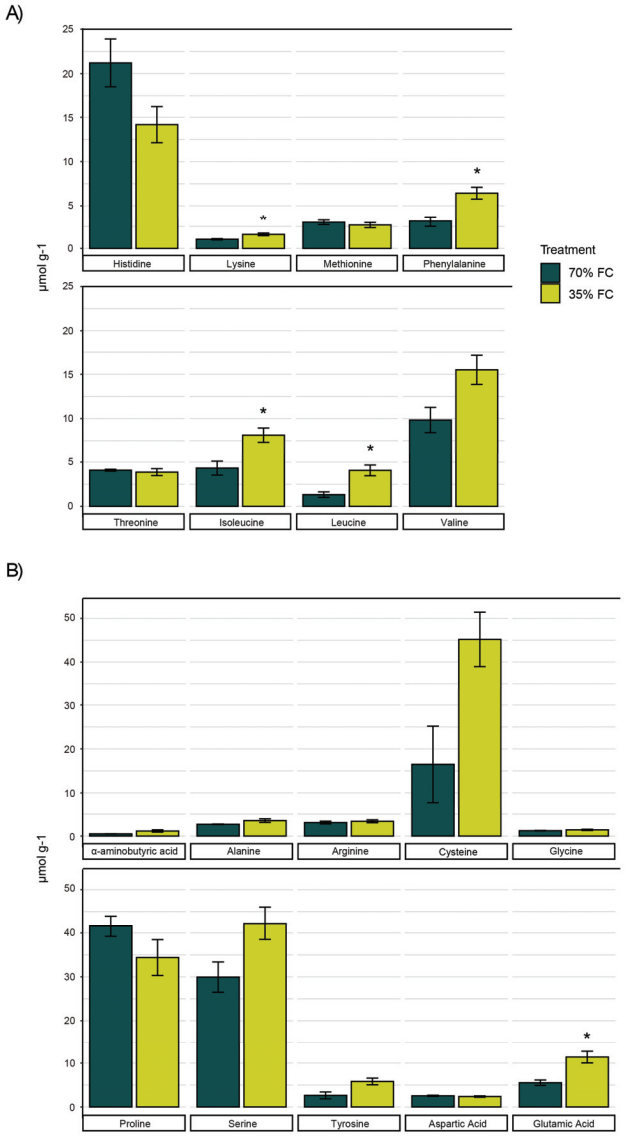


Figure 2. Concentrations of essential (**A**) and non-essential (**B**) amino acids in potato tubers ($n = 4$) sampled during the tuber bulking phase and the associated SEM ($p < 0.05$) when subjected to 70% and 35% field capacity. * significance ($p < 0.05$) between treatments.

3.3. Differential Gene Expression and Regulatory Cascades in Developing Tubers Under Drought Stress

Tubers subjected to soil moisture deficit showed both differential gene and transcript expression, with 75.2% and 68.8% being down-regulated, respectively. One-fifth of genes with differential expression were of unknown function. A full summary of the observed changes is listed in Figure 3A with a list of all differentially expressed genes provided in Supplementary Table S1. Down-regulated genes include those with functions in ABA, auxin and ethylene signaling as well as in auxin, carotenoid, and phenylpropanoid biosynthesis. Up-regulated genes have roles in amino acid biosynthesis, function as molecular chaperones and are involved in ubiquitin-driven proteolysis. Gene names, functional annotation, and the corresponding *Arabidopsis* homologs used for pathway mapping can be found in Table 2. Focusing on the regulation of the phenylpropanoid pathway, fifteen annotated MYB transcription factors were down-regulated by more than 4-fold under low-soil moisture conditions. Up-regulated transcription factors include two MYB transcription factors and one LIM transcription factor. No other MYB, KNOX or LIM transcription factors were above the cut-off values of 5% FDR and a log-fold change greater than two. Genes with key functions in the phenylpropanoid and carotenoid pathways that were down-regulated in potato tubers at 35% FC are highlighted in Figure 4. Raw read counts for gene expression analysis can be found in Supplementary Table S2.

A)

Differentially expressed (DE) gene summary	
No. DE genes	2600
No. DE transcripts	3898
Genes downregulated	75.2 %
Transcripts downregulated	68.8 %
Genes of unknown function	21.5 %
No. genes with At homologs	1026
Genes with At homologs mapped to a pathway	856

B)

Differentially expressed (DE) small RNA summary	
No. DE small RNA/ small RNA clusters	229
Small RNA downregulated	31.8 %
No. with identified gene targets	84
No. with target gene inhibition via cleavage	88.1 %
No. unique target genes	60
Unique target genes with unknown function	20.0 %

Figure 3. Summaries of differentially expressed genes (**A**) and small RNA (**B**) in potato tubers at 35% FC using a threshold of 4-fold difference in expression and a 5% FDR. Gene homologs in *A. thaliana* were considered if identity was greater than 50%. Functional annotation clustering to KEGG pathways was based on *Arabidopsis* gene IDs using the DAVID Bioinformatics online resource 6.8 (https://david.ncifcrf.gov/).

Table 2. List of differentially expressed genes in drought-stressed potato tubers, the corresponding homologs in *A. thaliana,* and the pathway in which they participate. Pathway mapping was done based on the *Arabidopsis* gene names and similarities between the original *S. tuberosum* gene and its homolog are expressed as the percentage of identical base pairs in the gene sequences (% ID).

Gene Regulation	Pathway	Gene Name	Description	Log2 Fold Change	At Homologs	Descriptor	% ID
Down-regulated	ABA signaling	PGSC0003DMG400002100	Abscisic acid receptor PYR1	−2.07	AT4G17,870	PYR1	72.2
					AT5G46,790	PYL1	61.0
	Auxin	PGSC0003DMG400001589	Amino acid transporter	−5.30	AT2G21050	LAX2	86.3
	biosynthesis and signaling	PGSC0003DMG400024978	Indole-3-acetic acid-amido synthetase GH3.3	−5.06	AT2G14960	GH3.1	77.3
					AT2G23170	GH3.3	74.1
					AT4G37390	GH3.2	73.3
					AT1G59500	GH3.4	69.8
		PGSC0003DMG400024997	Indole-3-acetic acid-amido synthetase GH3.6	−2.17	AT5G54510	GH3.6	70.8
		PGSC0003DMG400014707	Flavin monooxygenase	−3.42	AT4G28720	YUC8	68.3
					AT5G43890	YUC5	67.2
		PGSC0003DMG400026087	Flavin monooxygenase	−3.09	AT5G11320	YUC4	57.4
					AT4G32540	YUC	54.3
		PGSC0003DMG400003773	SAUR family protein	−8.34	AT1G75580	SAUR51	72.2
					AT1G19830	SAUR54	61.5
		PGSC0003DMG400001667	SAUR family protein	−7.40	AT4G38860	SAUR16	64.8
					AT4G34760	SAUR50	64.5
					AT2G21220	SAUR12	63.5
					AT2G16580	SAUR8	63.0
		PGSC0003DMG400001614	SAUR family protein	−3.75	AT4G34760	SAUR50	75.7
					AT4G38860	SAUR16	73.3
					AT2G16580	SAUR8	71.3
					AT2G21220	SAUR12	71.1
		PGSC0003DMG400001668	SAUR family protein	−3.71	AT4G38860	SAUR16	77.1
					AT4G34760	SAUR50	76.6
					AT2G21220	SAUR12	75.0
					AT2G16580	SAUR8	70.4
		PGSC0003DMG400001655	SAUR family protein	−2.98	AT4G34750	SAUR49	54.0
		PGSC0003DMG400022233	SAUR family protein ARG7	−2.93	AT3G12830	SAUR72	64.4
					AT1G16510	SAUR41	55.1
		PGSC0003DMG400001615	SAUR family protein	−2.06	AT4G34760	SAUR50	73.8
					AT4G38860	SAUR16	71.4
					AT2G21220	SAUR12	69.2
					AT2G16580	SAUR8	68.5
	Carotenoid biosynthesis	PGSC0003DMG400028180	Cytochrome P450-type monooxygenase 97C11	−2.07	AT3G53130	LUT1	77.2
		PGSC0003DMG400024063	Phytoene synthase 1, chloroplastic	−5.07	AT5G17230	PSY	64.3
	Ethylene signaling	PGSC0003DMG400014204	Transcription factor TSRF1	−3.57	AT3G23240	ERF1	51.4
	Phenylpropanoid	PGSC0003DMG400003605	Dihydroflavonol 4-reductase	−5.19	AT5G42800	DFR	59.2
	biosynthesis	PGSC0003DMG400014093	Flavonol synthase	−2.19	AT5G08640	FLS1	62.5
					AT5G63590	FLS3	50.3
		PGSC0003DMG400014152	Hydroxycinnamoyl transferase	−2.00	AT5G48930	HCT	77.8
		PGSC0003DMG400023458	Phenylalanine ammonia-	−4.68	AT3G10340	PAL4	79.9

Table 2. *Cont.*

Gene Regulation	Pathway	Gene Name	Description	Log2 Fold Change	At Homologs	Descriptor	% ID
			lyase		AT5G04230	PAL3	73.2
		PGSC0003DMG400014223	4-coumarate–CoA ligase 2	−2.30	AT3G21240 AT1G51680 AT3G21230	4CL2 4CL1 4CL4	68.5 67.9 58.9
		PGSC0003DMG400028929	4-coumarate–CoA ligase 2	−2.00	AT3G21240 AT1G51680 AT3G21230	4CL2 4CL1 4CL4	69.2 68.8 59.8
Up-regulated	Amino acid biosynthesis	PGSC0003DMG400034102	Acetolactate synthase	2.20	AT3G48560	CSR1	76.9
	Protein folding	PGSC0003DMG400008223	Heat shock factor protein HSF30	4.44	AT2G26150	HSFA2	51.0
		PGSC0003DMG400003219	Small heat shock protein, chloroplastic	4.11	AT4G27670	Heat shock protein 21	53.7
		PGSC0003DMG400030341	Small heat shock protein-Class I 17.6kD	3.99	AT2G29500	HSP17.6B	77.8
		PGSC0003DMG400024707	Small heat shock protein	2.90	AT1G09080	Heat shock protein 70	75.1
		PGSC0003DMG402028907	Small heat shock protein 90	2.72	AT5G52640	Heat shock protein 90	52.0
		PGSC0003DMG400030426	Small heat shock protein-Class I 17.6kD	2.50	AT2G29500	HSP17.6B	74.5
	Proteolysis	PGSC0003DMG400006185	Skp1 1	2.56	AT1G75950	SKP1	74.4
		PGSC0003DMG400006184	Skp1	2.20	AT1G75950	SKP1	75.0

Table 3. List of differentially expressed small RNA clusters in drought-stressed potato tubers that negatively correlate to target transcript expression. Target alignments, gene ID, expression and descriptions are included.

Small RNA Cluster	Log2 Fold Change	Target Alignment	Target Gene	Log2-Fold Change	Protein Description
Cluster 34023	5.03	AGCUCAUUAAUCUCUUCGAUA	PGSC0003DMG400009921	−6.24	Cysteine protease 14
Cluster 23921	4.68	AGGGUUCAAGAAAAUGCAUUA	PGSC0003DMG400029247	−4.75	Patatin group O
Cluster 15144	4.62	AGGGUUCAAGAAAAUGCAUUA			
Cluster 41775	4.49	ACCUCAGGGUUCAAGAAAAUG			
Cluster 83189	5.49	AGGCACUGGCACUACUUCAGA	PGSC0003DMG400017091	−4.25	Patatin-01; Probable lipolytic acyl hydrolase
Cluster 83175	4.98	AGCCAGUAAUAUUCACCAAGU			
Cluster 83174	3.45	AGGCACUGGCACUACUUCAGA			
Cluster 7920	4.95	GGCAGCAAGUUCUUACAUGAC	PGSC0003DMG400008749	−4.06	Patatin-05; Probable lipolytic acyl hydrolase
Cluster 68384	3.01	AUCAUUCCGGGUAUCAUUCUC			
Cluster 83190	2.87	UUCCGGGUAUCAUUCUCGAAU			
Cluster 83166	2.66	UCCGGGUAUCAUUCUCGAAU			
Cluster 68380	5.49	AGGCACUGGCACUAAUUCAGA	PGSC0003DMG400014104	−4.47	Patatin-2-Kuras 4; Probable lipolytic acyl hydrolase
Cluster 83164	5.49	AGGCAGCUAAAUGGGGUCCUC			
Cluster 20497	5.38	CUGUUGGUGAUCCGGCGUUA			
Cluster 68397	5.36	GUUGCUACUGUUGGUGAUCCG			
Cluster 83182	4.97	GGCACUACUUCAGAGUUUGAU	PGSC0003DMG401017090	−4.91	Patatin-3-Kuras 1

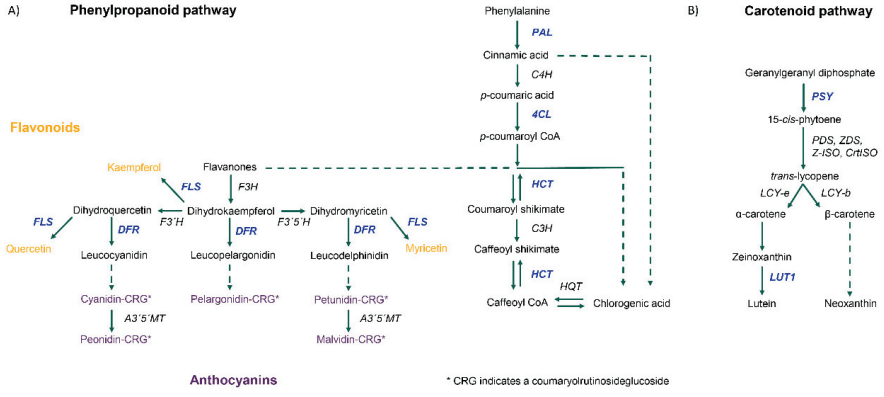


Figure 4. Diagrams depicting key enzymes in the phenylpropanoid (**A**) and carotenoid (**B**) biosynthetic pathways. Genes that are significantly down-regulated beyond cut-off values of 5% FDR and a log-fold change greater than 2 are written in blue.

The isolated small RNA were grouped into 87,213 clusters with an additional 10,209 unassigned sequences. Of these, 103 clusters and 126 unassigned sequences were differentially expressed. Additional summary statistics are listed in Figure 3B. Differentially expressed small RNA clusters with identified gene targets are listed in Supplementary Table S3. None of the small RNA clusters with target genes listed in Table 2 showed differential expression between the two treatments. Expression of target MYB transcription factor genes was also not correlated to small RNA cluster expression (Supplementary Table S4). Interestingly, the expression of small RNA clusters primarily targeting patatin genes was significantly up-regulated and negatively correlated to target gene expression ($r = -0.61$). These clusters and their targets can be found in Table 3.

4. Discussion

Optimal potato tuber growth occurs around 20 °C and plants are susceptible to losses in productivity under hot, arid conditions. Such conditions are expected to increase in the coming decade, therefore functional indicators of plant stress and the cascading effects on the developing tubers were evaluated. In this study, the Canadian potato cultivar "Vigor" was gradually exposed to increasing water deficit to a level of 35% FC beginning at the start of the tuber bulking phase (Figure 1). As seen previously in the literature, fluorescent measurements were not distinguishable between treatments [15] and thus were not dependable indicators of drought stress in potato plants (Table 1). At the plot level, there were no significant differences in yield or tuber number between the treatments (Table 1), however, discrepancies could become more prominent in commercial field production. Canopy temperature was considerably elevated in the drought treatments and has evidence supporting its use for drought stress assessments [30]. Average concentrations of ABA trended upwards in potato tubers exposed to water deficit compared to well-watered controls, suggesting drought responses had been initiated. Gene expression data further corroborated that drought signaling pathways had been activated as there was marked down-regulation of an ABA receptor PYR1, down-regulation of a series of small auxin up-regulated RNA (SAUR) genes involved in cell expansion and organ elongation in response to the environment [31] and the up-regulation of heat shock factor proteins [32] (Table 2). Notably, outside of common changes to the regulation of heat shock proteins, several genes previously identified as differentially expressed in severely drought-stressed potato leaves [32] were inversely regulated in the mildly stressed potato tubers collected in this study. These include the WRKY transcription factor (PGSC0003DMG400001434) and the developmental gene UPA16 (PGSC0003DMG400031742) [32] which had 2-fold and 78-fold increases in expression compared to well-watered control tubers. Lists of

genes implicated in drought stress response in potato leaf tissue [33] and potato stolons [34] have been compiled and here we provide those in developing tubers (Table 2; Supplementary Table S1).

Metabolic effects of drought on parameters such as free amino acids, soluble protein, and phenolics were assessed. The gradual drought stress to which the potato tubers were exposed resulted in no significant differences in total free amino acid concentrations, although treatment averages appeared to be divergent with 154.1 and 207.2 μmol g^{-1} in the control and drought treatments, respectively. These differences could be associated with the up-regulation of genes involved in proteolysis (Table 2). Elevated concentrations of proline have been shown to indicate stress in potato leaves [35], however, similar concentrations were observed in tubers irrespective of treatment. The largest changes occurred in the amino acid profile, where concentrations of branched-chain amino acids leucine and isoleucine increased (Figure 2). This indicated a greater proportion of dietary essential amino acids. Increases in branched-chain amino acids are likely attributable to the up-regulation of acetolactate synthase (Table 2), which is the first enzyme in the branched-chain amino acid synthesis pathway [36].

A major fraction (up to 40%) of the soluble protein in potato tubers consists of glycoproteins, known as patatins, that act both as storage proteins and show activity as non-specific lipid acyl hydrolases (LAH) with potential roles in plant defense against biotic stressors [37,38]. In the case of abiotic stress, it was observed that five patatin genes were down-regulated by at least 16-fold with the regulation of gene expression likely occurring via an increased presence of small RNA (Table 3). A possible consequence is reduced protein content in the resulting potato tubers.

As one of the major sources of plant phenolics in the human diet, potatoes have been targeted in breeding for greater total phenolics and antioxidant capacity [3,39]. Phenolic content is known to show a high degree of environmental plasticity with cooler temperatures during the growing period and storage, attributed to higher average accumulation [4,5]. Under drought conditions, expressions of key enzymes required for the biosynthesis of anthocyanins (DFR), flavonoids (FLS) and chlorogenic acid (HCT) were drastically reduced (Figure 4A). Initial flow into the phenylpropanoid pathway through PAL was also reduced, leading to the accumulation of phenylalanine observed in Figure 2. Similar results have been previously observed in the literature [5]. Key enzymes of the carotenoid pathway were also downregulated (Figure 4B). Environmental conditions leading to the repression of phenolic biosynthesis could minimize gains achieved in breeding programs. Regulation of the phenylpropanoid pathway can occur via MYB transcription factors [8], fifteen of which were significantly suppressed under drought (Supplementary Table S1). Unlike previous findings in potato leaves [10], expressions of small RNA and their target MYB transcription factors were not correlated in the drought-stressed tubers. There was therefore no evidence to suggest that small RNA played a role in regulating the phenylpropanoid or carotenoid pathways under drought conditions in potato tubers (Supplementary Table S4).

5. Conclusions

Potato is among one of the most important food crops, yet maintaining plant productivity in this drought-sensitive crop has become a challenge. From a nutrition perspective, decreasing soil water availability during tuber filling, as a function of a warming climate or as a production practice to induce senescence for an earlier harvest, may lead to a reduction in tuber quality. While mild drought increases the proportion of essential amino acids, potential losses in protein and phenolic content would outweigh the benefit. While MYB transcription factors may be targeted to reduce effects on the phenylpropanoid pathway, identification of small RNA as the regulator of patatin gene expression suggests it may be difficult to maintain patatin expression in drought-susceptible cultivars using current breeding techniques.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/8/864/s1; Figure S1: Experimental conditions; Table S1: DE genes; Table S2: Raw read counts; Table S3: DE small RNA; Table S4: small RNA, targets expression.

Author Contributions: R.S. designed the research; J.W., R.S., R.E., B.B. performed field study; S.K., L.D.R., R.S., M.K. analyzed transcriptome data; S.K. performed bioinformatics; S.P., R.S. performed the amino acid analysis; L.D.R., R.S., S.K. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: R.S., J.W., and B.B. received funding support from Agriculture and Agri-Food Canada.

Acknowledgments: The authors would like to thank Ken Achtymichuk and Greg Larson (AAFC Canada-Saskatchewan Irrigation Diversification Centre in Outlook, SK) for plot maintenance and Branimir Gjetvaj for assistance with physiological measurements.

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

References

1. King, J.C.; Slavin, J.L. White potatoes, human health, and dietary guidance. *Adv. Nutr.* **2013**, *4*, 393S–401S. [CrossRef]
2. Chun, O.K.; Kim, D.O.; Smith, N.; Schroeder, D.; Han, J.T.; Lee, C.Y. Daily consumption of phenolics and total antioxidant capacity from fruit and vegetables in the American diet. *J. Sci. Food Agric.* **2005**, *85*, 1715–1724. [CrossRef]
3. Andre, C.M.; Oufir, M.; Guignard, C.; Hoffmann, L.; Hausman, J.F.; Evers, D.; Larondelle, Y. Antioxidant profiling of native Andean potato tubers (*Solanum tuberosum* L.) reveals cultivars with high levels of β-carotene, α-tocopherol, chlorogenic acid, and petanin. *J. Agric. Food Chem.* **2007**, *55*, 10839–10849. [CrossRef]
4. Rosenthal, S.; Jansky, S. Effect of production site and storage on antioxidant levels in specialty potato (*Solanum tuberosum* L.) tubers. *J. Sci. Food Agric.* **2008**, *88*, 2087–2092. [CrossRef]
5. Payyavula, R.S.; Navarre, D.A.; Kuhl, J.C.; Pantoja, A.; Pillai, S.S. Differential effects of environment on potato phenylpropanoid and carotenoid expression. *BMC Plant Biol.* **2012**, *12*, 39. [CrossRef] [PubMed]
6. Ezekiel, R.; Singh, N.; Sharma, S.; Kaur, A. Beneficial phytochemicals in potato—A review. *Food Res.* **2013**, *50*, 487–496. [CrossRef]
7. Deng, Y.; Lu, S. Biosynthesis and regulation of phenylpropanoids in plants. *Crit. Rev. Plant Sci.* **2017**, *36*, 257–290. [CrossRef]
8. 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]
9. Chen, C.; Zeng, Z.; Liu, Z.; Xia, R. Small RNAs, emerging regulators critical for the development of horticultural traits. *Hort Res.* **2018**, *5*, 63–76. [CrossRef]
10. Zhang, N.; Yang, J.; Wang, Z.; Wen, Y.; Wang, J.; He, W.; Liu, B.; Si, H.; Wang, D. Identification of novel and conserved microRNAs related to drought stress in potato by deep sequencing. *PLoS ONE* **2014**, *9*, e95489. [CrossRef]
11. Pęksa, A.; Kita, A.; Kułakowska, K.; Aniołowska, M.; Hamouz, K.; Nemś, A. The quality of protein of coloured fleshed potatoes. *Food Chem.* **2013**, *141*, 2960–2966. [CrossRef]
12. Bártová, V.; Bárta, J.; Brabcová, A.; Zdráhal, Z.; Horáčková, V. Amino acid composition and nutritional value of four cultivated South American potato species. *J. Food Compos. Anal.* **2015**, *40*, 78–85. [CrossRef]
13. Galili, G.; Amir, R.; Fernie, A.R. The regulation of essential amino acid synthesis and accumulation in plants. *Annu. Rev. Plant Biol.* **2016**, *67*, 153–178. [CrossRef] [PubMed]
14. Rykaczewska, K. The effect of high temperature occurring in subsequent stages of plant development on potato yield and tuber physiological defects. *Am. Potato J.* **2015**, *92*, 339. [CrossRef]
15. Boguszewska-Mańkowska, D.; Pieczyński, M.; Wyrzykowska, A.; Kalaji, H.M.; Sieczko, L.; Szweykowska-Kulińska, Z.; Zagdańska, B. Divergent strategies displayed by potato (*Solanum tuberosum* L.) cultivars to cope with soil drought. *J. Agron. Crop Sci.* **2018**, *204*, 13–30. [CrossRef]
16. Wishart, J.; George, T.S.; Brown, L.K.; White, P.J.; Ramsay, G.; Jones, H.; Gregory, P.J. Field phenotyping of potato to assess root and shoot characteristics associated with drought tolerance. *Plant Soil* **2014**, *378*, 351–363. [CrossRef]
17. Pieczynski, M.; Wyrzykowska, A.; Milanowska, K.; Boguszewska-Mankowska, D.; Zagdanska, B.; Karlowski, W.; Jarmolowski, A.; Szweykowska-Kulinska, Z. Genomewide identification of genes involved in the potato response to drought indicates functional evolutionary conservation with *Arabidopsis* plants. *Plant Biotechnol. J.* **2018**, *16*, 603–614. [CrossRef]

18. Sprenger, H.; Erban, A.; Seddig, S.; Rudack, K.; Thalhammer, A.; Le, M.Q.; Walther, D.; Zuther, E.; Köhl, K.I.; Kopka, J.; et al. Metabolite and transcript markers for the prediction of potato drought tolerance. *Plant Biotechnol. J.* **2018**, *16*, 939–950. [CrossRef]
19. Inaba, K.; Fujiwara, T.; Hayashi, H.; Chino, M.; Komeda, Y.; Naito, S. Isolation of an *Arabidopsis thaliana* mutant, mto1 that overaccumulates soluble methionine (Temporal and spatial patterns of soluble methionine accumulation). *Plant Physiol.* **1994**, *104*, 881–887. [CrossRef]
20. Cohen, S.A.; Michaud, D.P. Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids via high-performance liquid chromatography. *Anal. Biochem.* **1993**, *211*, 279–287. [CrossRef]
21. Yan, D.; Easwaran, V.; Chau, V.; Okamoto, M.; Ierullo, M.; Kimura, M.; Endo, A.; Yano, R.; Pasha, A.; Gong, Y.; et al. NIN-like protein 8 is a master regulator of nitrate-promoted seed germination in *Arabidopsis*. *Nat. Commun.* **2016**, *7*, 13179. [CrossRef] [PubMed]
22. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinform.* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
23. 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] [PubMed]
24. Andrew, S. FastQC: A quality control tool for high throughput sequence data. 2015. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 5 September 2018).
25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* **2011**, *17*, 10–12. [CrossRef]
26. Johnson, N.R.; Yeoh, J.M.; Coruh, C.; Axtell, M.J. Improved placement of multi-mapping small RNAs. *G3 (Bethesda)* **2016**, *6*, 2103–2111. [CrossRef]
27. Dai, X.; Zhuang, Z.; Zhao, X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [CrossRef]
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
29. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, 15–28. [CrossRef]
30. Prashar, A.; Yildiz, J.; McNicol, J.W.; Bryan, G.J.; Jones, H.G. Infra-red thermography for high throughput field phenotyping in *Solanum tuberosum*. *PLoS ONE* **2013**, *8*, e65816. [CrossRef]
31. Ren, H.; Gray, W.M. SAUR proteins as effectors of hormonal and environmental signals in plant growth. *Mol. Plant* **2015**, *8*, 1153–1164. [CrossRef]
32. Sprenger, H.; Kurowsky, C.; Horn, R.; Erban, A.; Seddig, S.; Rudack, K.; Fischer, A.; Walther, D.; Zuther, E.; Köhl, K.; et al. The drought response of potato reference cultivars with contrasting tolerance. *Plant Cell Environ.* **2016**, *39*, 2370–2389. [CrossRef]
33. Obidiegwu, J.E.; Bryan, G.J.; Jones, H.G.; Prashar, A. Coping with drought: Stress and adaptive responses in potato and perspectives for improvement. *Front. Plant Sci.* **2015**, *6*, 1–23. [CrossRef] [PubMed]
34. Gong, L.; Zhang, H.; Gan, X.; Zhang, L.; Chen, Y.; Nie, F.; Shi, L.; Li, M.; Guo, Z.; Zhang, G.; et al. Transcriptome profiling of the potato (*Solanum tuberosum* L.) plant under drought stress and water-stimulus conditions. *PLoS ONE* **2015**. [CrossRef] [PubMed]
35. Vasquez-Robinet, C.; Mane, S.P.; Ulanov, A.V.; Watkinson, J.I.; Stromberg, V.K.; De Koeyer, D.; Schafleitner, R.; Willmot, D.B.; Bonierbale, M.; Bohnert, H.J.; et al. Physiological and molecular adaptations to drought in Andean potato genotypes. *J. Exp. Bot.* **2008**, *59*, 2109–2123. [CrossRef] [PubMed]
36. Dezfulian, M.; Foreman, C.; Jalili, E.; Pal, M.; Dhaliwal, R.; Roberto, D.; Imre, K.; Kohalmi, S.; Crosby, W. Acetolactate synthase regulatory subunits play divergent and overlapping roles in branched-chain amino acid synthesis and *Arabidopsis* development. *BMC Plant Biol.* **2017**, *17*, 71. [CrossRef]
37. Dhondt, S.; Geoffroy, P.; Stelmach, B.A.; Legrand, M.; Heitz, T. Soluble phospholipase A2 activity is induced before oxylipin accumulation in tobacco mosaic virus-infected tobacco leaves and is contributed by patatin-like enzymes. *Plant J.* **2000**, *23*, 431–440. [CrossRef]

38. Shewry, P.R. Tuber storage proteins. *Ann. Bot.* **2003**, *91*, 755–769. [CrossRef]
39. Reyes, L.F.; Miller, J.C.; Cisneros-Zevallos, L. Antioxidant capacity, anthocyanins and total phenolics in purple-and red-fleshed potato (*Solanum tuberosum* L.) genotypes. *Am. Potato. J.* **2005**, *82*, 271. [CrossRef]

genes

MDPI

Article

Comparative Transcriptomics and Co-Expression Networks Reveal Tissue- and Genotype-Specific Responses of *qDTYs* to Reproductive-Stage Drought Stress in Rice (*Oryza sativa* L.)

Jeshurun Asher Tarun [1,2], Ramil Mauleon [3], Juan David Arbelaez [1], Sheryl Catausan [2], Shalabh Dixit [2], Arvind Kumar [2], Patrick Brown [4], Ajay Kohli [2] and Tobias Kretzschmar [3,*]

1 Department of Crop Sciences, College of Agricultural, Consumer & Environmental Sciences, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA; tarun2@illinois.edu (J.A.T.); arbelaez@illinois.edu (J.D.A.)
2 International Rice Research Institute, Los Baños, Laguna 4031, Philippines; s.catausan@irri.org (S.C.); s.dixit@irri.org (S.D.); a.kumar@irri.org (A.K.); A.Kohli@irri.org (A.K.)
3 Southern Cross Plant Science, Southern Cross University, Military Rd, East Lismore NSW 2480, Australia; ramil.mauleon@scu.edu.au
4 Department of Plant Sciences, University of California Davis, One Shields Avenue, Davis, CA 95616, USA; pjbrown@ucdavis.edu
* Correspondence: tobias.kretzschmar@scu.edu.au; Tel.: +61-2-6626-3406

Received: 14 August 2020; Accepted: 22 September 2020; Published: 24 September 2020

Abstract: Rice (*Oryza sativa* L.) is more sensitive to drought stress than other cereals. To dissect molecular mechanisms underlying drought-tolerant yield in rice, we applied differential expression and co-expression network approaches to transcriptomes from flag-leaf and emerging panicle tissues of a drought-tolerant yield introgression line, DTY-IL, and the recurrent parent Swarna, under moderate reproductive-stage drought stress. Protein turnover and efficient reactive oxygen species scavenging were found to be the driving factors in both tissues. In the flag-leaf, the responses further included maintenance of photosynthesis and cell wall reorganization, while in the panicle biosynthesis of secondary metabolites was found to play additional roles. Hub genes of importance in differential drought responses included an expansin in the flag-leaf and two peroxidases in the panicle. Overlaying differential expression data with allelic variation in DTY-IL quantitative trait loci allowed for the prioritization of candidate genes. They included a differentially regulated auxin-responsive protein, with DTY-IL-specific amino acid changes in conserved domains, as well as a protein kinase with a DTY-IL-specific frameshift in the C-terminal region. The approach highlights how the integration of differential expression and allelic variation can aid in the discovery of mechanism and putative causal contribution underlying quantitative trait loci for drought-tolerant yield.

Keywords: co-expression network; drought-tolerant-yield; reproductive-stage drought; *qDTY*s; rice; transcriptomics

1. Introduction

Rice (*Oryza sativa* L.) is a staple crop feeding over half of the global population [1]. The Green Revolution accelerated the productivity of rice cultivation across Asia by focusing on irrigated, high input systems [2]. Intensification and expansion into new suboptimal cultivation areas, coupled with changing climatic conditions, however, necessitate a shift towards low input systems. Under resource- and water-limiting conditions, tolerance of both abiotic and biotic stress factors is crucial to ensure productivity.

Traditionally, rice was grown in areas naturally irrigated by seasonal floods [3]. Pre-domesticated rice is essentially a wetland species, making rice more sensitive to drought stress than most other staple crops [4]. Particularly, during the reproductive stage, drought typically causes yield reduction of 50% or more [5–10]. Consequently, water limitation is a major environmental constraint to rice production [11].

Successful strategies to identify factors contributing to drought tolerance involved mapping of quantitative trait loci (QTLs) for grain yield under drought conditions, so-called DTY (drought tolerant yield) QTLs [12]. By crossing the drought-tolerant donor N22 with Swarna, several major-effect DTY QTL, among them *qDTY1.1*, and *qDTY3.2* were identified as having consistent effects on grain yield under reproductive-stage drought stress (RDS) and no apparent yield or performance penalty under non-stress conditions [7]. These were subsequently introgressed into drought susceptible elite parents through backcrossing [13], resulting in the release of several drought-tolerant rice varieties. For example, "Bahuguni dhan-1", a sister line of DTY-IL used in this study, was recently released in Nepal [13]. In addition, several other large-effect DTY QTLs were identified from other populations and utilized for their potential to confer drought tolerance [14–17]. Gene discovery work in *qDTY12.1* resulted in the identification of a NAM transcription factor as an intra-QTL hub gene [18,19].

To unravel specific drought responses in rice, transcriptome studies across different cultivars and drought stress conditions identified hundreds of differentially expressed genes (DEGs) in an organ- and time-specific fashion [1,10,20–27]. Collectively these studies pinpointed key transcription factors (TFs) involved in ABA-dependent and ABA-independent pathways to be upregulated during water deficit stress, effecting osmolyte production, reactive oxygen species (ROS) scavenging and ion transportation [10,22,28].

After an initial focus on studying drought during vegetative stages, the importance of RDS was soon realized [8,23,29–33]. From an applied perspective, grain yield under drought is the key trait, making yield-associated developmental processes and responses under RDS a focus of drought research in rice [8]. Knowledge of broad level biological mechanisms governing RDS responses, however, is still limited [34]. Convincing concepts on how drought-stress-related genes are regulated are still in their infancy. DEG analysis, based on single-genes, often failed to result in meaningful biological interpretations [35,36], prompting the development of network-based techniques that consider complex relationships among genes [37,38].

Gene co-expression networks (GCNs) are increasingly employed to explore system-level functionality of genes and have been found useful for describing the pairwise relationships among genes [39]. GCNs provide a structured pathway for extracting modular responses from large datasets that are often missed by DEG or ANOVA approaches [40]. It shifts the focus from single candidate genes to groups of related genes that likely operate together within a tissue or in response to a stimulus [41]. Genes clustering in a module, provide insight into potential regulatory functions [40,42,43]. An essential application of GCN analysis is to identify functional gene modules, which are a group of nodes that have high topological overlap [44]. Weighted gene correlation network analysis (WGCNA) can be used for co-expression network analysis of gene expression data to find modules of highly correlated genes [45]. In rice GCN analysis provided some insights into gene regulation under drought stress, including (i) consensus modules of downregulated and upregulated genes [46]; (ii) a module enriched for genes involved in water homeostasis and embryonic development, including a heat shock TF [47] and (iii) new candidates involved in drought response [48].

While a number of major QTL for DTY have been discovered, knowledge regarding the underlying physiological mechanisms is largely lacking. On the other hand, while a number of transcriptome studies provided some general insights into drought responses of rice, they did not take presence of specific DTY QTL into account. In the 2014 and 2015 drought field trials at the International Rice Research Institute (IRRI), a DTY introgression line (DTY-IL) performed well under drought without showing a penalty under irrigated conditions. We decided to investigate this line further in a comparative transcriptomic approach against its drought susceptible recurrent parent Swarna.

Our rationale for this study was that combining a functional genomics with a classical genetics approach would improve resolution on drought tolerance mechanisms. By applying a gene co-expression network analysis and focusing on key source (flag-leaf) and sink (emerging panicle) tissues at reproductive stages, which had previously been demonstrated as critical for drought response [43], we speculated that adaptive mechanisms that drive yield under drought could be captured. By comparing the differential expression responses and overlaying genetic variation within the introgressed DTY QTL we further aimed to demonstrate that the differences at the genome-wide transcriptome level are modulated by the introgression segments. Our results support previous findings in respect to general mechanisms underlying drought response, and in addition suggest specific mechanisms underpinning DTY QTL.

2. Materials and Methods

2.1. Plant Materials and Experimental Treatments

Rice genotypes used in this study were: Swarna, a South Asian *indica* mega variety, susceptible to drought stress, and DTY-IL (designation IR 96321-1447-165-B-3-1-2), a highly drought-tolerant F_7 introgression line containing N22 fragments in a Swarna background. DTY-IL is sister line of IR 96321-1447-651-B1-1-2, which was recently released as a drought-tolerant variety in Nepal [13].

Field experiments were conducted in the 2014 and 2015 dry season (DS), and were laid out in an augmented RCB design. The 2014 trial had 420 entries and 6 checks with 5 blocks in a single row per plot while the 2015 trial, had 46 entries and 2 checks with 4 blocks in four rows per plot. For both trials, only checks were repeated based on the number of blocks. For the drought screening, water was removed from the field around 28–30 days after transplanting by opening all drainage canals around the field. PVC pipes measuring 1.1 to 1.2-m long were installed in different parts of the field for water table measurements. The PVC pipes were placed 1 m below the soil surface. The water table was measured regularly starting 1 day after draining the field.

A pot experiment was arranged in a randomized complete block design with two treatments (well-watered and 2-weeks drought-stressed), two genotypes, and six replications in a greenhouse at the International Rice Research Institute (Los Baños, Philippines) from July to November 2015. Three pre-germinated seeds of the two genotypes were initially seeded on white porcelain pots filled with 15 kg of clean puddled field soil (not sterilized). Upon seedling establishment, a healthy seedling was retained in each pot and grown under a well-watered condition in the greenhouse until the booting stage before initiating a dry-down experiment. A day before imposing stress, all the pots were saturated with water and allowed to drain excess water for 24 h to maintain the field capacity (FC) so that the soil moisture amount in each pot was uniform. Then each pot was weighed to know the amount of water at FC. The temperature inside the greenhouse during the stress induction was at a maximum of 30–34 °C and a minimum of 23–26 °C, and a day-time relative humidity of 69%–95% (Figure S1A). During the drought stress period, the pots were weighed daily, and the difference in weight on subsequent days was corrected by adding water to maintain the required FC [49]. For RDS, water was withheld at the reproductive R2 stage, on discrete morphological criteria as described by [50], until the soil moisture level dropped to 75% FC and was maintained for nine days, whereas control plants were well-watered. At day 10, FC was reduced to 50% for three more days (Figure S1B). Flag-leaf and whole panicle samples of well-watered and drought-stressed treatments were collected at the R3 stage [50] on the 12th day of RDS and immediately flash-frozen in liquid nitrogen. Four independent biological replicates for each tissue and each genotype sample were harvested.

2.2. RNA Extraction and Sequencing

Total RNA was extracted using the Qiagen RNeasy Plant Mini Kit (Qiagen, Limburg, The Netherlands). RNA concentration was quantified using Nanodrop spectrophotometer (ND-1000; Nanodrop Technologies, Wilmington, DE, USA), while purity and integrity were established using an

Agilent 2100 BioAnalyzer RNA 6000 Kit (Agilent Technologies, Santa Clara, CA, USA), with a RIN value of 8 used as quality threshold. Illumina library preparation and sequencing were completed following the standard protocols of Macrogen Inc. (Seoul, South Korea). Using Illumina HiSeq 2000 and HiSeq 4000 (Illumina, Inc., San Diego, CA, USA) platforms for the whole panicle and flag-leaf tissues, respectively, 101 bp aired-end sequencing was done. A quality check of raw RNA-Seq reads was performed using FastQC software (version 0.11.5) [51]. The sequencing data have been deposited in NCBI's Gene Expression Omnibus (GEO) database under the accession number GSE145870.

2.3. Transcriptome Assembly and Expression Level Quantification

Raw fastq reads were filtered using Trimmomatic software, version 0.36 [52] using default settings. An indexed transcriptome fasta file was built from the rice japonica genome (IRGSP1.0) and annotation gff3 file from Rice Genome Annotation release 7 (RGAP 7, http://rice.plantbiology.msu.edu/), using the "gffread" function of Cufflinks (version 2.2.1) [53], and Salmon (version 0.7.2) [54] "index" function. Salmon then quantified transcript abundance in quasi-mapping mode directly using the indexed transcriptome and the trimmed high-quality paired-end reads with parameters "-l A, -seqBias and -gcBias" to allow automatic inference of library type, learn and to correct for sequence-specific and fragment-level GC content biases, respectively. Gene expression levels were normalized using the transcripts per million mapped reads (TPM) method, exported as estimated transcript abundance, and aggregated to gene-level expression using the Bioconductor package tximport (version 1.2.0) [55] complemented with the reader package (version 1.1.1), within the R environment (version 3.3.3). Principal component analysis (PCA) was performed in R using ggplot2 (version 2.2.1) and gplots package (version 3.0.1) to determine relationships between samples.

2.4. Differential Expression Analysis

DESeq2 software (version 1.14.1) in R was used to identify DEGs in pairwise comparisons [56]. Two series of DE analysis using the following contrast arguments was performed with the reference assembly approach in flag-leaf and panicle samples: contrast 1—the condition effect for each genotype, in other words, IL_DvsIL_C, and SWA_DvsSWA_C, and contrast 2—the genotypic effect for each condition, in other words, IL_CvsSWA_C and IL_DvsSWA_D. Only genes that have at least ten reads in total were used for DE analysis. Differentially expressed genes (DEGs) were defined as those presenting an absolute fold change (FC) ≥ 2 or ≤ 0.5 and a False Discovery Rate (FDR) adjusted p-value ≤ 0.05 in any pairwise comparison. DEGs were then subjected to enrichment analysis of gene ontology (GO) terms, KEGG, and other metabolic pathways defined by a hypergeometric and Fisher's exact test using agriGO (version 2.0), KOBAS (version 3.0), and STRING database (version 10.5). The MapMan tool (http://MapMan.gabipd.org) was used to visualize the involvement of the DEGs in pathways of interest.

2.5. Gene Co-Expression Network Analysis

Gene co-expression network analysis to group genes into modules used the R package WCGNA (v1.6.1). A power value of 6 and 9 for flag-leaf and panicle, respectively, predicted a gene co-expression network that exhibited scale-free topology with inherent modular features (Figure S10A,D). The "blockwiseModules" function of WGCNA was used to detect and generate modules. Network interconnectedness was measured by calculating the topological overlap using the TOMdist function with a signed TOMType. Average hierarchical clustering using the "hclust" function was performed to group the genes based on the topological overlap dissimilarity measure (1-TOM) of their connection strengths. Network modules were identified using the dynamic tree cut algorithm (version 1.63.1) with minimum and maximum module size of 20 and 20,000, respectively, merging threshold function at 0.15, and deep split parameter set at level 2. The module "eigengenes" was used to calculate the correlation coefficient for the samples to identify biologically significant modules. To visualize the expression profiles of the modules, the eigengene (first PC) for each module

was plotted using a customized bar plot function in R. To identify hub genes within the module, the module membership (MM) for each gene also known as module eigengene-based connectivity (kME) was calculated, based on the Pearson correlation between actual expression values and the module eigengene. Incorporating the gene significance (GS) measures, which could also be defined by the −log10 (*p*-value) from IL_DvsSwa_D contrast in DE analysis as external information into the co-expression network, genes within a module with the highest MM and GS values are highly connected within that module. To identify modules shared between the flag-leaf and panicle networks, a consensus network was generated. An in-house R function was used for overlap counting and statistical testing. The consensus network matrix for the flag-leaf and panicle tissue networks was plotted to show the significant overlap in gene count of two modules based on Fisher's exact test with the −log10 (*p*-value).

2.6. Over-Representation Analysis (ORA)

The "kegga" function under the "goana" package in R utilized the user-supplied GO slim assignment and InterPro classification from RGAP 7 independently in the form of data.frame annotation alongside the multiple gene lists generated in each module. Gene.pathway was the data.frame linking genes to pathways and pathway.names was the data.frame giving full names of pathways. The universe was the vector specifying the set of unique gene identifiers used in WGCNA to be the background and not the whole genome. Finally, GO slim terms, and InterPro protein families and domains were called significantly over-represented in the gene set if the *p*-value is < 0.05.

2.7. Quantitative PCR Analysis

Primers were designed using QuantPrime (https://quantprime.mpimp-golm.mpg.de). cDNA was synthesized from 2 µg total RNA using the ImProm-II Reverse Transcription System (Promega, Madison, WI, USA), according to the manufacturer's protocol. qRT-PCR was performed using two independent biological replicates and three technical replicates. qRT-PCR was set up in 386-well PCR plates with 0.2 µM primers using SYBR Green PCR Master Mix kit (Applied Biosystems, Foster City, CA, USA), following the manufacturer's protocol in a reaction volume of 10 µL via a Roche LightCycler 480 Real-Time system (Rotkreuz, Switzerland). Reaction conditions were as follows: denaturation at 95 °C for 5 min, 45 cycles of 95 °C for 10 s, 60 °C for 15 s and 72 °C for 8 s, heating from 65 to 95 °C. Two internal reference genes ELF and ATU were designed to normalize the relative gene expression levels for flag-leaf and panicle tissue, respectively, using the $2^{-\Delta\Delta}$CT method with ΔCT = CTgene − CTreference gene [57]. For comparison of fold change, scatterplots were generated using the $\log_2$ fold change determined between RNA-Seq and qRT-PCR, which is defined as ΔΔCT (for comparative threshold cycle).

2.8. Genotyping

Using the C7AIR, 7098 SNPs were called for N22, Swarna and the DTY-IL [58]. Frequencies for each SNP across replicated samples were estimated and the most frequent genotype was considered true. Missing markers and monomorphic markers between the parents N22 and Swarna were discarded. For the remaining markers (1648 SNPs) genotypic calls from each parent were used to assign inheritance from N22 or Swarna. Fragments from the donor parent N22 were defined as consecutive SNPs with N22 homozygous genotypes. Markers that potentially represent miss-called double recombination were discarded if the probability of observing this event was smaller than 1 cM or 1 in 100 events. A graphical representation of the markers inherited from N22 and Swarna were graphed using the R package ggplot2. RNA-Seq reads of the candidate genes were aligned and visualized in IGV (version 2.8.2) relative to MSU7 (Nipponbare) and in the latest versions of the MH63 and N22. Clustal Omega (version 1.2.4) was used for multiple protein sequence alignments. FGENESH was used as the gene prediction tool with *O. sativa* vg. *indica* as the background organism. SNP-Seek (https://snp-seek.irri.org) was used to validate the NS SNPs across the 3 K genomes.

2.9. DNA Extraction and Sanger Sequencing

Genomic DNA was extracted from the leaves of N22, the DTY-IL *qDTY* donor, using the Qiagen DNeasy Kit (Qiagen, Limburg, The Netherlands) according to the manufacturer's instructions. LOC_Os01g67030 (auxin responsive protein) including a 1.8 kb upstream promoter region was amplified using the forward primer GAGCGTGCAGTCCACTAGGCATTATC and reverse primer GTGACACGTATTCTGATGTACTG. The amplicon was cloned into the pGEM-T Easy Vector (Promega, WI, USA), as per manufacturer's instructions and Sanger sequenced using Macrogen, Inc. South Korea as service provider.

3. Results

3.1. Phenotypes of Swarna and DTY-IL under Reproductive Stage Drought Stress

In the 2014 and 2015 dry season (DS) drought-stress field trials at IRRI, DTY-IL and Swarna, showed similar grain yield (Figure 1A) and plant height (Figure 1B) under irrigated conditions. Under reproductive-stage drought stress height was reduced in both genotypes to a similar degree (Figure 1B). While yield was reduced in both lines under drought, DTY-IL achieved about double the average grain yield compared to Swarna (Figure 1A). In addition, three weeks after draining the field, Swarna showed clear signs of leaf rolling. Leaf rolling was also observed for Swarna in the greenhouse experiment. While there was no visible difference in flag-leaf morphology between both genotypes under the well-watered condition (Figure 1C), a prominent leaf rolling phenotype in Swarna was observed after ten days of drought stress with complete leaf-rolling on the 12th day of drought stress (Figure 1D).

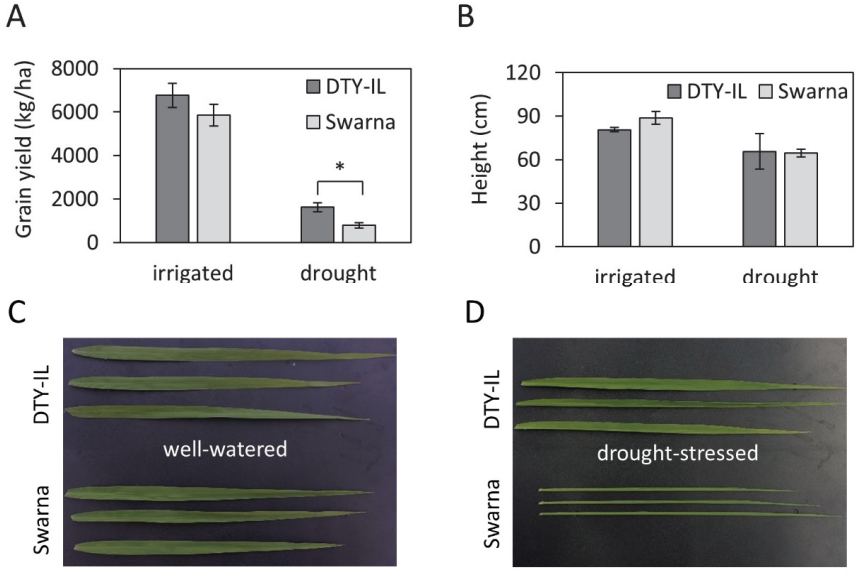


Figure 1. The effect of reproductive stage drought on yield, height and flag-leaf morphology. The average grain yield (**A**) and plant height (**B**) across the 2014 and 2015 dry season (DS) field trial of DTY-Il and Swarna under irrigated and drought condition ($N = 2–9$). Flag-leaf phenotypes of the DTY-IL and Swarna under well-watered conditions (**C**) and the leaf-rolling phenotype during drought-stress (**D**) under controlled greenhouse conditions. The asterisks (*) indicates a significant difference (Student's *t* test, * $p < 0.01$).

3.2. Generating a Transcriptional Map of the Moderate RDS Response in Rice

Mapping rates ranged from 77.7–92.9% for flag-leaf and 87.9–92.8% for panicle (Table S1), covering 28,283 and 33,698 genes, respectively. Sample clustering and heatmap visualization of $\log_2$-transformed, normalized count data demonstrated clear separation between genotypes and treatments for both flag-leaf and panicle samples (Figures S2A,B and S3A,B). Principal components (PC) analysis showed that the first and second PC explained 93% of the total variation for flag-leaf (Figure S2C) and 81% for panicle tissue (Figure S3C). Biological replicates of each genotype-treatment combination clustered together and the treatment effect was greater than the genotype effect for both tissues (Figures S2C and S3C). An exception was a single Swarna panicle sample, which was removed from all further analyses as an outlier (Figure S4). The sum of the non-redundant/unique $\log_2$ normalized genes with significant changes of expression at FDR adjusted p-value ≤ 0.05 across four different contrasts were 17,616 and 18,614 for flag-leaf and panicle tissue, respectively. Pairwise DE analysis for all genotype-treatment combinations identified DEGs of significance for flag-leaf and panicle (Figure S5), and a Venn diagram was used to visualize three categories of unique and common responses in flag-leaf (Figure 2A and Figure S6A,B, Table 1) and panicle (Figure 2B and Figure S6C,D, Table 1).

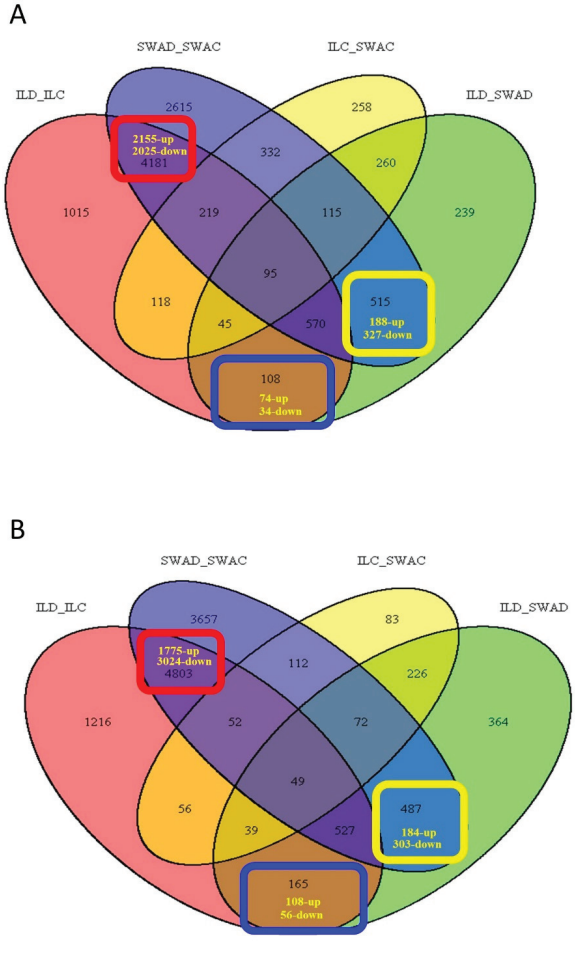


Figure 2. Gene differential expression and identification. Venn diagram of differentially expressed

genes (DEGs) for flag-leaf (**A**) and panicle (**B**) tissues in Swarna and DTY-IL under reproductive drought stress (RDS) at a false discovery rate (FDR) adjusted p-value < 0.05 and $-1 \leq \log_2$-ratio $\leq +1$ (fold change ≥ 2 and ≤ 0.5). The three highlighted boxes for each tissue represent the common DEGs (red), unique to Swarna (yellow), and unique to DTY-IL (blue). SWAC = Swarna control, SWAD = Swarna under RDS, ILC = DTY-IL control, ILD = DTY-IL under RDS.

Table 1. A summary of GO enrichment analysis of differentially expressed genes (DEGs) common to both genotypes, and unique to DTY-IL and Swarna for flag-leaf and panicle tissues under RDS.

DEGs of Interest	Expression	Genes		GO Terms		Significant GO Terms	
		Flag-Leaf	Panicle	Flag-Leaf	Panicle	Flag-Leaf	Panicle
Common responses of DTY-IL and Swarna DEGs to drought	Upregulated	2155	1775	409	324	88	110
	Downregulated	2025	3024	482	525	171	135
	Sub-total	4180	4799	891	849	259	245
Unique responses of Swarna DEGs to drought	Upregulated	188	184	92	89	0	0
	Downregulated	327	303	134	141	35	13
	Sub-total	515	487	226	230	35	13
Unique responses of DTY-IL DEGs to drought	Upregulated	74	108	31	58	4	36
	Downregulated	34	56	19	22	0	0
	Sub-total	108	164	50	80	4	36

3.2.1. Expression Profiles of Drought-Responsive Genes in the Flag-Leaf Tissue under RDS

A total of 4180 genes were found to be drought-responsive in flag-leaves of both Swarna and DTY-IL (Figure 2A, Table 1). Gene ontology (GO) enrichment analysis showed functional enrichment among upregulated DEG for transcription, regulation of biological processes, and oxidation-reduction (Table S2-1), while the shared downregulated DEGs were mainly associated with transmembrane transport, localization, and post-translational protein modification (Table S2-2).

In Swarna flag-leaves a total of 515 (188 up- and 327 downregulated) genes were uniquely drought-responsive (Table 1). While no significant GO terms were detected for upregulated DEGs (Table 1), significant GO terms for uniquely downregulated DEGs were largely related to post-translational protein modification, photosynthesis, defense response, and programmed cell death (Figure S7A). Pathway enrichment suggested photosynthesis, ubiquinone, and other terpenoid-quinone biosynthesis, as well as glutathione-mediated detoxification II and tyrosine biosynthesis (Table S3), to be significantly downregulated. MapMan visualization supported the downregulation of cell wall, carbon metabolism, secondary metabolism, and light reaction in Swarna flag-leaves (Figure 3A and Figure S9A).

In DTY-IL flag-leaves 108 (74 up- and 34 downregulated) DEGs were uniquely drought-responsive. While no significant GO terms could be associated with downregulated DEGs, oxidation-reduction, response to stress, and response to stimulus were among the significant GO terms in upregulated DEGs (Figure S7B). Pathway enrichment suggested phenylpropanoid biosynthesis, dhurrin, xylan, and scopoletin biosynthesis, as well as detoxification of reactive carbonyls in chloroplasts to be uniquely upregulated under RDS (Table S3). This was supported by MapMan visualization, showing upregulation of cell wall, lipids, and secondary metabolism (Figure 3B and Figure S9B).

Numerous overrepresented *cis*-elements were found in the group of 327 promoters of uniquely downregulated DEGs in Swarna in flag-leaf, which are mostly involved as binding sites for dehydration responsive genes. These include ACGTATERD1, IBOX, PREATPRODH, MYCATERD1, MYCATRD22, CCAATBOX1, and MYB2AT (Table S4-1). The overrepresented motifs in a set of 74 promoters in the uniquely upregulated DEGs in IL in flag-leaf were mostly functioning upon induction of dehydration stress through the ACGTATERD1 motif (Table S4-2).

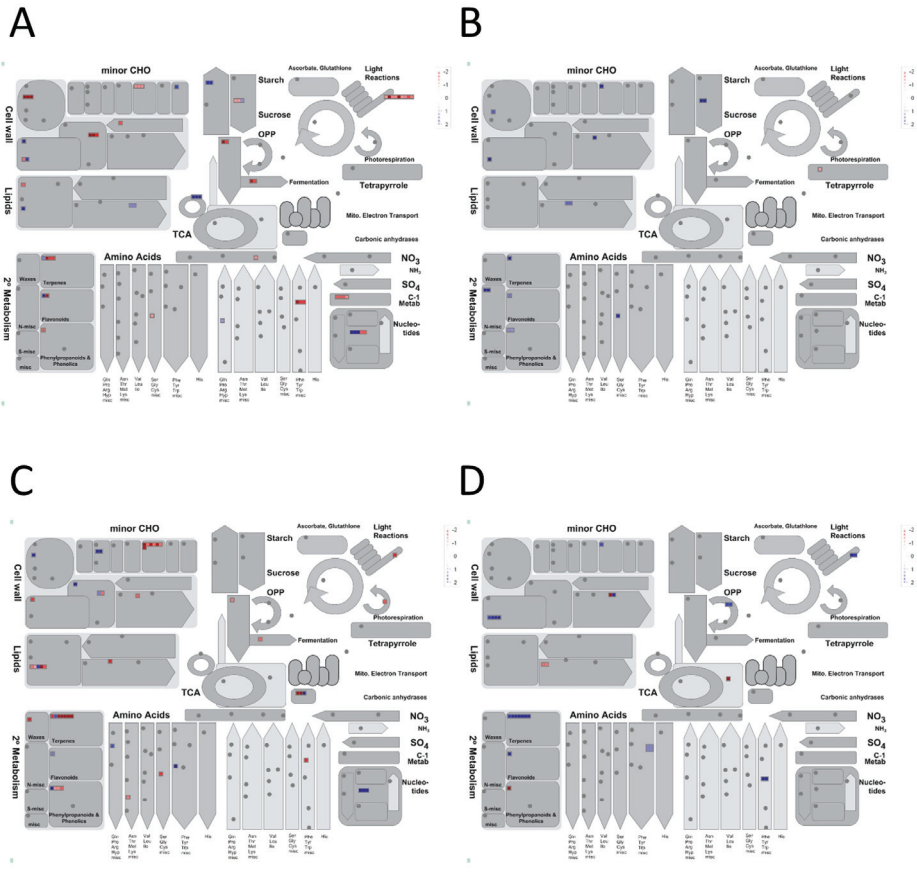


Figure 3. Mapman overview of the DEGs of interest in DTY-IL and Swarna under RDS. The metabolism overview in in the flag-leaf tissue for Swarna (**A**) and DTY-IL (**B**), as well as for the panicle tissue in Swarna (**C**) and DTY-IL (**D**). The DEGs were binned to the MapMan functional categories. The values are the $\log_2$ fold changes. Upregulated categories are represented in blue and downregulated categories in red.

3.2.2. Expression Profiles of Drought-Responsive Genes in the Panicle Tissue under RDS

A total of 4799 genes were found to be drought-responsive in panicles of both Swarna and DTY-IL (Figure 2B; Table 1). Enriched GO categories of the upregulated DEGs related to post-translational protein modification and response to stress (Table S5-1), while enriched GO terms of the downregulated DEGs were mostly related to transmembrane transport, carbohydrate metabolic process, and localization (Table S5-2).

In Swarna panicles, a total of 487 (184 up and 303 downregulated) genes were found uniquely drought-responsive (Table 1). No significant GO terms were identified within the uniquely upregulated DEGs. For the uniquely downregulated genes, significant GO terms included oxidation-reduction as well as monooxygenase activity, tetrapyrrole, and heme-binding (Figure S8A). Significantly enriched pathways associated with downregulated DEGs were related to DNA replication, diterpenoid biosynthesis, phenylpropanoid biosynthesis, and photosynthesis (Table S6). MapMan visualization supported the downregulation of cell wall, lipids, secondary metabolism, amino acids, as well as carbohydrate metabolism (Figure 3C and Figure S9C).

In DTY-IL panicle 164 (108 up and 56 downregulated) DEGs were uniquely drought-responsive (Table 1). No significant GO enrichment was identified among the uniquely downregulated DEGs of DTY-IL. Prevalent GO terms of upregulated DEGs were related to protein amino acid phosphorylation as well as oxidation-reduction and carbohydrate metabolic processes (Figure S8B). The most significantly enriched pathways in the panicle tissue of DTY-IL upregulated DEGs were related to propanoate metabolism, methylerythritol phosphate pathway, diterpenoid biosynthesis, camalexin biosynthesis, and circadian rhythm in plants (Table S6). This was supported by MapMan visualization, showing upregulation of cell wall, lipids, secondary metabolism as well as amino acids (Figure 3D and Figure S9D).

Overrepresented *cis*-acting elements in the group of 303 promoters of the uniquely downregulated DEGs in Swarna in panicle tissue mostly involved in dehydration response like the ACGTATERD1, MYBCOREATCYB1, ABRELATEDRD1, and IBOX motifs (Table S7-1). Overrepresented *cis*-elements in a set of 108 promoters of uniquely upregulated DEGs in DTY-IL in panicle tissue were mostly related to dehydration response like the MYB2AT, MYBCOREATCYCB1, and ACGTABOX (Table S7-2).

3.3. Drought-Stress-Related Gene Modules within the Transcriptional Map

WGCNA identified 21 distinct co-expressed modules with different expression patterns in flag-leaf (designated as FL-M1 to FL-M21, capturing 17,616 genes) (Figure S10B,C), and 23 distinct modules for panicle network (designated as P-M1 to P-M23, capturing 18,614 genes) (Figure S10E,F).

More than 70% of genes were distributed in the FL-M1 and FL-M2, and P-M1 and P-M2 (common response between DTY-IL and Swarna under RDS and control) for flag-leaf and panicle networks, respectively (Table S8-1,S8-2; Figure S11), signifying a common response shared between Swarna and DTY-IL. Modules FL-M1 and FL-M2 in the flag-leaf network showed "localization" and "transport" as the most enriched GO terms (Table S9-1). In the panicle network, P-M1 genes were enriched for functions related to "RNA processing" while the P-M2 module was linked with "localization" and "transport" (Table S9-2).

3.3.1. Flag-Leaf Specific Modules

Two specific flag-leaf modules with a contrasting pattern of expression were investigated in more detail for implications in the differential performance of the two genotypes under RDS. FL-M14 consisting of 140 genes (Table S10-1) had significantly higher expression profiles in all samples of the DTY-IL genotype under RDS, whereas a lower expression across the three other groups of samples was observed (Figure 4A). FL-M16 consisting of 102 genes (Table S10-2) had significantly lower expression in Swarna under RDS, whereas the three other groups of samples had a higher expression (Figure 4B).

FL-M14 enriched BP GO terms were "cellular amino acid biosynthetic process" as well as "cell wall organization or biogenesis" (Figure S12A). The enriched pathway included "biosynthesis of secondary metabolites", "fatty acid elongation", and "phagosome" while the overrepresented Interpro domains included "Expansin" as well as "Glycosyl transferase, family 43", and "Plant peroxidase" (Figure 5A). FL-M14 hub genes included amidase and expansin (Figure S13A; Table S11-1). Interestingly, numerous cell wall organization and biogenesis genes are upregulated in DTY-IL and downregulated in Swarna under RDS (Figure 6).

FL-M16 was enriched in "oxidation-reduction process" as well as "cellular carbohydrate metabolic process" (Figure S12B). The enriched pathways include "photosynthesis", "folate biosynthesis", and "vitamin B6 metabolism" while the overrepresented Interpro domains in FL-M16 included "Ferrodoxin—NAP reductase", "Multicopper oxidase", and "Photosystem antenna protein-like" (Figure 5B). FL-M16 hub genes included OsSub37—putative subtilisin homologue and cytochrome P450 (Figure S13B; Table S11-2).

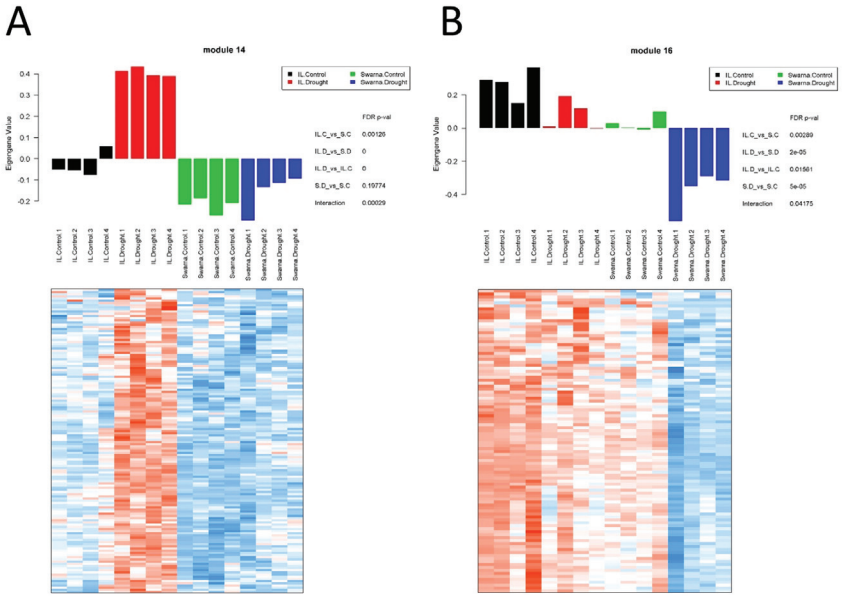


Figure 4. Gene co-expression network analysis in flag-leaf under RDS. Bar plots of the module eigengene as representatives of gene expression profiles across samples in FL-M14 (module 14) (**A**) and FL-M16 (module 16) (**B**). *X*-axis represents 16 different samples across four different groups, while *Y*-axis corresponds to the eigengene value. Heatmaps showing gene expression levels of genes in FL-M14 and FL-M16. Columns represent samples, while rows correspond to genes in the module. Red indicates positive and blue negative expression profile. S.C = Swarna control, S.D = Swarna under RDS, IL.C = DTY-IL control, IL.D = DTY-IL under RDS.

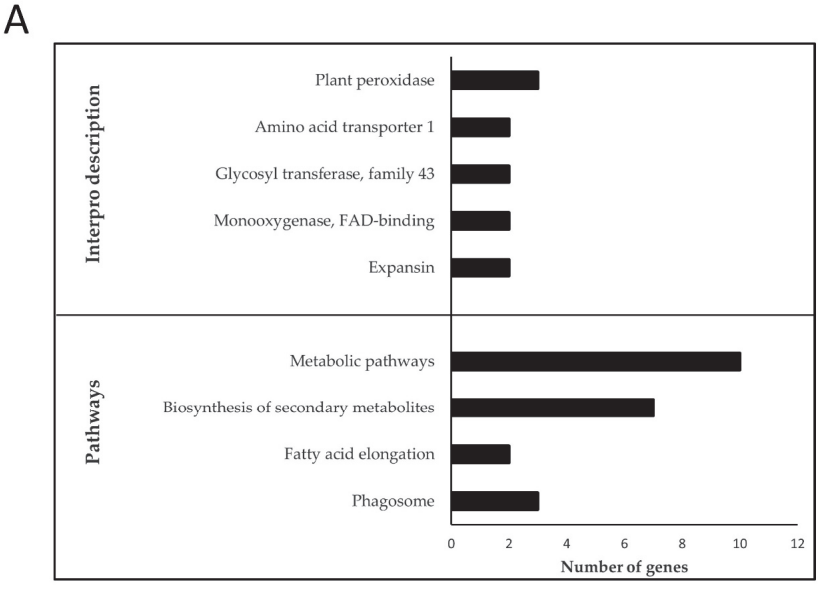


Figure 5. *Cont.*

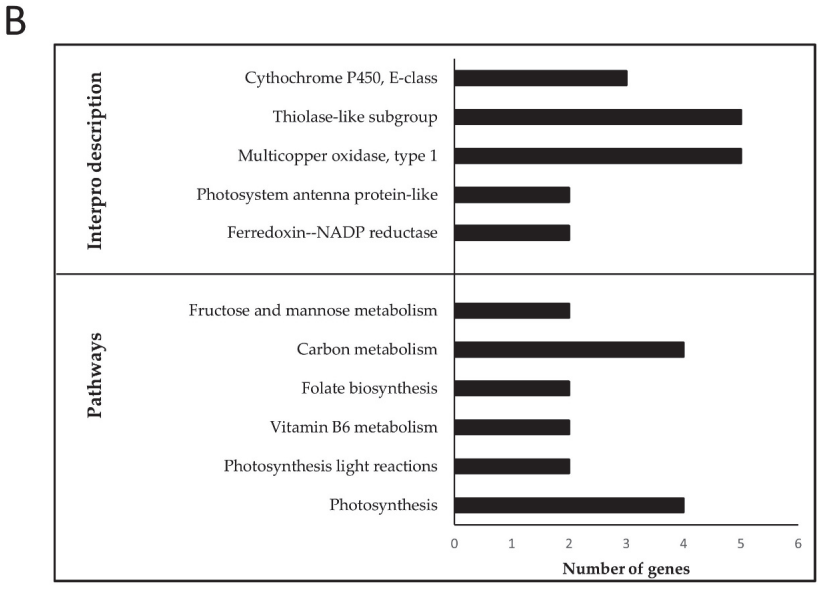


Figure 5. Enrichment analysis of the functional categories in FL-M14 and FL-M16. Over-represented Interpro domains and enriched pathways in FL-M14 (**A**) and FL-M16 (**B**) in flag-leaf under RDS. Top significant pathways and Interpro domains are shown in *Y*-axis with the number of represented genes on the *X*-axis.

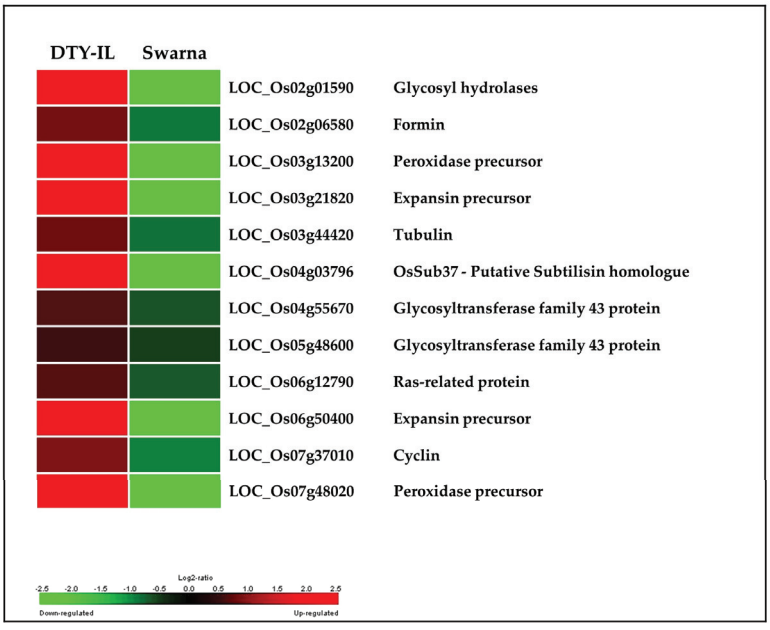


Figure 6. Cell wall organization and biogenesis during RDS in the flag-leaf tissue. Expression of cell wall organization or biogenesis related genes in drought-responsive modules between DTY-IL and Swarna are shown as $\log_2$ fold change heat-maps.

3.3.2. Panicle Specific Modules

Two contrasting panicle modules (P-M10 and P-M15) were investigated in more detail. P-M10 consisting of 138 genes (Table S12-1) and P-M15 consisting of 73 genes (Table S12-2) had significant interaction with drought response in a genotype-specific manner. P-M10 had significantly higher expression profiles in all samples of DTY-IL under RDS and a lower expression pattern across the other three groups of samples (Figure 7A). P-M15 had significantly lower expression profiles in all samples in Swarna under RDS, whereas the three other groups of samples had a higher expression (Figure 7B).

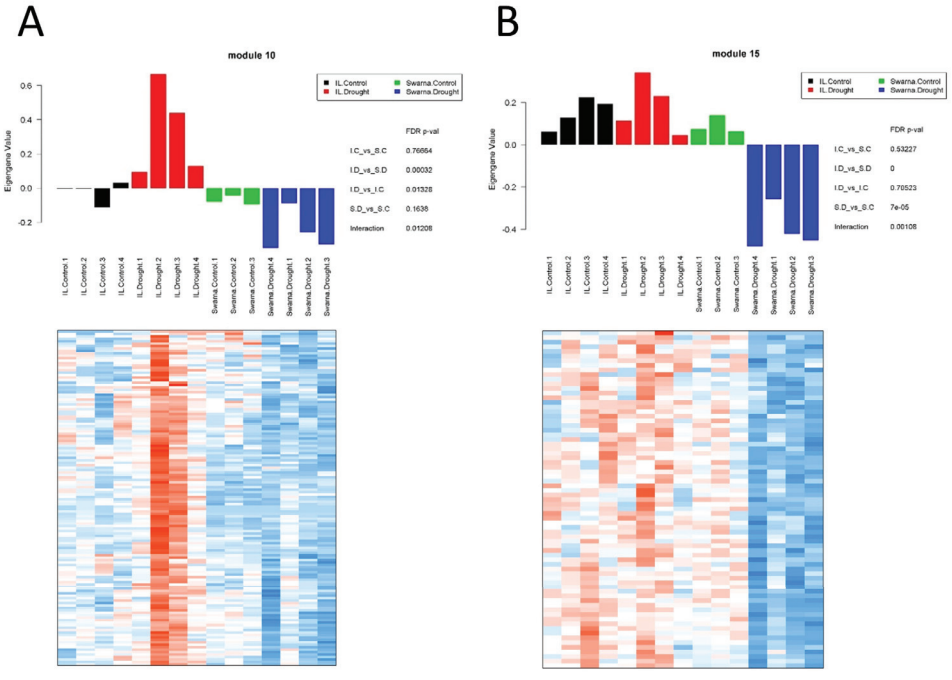


Figure 7. Gene co-expression network analysis in panicle under RDS. Bar plot of the module eigengene as representatives of the gene expression profiles across different samples in P-M10 (module 10) (**A**) and P-M15 (module 15) (**B**). *X*-axis represents 15 different samples across four different groups, while *Y*-axis corresponds to the eigengene value. Heatmaps showing gene expression levels of genes in P-M10 and P-M15. Columns represent samples, while rows correspond to genes in the module. Red indicates positive and blue negative expression profile. S.C = Swarna control, S.D = Swarna under RDS, IL.C = DTY-IL control, IL.D = DTY-IL under RDS.

P-M10 enriched BP GO terms were "calcium ion transmembrane" as well as "lipid biosynthetic process" (Figure S14A). Pathway enrichment analysis included "glutathione metabolism", "phenylpropanoid biosynthesis", and "ribosome" while the overrepresented Interpro domains were "Protein kinase" as well as "Plant peroxidase" (Figure 8A). P-M10 hub genes included OsWAK receptor-like cytoplasmic kinase and a serine-type peptidase (Figure S15A; Table S13-1). Analysis of genome-scale metabolic pathways in the DTY-IL revealed the up-regulation of genes involved in the biosynthesis of antioxidant enzymes and metabolites (Figure 9 and Figure S14).

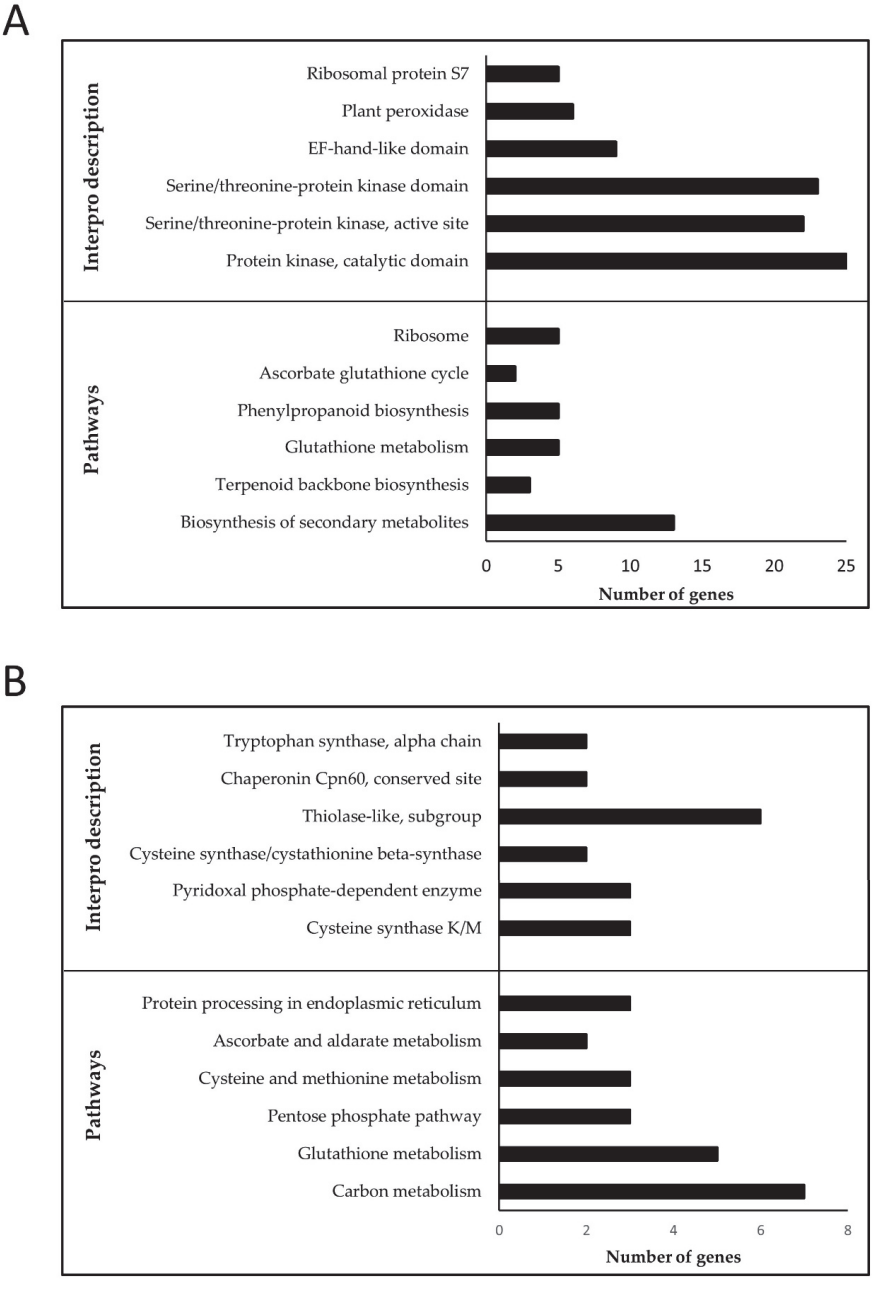


Figure 8. Enrichment analysis of the functional categories in P-M10 and P-M15. Over-represented Interpro domains and enriched pathways in P-M10 (**A**) and P-M15 (**B**) in panicle under RDS. Top significant pathways and Interpro domains are shown in *Y*-axis with the number of represented genes on the *X*-axis.

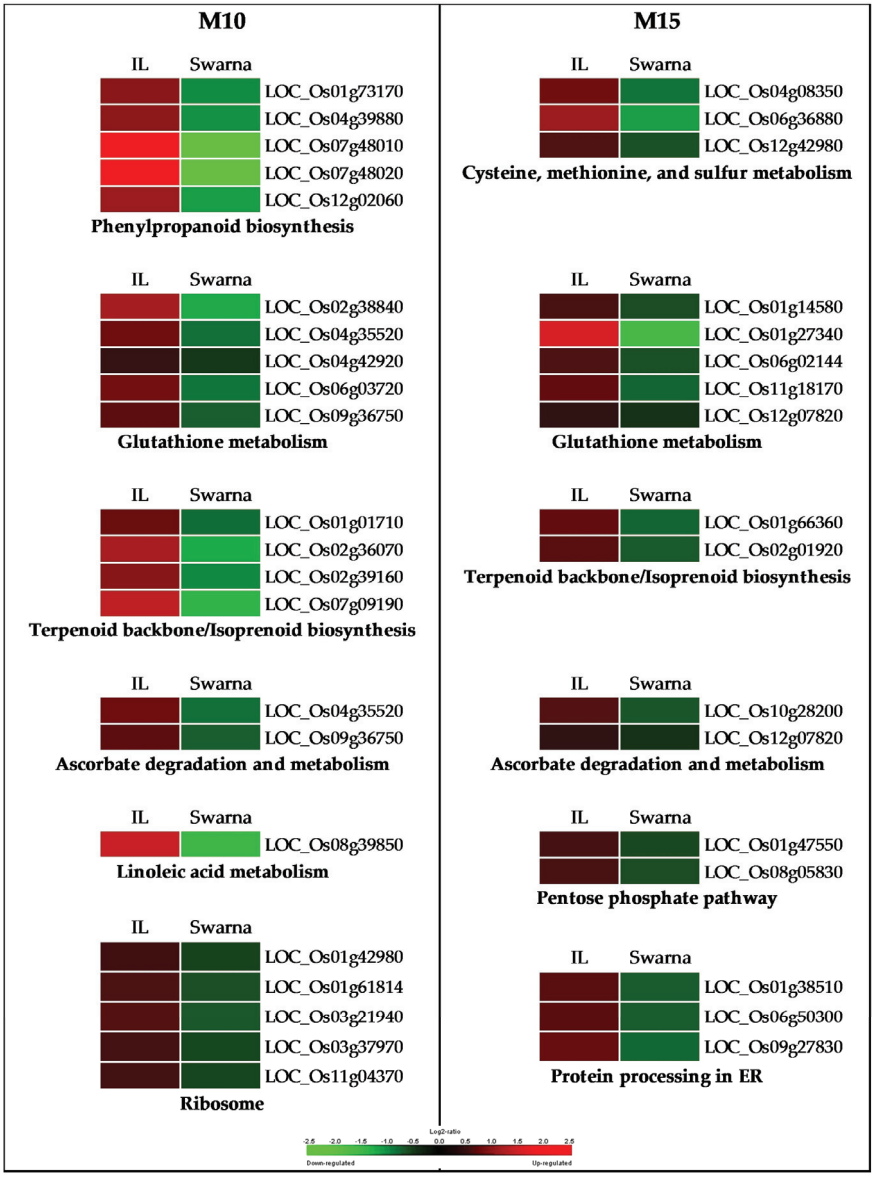


Figure 9. Regulation of metabolic pathways during RDS in panicle. The metabolic pathways enriched in the drought-responsive modules P-M10 (M10) and P-M15 (M15) between DTY-IL and Swarna under RDS are shown in heat-maps representing their expression profile. The scale represents a $\log_2$ fold change in expression. IL = DTY-IL.

P-M15 were GO enriched for "carboxylic acid metabolic process" as well as "coenzyme metabolic process" (Figure S14B). The enriched pathways includes "carbon metabolism", "glutathione metabolism", and "pentose phospate pathway" while the overrepresented Interpro domains included "Cysteine synthase" and "Thiolase-like, subgroup" (Figure 8B). P-M15 hub genes included the glycosyl hydrolase family 29, and dehydrogenase/reductase (Figure S15B; Table S13-2).

The consensus network from the flag-leaf and panicle transcriptomes contained few significant overlaps in module classifications between the flag-leaf and panicle networks, consistent with the tissue-specific expression under RDS [23] (Figure S16). Similarly, colored modules between the flag-leaf and panicle networks contained a few significant overlaps of genes with a common consensus network module, consistent with their similar eigengenes profiles (Figures S10 and S11). The significant overlap portrays a common expression pattern for each condition of both genotypes (FL-M1 and P-M1, and FL-M2 and P-M2) and a common expression pattern for each genotype on both conditions (FL-M6 and FL-M7, and P-M5 and P-M6) in the flag-leaf and panicle networks, respectively.

3.4. Validation of Differential Gene Expression

RNA-Seq results were validated using ten genes from different response categories (increased, decreased, and non-differentially expressed genes upon treatment in both flag-leaf and panicle) for qRT-PCR (Figure S17A,B). Three selected genes from within the *qDTY1.1* region (LOC_Os01g66120, LOC_Os01g66820, and LOC_Os01g67030) showed differential expression. LOC_Os01g66120 (no apical meristem protein) was upregulated in both genotypes and both tissues under drought. LOC_Os01g66820 (inactive receptor kinase At1g27190 precursor) was downregulated in DTY-IL flag-leaves under RDS. LOC_Os01g67030 (auxin responsive protein) was upregulated under RDS in DTY-IL panicles and downregulated in Swarna panicles but not affected by RDS in flag-leaves. A high correlation between qRT-PCR results and RNA-Seq results was observed for flag-leaf (R^2 = 0.88) and panicle (R^2 = 0.91) tissues (Figure S17C,D), supporting RNA-Seq-based findings and interpretations. Targeted transcripts and respective primer sets used are shown in Table S14.

3.5. Colocalization of DEGs in the Introgression Fragments

SNP genotyping revealed 16 N22-derived fragments in DTY-IL (Figure 10). The largest fragment was found on chromosome 1, encompassing *qDTY*1.1, followed by an introgression on chromosome 3, containing parts of *qDTY*3.2, which was also reported as N22-derived in an N22 by Swarna population [7]. Additional introgressions on chromosomes 4, 8, 9, and 10 did not overlap with major DTY QTL. Overlaying DEGs on the N22 introgressions identified 463 DEGs in the flag-leaf (Table S15-1) and 433 DEGs in the panicle (Table S15-2), of which 6 overlapped with the fine mapped region of *qDTY*1.1 [59], while 5 overlapped with the *qDTY*3.2 region.

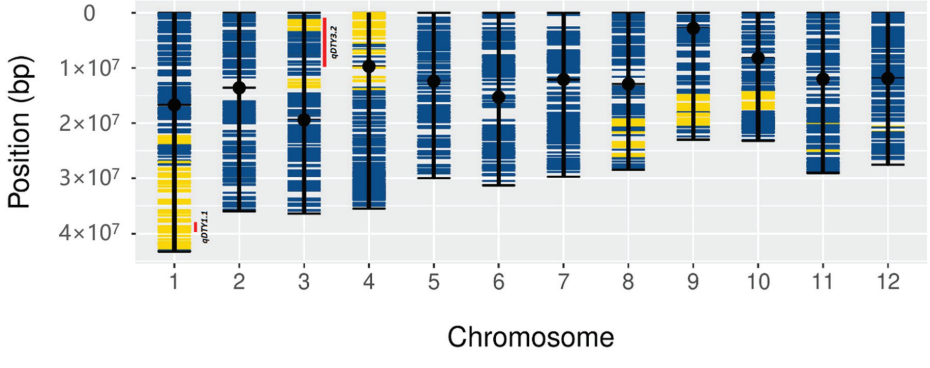


Figure 10. Physical map of DTY-IL. Genome-wide physical position distribution of 1648 SNPs from the 7K genotyping assay across all rice chromosomes. Swarna SNP's alleles are represented in blue. N22 SNP alleles are represented in yellow. Names and ranges of N22-derived DTY QTLs (*qDTY*1.1 and *qDTY*3.2) are shown as red bars on the sides of the chromosome, more details are provided in Supplementary Table S15.

Of the 6 DEGs in the *qDTY*1.1 region (Table S16-2), LOC_Os01g67030 was upregulated in the panicle (Log_2 fold change = 3.1), was annotated as an auxin-responsive protein of 418 amino acids (AA) in Nipponbare. While LOC_Os01g67030 was annotated in the *indica* reference MH63v2, it was missing in the N22v2 reference genome. FGENESH-based gene prediction in N22v2 revealed a putative homologue showing 91.4% identity with Nipponbare and 92.3% identity with MH63v2. Differences were found in the 5′UTR, resulting in the loss of coding sequence for the first 37 AA in the MH63v2 and N22v2 alleles, as well as a number of nonsynonymous (NS) SNPs (Figure S18). NS SNPs unique to N22 corresponded to four AA changes (P52L, C60F, S80T, Q137P) and an AA deletion (G253del) (Figure S19). While the alignment of the 2 kb upstream *cis*-regulatory region of LOC_Os01g67030 in Nipponbare and MH63v2 showed a 99.1% identity, N22v2 displayed a large deletion, including the 5′UTR. With an identity of less than 15% to either Nipponbare or MH63v2, the N22v2 *cis*-regulatory region of LOC_Os01g67030 was distinct with unique regulatory elements (Table S17) that could explain the observed expression differences.

Of the 5 DEG within *qDTY*3.2, LOC_Os03g03510, downregulated in the panicle (Log_2 fold change = −1.11617), was annotated as CAMK_KIN1/SNF1/Nim1_like.15-calcium/calmodulin-dependent protein kinase in Nipponbare, with corresponding annotations in MH63v2 and predictions in N22v2. LOC_Os03g03510 contains a sucrose non-fermenting 1-related kinase 3 (SnRK3) domain and a CBL-interacting serine/threonine-protein kinase 3 (CIPK3) domain. While the coding sequences were largely conserved in multiple sequence alignment between Nipponbare, MH63v2, and N22v2 alleles, the N22v2 allele featured an altered stop codon resulting in a 35 AA C-terminal extension (Figure S20).

4. Discussion

Source-sink relationships largely determine the grain yield of cereal crops, with developing grains being primary sinks, while the top two leaves, the flag-leaf, in particular, serves as the primary source [60,61]. Source sink regulation is orchestrated through intricate metabolic signaling [62], of which key players in sucrose sensing (e.g., trehalose-6-phosphate) and signal integration (e.g., SnRK1) are beginning to emerge [62]. Drought stress affects these relationships by reducing both source and sink strengths. In source organs, limitations in carbon fixation and primary metabolism lead to reduced resource allocation to developing sinks, causing yield reduction characterized by suboptimal grain filling [63]. In sink organs, drought reduces fertility, causing yield reductions through suboptimal seed setting [64].

While DTY-QTLs have demonstrated effects to improve rice grain yields under RDS, knowledge about underlying molecular mechanisms is limited. Functional studies of *qDTY12.1* suggested an intricate pattern of below-ground contributions [18], while physiological studies of *qDTY1.1* suggested above-ground implications [59]. Though confined to a single time point at the late booting stage (close to anthesis) after two weeks of RDS, our study suggested that DTY controlled mechanisms improve yield under drought by acting at both source and sink organs. At the flag-leaf, a coordinated response to sustain primary metabolism through cell wall loosening and maintained photosynthetic rates seems to allow for sufficient carbon and energy allocation to the developing panicle, which in turn enable reproductive structures to invest in protective mechanisms, including protein stabilization and turnover, ROS scavenging and production of protective secondary metabolites. While proximate effects in the panicle are suggested as improved male fertility and improved sink strength under RDS, the ultimate effects are improved seed setting and grain filling, and consequently, drought-tolerant yield (DTY) (Figure S21).

4.1. Source Effects—Flag-Leaf-Specific Differences between DTY-IL and Swarna

Collectively WGCNA and DE analyses suggested a complex interplay of a range of processes in the flag-leaf to contribute to the observed differences in RDS tolerance between DTY-IL and

Swarna. These included specific protein turn-over, cell wall loosening, efficient ROS scavenging, and maintenance of photosynthesis (Figures 4–6 and Figure S12).

A direct consequence of drought is impaired cell turgor [65], which is countered by the stiffening of cell walls to provide structural resistance [66,67]. Prolonged drought stress challenges plants to modify their cell walls, resulting in both cell wall tightening and loosening. Tightening occurs in tissues that are of relevance to structural integrity, while loosening occurs in tissues that need to be maintained in a growing and metabolically active mode [66].

Leaf rolling, a common indicator of drought stress in rice [68] was prominent in Swarna under drought but nearly absent in DTY-IL (Figure 1D). Leaf rolling likely relates to aberrant cell turgor and cell wall homeostasis and phenotypically reflects findings in the flag-leaf specific module M14.

Cell wall organization or biogenesis genes showed an increase in expression in FL-M14. A total of 12 cell wall-related genes were significantly upregulated in the DTY-IL and significantly downregulated in Swarna (Figure 6). These included two glycosyltransferase family 43 proteins, previously reported being involved in the synthesis of glucuronoxylan hemicellulose of secondary cell walls [69] and two expansin genes. Expansins facilitate loosening and extension of plant cell walls by disrupting non-covalent bonding between cellulose microfibrils and matrix glucans [70] and implications in response to dehydration are well documented [71–75] and rose [76].

Higher expression of cytoskeleton and cell cycle-related genes in DTY-IL (Figure 6) further supported the concept of maintained cell growth and stability in the tolerant flag-leaf tissue. Contrastingly, cytoskeletal genes (tubulin and formin) and a cell cycle gene (cyclin) were significantly downregulated in Swarna (Table S10-1).

Several classes of enzymes control ROS production in the cell wall, most prominently plasma membrane NADPH oxidases [77] and class III peroxidases (CIII Prxs) [78]. CIII Prxs are secreted in the extracellular space, where they perform either cell wall stiffening through the peroxidative cycle [79] or cell wall loosening through the hydroxylic cycle [80,81]. In the present study, three CIII Prxs (LOC_Os03g13200, LOC_Os07g01370, and LOC_Os07g48020) were present in FL-M14 (Table S10-1), with LOC_Os03g13200 and LOC_Os07g48020 significantly upregulated in DTY-IL and significantly downregulated in Swarna (Figure 6). High CIII Prxs expression in DTY-IL could support the generation of $^{\bullet}OH$ for cell-wall loosening through cleavage of cell wall polymers [67]. Interestingly, decreased expression of a calcium-dependent NADPH oxidase in Swarna and increased in activity of the DTY-IL in FL-M16 (Table S10-2) was also observed. It is also known as respiratory burst oxidase, and is a well-studied enzymatic source of superoxide [82,83], which had previously been implicated in drought and high-temperature stability [83]. Hence, loosening of the cell wall and synthesis of structural constituents together is suggested to contribute to tolerance of water-deficit in the flag-leaf of DTY-IL.

More effective ROS scavenging, in general, seemed to be an important mechanism differentiating drought responses of Swarna and DTY-IL. Higher expression of peroxiredoxin in DTY-IL (Table S10-2) suggests increased reduction capacity for H_2O_2, indicating enhanced detoxification in drought-stressed leaves.

A primary detrimental effect of water stress in source tissues is impaired photosynthesis [84]. Reduced abundance of photosynthesis-related proteins in response to RDS had previously been reported [85] and was indeed reflected in drought-stressed leaves of RDS-susceptible Swarna (FL-M16). Components of the light reaction (two photosystem genes, components of the core complex of photosystem II (PSII) involved the primary light-induced photochemical processes), the dark reaction (ribulose bisphosphatase and the fructose-1,6 bisphosphatase), and photorespiration (ribulose bisphosphate carboxylase large chain precursor) were found to be consistently downregulated in Swarna (Table S10-2), suggesting functional impairments of general photosynthesis. Protection of photosynthesis from photoinhibition through photorespiration is a well-characterized drought response and furthermore prevents ROS accumulation in green tissues [86]. In addition, Swarna showed a lower expression of two ferredoxin-NADP genes, involved in thylakoid electron transport, suggesting reduced capacity in regulating the relative amounts of cyclic and non-cyclic electron for

ATP and redox homeostasis [69]. Consequently, it is argued that the physiological environment in DTY-IL under RDS supports relatively higher rates of photosynthesis, which in turn, might sustain higher rates of energy and carbon production to support primary metabolism and source strength, ultimately leading to improved grain filling.

4.2. Sink Effects—Panicle Specific Differences between DTY-IL and Swarna

Collectively WGCNA and DE suggested a number of distinct mechanisms to contribute to differences in RDS tolerance between DTY-IL and Swarna in panicles. They included protein stabilization and turnover, ROS scavenging, biosynthesis of secondary metabolites for protection of reproductive organs, and hormonal signaling presumably to adapt reproductive developmental processes to drought. Under field conditions, they resulted in an approximate doubling of yield under drought for DTY-IL as compared to Swarna (Figure 1A), at no significant difference in plant height (Figure 1B).

Dehydration stress enhances the production of ROS and ROS-associated peroxidation causing damage to cellular structures [87]. Being essential for cellular signaling, ROS homeostasis depends on the balance between ROS production and scavenging [82]. Analysis of genome-scale metabolic pathways in the DTY-IL revealed the up-regulation of genes involved in the biosynthesis of antioxidant enzymes and metabolites (Figures 8 and 9 and Figure S14).

Secondary metabolite production is crucial in stress-adaptive mechanisms [88]. Genes in pathways involved in secondary metabolite biosynthesis, lipid biosynthesis, redox homeostasis, amino acid metabolism, carbohydrate metabolism, and protein metabolism were upregulated at the maximum booting stage under RDS in DTY-IL and downregulated in Swarna for P-M10 and P-M15 (Figure 9). Several metabolic pathways found to be shared between the two modules include glutathione, terpenoid, and ascorbate metabolism.

De novo protein synthesis and turnover is fundamental for plants to cope with drought stress [85]. Translational efficiency is affected by ribosome composition, thus relative ribosomal protein abundance can modulate plant environmental responses [89]. Similarly, drought-responsive peptidases and heat shock proteins can alter the active proteome to cope with stress [85,90]. In P-M10 five ribosomal protein-related genes, six protein degradation-related genes (among them four peptidases), and two genes related to protein folding and repair displayed higher expression in DTY-IL (Figure 9; Table S12-1). Congruently reduced expression of two peptidases and three genes related to protein processing, including a heat shock protein was observed for Swarna in P-M15 (Figure 9; Table S12-2). Collectively this suggested that panicles of DTY-IL were more responsive and had the necessary energy to adapt its proteome to drought conditions than Swarna.

In P-M10, six genes involved in the ROS scavenging (two ascorbate peroxidases and four peroxidase precursors) had elevated expression profiles in the DTY-IL (Figure 9; Table S12-1). Efficient reduction of H_2O_2 by peroxidases had previously been implicated with drought-tolerance in rice [19]. Specifically, plant ascorbate peroxidases (APXs) are crucial for ROS homeostasis [91] and free radical detoxification though the ascorbate-glutathione cycle [92], and their functional role in rice drought tolerance was demonstrated through transgenic approaches [91]. In P-M15 three ROS scavenging genes (1 glutathione *S*-transferase, 1 glutathione peroxidase, and 1 stromal ascorbate peroxidase) had a lower expression profile in Swarna (Figure 9; Table S12-2). Glutathione peroxidase (GPX) catalyzes the reduction of H_2O_2 using thioredoxin (Trx), while glutathione *S*-transferases (GSTs) are key to the removal of xenobiotic compounds [85]. Ectopic expression of a GST in *Arabidopsis* [93] and a GPX in rice [94] resulted in enhanced drought tolerance, suggesting functional implications.

Interestingly, an ABC function gene with AP2 domain-containing protein (LOC_Os07g22770) controlling floral organ identity was downregulated in Swarna RDS under P-M15 (Table S12-2), suggesting a link to aberrant Swarna floral development under drought [95]. TFs belonging to AP2 and MYB family are involved in panicle development as well as water-deficit stress response, implying that they may represent a crosstalk component between redevelopment and stress.

A negative regulator of plant drought tolerance in abscisic acid (ABA) signaling, protein phosphatase 2C (PP2C) [96], was upregulated in the DTY-IL in P-M10 (Table S12-1). PP2C inhibits the activity of sucrose non fermenting 1 related kinase 1 (SnRK1) [97], a central integrator of metabolic signaling and regulator of starvation response. Thus, the higher expression of PP2C in DTY-IL might correlate with reduced SnRK1 activity, indicative of anabolism rather than catabolism and thus growth rather than the starvation mode in panicles of DTY-IL.

Brassinosteroids (BRs) are growth-promoting steroid hormones important for male fertility and pollen development [98]. BR catabolism is controlled by BAS1, a cytochrome P450 monooxygenase [99]. BRs bind to the extracellular domain of a cell-surface receptor kinase, BRASSINOSTEROID INSENSITIVE1 (BRI1) to activate kinase activity [100,101]. In P-M10, a BAS1-orthologue and two BRI1 genes were found to be upregulated in DTY-IL (Supplementary Table S12-1), suggesting a role for BR signaling in the maintenance of male fertility as part of the *qDTY*1.1-mediated RDS responses.

4.3. qDTY1.1-Specific Contributions to Drought Tolerance

Of the 16 N22-derived introgressions in DTY-IL, two overlapped with known *qDTY*s for which N22 was a known donor [21]. While the large introgressions on chromosome 1 contain the full *qDTY*1.1 region, a smaller introgression on chromosome 3 partially overlaps with *qDTY*3.2 (Figure 10). Working under the assumption that the genome-wide changes in transcriptomes between the drought-tolerant and drought susceptible genotypes in both flag-leaves and panicle, would at least partially trackback to either transcriptional or allelic differences of specific loci within the *qDTY* regions we took a closer look at the *qDTY*1.1 fine mapped region and *qDTY*3.2 overlap.

Based on the Nipponbare reference, the fine-mapped *qDTY*1.1 region encompasses 79 genes, of which six were differentially expressed between DTY-IL and Swarna under RDS in either tissue (Table S15). An auxin-responsive protein (LOC_Os01g67030), specifically upregulated in DTY-IL panicle under RDS was considered as a likely causative candidate. Generally, auxin has been shown to negatively regulate drought adaptation in plants [63]. Notably, the DEEPER ROOTING 1 (DRO1) promoter was shown to contain auxin-responsive elements (AuxRes) and negatively regulated by auxin [102]. While two AuxRes were found in the LOC_Os01g67030 promoter of Nipponbare and MH63v2 (-513 bp and -1606 bp) only one was found at position -1573 bp in N22v2. The absence of the proximal AuxRe motif, in addition to the presence of novel, putative drought-responsive elements (Figure S18) could explain the observed differential regulation of LOC_Os01g67030 under RDS in DTY-IL.

LOC_Os01g67030.1 contains a cytochrome b561 (Cyt_b561) and a dopamine β-monooxygenase (DOMON) domain (Figure S19). Cyt_b561 proteins are involved in the regeneration of ascorbate through transmembrane electron transport [103,104] and have previously been implicated in drought tolerance through redox homeostasis [105]. The functionally uncharacterized β sheet-rich DOMON domain has been implicated in sugar and heme recognition [106] and predicted to be involved in protein-protein interactions, putatively functioning in metabolic signaling, in redox reactions, or both. Interestingly, LOC_Os01g67030 thus has the potential to link sugar signaling and ROS signaling, both of which have emerged as essential in the DTY-IL-specific drought response.

The 5 AA differences in the N22v2 prediction of LOC_Os01g67030 sequence fall under the two conserved domains (Figure S19). Notably they include two proline conversions and a glycine deletion, with potential structural implications, particularly in the context of transmembrane domains and β sheets [103,104]. This could affect the ability of the Cyt_b561 domain to mediate transmembrane transport and the DOMON domain to mediate protein-protein interactions or ligand binding.

Efficient ROS scavenging was identified as a key mechanism of RDS tolerance in both panicles and flag-leaves of DTY-IL. In the panicle, LOC_Os01g67030 could directly contribute to ROS homeostasis and with ROS being increasingly implicated in stress signaling including modulation of gene expression [107,108], LOC_Os01g67030 activity could be responsible for the some of the expression changes observed in P-M10 and P-M15. LOC_Os01g67030, however, was not expressed in flag-leaves.

It is possible that the ultimate positive effects of the N22 allele of LOC_Os01g67030 on seed setting in DTY-IL panicles could increase sink strength in a way that it positively affects the source strength of flag-leaves, which could contribute to maintained photosynthetic rates. A similar sink on source effects has been demonstrated by manipulation of SnRK1 dependent metabolic signaling in maize under control and drought [109].

LOC_Os03g03510 was found downregulated in DTY-IL. Both CIPK_C domain [110] and SnRK3 domain [111,112] have been implicated in abiotic stress responses, including drought tolerance. In addition, SnRK3, like SnRK1 [62,109] has been demonstrated to function in metabolic signaling and source-sink relationships. In sinks, SnRK1 activity has detrimental effects on grain filling [109]. A 35 AA C-terminal extension in the N22 allele could have functional implications, which, in addition to its observed downregulation could reduce its efficiency in DTY-IL. The postulated effect would be altered downstream phosphorylation responses with potential transcriptional changes that reflect some of the differences seen between Swarna and DTY-IL Ultimately this could contribute to maintained sink strength of the panicle with putative effects on flag-leaf source metabolism.

In theory the postulated functions of both candidates could have synergistic effects that could explain a range of the observed DTY-IL specific drought responses. Gene validation studies expressing the N22 allele of LOC_Os01g67030, LOC_Os03g03510, or both, under control of their native promoters in drought susceptible *indica* background, respective knock-outs in DTY-IL, or both, are needed to confirm their postulated roles.

5. Conclusions

This study provided novel insight into global transcriptional responses in rice under moderate RDS in a DTY-dependent manner and highlighted associated physiological mechanisms that allow DTY-IL to better cope with RDS. In DTY-IL flag-leaves, structural and metabolic integrity associated with cell wall re-organization and active ROS metabolism prevent leaf rolling and allow for maintenance of cellular growth and homeostasis under RDS, which supports sustained rates of photosynthetic activity and consequently provisioning of energy and carbon to developing sinks. In the developing panicles close to anthesis, sustained energy and carbon allocation enables the minimizing of damage to reproductive structures due to RDS through protective mechanisms, including ROS homeostasis, post-transcriptional modifications, detoxification, and secondary metabolite production; ultimately this results in improved fertility and yield under moderate RDS (Figure S21). Assessment of DTY-specific allelic variation within the *qDTY1.1* and *qDTY3.2* regions prioritized two candidate genes in DTY-IL, a predicted auxin-responsive protein with a DOMON_DOH and a Cyt_*b561* domain, and a CIPK_C and SnRK3-domain-containing protein, which might positively affect source-sink regulation under drought.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/10/1124/s1, Table S1. Mapping results of RNA-Seq reads of Swarna and DTY-IL in flag-leaf and emerging panicle tissues under RDS and control conditions, Table S2. GO enrichment analysis of the commonly down-regulated genes in DTY-IL and Swarna under RDS in the flag-leaf and emerging panicle tissues, Table S3. Pathway enrichment analysis of Swarna downregulated and DTY-IL upregulated DEGs in flag-leaf under RDS, Table S4. Overrepresented *cis*-acting elements of Swarna downregulated and DTY-IL upregulated DEGs in flag-leaf under RDS, Table S5. GO enrichment analysis of the commonly up-regulated genes in DTY-IL and Swarna under RDS in the flag-leaf and emerging panicle tissues, Table S6. Pathway enrichment analysis of Swarna downregulated and DTY-IL upregulated DEGs in emerging panicle under RDS, Table S7. Overrepresented *cis*-acting elements of Swarna downregulated and DTY-IL upregulated DEGs in emerging panicle under RDS, Table S8. List of different module color and sizes generated in flag-leaf and emerging panicle tissue using WGCNA, Table S9. The top 10 Go terms in biological process, molecular function, and cellular component categories in M1 and M2 in flag-leaf and emerging panicle tissues under RDS, Table S10. The list of 140 and 102 genes with their putative functions and expression profiles in the DTY-IL and Swarna under RDS in the M14 and M16 modules in flag-leaf tissue, Table S11. Identified hub genes in M14 and M16 in flag-leaf with their putative functions and expression profiles between the DTY-IL and Swarna under RDS, Table S12. The list of 138 and 73 with their putative functions and expression profiles in the DTY-IL and Swarna under RDS in the M10 and M15 modules in panicle tissue, Table S13. Identified hub genes in M10 and M15 in panicle with their putative functions and expression profiles between the DTY-IL and Swarna under RDS, Table S14. qRT-PCR primers used in the study, Table S15. Information on the introgressed chromosome segments and the differentially expressed genes in the DTY1.1-IL compared with

Swarna under control and reproductive-drought stress conditions and overlapping *qDTY's* in flag-leaf and panicle tissues, Table S16. List of DEGs in the *qDTY1.1* region across the different pairwise comparisons in flag-leaf and emerging panicle transcriptomes under RDS, Table S17. Distinct *cis* regulatory elements in the 2 kb upstream promoter region of LOC_Os01g67030 in Nipponbare, MH63, and N22 sequences, Figure S1. Climate data and soil water content during water stress treatment. Figure S2. Quality control assessment for 16 flag-leaf tissue samples, Figure S3. Quality control assessment for 15 panicle tissue samples, with one Swarna control outlier removed, Figure S4. Quality control assessment for 16 panicle tissue samples, Figure S5. DEGs identification, Figure S6. Venn diagrams of differentially expressed genes (DEGs), Figure S7. Flag-leaf significant GO terms of biological processes that are down-regulated in Swarna (A) and up-regulated in DTY-IL (B) under RDS, Figure S8. Panicle significant GO terms of biological processes that are down-regulated in Swarna (A) and up-regulated in DTY-IL (B) under RDS, Figure S9. Mapman overview of the DEGs of interest in DTY-IL and Swarna under RDS, Figure S10. Identification of gene co-expression modules in the flag-leaf (A-C) and panicle (D-F) transcriptome under RDS, Figure S11. Bar graphs and Heatmaps of M1 and M2 in flag-leaf and panicle tissues under RDS, Figure S12. Major biological processes and over-represented GO Slim descriptions of drought-responsive M14 (A) and M16 (B) in flag-leaf tissue under RDS, Figure S13. Hub genes from the drought-responsive modules in the flag-leaf, Figure S14. Major biological processes and over-represented GO Slim descriptions of drought responsive M10 (A) and M15 (B) in panicle tissue under RDS, Figure S15. Hub genes from the drought-responsive modules in the panicles, Figure S16. The tissue-specific expression under RDS, Figure S17. qRT-PCR validation of candidate genes, Figure S18. Gene structure including the 2 kb upstream region of LOC_Os01g67030 in the three published reference genomes representing Nipponbare, *indica*, and *ausboro*, Figure S19. Multiple peptide sequence alignment of LOC_Os01g67030 in Nipponbare, MH63, and N22 sequences, Figure S20. Multiple peptide sequence alignment of LOC_Os03g03510 in Nipponbare, MH63, and N22 sequences. Figure S21. Model of suggested DTY-IL dependent drought tolerance mechanisms.

Author Contributions: J.A.T. and T.K. conceptualized and designed the experiment; J.A.T. and S.C. performed the experiments; J.A.T. conducted the data analyses; R.M., J.D.A., and P.B. supervised on some of the data analyses; J.A.T. and T.K. wrote the manuscript; R.M. and A.K. (Ajay Kohli) contributed to reviewing and editing; S.D. gave the seeds; A.K. (Arvind Kumar) was involved in funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Bill and Melinda Gates Foundation.

Acknowledgments: We thank the Bill and Melinda Gates Foundation for the project support.

Conflicts of Interest: The authors declare no conflict of interest and 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. Borah, P.; Sharma, E.; Kaur, A.; Chandel, G.; Mohapatra, T.; Kapoor, S.; Khurana, J.P. Analysis of drought-responsive signalling network in two contrasting rice cultivars using transcriptome-based approach. *Sci. Rep.* **2017**, *7*, 42131. [CrossRef] [PubMed]
2. Peng, S.; Cassman, K.G.; Virmani, S.S.; Sheehy, J.; Khush, G.S. Yield Potential Trends of Tropical Rice since the Release of IR8 and the Challenge of Increasing Rice Yield Potential. *Crop Sci.* **1999**, *39*, 1552–1559. [CrossRef]
3. Todaka, D.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front. Plant Sci.* **2015**, *6*, 84. [CrossRef] [PubMed]
4. Lafitte, H.R.; Price, A.H.; Courtois, B. Yield response to water deficit in an upland rice mapping population: Associations among traits and genetic markers. *Theor. Appl. Genet.* **2004**, *109*, 1237–1246. [CrossRef]
5. Venuprasad, R.; Lafitte, H.R.; Atlin, G.N. Response to Direct Selection for Grain Yield under Drought Stress in Rice. *Crop Sci.* **2007**, *47*, 285–293. [CrossRef]
6. Liu, J.; Bennett, J. Reversible and Irreversible Drought-Induced Changes in the Anther Proteome of Rice (*Oryza sativa* L.) Genotypes IR64 and Moroberekan. *Mol. Plant* **2011**, *4*, 59–69. [CrossRef]
7. Vikram, P.; Swamy, B.P.; Dixit, S.; Ahmed, H.; Cruz, M.T.; Singh, A.; Kumar, A. QDTY1.1, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC Genet.* **2011**, *12*, 89. [CrossRef]
8. Jin, Y.; Yang, H.; Wei, Z.; Ma, H.; Ge, X. Rice Male Development under Drought Stress: Phenotypic Changes and Stage-Dependent Transcriptomic Reprogramming. *Mol. Plant* **2013**, *6*, 1630–1645. [CrossRef]
9. Guo, C.; Ge, X.; Ma, H. The rice OsDIL gene plays a role in drought tolerance at vegetative and reproductive stages. *Plant Mol. Biol.* **2013**, *82*, 239–253. [CrossRef]

10. Shankar, R.; Bhattacharjee, A.; Jain, M. Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep.* **2016**, *6*, 23719. [CrossRef]
11. Pandey, S. Economic costs of drought and rice farmers' coping mechanisms. *Int. Rice Res. Notes* **2009**, *32*, 1. [CrossRef]
12. Kumar, A.; Dixit, S.; Ram, T.; Yadaw, R.B.; Mishra, K.K.; Mandal, N.P. Breeding high-yielding drought-tolerant rice: Genetic variations and conventional and molecular approaches. *J. Exp. Bot.* **2014**, *65*, 6265–6278. [CrossRef] [PubMed]
13. Kumar, A.; Sandhu, N.; Dixit, S.; Yadaw, S.; Swamy, M.P.; Shamsudin, N.A. Marker-assisted selection strategy to pyramid two or more QTLs for quantitative trait-grain yield under drought. *Rice* **2018**, *11*, 35. [CrossRef] [PubMed]
14. Bernier, J.; Kumar, A.; Ramaiah, V.; Spaner, D.; Atlin, G.A. Large-Effect QTL for Grain Yield under Reproductive-Stage Drought Stress in Upland Rice. *Crop Sci.* **2007**, *47*, 507. [CrossRef]
15. Venuprasad, R.; Bool, M.E.; Quiatchon, L.; Atlin, G.N. A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. *Theor. Appl. Genet.* **2011**, *124*, 323–332. [CrossRef]
16. Ghimire, K.H.; Quiatchon, L.A.; Vikram, P.; Swamy, B.M.; Dixit, S.; Ahmed, H.; Hernandez, J.; Borromeo, T.; Kumar, A. Identification and mapping of a QTL (qDTY1.1) with a consistent effect on grain yield under drought. *Field Crop. Res.* **2012**, *131*, 88–96. [CrossRef]
17. Yadaw, R.B.; Dixit, S.; Raman, A.; Mishra, K.K.; Vikram, P.; Swamy, B.M.; Sta Cruz, M.T.; Maturan, P.; Pandey, M.; Kumar, A. A QTL for high grain yield under lowland drought in the background of popular rice variety Sabitri from Nepal. *Field Crop. Res.* **2013**, *144*, 281–287. [CrossRef]
18. Dixit, S.; Biswal, A.K.; Min, A.; Henry, A.; Oane, R.H.; Raorane, M.L.; Longkumer, T.; Pabuayon, I.M.; Mutte, S.K.; Vardarajan, A.R.; et al. Action of multiple intra-QTL genes concerted around a co-localized transcription factor underpins a large effect QTL. *Sci. Rep.* **2015**, *5*, 15183. [CrossRef]
19. Raorane, M.L.; Pabuayon, I.M.; Varadarajan, A.R.; Mutte, S.K.; Kumar, A.; Treumann, A.; Kohli, A. Proteomic insights into the role of the large-effect QTL qDTY 12.1 for rice yield under drought. *Mol. Breed.* **2015**, *35*, 6. [CrossRef]
20. Degenkolbe, T.; Do, P.T.; Zuther, E.; Repsilber, D.; Walther, D.; Hincha, D.K.; Köhl, K.I. Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. *Plant Mol. Biol.* **2009**, *69*, 133–153. [CrossRef]
21. Lenka, S.K.; Katiyar, A.; Chinnusamy, V.; Bansal, K.C. Comparative analysis of drought-responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant Biotechnol. J.* **2011**, *9*, 315–327. [CrossRef] [PubMed]
22. Ray, S.; Dansana, P.K.; Giri, J.; Deveshwar, P.; Arora, R.; Agarwal, P.; Khurana, J.P.; Kapoor, S.; Tyagi, A.K. Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. *Funct. Integr. Genom.* **2011**, *11*, 157–178. [CrossRef] [PubMed]
23. Wang, D.; Pan, Y.; Zhao, X.; Zhu, L.; Fu, B.; Li, Z. Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *BMC Genom.* **2011**, *12*, 149. [CrossRef] [PubMed]
24. Huang, L.; Zhang, F.; Zhang, F.; Wang, W.; Zhou, Y.; Fu, B.; Li, Z. Comparative transcriptome sequencing of tolerant rice introgression line and its parents in response to drought stress. *BMC Genom.* **2014**, *15*, 1–26. [CrossRef]
25. Hu, H.; Xiong, L. Genetic Engineering and Breeding of Drought-Resistant Crops. *Annu. Rev. Plant Biol.* **2014**, *65*, 715–741. [CrossRef]
26. Baldoni, E.; Bagnaresi, P.; Locatelli, F.; Mattana, M.; Genga, A. Comparative Leaf and Root Transcriptomic Analysis of two Rice Japonica Cultivars Reveals Major Differences in the Root Early Response to Osmotic Stress. *Rice* **2016**, *9*, 25. [CrossRef]
27. Zhang, Z.; Li, Y.; Xiao, B. Comparative transcriptome analysis highlights the crucial roles of photosynthetic system in drought stress adaptation in upland rice. *Sci. Rep.* **2008**, *6*, 19349. [CrossRef]
28. Arvidsson, S.; Kwasniewski, M.; Riano-Pachon, D.M.; Mueller-Roeber, B. QuantPrime-a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinform.* **2016**, *9*, 465. [CrossRef]
29. Umezawa, T.; Fujita, M.; Fujita, Y.; Yamaguchi-Shinozaki, K.; Shinozaki, K. Engineering drought tolerance in plants: Discovering and tailoring genes to unlock the future. *Curr. Opin. Biotechnol.* **2006**, *17*, 113–122. [CrossRef]

30. Moumeni, A.; Satoh, K.; Kondoh, H.; Asano, T.; Hosaka, A.; Venuprasad, R.; Serraj, R.; Kumar, A.; Leung, H.; Kikuchi, S. Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC Plant Biol.* **2011**, *11*, 174. [CrossRef]
31. Ding, X.; Li, X.; Xiong, L. Insight into Differential Responses of Upland and Paddy Rice to Drought Stress by Comparative Expression Profiling Analysis. *Int. J. Mol. Sci.* **2013**, *14*, 5214–5238. [CrossRef]
32. Weng, X.; Wang, L.; Wang, J.; Hu, Y.; Du, H.; Xu, C.; Xing, Y.; Li, X.; Xiao, J.; Zhang, Q. Grain Number, Plant Height, and Heading Date7 Is a Central Regulator of Growth, Development, and Stress Response. *Plant Physiol.* **2014**, *164*, 735–747. [CrossRef]
33. Moumeni, A.; Satoh, K.; Venuprasad, R.; Serraj, R.; Kumar, A.; Leung, H.; Kikuchi, S. Transcriptional profiling of the leaves of near-isogenic rice lines with contrasting drought tolerance at the reproductive stage in response to water deficit. *BMC Genom.* **2015**, *16*, 1110. [CrossRef] [PubMed]
34. Wei, H.; Chen, C.; Ma, X.; Zhang, Y.; Han, J.; Mei, H.; Yu, S. Comparative Analysis of Expression Profiles of Panicle Development among Tolerant and Sensitive Rice in Response to Drought Stress. *Front. Plant Sci.* **2017**, *8*, 437. [CrossRef] [PubMed]
35. Hadiarto, T.; Tran, L.P. Progress studies of drought-responsive genes in rice. *Plant Cell Rep.* **2010**, *30*, 297–310. [CrossRef] [PubMed]
36. Shojaie, A.; Michailidis, G. Analysis of Gene Sets Based on the Underlying Regulatory Network. *J. Comput. Biol.* **2009**, *16*, 407–426. [CrossRef] [PubMed]
37. Dela Fuente, A. From 'differential expression' to 'differential networking'–identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **2010**, *26*, 326–333. [CrossRef] [PubMed]
38. Gitter, A.; Carmi, M.; Barkai, N.; Bar-Joseph, Z. Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Res.* **2013**, *23*, 365–376. [CrossRef] [PubMed]
39. Ma, C.; Xin, M.; Feldmann, K.A.; Wang, X. Machine Learning-Based Differential Network Analysis: A Study of Stress-Responsive Transcriptomes in Arabidopsis. *Plant Cell* **2014**, *26*, 520–537. [CrossRef]
40. Zhang, B.; Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat. Appl. Genet. *Mol. Biol.* **2005**, *4*, 17.
41. Gehan, M.A.; Greenham, K.; Mockler, T.C.; Mcclung, C.R. Transcriptional networks—crops, clocks, and abiotic stress. *Curr. Opin. Plant Biol.* **2015**, *24*, 39–46. [CrossRef] [PubMed]
42. Liseron-Monfils, C.; Ware, D. Revealing gene regulation and associations through biological networks. *Curr. Plant Biol.* **2015**, *3–4*, 30–39. [CrossRef]
43. Segal, E.; Shapira, M.; Regev, A.; Peer, D.; Botstein, D.; Koller, D.; Friedman, N. Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **2003**, *34*, 166–176. [CrossRef] [PubMed]
44. Sircar, S.; Parekh, N. Functional characterization of drought-responsive modules and genes in Oryza sativa: A network-based approach. *Front. Genet.* **2015**, *6*, 256. [CrossRef] [PubMed]
45. Horvath, S.; Dong, J. Geometric Interpretation of Gene Coexpression Network Analysis. *PLoS Comput. Biol.* **2008**, *4*, 1–27. [CrossRef]
46. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]
47. Shaik, R.; Ramakrishna, W. Genes and Co-Expression Modules Common to Drought and Bacterial Stress Responses in Arabidopsis and Rice. *PLoS ONE* **2013**, *8*, 10. [CrossRef]
48. Zhang, L.; Yu, S.; Zuo, K.; Luo, L.; Tang, K. Identification of Gene Modules Associated with Drought Response in Rice by Network-Based Analysis. *PLoS ONE* **2012**, *7*, 5. [CrossRef]
49. Counce, P.A.; Keisling, T.C.; Mitchell, A. A Uniform, Objective, and Adaptive System for Expressing Rice Development. *Crop Sci.* **2000**, *40*, 436. [CrossRef]
50. Krishnan, A.; Gupta, C.; Ambavaram, M.M.R.; Pereira, A. RECoN: Rice Environment Coexpression Network for Systems Level Analysis of Abiotic-Stress Response. *Front. Plant Sci.* **2017**, *8*, 1–15. [CrossRef]
51. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 23 September 2020).
52. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

53. Trapnell, C.; Williams, B.; Pertea, G. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [CrossRef] [PubMed]
54. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef] [PubMed]
55. Soneson, C.; Love, M.I.; Robinson, M.D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000 Res.* **2016**, *4*, 1521. [CrossRef]
56. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]
57. 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]
58. Morales, K.Y.; Singh, N.; Perez, F.A.; Ignacio, J.C.; Thapa, R.; Arbelaez, J.D.; Tabien, R.E.; Famoso, A.; Wang, D.R.; Septiningsih, E.M.; et al. An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS ONE* **2020**, *15*, e0232479. [CrossRef]
59. Vikram, P.; Swamy, B.P.; Dixit, S.; Singh, R.; Singh, B.P.; Miro, B.; Kohli, A.; Henry, A.; Singh, N.K.; Kumar, A. Drought susceptibility of modern rice varieties: An effect of linkage of drought tolerance with undesirable traits. *Sci. Rep.* **2015**, *5*, 14799. [CrossRef]
60. Hirota, O.; Oka, M.; Takeda, T. Sink Activity Estimation by Sink Size and Dry Matter Increase During the Ripening Stage of Barley (*Hordeum vulgare*) and Rice (*Oryza sativa*). *Ann. Bot.* **1990**, *65*, 349–353. [CrossRef]
61. Biswal, A.K.; Kohli, A. Cereal flag leaf adaptations for grain yield under drought: Knowledge status and gaps. *Mol. Breed.* **2013**, *31*, 749–766. [CrossRef]
62. Lawlor, D.W.; Paul, M.J. Source/sink interactions underpin crop yield: The case for trehalose 6-phosphate/SnRK1 in improvement of wheat. *Front. Plant Sci.* **2014**, *5*, 418. [CrossRef] [PubMed]
63. Basu, S.; Ramegowda, V.; Kumar, A.; Pereira, A. Plant adaptation to drought stress. *F1000 Res.* **2016**, *5*, 1554. [CrossRef] [PubMed]
64. Guo, C.; Yao, L.; You, C.; Wang, S.; Cui, J.; Ge, X.; Ma, H. MID1 plays an important role in response to drought stress during reproductive development. *Plant J.* **2016**, *88*, 280–293. [CrossRef] [PubMed]
65. Farooq, M.; Wahid, A.; Kobayashi, N.; Fujita, D.; Basra, S.M. Plant Drought Stress: Effects, Mechanisms and Management. *Sustain. Agric.* **2009**, *29*, 153–188.
66. Moore, J.P.; Vicré-Gibouin, M.; Farrant, J.M.; Driouich, A. Adaptations of higher plant cell walls to water loss: Drought vs. desiccation. *Physiol. Plant.* **2008**, *134*, 237–245. [CrossRef]
67. Tenhaken, R. Cell wall remodeling under abiotic stress. *Front. Plant Sci.* **2015**, *5*, 771. [CrossRef]
68. Cal, A.; Sanciangco, M.; Rebolledo, M.B.; Luquet, D.; Torres, R.; McnNaly, K.; Henry, A. Leaf morphology, rather than plant water status, underlies, genetic variation of rice leaf rolling under drought. *Plant Cell Environ.* **2019**, *42*, 1532–1544. [CrossRef]
69. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45*, D362–D368. [CrossRef]
70. Cosgrove, D.J. Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 850–861. [CrossRef]
71. Wu, Y.; Sharp, R.E.; Durachko, D.M.; Cosgrove, D.J. Growth Maintenance of the Maize Primary Root at Low Water Potentials Involves Increases in Cell-Wall Extension Properties, Expansin Activity, and Wall Susceptibility to Expansins. *Plant Physiol.* **1996**, *111*, 765–772. [CrossRef]
72. Jones, L.; Mcqueen-Mason, S. A role for expansins in dehydration and rehydration of the resurrection plantCraterostigma plantagineum. *FEBS Lett.* **2004**, *559*, 61–65. [CrossRef]
73. Harb, A.; Krishnan, A.; Ambavaram, M.M.; Pereira, A. Molecular and Physiological Analysis of Drought Stress in Arabidopsis Reveals Early Responses Leading to Acclimation in Plant Growth. *Plant Physiol.* **2010**, *154*, 1254–1271. [CrossRef] [PubMed]
74. Guo, W.; Zhao, J.; Li, X.; Qin, L.; Yan, X.; Liao, H. A soybean β-expansin gene GmEXPB2 intrinsically involved in root system architecture responses to abiotic stresses. *Plant J.* **2011**, *66*, 541–552. [CrossRef] [PubMed]
75. Li, F.; Han, Y.; Feng, Y.; Xing, S.; Zhao, M.; Chen, Y.; Wang, W. Expression of wheat expansin driven by the RD29 promoter in tobacco confers water-stress tolerance without impacting growth and development. *J. Biotechnol.* **2013**, *163*, 281–291. [CrossRef]

76. Dai, F.; Zhang, C.; Jiang, X.; Kang, M.; Yin, X.; Lu, P.; Zhang, X.; Zheng, Y.; Gao, J. RhNAC2 and RhEXPA4 Are Involved in the Regulation of Dehydration Tolerance during the Expansion of Rose Petals. *Plant Physiol.* **2012**, *160*, 2064–2082. [CrossRef]
77. Torres, M.A.; Dangl, J.L. Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Curr. Opin. Plant Biol.* **2005**, *8*, 397–403. [CrossRef]
78. Shigeto, J.; Tsutsumi, Y. Diverse functions and reactions of class III peroxidases. *New Phytol.* **2016**, *209*, 1395–1402. [CrossRef]
79. Raggi, S.; Ferrarini, A.; Delledonne, M.; Dunand, C.; Ranocha, P.; Lorenzo, G.D.; Cervone, F.; Ferrari, S. The Arabidopsis thaliana Class III Peroxidase AtPRX71 Negatively Regulates Growth under Physiological Conditions and in Response to Cell Wall Damage. *Plant Physiol.* **2015**, *169*, 2513–2525. [CrossRef]
80. Passardi, F.; Penel, C.; Dunand, C. Performing the paradoxical: How plant peroxidases modify the cell wall. *Trends Plant Sci.* **2004**, *9*, 534–540. [CrossRef]
81. Kunieda, T.; Shimada, T.; Kondo, M.; Nishimura, M.; Nishitani, K.; Hara-Nishimura, I. Spatiotemporal Secretion of PEROXIDASE36 Is Required for Seed Coat Mucilage Extrusion in Arabidopsis. *Plant Cell* **2013**, *25*, 1355–1367. [CrossRef]
82. Miller, G.; Suzuki, N.; Ciftci-Yilmaz, S.; Mittler, R. Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant Cell Environ.* **2010**, *33*, 453–467. [CrossRef]
83. You, J.; Chan, Z. ROS Regulation During Abiotic Stress Responses in Crop Plants. *Front. Plant Sci.* **2015**, *6*, 1092. [CrossRef] [PubMed]
84. Patro, L.; Mohapatra, P.K.; Biswal, U.C.; Biswal, B. Dehydration induced loss of photosynthesis in Arabidopsis leaves during senescence is accompanied by the reversible enhancement in the activity of cell wall β-glucosidase. *J. Photochem. Photobiol. B Biol.* **2014**, *137*, 49–54. [CrossRef] [PubMed]
85. Wang, X.; Cai, X.; Xu, C.; Wang, Q.; Dai, S. Drought-Responsive Mechanisms in Plant Leaves Revealed by Proteomics. *Int. J. Mol. Sci.* **2016**, *17*, 1706. [CrossRef] [PubMed]
86. Voss, I.; Sunil, B.; Scheibe, R.; Raghavendra, A.S. Emerging concept for the role of photorespiration as an important part of abiotic stress response. *Plant Biol.* **2013**, *15*, 713–722. [CrossRef] [PubMed]
87. Noctor, G.; Mhamdi, A.; Foyer, C.H. The Roles of Reactive Oxygen Metabolism in Drought: Not So Cut and Dried. *Plant Physiol.* **2014**, *164*, 1636–1648. [CrossRef]
88. Nakabayashi, R.; Yonekura-Sakakibara, K.; Urano, K.; Suzuki, M.; Yamada, Y.; Nishizawa, T.; Matsuda, F.; Kojima, F.; Sakakibara, H.; Shinozaki, K.; et al. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.* **2014**, *77*, 367–379. [CrossRef]
89. Wang, J.; Lan, P.; Gao, H.; Zheng, L.; Li, W.; Schmidt, W. Expression changes of ribosomal proteins in phosphate-and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition. *BMC Genom.* **2013**, *14*, 783. [CrossRef]
90. Miazek, A.; Zagdańska, B. Involvement of exopeptidases in dehydration tolerance of spring wheat seedlings. *Biologia Plant.* **2008**, *52*, 687–694. [CrossRef]
91. Zhang, Z.; Zhang, Q.; Wu, J.; Zheng, X.; Zheng, S.; Sun, X.; Qiu, Q.; Lu, T. Gene Knockout Study Reveals That Cytosolic Ascorbate Peroxidase 2(OsAPX2) Plays a Critical Role in Growth and Reproduction in Rice under Drought, Salt and Cold Stresses. *PLoS ONE* **2013**, *8*, 2. [CrossRef]
92. Cramer, G.R.; Van Sluyter, S.C.; Hopper, D.W.; Pascovici, D.; Keighley, T.; Haynes, P.A. Proteomic analysis indicates massive changes in metabolism prior to the inhibition of growth and photosynthesis of grapevine (*Vitis vinifera* L.) in response to water deficit. *BMC Plant Biol.* **2013**, *13*, 49. [CrossRef]
93. Xu, J.; Xing, X.; Tian, Y.; Peng, R.; Xue, Y.; Zhao, W.; Yao, Q. Transgenic Arabidopsis Plants Expressing Tomato Glutathione S-Transferase Showed Enhanced Resistance to Salt and Drought Stress. *PLoS ONE* **2015**, *10*, 9. [CrossRef] [PubMed]
94. Islam, T.; Manna, M.; Reddy, M.K. Glutathione Peroxidase of Pennisetum glaucum (PgGPx) Is a Functional Cd2 Dependent Peroxiredoxin that Enhances Tolerance against Salinity and Drought Stress. *PLoS ONE* **2015**, *10*, 1–18. [CrossRef] [PubMed]
95. Su, Z.; Ma, X.; Guo, H.; Sukiran, N.L.; Guo, B.; Assmann, S.M.; Ma, H. Flower Development under Drought Stress: Morphological and Transcriptomic Analyses Reveal Acute Responses and Long-Term Acclimation in Arabidopsis. *Plant Cell* **2013**, *25*, 3785–3807. [CrossRef] [PubMed]
96. Bhaskara, G.B.; Nguyen, T.T.; Verslues, P.E. Unique Drought Resistance Functions of the Highly ABA-Induced Clade A Protein Phosphatase 2Cs. *Plant Physiol.* **2012**, *160*, 379–395. [CrossRef] [PubMed]

97. Gosti, F. ABI1 Protein Phosphatase 2C Is a Negative Regulator of Abscisic Acid Signaling. *Plant Cell Online* **1999**, *11*, 1897–1910. [CrossRef] [PubMed]
98. Mandava, N.B. Plant Growth-Promoting Brassinosteroids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1988**, *39*, 23–52. [CrossRef]
99. Neff, M.M.; Nguyen, S.M.; Malancharuvil, E.J.; Fujioka, S.; Noguchi, T.; Seto, H.; Tsukubi, M.; Takatsuto, S.; Yoshida, S.; Chory, J. BAS1: A gene regulating brassinosteroid levels and light responsiveness in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 15316–15323. [CrossRef] [PubMed]
100. Belkhadir, Y.; Chory, J. Brassinosteroid Signaling: A Paradigm for Steroid Hormone Signaling from the Cell Surface. *Science* **2006**, *314*, 1410–1411. [CrossRef] [PubMed]
101. Clouse, S.D. Brassinosteroid Signal Transduction: From Receptor Kinase Activation to Transcriptional Networks Regulating Plant Development. *Plant Cell* **2011**, *23*, 1219–1230. [CrossRef] [PubMed]
102. Uga, Y.; Sugimoto, K.; Ogawa, S.; Rane, J.; Ishitani, M.; Hara, N.; Kitomi, Y.; Inukai, Y.; Ono, K.; Kanno, N.; et al. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nat. Genet.* **2013**, *45*, 1097–1102. [CrossRef]
103. Verelst, W.; Asard, H. A phylogenetic study of cytochrome b561 proteins. *Genome Biol.* **2003**, *4*, 6. [CrossRef] [PubMed]
104. Asard, H.; Barbaro, R.; Trost, P.; Bérczi, A. Cytochromes b561: Ascorbate-mediated trans-membrane electron transport. *Antioxid. Redox Signal.* **2013**, *19*, 1026–1035. [CrossRef] [PubMed]
105. Nanasato, Y.; Akashi, K.; Yokota, A. Co-expression of Cytochrome b561 and Ascorbate Oxidase in Leaves of Wild Watermelon under Drought and High Light Conditions. *Plant Cell Physiol.* **2005**, *46*, 1515–1524. [CrossRef]
106. Lyer, L.M.; Anantharaman, V.; Aravind, L. The DOMON domains are involved in heme and sugar recognition. *Bioinformatics* **2007**, *23*, 2660–2664.
107. Choudhury, F.K.; Rivero, R.M.; Blumwald, E.; Mittler, R. Reactive oxygen species, abiotic stress and stress combination. *Plant J.* **2016**, *90*, 856–867. [CrossRef] [PubMed]
108. Mittler, R. ROS Are Good. *Trends Plant. Sci.* **2017**, *22*, 11–19. [CrossRef]
109. Oszvald, M.; Primavesi, L.F.; Griffiths, C.A.; Cohn, J.; Basu, S.S.; Nuccio, M.L.; Paul, M.J. Trehalose 6-Phosphate Regulates Photosynthesis and Assimilate Partitioning in Reproductive Tissue. *Plant Physiol.* **2018**, *176*, 2623–2638. [CrossRef]
110. Sánchez-Barrena, M.J.; Fujii, H.; Angulo, I.; Martínez-Ripoll, M.; Zhu, J.; Albert, A. The Structure of the C-terminal Domain of the Protein Kinase AtSOS2 Bound to the Calcium Sensor AtSOS3. *Mol. Cell* **2007**, *26*, 427–435. [CrossRef]
111. Hirabak, 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]
112. Wang, Y.; Yan, H.; Qiu, Z.; Hu, B.; Zeng, B.; Zhong, C.; Fan, C. Comprehensive Analysis of SnRK Gene Family and their Responses to Salt Stress in Eucalyptus grandis. *Int. J. Mol. Sci.* **2019**, *20*, 2786. [CrossRef] [PubMed]

genes

MDPI

Article

PEG-Delivered CRISPR-Cas9 Ribonucleoproteins System for Gene-Editing Screening of Maize Protoplasts

Rodrigo Ribeiro Arnt Sant'Ana [1], Clarissa Alves Caprestano [1], Rubens Onofre Nodari [1] and Sarah Zanon Agapito-Tenfen [2,*]

[1] CropScience Department, Federal University of Santa Catarina, Florianópolis 88034000, Brazil; rodrigoarnt@hotmail.com (R.R.A.S.); clarissacapre@gmail.com (C.A.C.); rubens.nodari@ufsc.br (R.O.N.)
[2] GenØk—Centre for Biosafety, Siva innovasjonssenter Postboks 6418, 9294 Tromsø, Norway
* Correspondence: sarah.agapito@genok.no

Received: 18 August 2020; Accepted: 30 August 2020; Published: 2 September 2020

Abstract: Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 technology allows the modification of DNA sequences in vivo at the location of interest. Although CRISPR-Cas9 can produce genomic changes that do not require DNA vector carriers, the use of transgenesis for the stable integration of DNA coding for gene-editing tools into plant genomes is still the most used approach. However, it can generate unintended transgenic integrations, while Cas9 prolonged-expression can increase cleavage at off-target sites. In addition, the selection of genetically modified cells from millions of treated ones, especially plant cells, is still challenging. In a protoplast system, previous studies claimed that such pitfalls would be averted by delivering pre-assembled ribonucleoprotein complexes (RNPs) composed of purified recombinant Cas9 enzyme and in vitro transcribed guide RNA (gRNA) molecules. We, therefore, aimed to develop the first DNA-free protocol for gene-editing in maize and introduced RNPs into their protoplasts with polyethylene glycol (PEG) 4000. We performed an effective transformation of maize protoplasts using different gRNAs sequences targeting the inositol phosphate kinase gene, and by applying two different exposure times to RNPs. Using a low-cost Sanger sequencing protocol, we observed an efficiency rate of 0.85 up to 5.85%, which is equivalent to DNA-free protocols used in other plant species. A positive correlation was displayed between the exposure time and mutation frequency. The mutation frequency was gRNA sequence- and exposure time-dependent. In the present study, we demonstrated that the suitability of RNP transfection was proven as an effective screening platform for gene-editing in maize. This efficient and relatively easy assay method for the selection of gRNA suitable for the editing of the gene of interest will be highly useful for genome editing in maize, since the genome size and GC-content are large and high in the maize genome, respectively. Nevertheless, the large amplitude of mutations at the target site require scrutiny when checking mutations at off-target sites and potential safety concerns.

Keywords: gene editing; mutagenesis; genetically modified; genetically modified organism (GMO); crop breeding; ribonucleoprotein complex (RNP); genetic screening

1. Introduction

The development of technologies related to genetic improvement, such as transgenesis and more recently, genome editing, have changed the way humans have grown food for thousands of years. Today, the most promising tool to DNA manipulation is CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), a gene-editing technology that has been adapted from the bacterial immune system against viral infections [1]. Currently, a target site in the host genome can be reached by an endonuclease enzyme (i.e., Cas9) led by a guide RNA molecule (gRNA) that contains the target-specific

sequence. This protein complex forms the CRISPR-Cas9 ribonucleoproteins (RNPs) gene-editing system. Cas9 introduces a site-specific double-stranded DNA break (DSB) followed by the natural cell repair of disrupted genome integrity by error-prone non-homologous end-joining (NHEJ) or homology-directed repair (HDR) [2]. Therefore, this tool allows the in vivo modification of the DNA at the gene sequence of interest, with unprecedented speed, and has made it a milestone in manipulating and producing living modified organisms.

Although much is already known about the principles of CRISPR-Cas9 genome editing, the likelihood of different outcomes in terms of resolution, efficiency, accuracy, and DNA modification structure has shown to be species-dependent. Various factors, including target site choice, gRNA design, the properties of the endonuclease, the type of DSB introduced, whether or not the DSB is unique, the quantity of the endonuclease and gRNA, and the intrinsic differences in DNA repair pathways in different species, tissues, and cells will result in differences in the mutation signatures generated in plant species [3].

In the case of maize, most researchers acknowledge that *Zea mays* L. is an ancient amphidiploid species with a duplicated chromosome number of $n = 10$ [4]. Multiple independent domestication events have also contributed to the genetic variability encountered in modern maize [4]. In addition, the genetic variants and the underlying mechanisms influencing variance heterogeneity in maize have so far hidden additive genetic effects and epistatic interaction effects in elite varieties [5]. Therefore, CRISPR can be a useful tool to access genomic regions in the maize genome, which are difficult to achieve by conventional breeding.

While CRISPR technology has already been tested on commercial crops to increase yield, drought tolerance and growth under limited nutrient conditions, improve nutritional properties and develop resistance to plant pathogens [6]; breeding and research of major monocotyledon species, more specifically maize, are still at its infancy. Maize has shown to be an exemption in the plant portfolio for the in silico analysis of potential Cas9 target sites as only 29.5% of annotated transcripts matched a specific gRNA [7]. Among eight analyzed plant species, maize had the largest genome, the highest GC content, and the greatest number of annotated transcripts. Thus, reflecting the abundance of highly repetitive DNA and dispersed repeats, which may be challenging to develop unique target sites for the majority of genes in maize [7].

Despite such challenges, CRISPR technology opens up the possibility for genome changes without foreign introgression of DNA vectors. CRISPR-Cas9 technology can be used as RNP complexes without the introgression and expression of a transgenic cassette in the host genome [8]. Such an approach would avoid a number of generations of backcrossing, expression vectors, and other invasive methods of cell penetration (e.g., biolistics) that can lead to gene disruption, including large deletions, partial trisomy, genome shattering events, and plant mosaicism [9]. Overall, these side effects can mask or interfere in the target gene functional analysis, and further additional biosafety concerns prior to commercial release.

CRISPR RNPs can be delivered directly to plant cells without the cell wall. Therefore, prior to transfection, the cell wall must be removed through enzymatic digestion reactions. Protoplast cells are viable in vivo biological material for DNA-free CRISPR delivery in plants. In addition, protoplasts are viable after transfection, which allows further tissue cultivation and propagation.

Delivery of preassembled Cas9 protein-gRNA RNPs or plant DNA-free genome-editing techniques are not exempt from off-target effects but represent an approach in which the effects of Cas9 could be isolated from other more invasive techniques [8,10,11]. This approach was first demonstrated in *Arabidopsis thaliana*, tobacco, lettuce, and rice protoplasts, including the regeneration of gene-edited lettuce [12]. After that, a few successful attempts were also accomplished on grapevine and apple [13], *Petunia* × *hybrida* [14], potato [15], and on soybeans and tobacco using CRISPR-Cpf1 (CRISPR from *Prevotella* and *Francisella*), recently named Cas12a (for review, please read Metje-Sprink et al. [8]). Maize and wheat plants with targeted mutations have also been successfully obtained by delivering gold particles coated with the RNPs into embryo cells (biolistics), followed by post-bombardment culture

and plant regeneration [16]. However, the frequency of obtaining genome-edited plants was relatively low, since only 0.3–0.9% of regenerated maize plants possessed biallelic mutations [16].

The few studies published on maize genome editing rely mostly on the stable transformation [17–20]. In the manuscript, we delivered Cas9-gRNA RNP into maize leaf protoplasts via polyethylene glycol (PEG)-calcium mediated transfection, and indicated that In/Del mutations occurred with relatively high efficiency of 0.85–5.85% among the PEG-calcium-treated protoplast. We targeted the *inositol phosphate kinase* gene (IPK) involved in the phytic acid biosynthetic pathway. To develop a standard protocol for different maize varieties, we designed gRNAs and primers complementary to coding regions in exon three that are conserved in the species, in order to evaluate the efficiency and spectrum of DNA changes generated by CRISPR-Cas9 technology in maize, and also add relevant information to the safety of gene-edited organisms. This efficient and relatively easy assay method for the selection of gRNA suitable for editing of genes of interest will be highly useful for genome editing in maize, since the genome size and GC-content are large and high in the maize genome, respectively.

2. Materials and Methods

2.1. Target Site Selection and In Vitro Cleavage Assay

The *Zea mays* IPK gene sequence data was obtained from the NCBI GenBank (accession B73RefGen_v3). In vitro tests were performed to confirm the RNP complex efficiency to cleave target DNA. The target site was amplified using specific primers. The crispr-RNAs (crRNAs) were designed for the third exon of the IPK gene in maize using the platform CRISPR-Cas9 guide RNA Design Checker (Integrated DNA Technologies Inc. IDT, Coralville, IA, SUA). Commercially available Cas9 protein (160 kDa) was also purchased from IDT (Table 1).

Table 1. List of primers and crispr-RNAs (crRNAs) used for the amplification and mutation of the *inositol phosphate kinase* gene (IPK) target locus in maize via CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9.

Primer	Sequence (5′–3′)	Amplicon Size (bp)
ZmIPK-F	GAAGAAGCAGCAGAGCTTCA	876
ZmIPK-R	CAGAAGAAATCCGTGAGGACAG	
crRNA	**Sequence (5′–3′)**	**Cleavage Fragments (bp)**
crRNA1	AGCTCGACCACGCCGCCGAC	279 | 597
crRNA2 *	GGGATCCGTCTCCTTCTCCC	617 | 259
crRNA3 *	ATCTTCAAGGTCTACGTCGT	525 | 351
crRNA4 *	CAGGAGTTCGTCAACCATGG	498 |378
crRNA5	ACAAGCTCTACGGAGACGAC	141 | 735

Note: * Selected crRNAs used for protoplasts transfection.

gRNA, crRNA (100 nM), and trans-activating crispr-RNA (tracrRNA) (100 nM) were incubated for 5 min at 95 °C, according to the manufacturer's instructions. Cas9 (100 nM), gRNA, and 1 × NEBuffer 3 (New England Biolabs Inc, NEB) were incubated for 10 min at 25 °C to form the RNP complex. Amplified PCR products (300 ng) were then incubated for 60 min at 37 °C with the RNP complex. Proteinase K (800U/mL) was added to stop the reaction. The products were visualized using 1% agarose gel electrophoresis [13].

2.2. Maize Protoplast Isolation and Fluorescent Transfection Assay

Mesophyll protoplasts were isolated following the protocol described by Sheen et al. [21] with some modifications. Etiolated maize seedlings were grown in vermiculite after disinfestation of the seeds with 70% alcohol (60 s), NaOCl2% (twice of 15 min), and triple washing with sterile water. Ten-days-old seedlings were used (three days under 16 h light/day and seven days in darkness).

The middle portion of the second leaves were cut into thin strips (0.5–1 mm), and immersed in a cell-wall digestion enzyme solution (0.3% macerozyme R-10, 1.5% cellulase R-10, 10 mM of MES pH 5.7, 0.6 M mannitol, 10 mM $CaCl_2$, 5 mM ß-mercapto, 0.1% BSA). The material was left in a vacuum for 30 min and gentle shaking at 40 rpm in the dark for 4 h. The protoplasts were released thoroughly by shaking at 80 rpm for 5 min After digestion, the protoplasts were diluted with the same volume of cold W5 solution [2 mM MES (pH 5.7), 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl] and filtered through a double filter (14 and 40 µM Nylon mesh). Protoplasts were collected after centrifugation at 100 g for 3 min and washed two times in 10 mL of W5 solution. Protoplasts were resuspended in cold MMG solution [0.4 M mannitol, 4 mM MES (pH 5.7), 15 mM $MgCl_2$]. Its viability and concentration were determined using Fluorescein Diacetate (FDA; 0.2%) dye in the hemocytometer. To confirm the internalization of the RNP complex inside cells, an assay was performed using fluorescent-labeled tracrRNA molecules (ATTO 550, IDT) [9]. Microphotographs were taken using an inverted optic microscope ix80 Olympus.

2.3. Maize Protoplast Transformation

Maize protoplasts were gene-edited by introducing CRISPR-Cas9 RNP complex (no integration of exogenous DNA) via PEG-mediated transfection (Figure 1). Protoplast transformation was adapted from Woo et al. [12] and Malnoy et al. [13]. First, 15 µg of the two components of the gRNA (crRNA and trans-activating crispr RNA-tracrRNA)) were incubated at 95 °C for 5 min. After allowing them to cool at room temperature, 45 µg of Cas9 and 1 × NEBuffer 3 were added, then mixed, and incubated at 25 °C for 10 min. Finally, the RNP complex was mixed with 100 µL of protoplasts (1×10^5 protoplasts), 250 µL of PEG solution (40% PEG 4000, 0.2 M mannitol, 0.1 M $CaCl_2$) (pH 6.0), and incubated at 25 °C in the dark. Two incubation times were tested: T1 = 20 min and T2 = 40 min W5 solution (950 ul) was added and the tubes were centrifuged at 100 g for 3 min. Protoplasts were resuspended in 1 mL W1 solution [4 mM MES (pH 5.7), 0.5 M mannitol, 20 mM KCl], and then transferred to multi-well plates for 24 h under gentle agitation (40 rpm) in the dark at 25 °C.

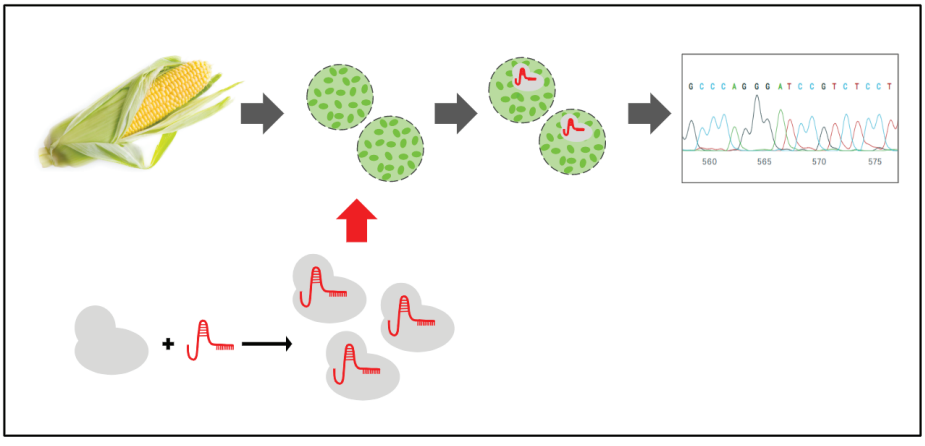


Figure 1. Methodological approach of the present study. Maize seeds were germinated *in vitro*; the second leaves of the seedlings were used to obtain the protoplasts. The protoplasts were exposed to the CRISPR-Cas9 ribonucleoproteins complex (RNP), and after 24 h, the DNA was extracted from the samples. PCR fragments were amplified and sequenced. Maize image was taken from the free repository from pexel.com.

2.4. Gene-Editing Efficiency Analysis by Sanger Sequencing

In order to characterize the spectrum and frequency of DNA changes at the target gene, genomic DNA was isolated using DNeasy® Plant Mini Kit (QIAGEN® Biotecnologia Brasil Ltd., Sao Paulo, SP, Brazil), followed by amplification of the target region by PCR using Taq Q5 High-Fidelity DNA Polymerase (NEB®, Ipswich, MA, USA) and primers listed in Table 1. PCR samples were purified and sequenced using the BigDye Terminator 3.1v Kit (ThermoFisher Scientific, Sao Paulo, SP, Brazil). Samples were resuspended in formamide, denatured at 95 °C for 5 min, and incubated on ice for 3 min. Sequencing was performed using the Sanger [22] automated sequencer from the 3500xL Dx Genetic Analyzer for Sequencing (Applied Biosystems™-ThermoFisher Scientific, Sao Paulo, SP, Brazil).

CRISPR-Cas9 DNA changes were calculated based on the insertions and deletions (In/Dels) around the cleavage site (3 bp upstream of the PAM sequence) using the Inference of CRISPR Editing Software—ICE software v.2 (Synthego Corporation, Palo Alto, CA, USA). It has been previously shown that the ICE software results are comparable to the Next Generation Sequencing (NGS) results [23].

3. Results

3.1. In Vitro Cleavage Assay

Cleavage activity of gRNAs 1 to 5 was tested using 0.5 µg of crRNA and 1.5 µg of Cas9 enzyme to 300 ng of DNA. While all the designed gRNAs were able to cleave PCR products of the IPK gene in our study, the different gRNA sequences varied in their cleavage efficiency (Figure 2). gRNA2, 3, and 4 showed the highest activity and were, therefore, chosen for subsequent experiments on the transfection of maize protoplasts.

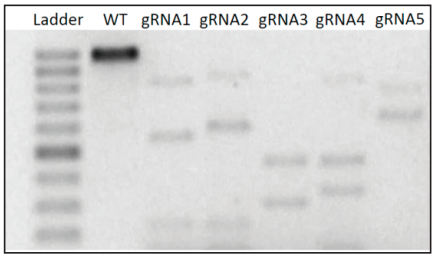


Figure 2. Schematic diagram of the Maize IPK gene locus with the guide RNAs (gRNAs) target sites. The in vitro CRISPR-Cas9 assay showing the original and the cleaved fragments of the IPK gene in maize that were submitted to the RNP complex with the crRNA1, 2, 3, 4, and 5. Note: WT = Wild Type (control).

3.2. Targeted Mutagenesis in Maize Using CRISPR-Cas9 Ribonucleoproteins

Protoplasts viability was checked before and 1 h after transformation. The viability of protoplasts at the isolation step was on average 75% and on average more than 65% after transformation. Therefore, these protoplasts were considered stable after transformation, which would allow further cell culture propagation.

Frequently, results indicating low efficiency of CRISPR-Cas9 editing using RNPs delivery cannot discriminate low transfection rates from poor DNA cleavage and repair activity. In order to overcome this limitation and confirm the internalization of Cas9-gRNA RNPs, we performed a fluorescent microscopy assay. Labeled tracrRNA molecules confirmed the internalization of the RNP complex (Figure 3). Although this is not a quantitative method, it showed that at least one-third of the labeled molecules were internalized.

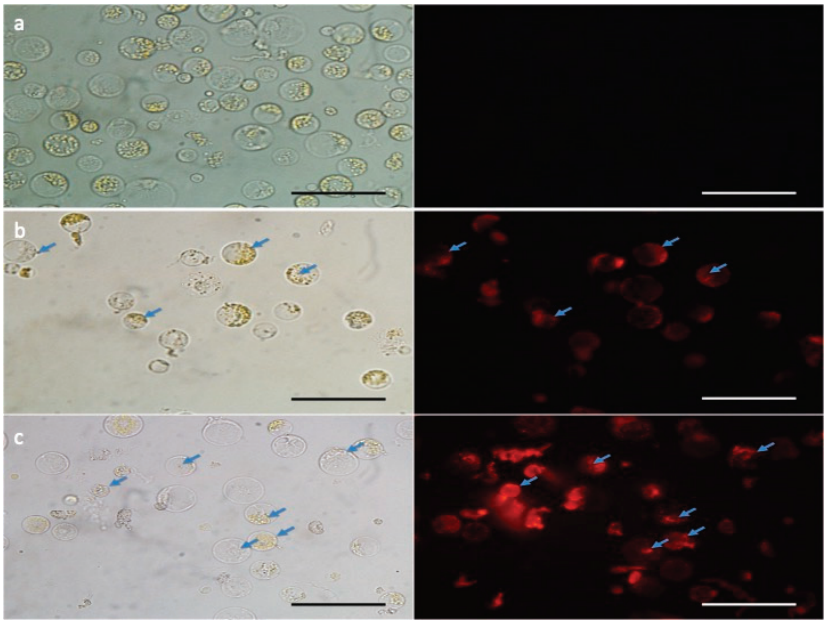


Figure 3. Microscope photographs of maize protoplast cells transfected with polyethylene glycol (PEG)-mediated CRSIPR-Cas9 RNP. The RNP containing the ATTO550-conjugated trans-activating crispr RNA (tracrRNA) was transfected into maize protoplasts and cells were monitored under fluorescent microscopy to check the transfection efficiency. Comparing to the cells without CRISPR-Cas9 delivery (**a**—control sample), the red fluorescent signal was detected in CRISPR-Cas9 treated samples (**b** and **c**—right). Photographs were taken with white light (left side) and fluorescent light (right side). Blue arrows indicate the internalization of the RNP complex. Scale bars: 100 μm.

RNPs containing gRNAs 2, 3, and 4 were transfected into isolated protoplasts with PEG 4000 and the results are displayed as the percentage of indels detected at the cleavage site based on Sanger sequencing and analysis with the ICE software v.2 (Figure 4). The concentration of 45 ug of Cas9 and 15 ug of gRNA to 100 ul protoplasts in a 3:1 ratio resulted in the best cost-efficiency correlation. This result is also in agreement with previous reports on RNPs delivery into protoplasts, which ranged from 30–60 ug of Cas9 in a 1:1 and 3:1 ratio [12,13,24]. Exposure time was tested for all three gRNA sequences in 20- and 40-min of exposure. While a longer exposure time to the RNP complex led to a higher mutation index for all gRNAs tested, the increase in mutation rate was not consistent among all gRNA sequences (approx. seven-fold, three-fold, and one-fold for the gRNAs 2, 3, and 4, respectively). Deletions were shown more frequent than insertions in this model system. A higher insertion rate was only observed for gRNA 2 at 40 min time of exposure.

DNA sequences from the universal primer used to amplify the IPK gene in maize (876 bp) from treated samples were compared to the same fragment in negative controls (no RNP delivered). Other negative controls (Cas9 or gRNAs delivered alone) were also tested against the first negative control and showed no DNA changes. The obtained gene-edited sequences for each gRNA at 40 min of exposure are displayed in Figure 5. About six sequence variants were of major contribution in gene-editing efficiency for the three selected gRNA sequences.

Results on the percentage of mutated sequences (technology efficiency), the size of the DNA change (number of base pair change), the type of DNA change (deletions or insertions), and a theoretical knockout (KO) score are summarized in Table 2. The indel percentage at 20 min of exposure was, on average, 1.63% in contrast to 4.37% at 40 min-exposure time. Overall, gRNA 4 was most efficient and

consistent at both exposure time. Intriguingly, gRNA 2 showed the lowest efficiency at 20 min (0.85%) but the highest efficiency at 40 min (5.85%). The Knockout Score accounts for the reads containing an amino acid frameshift change or 21 + bp indel. Thus, indicating the contributing indels that are likely to result in a functional knockout of the targeted gene. In the present study, the KO score was, on average, 0.83% for the 20 min and 3.17% for the 40 min of exposure treatment, which suggests that the majority of indels were frameshift modifications. In addition, only one gRNA at one time point presented a single base pair change as the most frequent mutation (gRNA 2 with a −1 bp). Notably, a deletion of 19 bp was the most abundant DNA change for gRNA 3 (0.9%). Moreover, all other gRNAs and exposure times showed a 2 bp deletion as the most frequent DNA change. Overall, the DNA change ranged from −28 bp to +12 bp change.

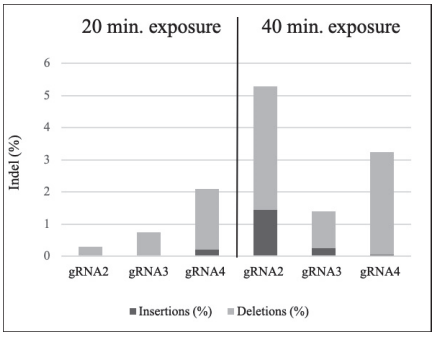


Figure 4. Frequency of mutations in maize protoplasts generated by CRISPR-Cas9 and measured by the Inference of CRISPR Editing Software—ICE software v.2. Different gRNAs and exposure time of the protoplasts to the RNP complex are represented. Percentages of insertions and deletions are represented in dark grey and light grey, respectively.

Table 2. Mutation rates in *Zea mays* IPK gene target region produced by CRISPR-Cas RNP delivery and analyzed by Sanger sequencing and ICE software analysis.

Sample	Incubation Time (min)	% of Indel	Model Fit (R^2)	KO Score	Mutation Range (bp)	Greater Contribution (bp)
Cas9 only	20	0		0	0	0
gRNA2 only	20	0		0	0	0
gRNA3 only	20	0		0	0	0
gRNA4 only	20	0		0	0	0
Cas9 + gRNA2 rep1	20	0	0.99	0	0	0
Cas9 + gRNA2 rep2	20	1	1	1	−4 to −2	−2
Cas9 + gRNA3 rep1	20	0	1	0	−7 to −1	−2
Cas9 + gRNA3 rep2	20	1	1	1	−7 to−1	−2
Cas9 + gRNA4 rep1	20	3	0.99	2	−7 to +12	−2
Cas9 + gRNA4 rep2	20	1	0.99	1	−2 to +3	−2
Cas9 + gRNA2 rep1	40	4	0.99	4	−7 to +1	−1
Cas9 + gRNA2 rep2	40	6	1	6	-8 to +1	+1
Cas9 + gRNA3 rep1	40	1	1	1	−19 to−1	−1
Cas9 + gRNA3 rep2	40	2	0.96	2	−28 to +8	−19
Cas9 + gRNA4 rep1	40	2	0.99	2	−3 to −2	−2
Cas9 + gRNA4 rep2	40	5	1	4	−16 to +12	−2

Note: KO = Knockout.

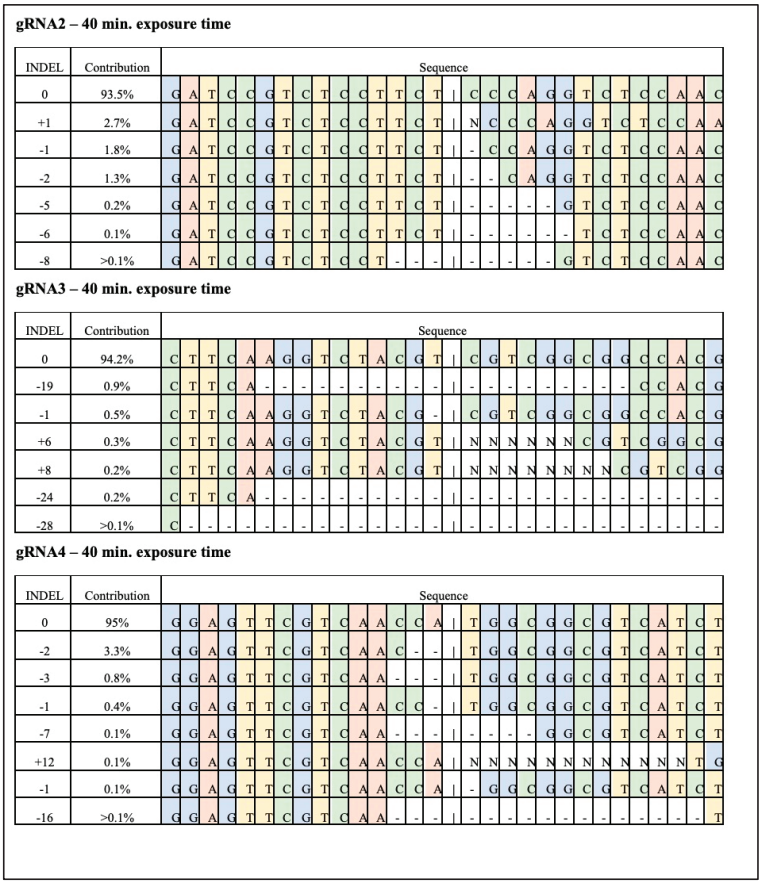


Figure 5. Sequence distribution of most efficient mutations identified with the ICE software around the IPK gene target site in *Zea mays*. Edited sequences were obtained after CRISPR-Cas9 RNP transfection to maize protoplasts. Forty-five μg of Cas9 preassembled with 15 μg of each gRNA were used in the protoplast transformation. Different exposure times of the RNP complex with the protoplast are presented. Cut sites are represented by black vertical dotted lines, insertions are represented by 'N' and deletions by black horizontal dotted lines.

4. Discussion

CRISPR-Cas9 technology is a powerful tool for plant breeding and research. While still evolving as a technology to determine the rules for gRNA design and the algorithms to predict target and 'off-target' sequences, CRISPR applications still rely on empirical results to test the performance of new systems [25]. Notably, gene-editing results outcomes are frequently species-dependent [26]. Therefore, a CRISPR platform for different species with a rapid and efficient evaluation protocol is needed before commercialization.

Our experiments demonstrated the suitability of the PEG-delivered CRISPR-Cas9 RNPs system for gene-editing screening in maize. We showed that high-efficiency gene-edited maize cells could be obtained using less time-consuming (15 days) and labor-intensive procedures (PCR, agarose gel electrophoresis, and Sanger sequencing). In addition, the advantage of our system in relation to the use of vectors is that it prevents the integration and expression of exogenous DNA sequences, isolating

the effect of gene-editing modification and avoiding transgene-introgressed side effects. Although non-integrating plasmids could be transfected into plant cells to deliver programmable nucleases, transfected plasmids are degraded in cells by endogenous nucleases, and the resulting small DNA fragments can be inserted at both on-target and off-target sites in host cells (Kim et al. 2014) [27]. For example, Braatz et al. [28] performed whole-genome sequencing after the transfection of the expression construct CRISPR-Cas9 in *Brassica napus* and found that the transformation resulted in at least five independent insertions of the vector backbone sequences in the plant genome.

4.1. Ribonucleoprotein Delivery in Plants

In the present report, we show a positive correlation between the time of exposure to RNPs and the efficiency of site-directed mutagenesis in maize, as ascertained with Sanger sequencing. In previous reports referring to the use of RNPs in plant protoplasts (Table 3), the authors used one or more different concentrations of RNPs and Cas9:gRNA ratios, but the effect of exposure time on mutation frequencies was not tested [29,30]. CRISPR RNPs were delivered to apple, grapevine, brassica sp., lettuce, tobacco, and rice plants at less or equal than 20 min exposure time and their efficiencies ranged between 0.1 and 40% [12,13,31]. In petunia and wheat protoplasts exposure for 30 min granted 0.2 up to 45% efficiency [14,17]; thus, suggesting that time of exposure might not alone explain indel frequency in different plant systems. In our system, when all other factors are maintained, exposure time consistently increased indel frequency for all three gRNA sequences tested (up to approx. seven-fold change increase).

The Cas9:gRNA ratio also influences target efficiency in a species-specific manner. Three different Cas9:gRNA ratios were tested in apple and grapevine protoplasts [13]. While the 1:1, 1:3, and 3:1 ratios did not differ in mutation frequency for grapevine (0.1%), the 1:1 and 3:1 ratios increased indel frequency in two (6.6 and 2.6 fold change respectively) out of three gRNA sequences in apple. Overall, the results obtained for the 3:1 ratio are equivalent to our results applying the same ratio (from 3.3% up to 6.7% efficiency). Cas9 concentration has been shown to be of major factor influencing the delivery of RNPs to plant cells. Woo et al. [12] tested 20 and 60 ug of Cas9 to *Arabidopsis* protoplasts and found that the editing efficiency was not directly related to the Cas9 concentration but also dependent on the time course of the analysis. At 24 h after delivery, more efficiency was observed when applying 20 instead of 60 ug of Cas9 (71 in contrast to 54%). Opposite results were obtained at 72 h after delivery. Nevertheless, increasing the amount of Cas9 (7.5, 15. 30, and 60 ug) was consistent with a crescent indel frequency in brassica sp. protoplasts [31]. The efficiency results obtained in our study were similar to those obtained applying approximately 60–90 ug of Cas9, thus, indicating that a lower amount of Cas9 (45 ug) but a higher exposure time (40 min) might have similar cleavage levels. Overall, it is clear that the limited amount of studies investigating the RNP delivery into plant cells is insufficient to draw definitive conclusions for increasing gene-editing efficiency using this system.

Table 3. Publications with DNA-free gene editing in plants using CRISPR-Cas9 RNPs and other delivery methods for maize.

Reference	Plant Species	Plant Material	Transfection Method	Gene-Editing Efficiency
		RNP delivered in plants		
[12]	*Arabidopsis thaliana, Lactuca sativa, Nicotiana attenuata, Oryza sativa*	Protoplasts	PEG-mediated	5.7–40.0%
[13]	*Malus domestica, Vitis vinifera*	Protoplasts	PEG-mediated	0.1–6.9%
[14]	*Petunia x hybrid*	Protoplasts	PEG-mediated	2.4–21.0%
[31]	*Triticum aestivum*	Protoplasts, immature embryos	PEG-mediated, Biolistics	0.2–45.3%
[15]	*Solanum tuberosum*	Protoplasts	PEG-mediated	1.0–25.0%
[32]	*Brassica oleracea, Brassica rapa*	Protoplasts	PEG-mediated	0.1–24.5%
[33]	*Oryza sativa*	Zygotes	PEG-mediated	14.0–64.0%

Table 3. *Cont.*

Reference	Plant Species	Plant Material	Transfection Method	Gene-Editing Efficiency
Other delivery methods in maize				
[34]	*Zea mays*	Protoplasts	Vector via PEG-mediated	13.1%
		Immature embryos	*Agrobacterium*-mediated	16.4–19.1%
[35]	*Zea mays*	Protoplasts	Vector via PEG-mediated	N.A
[24]	*Zea mays*	Immature embryos	Vector via biolistics	1.3–4.6%
[17]	*Zea mays*	Protoplasts	Vector via biolistics	80.0–90.0%
[36]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	57.1–71.4%
[19]	*Zea mays*	Protoplasts	Vector via PEG-mediated	2.8–27.0%
		Immature embryos	*Agrobacterium*-mediated	19.0–31.0%
[37]	*Zea mays*	Protoplasts	Vector via PEG-mediated	4.0–11.9%
		Immature embryos	*Agrobacterium*-mediated	65.8–86.9%
[16]	*Zea mays*	Immature embryos	RNP via Biolistics	0.01–0.7%
[18]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	12.0–74.0%
[38]	*Zea mays*	Immature embryos	Vector via Biolistics	60.0–98.0%
[39]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[19]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	5.0–100%
[20]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[40]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	90.0–100%
[41]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[42]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[43]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[44]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[45]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
[46]	*Zea mays*	Immature embryos	Vector via biolistics	N.A
[47]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	25.0–100%
[48]	*Zea mays*	Immature embryos	*Agrobacterium*-mediated	N.A
Present study	*Zea mays*	Protoplasts	RNA via PEG-mediated	0.85–5.85%

Note: N.A is 'not applicable' because transgenic plants were either selected using antibiotic marker genes or the analysis was not performed.

4.2. CRISPR Delivery Methods in Maize

Other delivery methods have been tested for maize as screening methods or gene-editing breeding methods, and these include PEG-mediated vector transfection, *Agrobacterium*-mediated, and biolistics (Table 3). The vast majority of studies still rely on vector-based transformation delivery of CRISPR. None of the listed studies have provided a cell-based screening method without the insertion of foreign vector-based DNA. On average, PEG-mediated vector transfection reached an average of about 12% efficiency, whereas *Agrobacterium* and biolistics reached 44% and 83%, respectively. In addition, most of these methods used embryogenic callus as explant material. Callus-based methods harbor chimeric tissues, thus, requiring subsequent genetic fixation to allow the stable inheritance of the edited traits. Therefore, these are not suitable material for genetic screening of successful gene-edited plants. Currently, many protocols are available for regeneration of whole plants from protoplasts. These include lettuce, tobacco, and rice, petunia, wheat, apple, and soybean. It is also suggested as a future choice for gene-edited maize and the list seems to expand because of the capabilities of the RNP technology [16,49]. The trade-off in efficiency percentage plays in return for avoiding unintended DNA integration and the potential undesirable biosafety risks.

4.3. Analytical Platforms for Gene-Editing Detection

Different analytical platforms for the detection and identification of CRISPR outcomes are reflected in frequency results as they show different analytical sensitivity. Woo et al. [12] showed a 40% transformation efficiency in tobacco when samples were analyzed by the Illumina sequencing platform. In contrast, the same samples showed a much lower efficiency rate (17%) when analyzed by the T7 cleavage assay. While T7 cleavage assays are nowadays being limited to a qualitative rather than quantitative detection methods, high-throughput sequencing platforms are time-consuming and costly

options for screening protocols. Usually, such platforms are available in other labs or through service providers, which require long processing time and high costs for samples and assays that are still at the screening stage. In order to overcome such problems, we have proposed a model that analyzes gene-edited cell pools using data from common Sanger sequencing analysis. More specifically, we use the ICE: *Inference of CRISPR Edits* software, which enable the analysis of mixed populations and strongly correlates with next-generation sequencing of amplicons using Sanger sequencing data [23].

However, the proposed model is focused on providing a simple, cost-efficient analysis of gene-editing outcomes at the screening stage. The model is currently limited to detect mutations with more than 30 bp deletions or more than 14 bp additions. The analysis also does not account for very small mutant populations (<0.1%) neither presents mutations with a substitution of bp [23]. Although our platform does not apply exogenous DNA generating potential integration, it cannot be ruled out that microhomologies with gRNA sequences produce spurious cleavage or larges genomic rearrangements [50]. Therefore, long-run high throughput sequencing analysis is recommended for follow up breeding programs and safety tests [51].

5. Conclusions

We have shown that RNPs could be used for targeted CRISPR-Cas9 via PEG delivery as a model system to screen for gene-editing outcomes in maize. Target insertion and deletion DNA changes (In/Dels) were induced using 45 ug of Cas9 and gRNA used at a 3:1 ratio, and a positive correlation between exposure time (20 and 40 min) and indel frequency was observed. By targeting preserved coding regions, we can anticipate that the model could be applied to several maize varieties by validation using IPK gene. However, in vivo In/Del frequencies differed among gRNA sequences. In addition, the proposed method for sequencing analysis is also restricted to a window of 30 bp deletions and 14 bp additions. Further studies will be focused on CRISPR-Cas9 off-target activity and on the regeneration of edited protoplasts.

Author Contributions: Conceptualization, R.O.N. and S.Z.A.-T.; Methodology, R.R.A.S., C.A.C. and S.Z.A.-T.; Software, R.R.A.S.; Validation, R.R.A.S. and C.A.C.; Formal Analysis, R.R.A.S., C.A.C. R.O.N. and S.Z.A.-T.; Investigation, R.R.A.S., C.A.C. and S.Z.A.-T.; Resources, R.O.N. and S.Z.A.-T.; Data Curation, R.R.A.S. and S.Z.A.-T.; Writing—Original Draft Preparation, R.R.A.S. and S.Z.A.; Writing—Review & Editing, R.R.A.S., C.A.C. R.O.N. and S.Z.A.-T.; Visualization, R.R.A.S. and S.Z.A.; Supervision, R.O.N. and S.Z.A.-T.; Project Administration, S.Z.A.-T.; Funding Acquisition, S.Z.A.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by GenØk Centre for Biosafety through its funds from The Ministry of Climate and Environment in Norway. Scholarship from National Council for Scientific and Technological Development (CNPq) was provided to R.R.A.S. and R.O.N.

Acknowledgments: We would like to thank Lilian Machado (UFSC) for taking care of the Sanger sequencing. The authors are also grateful to Vinicius Vilperte for insightful comments and suggestions on the sequencing analysis and also to Miguel Guerra for his support in using his laboratory. This work was financed by GenØk Centre for Biosafety through its funds from The Ministry of Climate and Environment in Norway, whose support is gratefully acknowledged.

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

References

1. 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] [PubMed]
2. Hsu, P.D.; Scott, D.; 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] [PubMed]
3. Bortesi, L.; Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **2015**, *33*, 41–52. [CrossRef] [PubMed]
4. Grobman, A.; Bonavia, D. *Origin, Domestication, and Evolution of Maize*; Cambridge University Press (CUP): Cambridge, UK, 2013; pp. 329–486.

5. Li, H.; Wang, M.; Li, W.; He, L.; Zhou, Y.; Zhu, J.; Che, R.; Warburton, M.L.; Yang, X.; Yan, J. Genetic variants and underlying mechanisms influencing variance heterogeneity in maize. *Plant J.* **2020**, *103*, 1089–1102. [CrossRef] [PubMed]
6. Barrangou, R.; Doudna, J. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **2016**, *34*, 933–941. [CrossRef]
7. 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*, e100448. [CrossRef]
8. Metje-Sprink, J.; Menz, J.; Modrzejewski, D.; Sprink, T. DNA-free genome editing: Past, present and future. *Front. Plant Sci.* **2019**, *9*. [CrossRef]
9. Liu, Q.; Yuan, Y.; Zhu, F.; Hong, Y.; Ge, R. Efficient genome editing using CRISPR/Cas9 ribonucleoprotein approach in cultured Medaka fish cells. *Biol. Open* **2018**, *7*, bio035170. [CrossRef]
10. Kanchiswamy, C.N. DNA-free genome editing methods for targeted crop improvement. *Plant Cell Rep.* **2016**, *35*, 1469–1474. [CrossRef]
11. Agarwal, A.; Yadava, P.; Kumar, K.; Singh, I.; Kaul, T.; Pattanayak, A.; Agrawal, P.K. Insights into maize genome editing via CRISPR/Cas9. *Physiol. Mol. Biol. Plants* **2018**, *24*, 175–183. [CrossRef]
12. Woo, J.W.; Kim, J.; Kwon, S.I.; Corvalan, 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] [PubMed]
13. Malnoy, M.; Viola, R.; Jung, M.-H.; Koo, O.-J.; Kim, S.; Kim, J.-S.; Velasco, R.; Kanchiswamy, C.N. DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 Ribonucleoproteins. *Front. Plant Sci.* **2016**, *7*. [CrossRef] [PubMed]
14. Subburaj, S.; Chung, S.J.; Lee, C.; Ryu, S.-M.; Kim, D.H.; Kim, J.-S.; Bae, S.; Lee, G. Site-directed mutagenesis in Petunia × hybrida protoplast system using direct delivery of purified recombinant Cas9 ribonucleoproteins. *Plant Cell Rep.* **2016**, *35*, 1535–1544. [CrossRef] [PubMed]
15. Andersson, M.; Turesson, H.; Olsson, N.; Fält, A.-S.; Ohlsson, P.; González, M.N.; Samuelsson, M.; Hofvander, P.; Olsson, P. Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. *Physiol. Plant.* **2018**, *164*, 378–384. [CrossRef]
16. Svitashev, S.; Schwartz, C.; Lenderts, B.; Young, J.K.; Cigan, A.M. Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nat. Commun.* **2016**, *7*, 13274. [CrossRef]
17. Liang, Z.; Zong, Y.; Gao, C. An efficient targeted mutagenesis system using CRISPR/Cas in monocotyledons. *Curr. Protoc. Plant Biol.* **2016**, *1*, 329–344. [CrossRef]
18. Char, S.N.; Neelakandan, A.; 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.* **2016**, *15*, 257–268. [CrossRef]
19. Feng, C.; Su, H.; Bai, H.; Wang, R.; Liu, Y.; Guo, X.; Liu, C.; Zhang, J.; Yuan, J.; Birchler, J.A.; et al. High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnol. J.* **2018**, *16*, 1848–1857. [CrossRef]
20. Dong, L.; Qi, X.; Zhu, J.; Liu, C.; Zhang, X.; Cheng, B.; Mao, L.; Xie, C. Supersweet and waxy: Meeting the diverse demands for specialty maize by genome editing. *Plant Biotechnol. J.* **2019**, *17*, 1853–1855. [CrossRef]
21. Sheen, J. Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. *Plant Cell* **1991**, *3*, 225–245.
22. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef] [PubMed]
23. Hsiau, T.; Conant, D.; Rossi, N.; Maures, T.; Waite, K.; Yang, J.; Joshi, S.; Kelso, R.; Holden, K.; Enzmann, B.L.; et al. Inference of crispr edits from sanger trace data. *bioRxiv* **2019**, 251082. [CrossRef]
24. Svitashev, S.; Young, J.K.; Schwartz, C.; Gao, H.; Falco, S.C.; Cigan, A.M. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol.* **2015**, *169*, 931–945. [CrossRef] [PubMed]
25. Lin, C.-S.; Hsu, C.; Yang, L.; Lee, L.; Fu, J.; Cheng, Q.; Wu, F.; 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] [PubMed]

26. Bortesi, L.; Zhu, C.; Zischewski, J.; Perez, L.; Bassié, L.; Nadi, R.; Forni, G.; Lade, S.B.; Soto, E.; Jin, X.; et al. Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol. J.* **2016**, *14*, 2203–2216. [CrossRef]
27. Kim, S.; Kim, D.; Cho, S.W.; Kim, J.; 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]
28. Braatz, J.; Harloff, H.-J.; Mascher, M.; Stein, N.; Himmelbach, A.; Harloff, H.-J. 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]
29. Sandhya, D.; Jogam, P.; Allini, V.R.; Abbagani, S.; Alok, A. The present and potential future methods for delivering CRISPR/Cas9 components in plants. *J. Genet. Eng. Biotechnol.* **2020**, *18*, 1–11. [CrossRef]
30. Wada, N.; Ueta, R.; Osakabe, Y.; Osakabe, K. Precision genome editing in plants: State-of-the-art in CRISPR/Cas9-based genome engineering. *BMC Plant Biol.* **2020**, *20*, 1–12. [CrossRef]
31. 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]
32. Murovec, J.; Guček, K.; Bohanec, B.; Avbelj, M.; Jerala, R. DNA-free genome editing of brassica oleracea and B. rapa protoplasts using CRISPR-Cas9 Ribonucleoprotein complexes. *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
33. Toda, E.; Koiso, N.; Takebayashi, A.; Ichikawa, M.; Kiba, T.; Osakabe, K.; Osakabe, Y.; Sakakibara, H.; Kato, N.; Okamoto, T. An efficient DNA- and selectable-marker-free genome-editing system using zygotes in rice. *Nat. Plants* **2019**, *5*, 363. [CrossRef] [PubMed]
34. Liang, Z.; Zhang, K.; Chen, K.; Gao, C. Targeted Mutagenesis in Zea mays Using TALENs and the CRISPR/Cas System. *J. Genet. Genomics* **2014**, *41*, 63–68. [CrossRef]
35. Xing, H.; Dong, L.; Wang, Z.; Zhang, H.; Han, C.; Liu, B.; Wang, X.; Chen, Q. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **2014**, *14*, 327. [CrossRef]
36. 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]
37. Zhu, J.; Song, N.; Sun, S.; Yang, W.; Zhao, H.; Song, W.; Lai, J. Efficiency and inheritance of targeted mutagenesis in maize using CRISPR-Cas9. *J. Genet. Genomics* **2016**, *43*, 25–36. [CrossRef] [PubMed]
38. 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] [PubMed]
39. Chen, R.; Xu, Q.; Liu, Y.; Zhang, J.; Ren, D.; Wang, G.; Liu, Y. Generation of transgene-free maize male sterile lines using the CRISPR/Cas9 system. *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
40. Lee, K.; Zhang, Y.; Kleinstiver, B.P.; Guo, J.A.; Aryee, J.; Miller, J.; Malzahn, A.; Zarecor, S.; Lawrence-Dill, C.J.; Joung, J.L.; et al. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol. J.* **2019**, *17*, 362–372. [CrossRef]
41. Kelliher, T.; Starr, D.; Su, X.; Tang, G.; Chen, Z.; Carter, J.; Wittich, P.E.; Dong, S.; Green, J.; Burch, E. One-step genome editing of elite crop germplasm during haploid induction. *Nat. Biotechnol.* **2019**, *37*, 287–292. [CrossRef]
42. Doll, N.M.; Gilles, L.M.; Gérentes, M.F.; Richard, C.; Just, J.; Fierlej, Y.; Borrelli, V.G.M.; Gendrot, G.; Ingram, C.G.; Rogowsky, P.M. Single and multiple gene knockouts by CRISPR–Cas9 in maize. *Plant Cell Rep.* **2019**, *38*, 487–501. [CrossRef] [PubMed]
43. Wu, Q.; Xu, F.; Liu, L.; Char, S.N.; Ding, Y.; Je, B.I.; Schmelz, E.; Yang, B.; Jackson, D. The maize heterotrimeric G protein β subunit controls shoot meristem development and immune responses. *Proc. Natl. Acad. Sci. USA* **2019**, *117*, 1799–1805. [CrossRef] [PubMed]
44. Liu, H.; Jian, L.; Xu, J.; Zhang, Q.; Zhang, M.; Jin, M.; Peng, Y.; Yan, J.; Han, B.; Liu, J.; et al. High-throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. *Plant Cell* **2020**, *32*, 1397–1413. [CrossRef] [PubMed]
45. Gao, H.; Gadlage, M.J.; Lafitte, H.R.; Lenderts, B.; Yang, M.; Schroder, M.; Farrell, J.; Snopek, K.; Peterson, D.; Feigenbutz, L.; et al. Superior field performance of waxy corn engineered using CRISPR–Cas9. *Nat. Biotechnol.* **2020**, *38*, 579–581. [CrossRef]

46. Gao, H.; Mutti, J.; Young, J.K.; Yang, M.; Schroder, M.; Lenderts, B.; Wang, L.; Peterson, D.; Clair, G.S.; Jones, S.; et al. Complex trait loci in maize enabled by CRISPR-Cas9 mediated gene insertion. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
47. Barone, P.; Wu, E.; Lenderts, B.; Anand, A.; Gordon-Kamm, W.; Svitashev, S.; Kumar, S. Efficient gene targeting in maize using inducible CRISPR-Cas9 and marker-free donor template. *Mol. Plant* **2020**, *13*. [CrossRef]
48. Zhang, J.; Zhang, X.; Chen, R.; Yang, L.; Fan, K.; Liu, Y.; Wang, G.; Ren, Z.; Liu, Y. Generation of transgene-free Semidwarf maize plants by gene editing of gibberellin-oxidase20-3 using CRISPR/Cas9. *Front. Plant Sci.* **2020**, *11*. [CrossRef]
49. Park, J.; Choe, S. DNA-free genome editing with preassembled CRISPR/Cas9 ribonucleoproteins in plants. *Transgenic Res.* **2019**, *28*, 61–64. [CrossRef]
50. Vu, G.T.H.; Cao, X.H.; Fauser, F.; Reiss, B.; Puchta, H.; Schubert, I. Endogenous sequence patterns predispose the repair modes of CRISPR /Cas9-induced DNA double-stranded breaks in Arabidopsis thaliana. *Plant J.* **2017**, *92*, 57–67. [CrossRef]
51. Agapito-Tenfen, S.Z.; Okoli, A.S.; Bernstein, M.J.; Wikmark, O.-G.; Myhr, A.I. Revisiting risk governance of GM plants: The need to consider new and emerging gene-editing techniques. *Front. Plant Sci.* **2018**, *9*. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Genes Editorial Office
E-mail: genes@mdpi.com
www.mdpi.com/journal/genes

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-0365-0977-8