International Journal of
Molecular Sciences

Special Issue Reprint

Molecular World Today and Tomorrow

Recent Trends in Biological Sciences 2.0

Edited by
Wajid Zaman

mdpi.com/journal/ijms

MDPI

Molecular World Today and Tomorrow: Recent Trends in Biological Sciences 2.0

Molecular World Today and Tomorrow: Recent Trends in Biological Sciences 2.0

Editor

Wajid Zaman

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editor
Wajid Zaman
Department of Life Sciences
Yeungnam University
Gyeongsan
Korea, South

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

This is a reprint of articles from the Special Issue published online in the open access journal *International Journal of Molecular Sciences* (ISSN 1422-0067) (available at: https://www.mdpi.com/journal/ijms/special_issues/M171QV56L4).

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

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

ISBN 978-3-7258-0600-3 (Hbk)
ISBN 978-3-7258-0599-0 (PDF)
doi.org/10.3390/books978-3-7258-0599-0

Contents

International Journal of
Molecular Sciences

MDPI

Editorial

Molecular World Today and Tomorrow: Recent Trends in Biological Sciences 2.0

Wajid Zaman

Department of Life Sciences, Yeungnam University, Gyeongsan 38541, Republic of Korea; wajidzaman@yu.ac.kr

Molecular techniques have become influential instruments in biological study, transforming our comprehension of life at the cellular and genetic levels [1]. The Special Issue "Molecular World Today and Tomorrow: Recent Trends in Biological Sciences 2.0" examines current trends in biological sciences and the impact of advanced techniques on scientific advancements. The problem highlights how molecular techniques can greatly aid in the advancement of focused medication creation and disease diagnostics. Combining molecular phylogenetic and omics approaches with expression and pathway analysis, we seek to improve our comprehension of molecular mechanisms and stress responses in biological systems [2]. This collaboration represents significant progress in genomics, proteomics, and phylogenetic methods, providing new opportunities for study and innovation.

Advancements in omics technologies, including transcriptomics, proteomics, and genomics, have significantly improved our capacity to recognize and describe biological entities [2,3]. Researchers currently use systematic methods to uncover genetic variations, forecast ecological habitats, and clarify taxonomic intricacies [4,5]. Omics data offer a thorough perspective on the molecular world, ranging from analyzing microbial populations to uncovering the genetic foundations of diseases. Integrating molecular phylogenetics with omics techniques enhances our comprehension of evolutionary connections. Researchers investigate molecular pathways across many taxonomic groups by integrating sequence data with expression profiles [6,7]. This synergy influences both evolutionary history and functional adaptations, as well as stress responses.

Molecular data drives precision medicine [8,9]. Genomics and proteomics play a crucial role in guiding individualized treatments by identifying therapeutic targets and predicting medication responses [10,11]. Furthermore, diagnostic tests utilizing molecular markers improve the diagnosis and prediction of diseases. As we decode the complexities of individual genomes, the future shows potential for personalized interventions. The study in this Special Issue highlights the significance of interdisciplinary approaches in addressing intricate biological issues through collaboration. It unites specialists from several disciplines, such as molecular biology, genetics, biochemistry, biophysics, and computational biology, to promote a comprehensive comprehension of life at the molecular scale. This collaborative effort is crucial for the integration of knowledge and the acceleration of scientific discoveries (Figure 1).

The edition features 20 papers, including 3 comprehensive review articles, a brief report, 1 communication, and 15 research articles. The articles in the Special Issue emphasize the importance of molecular data in promoting targeted medicines, highlighting the vital significance of molecular sciences in medicine. This interdisciplinary method highlights the significant impact of molecular approaches in research and clinical environments, presenting new opportunities for customized medicine and treatments.

In conclusions, this Special Issue demonstrates the dynamic and interdisciplinary nature of molecular sciences. It motivates researchers to expand the current knowledge and investigate new areas, assuring the ongoing development of the biological sciences field and its contribution to our understanding of life's molecular foundation. The compilation

Citation: Zaman, W. Molecular World Today and Tomorrow: Recent Trends in Biological Sciences 2.0. *Int. J. Mol. Sci.* **2024**, *25*, 3070. https://doi.org/10.3390/ijms25053070

Received: 19 February 2024
Accepted: 28 February 2024
Published: 6 March 2024

showcases the present status of molecular sciences and paves the way for future progress, encouraging further inquiry and invention in this captivating discipline.

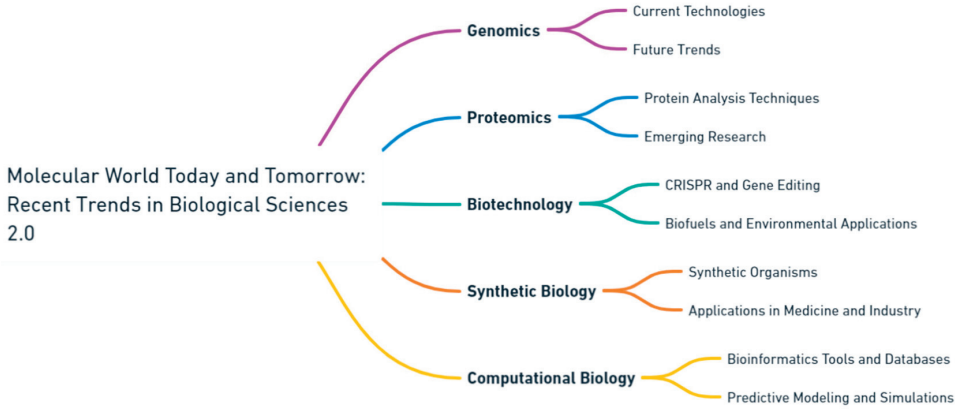


Figure 1. Navigating the frontier: current and future perspectives in biological sciences (generated by OpenAI's Whimsical Diagrams GPT).

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

List of Contributions:

1. Arai, I.; Tsuji, M.; Takahashi, K.; Saito, S.; Takeda, H. Analyzing the Antinociceptive Effect of Interleukin-31 in Mice. *Int. J. Mol. Sci.* **2023**, *24*, 11563. https://doi.org/10.3390/ijms241411563.
2. Arora, D.; Hernandez, A.G.; Walden, K.K.O.; Fields, C.J.; Yan, G. First Draft Genome Assembly of Root-Lesion Nematode Pratylenchus scribneri Generated Using Long-Read Sequencing. *Int. J. Mol. Sci.* **2023**, *24*, 7311. https://doi.org/10.3390/ijms24087311.
3. Arshad, S.; Wei, M.; Ali, Q.; Mustafa, G.; Ma, Z.; Yan, Y. Paclitaxel and Caffeine–Taurine, New Colchicine Alternatives for Chromosomes Doubling in Maize Haploid Breeding. *Int. J. Mol. Sci.* **2023**, *24*, 14659. https://doi.org/10.3390/ijms241914659.
4. Barreda-Castillo, J.M.; Monribot-Villanueva, J.L.; Velázquez-Rosas, N.; Bayman, P.; Guerrero-Analco, J.A.; Menchaca-García, R.A. Morphological and Physio-Chemical Responses to PEG-Induced Water Stress in *Vanilla planifolia* and *V. pompona* Hybrids. *Int. J. Mol. Sci.* **2023**, *24*, 4690. https://doi.org/10.3390/ijms24054690.
5. Bhujel, B.; Oh, S.-H.; Kim, C.-M.; Yoon, Y.-J.; Kim, Y.-J.; Chung, H.-S.; Ye, E.-A.; Lee, H.; Kim, J.-Y. Mesenchymal Stem Cells and Exosomes: A Novel Therapeutic Approach for Corneal Diseases. *Int. J. Mol. Sci.* **2023**, *24*, 10917. https://doi.org/10.3390/ijms241310917.
6. Costantini, M.; Esposito, R.; Ruocco, N.; Caramiello, D.; Cordella, A.; Ventola, G.M.; Zupo, V. De Novo Assembly of the Genome of the Sea Urchin Paracentrotus lividus (Lamarck 1816). *Int. J. Mol. Sci.* **2024**, *25*, 1685. https://doi.org/10.3390/ijms25031685.
7. Dzidek, A.; Czerwińska-Ledwig, O.; Żychowska, M.; Pilch, W.; Piotrowska, A. The Role of Increased Expression of Sirtuin 6 in the Prevention of Premature Aging Pathomechanisms. *Int. J. Mol. Sci.* **2023**, *24*, 9655. https://doi.org/10.3390/ijms24119655.
8. Félix, J.W.; Granados-Alegría, M.I.; Gómez-Tah, R.; Tzec-Simá, M.; Ruíz-May, E.; Canto-Canché, B.; Zamora-Briseño, J.A.; Bojórquez-Velázquez, E.; Oropeza-Salín, C.; Islas-Flores, I. Proteome Landscape during Ripening of Solid Endosperm from Two Different Coconut Cultivars Reveals Contrasting Carbohydrate and Fatty Acid Metabolic Pathway Modulation. *Int. J. Mol. Sci.* **2023**, *24*, 10431. https://doi.org/10.3390/ijms241310431.
9. He, R.; Liu, K.; Zhang, S.; Ju, J.; Hu, Y.; Li, Y.; Liu, X.; Liu, H. Omics Analysis Unveils the Pathway Involved in the Anthocyanin Biosynthesis in Tomato Seedling and Fruits. *Int. J. Mol. Sci.* **2023**, *24*, 8690. https://doi.org/10.3390/ijms24108690.
10. Jia, P.; Liu, J.; Yan, R.; Yang, K.; Dong, Q.; Luan, H.; Zhang, X.; Li, H.; Guo, S.; Qi, G. Systematical Characterization of the AT-Hook Gene Family in *Juglans regia* L. and the Functional Analysis of the JrAHL2 in Flower Induction and Hypocotyl Elongation. *Int. J. Mol. Sci.* **2023**, *24*, 7244. https://doi.org/10.3390/ijms24087244.
11. Kim, H.-J.; Kim, K.E.; Kim, Y.J.; Kang, H.; Shin, J.W.; Kim, S.; Lee, S.H.; Jung, S.W.; Lee, T.-K. Marine Bacterioplankton Community Dynamics and Potentially Pathogenic Bacteria in Seawater around Jeju Island, South Korea, via Metabarcoding. *Int. J. Mol. Sci.* **2023**, *24*, 13561. https://doi.org/10.3390/ijms241713561.

12. Kim, Y.; Wang, J.; Ma, C.; Jong, C.; Jin, M.; Cha, J.; Wang, J.; Peng, Y.; Ni, H.; Li, H.; et al. GmTCP and GmNLP Underlying Nodulation Character in Soybean Depending on Nitrogen. *Int. J. Mol. Sci.* **2023**, *24*, 7750. https://doi.org/10.3390/ijms24097750.
13. Kowalski, T.W.; Feira, M.F.; Lord, V.O.; Gomes, J.D.; Giudicelli, G.C.; Fraga, L.R.; Sanseverino, M.T.; Recamonde-Mendoza, M.; Schuler-Faccini, L.; Vianna, F.S. A New Strategy for the Old Challenge of Thalidomide: Systems Biology Prioritization of Potential Immunomodulatory Drug (IMiD)-Targeted Transcription Factors. *Int. J. Mol. Sci.* **2023**, *24*, 11515. https://doi.org/10.3390/ijms241411515.
14. Rocchetti, M.T.; Bellanti, F.; Zadorozhna, M.; Fiocco, D.; Mangieri, D. Multi-Faceted Role of Luteolin in Cancer Metastasis: EMT, Angiogenesis, ECM Degradation and Apoptosis. *Int. J. Mol. Sci.* **2023**, *24*, 8824. https://doi.org/10.3390/ijms24108824.
15. Shi, Y.; Yu, B.; Cheng, S.; Hu, W.; Liu, F. The Change in Whole-Genome Methylation and Transcriptome Profile under Autophagy Defect and Nitrogen Starvation. *Int. J. Mol. Sci.* **2023**, *24*, 14047. https://doi.org/10.3390/ijms241814047.
16. Su, H.-H.; Huang, Y.-H.; Lien, Y.; Yang, P.-C.; Huang, C.-Y. Crystal Structure of DNA Replication Protein SsbA Complexed with the Anticancer Drug 5-Fluorouracil. *Int. J. Mol. Sci.* **2023**, *24*, 14899. https://doi.org/10.3390/ijms241914899.
17. Vásquez, C.; Leyton-Carcaman, B.; Cid-Alda, F.P.; Segovia, I.; Pinto, F.; Abanto, M. Physical Pretreatments Applied in Three Commercial Kits for the Extraction of High-Quality DNA from Activated Sewage Sludge. *Int. J. Mol. Sci.* **2023**, *24*, 15243. https://doi.org/10.3390/ijms242015243.
18. Wang, Z.; Yadav, V.; Chen, X.; Zhang, S.; Yuan, X.; Li, H.; Ma, J.; Zhang, Y.; Yang, J.; Zhang, X.; et al. Multi-Omics Analysis Reveals Intricate Gene Networks Involved in Female Development in Melon. *Int. J. Mol. Sci.* **2023**, *24*, 16905. https://doi.org/10.3390/ijms242316905.
19. Zhang, J.; Zhang, Y.; Feng, C. Genome-Wide Analysis of MYB Genes in Primulina eburnea (Hance) and Identification of Members in Response to Drought Stress. *Int. J. Mol. Sci.* **2024**, *25*, 465. https://doi.org/10.3390/ijms25010465.
20. Zheng, S.; Shin, K.; Lin, W.; Wang, W.; Yang, X. Identification and Characterization of PRE Genes in Moso Bamboo (Phyllostachys edulis). *Int. J. Mol. Sci.* **2023**, *24*, 6886. https://doi.org/10.3390/ijms24086886.

References

1. Hartwell, L.H.; Hopfield, J.J.; Leibler, S.; Murray, A.W. From molecular to modular cell biology. *Nature* **1999**, *402*, C47–C52. [CrossRef] [PubMed]
2. Misra, B.B.; Langefeld, C.; Olivier, M.; Cox, L.A. Integrated omics: Tools, advances and future approaches. *J. Mol. Endocrinol.* **2019**, *62*, R21–R45. [CrossRef] [PubMed]
3. Veenstra, T.D. Omics in systems biology: Current progress and future outlook. *Proteomics* **2021**, *21*, 2000235. [CrossRef] [PubMed]
4. Johnson, M.D.; Freeland, J.R.; Parducci, L.; Evans, D.M.; Meyer, R.S.; Molano-Flores, B.; Davis, M.A. Environmental DNA as an emerging tool in botanical research. *Am. J. Bot.* **2023**, *110*, e16120. [CrossRef] [PubMed]
5. Waldvogel, A.-M.; Feldmeyer, B.; Rolshausen, G.; Exposito-Alonso, M.; Rellstab, C.; Kofler, R.; Mock, T.; Schmid, K.; Schmitt, I.; Bataillon, T. Evolutionary genomics can improve prediction of species' responses to climate change. *Evol. Lett.* **2020**, *4*, 4–18. [CrossRef] [PubMed]
6. Borisov, N.; Sorokin, M.; Garazha, A.; Buzdin, A. Quantitation of molecular pathway activation using RNA sequencing data. In *Nucleic Acid Detection and Structural Investigations: Methods and Protocols*; Humana New York: New York, NY, USA, 2020; pp. 189–206.
7. Graw, S.; Chappell, K.; Washam, C.L.; Gies, A.; Bird, J.; Robeson, M.S.; Byrum, S.D. Multi-omics data integration considerations and study design for biological systems and disease. *Mol. Omics* **2021**, *17*, 170–185. [CrossRef] [PubMed]
8. Subramanian, M.; Wojtusciszyn, A.; Favre, L.; Boughorbel, S.; Shan, J.; Letaief, K.B.; Pitteloud, N.; Chouchane, L. Precision medicine in the era of artificial intelligence: Implications in chronic disease management. *J. Transl. Med.* **2020**, *18*, 472. [CrossRef] [PubMed]
9. Denny, J.C.; Collins, F.S. Precision medicine in 2030—Seven ways to transform healthcare. *Cell* **2021**, *184*, 1415–1419. [CrossRef] [PubMed]
10. Su, M.; Zhang, Z.; Zhou, L.; Han, C.; Huang, C.; Nice, E.C. Proteomics, personalized medicine and cancer. *Cancers* **2021**, *13*, 2512. [CrossRef] [PubMed]
11. Park, D.I. Genomics, transcriptomics, proteomics and big data analysis in the discovery of new diagnostic markers and targets for therapy development. *Prog. Mol. Biol. Transl. Sci.* **2020**, *173*, 61–90. [PubMed]

International Journal of
Molecular Sciences

MDPI

Article

De Novo Assembly of the Genome of the Sea Urchin *Paracentrotus lividus* (Lamarck 1816)

Maria Costantini [1,*], Roberta Esposito [1], Nadia Ruocco [2], Davide Caramiello [3], Angela Cordella [4,5], Giovanna Maria Ventola [4] and Valerio Zupo [6,*]

1 Stazione Zoologica Anton Dohrn, Department of Ecosustainable Marine Biotechnology, Via Ammiraglio Ferdinando Acton n. 55, 80133 Napoli, Italy; roberta.esposito@szn.it
2 Stazione Zoologica Anton Dohrn, Department of Ecosustainable Marine Biotechnology, Calabria Marine Centre, C.da Torre Spaccata, 87071 Amendolara, Italy; nadia.ruocco@szn.it
3 Stazione Zoologica Anton Dohrn, Department of Marine Animal Conservation and Public Engagement, Villa Comunale, 1, 80121 Naples, Italy; davide.caramiello@szn.it
4 Genomix4Life S.r.l., Baronissi, 84081 Salerno, Italy; angela.cordella@genomix4life.com (A.C.); giovanna.ventola@genomix4life.com (G.M.V.)
5 Genome Research Center for Health-CRGS, Baronissi, 84081 Salerno, Italy
6 Stazione Zoologica Anton Dohrn, Department of Ecosustainable Marine Biotechnology, Ischia Marine Centre, 80121 Naples, Italy
* Correspondence: maria.costantini@szn.it (M.C.); vzupo@szn.it (V.Z.)

Abstract: The Mediterranean purple sea urchin *Paracentrotus lividus* (Lamarck 1816) is a remarkable model system for molecular, evolutionary and cell biology studies, particularly in the field of developmental biology. We sequenced the genome, performed a *de novo* assembly, and analysed the assembly content. The genome of *P. lividus* was sequenced using Illumina NextSeq 500 System (Illumina) in a 2 × 150 paired-end format. More than 30,000 open reading frames (ORFs), (more than 8000 are unique), were identified and analysed to provide molecular tools accessible for the scientific community. In particular, several genes involved in complex innate immune responses, oxidative metabolism, signal transduction, and kinome, as well as genes regulating the membrane receptors, were identified in the *P. lividus* genome. In this way, the employment of the Mediterranean sea urchin for investigations and comparative analyses was empowered, leading to the explanation of cis-regulatory networks and their evolution in a key developmental model occupying an important evolutionary position with respect to vertebrates and humans.

Keywords: genes; genomic resources; sea urchin

Citation: Costantini, M.; Esposito, R.; Ruocco, N.; Caramiello, D.; Cordella, A.; Ventola, G.M.; Zupo, V. *De Novo* Assembly of the Genome of the Sea Urchin *Paracentrotus lividus* (Lamarck 1816). *Int. J. Mol. Sci.* **2024**, *25*, 1685. https://doi.org/10.3390/ijms25031685

Academic Editor: Wajid Zaman

Received: 4 January 2024
Revised: 24 January 2024
Accepted: 25 January 2024
Published: 30 January 2024

1. Introduction

The introduction of the Sanger method of polymerase-based sequencing revolutionized molecular/genomic studies in the early 1970s, permitting the definition of individual genomes and their regulation [1–3]. Next-generation sequencing (NGS) platforms increased the power of massive DNA sequencing to digitally interrogate genomes on a revolutionary scale, allowing functional genomic studies of gene expression profiling, genome annotation, and epigenetic modifications of histones or DNA methylations [4]. Availability of these omics approaches directly improved ecological genomics and/or molecular ecology studies [5]. Understanding the molecular responses of organisms to environmental stress is critical to current research on the environmental effects of global warming, ocean acidification, and increasing pollution. In this view, sea urchins are ideal models for monitoring marine environmental hazards [6], and as deuterostomes [7–12] (Figure 1), they are a perfectly positioned outgroup to the chordates [13].

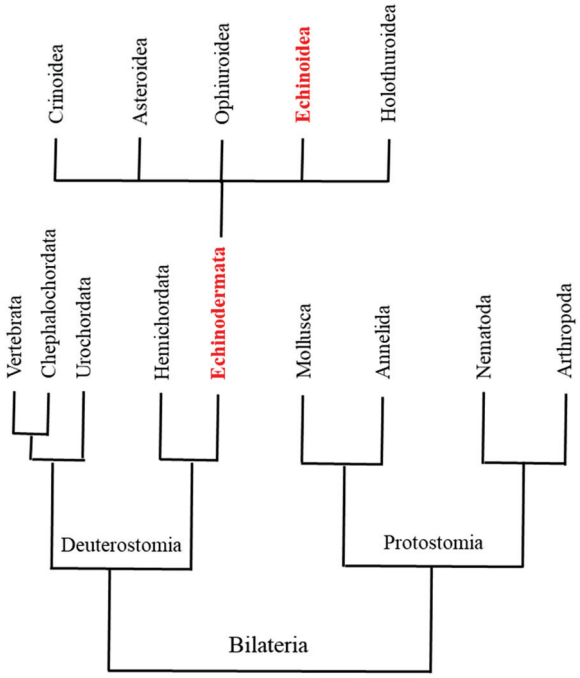


Figure 1. A simplified phylogenetic tree for Deuterostomia and Protostomia (according to information in Sodergren et al. [7]). Deuterostomia Echinodermata are highlighted in red font. The five classes of Echinodermata are also reported; Echinoidea class is highlighted in red.

Among sea urchins, *Strongylocentrotus purpuratus* and *Paracentrotus lividus* are well-established model organisms for developmental and ecotoxicological studies; the genome of the former was sequenced in 2006 [7], yielding important insights into the evolution of deuterostomes. *Paracentrotus lividus* (Parechinidae) [8], in particular, has wide geographical distribution, inhabiting shallow marine environments in the Mediterranean Sea and the eastern Atlantic Ocean [12], and is a keystone herbivore often controlling the standing crop of algal turfs and seagrass meadows [9]. In some areas, sea urchins transform macrophyte communities into barren areas, reducing biodiversity and altering ecosystem function [7–10,12,13]. Gonads of *Paracentrotus lividus* are considered a food delicacy [10,11] and it is thus intensively exploited in many Mediterranean areas. Moreover, *Paracentrotus lividus* is a well-established model for evo-devo and toxicology investigations because of the peculiar transparency of its embryos, which follow well-defined temporal patterns of development. In spite of the importance of *Paracentrotus lividus* for the ecology of coastal areas and as a model for scientific research, it was only very recently that a chromosome-scale genome assembly for this sea urchin was published [14], also reporting extensive gene expression and the epigenetic profiles of embryonic development. In addition, several SRA experiments and RNA sequences are already available on the website of the National Centre for Biotechnology Information (NCBI), which represents a relevant genomic resource.

We present here a draft of the genome sequence of the sea urchin *Paracentrotus lividus*, with a *de novo* assembly and analysis of its content. Our findings provide a remarkable resource to elucidate the genetic mechanisms underlying the adaptation and resilience of this key grazer and, consequently, they will be of great significance for theoretical and applied research. The genome presented here will provide a paradigm for studying novel features in model animals, such as molecular pathways underlying important physiological processes, and will represent an additional resource for the conservation and management of this widely distributed marine resource.

2. Results and Discussion

2.1. Sequencing and Annotation of Paracentrotus Lividus Genome

An important node in the deuterostome phylogeny is represented by the position of Echinodermata as an early branch, implying information on vertebrate biology [13,15,16]. Among sea urchins, the family Strongylocentrotidae represents the best-studied group [17], including the species for which the genome is available, *Strongylocentrotus purpuratus* [7]. Strongylocentrotidae contain several species of marine echinoids, including four genera: *Strongylocentrotus*, *Mesocentrotus*, *Hemicentrotus* and *Pseudocentrotus* [18–20] (Table 1).

Table 1. Species belonging to the four genera of Strongylocentrotidae.

Genera	Species
Strongylocentrotus	*S. purpuratus* *S. pallidus* *S. droebachiensis* *S. intermedius* *S. fragilis* *S. polyacanthus*
Mesocentrotus	*M. franciscanus* *M. nudus*
Hemicentrotus	*H. pulcherrimus*
Pseudocentrotus	*P. depressus*

The phylogenetic tree reported in Figure 2 was based on complete mitochondrial genomes available for 12 sea urchin species, and *Paracentrotus lividus* was selected to root the mitochondrial trees because it is strongly supported as an appropriate outgroup for Strongylocentrotidae, with an estimate of 35–50 Myr used as a reference time point for the split between the Strongylocentrotid species and Parechinids [21].

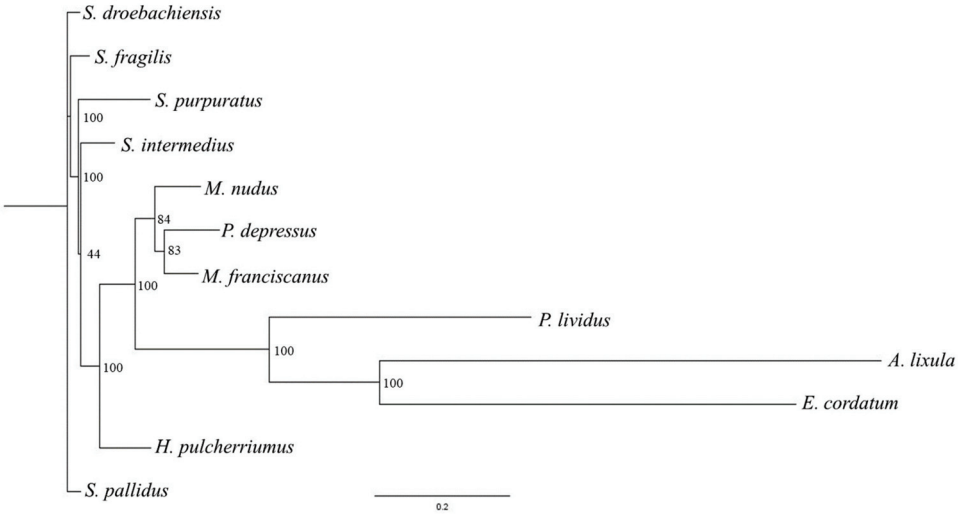


Figure 2. Phylogenetic tree (produced using RAxML, ML + rapid bootstrap) based on complete mitochondrial genomes available for 12 sea urchin species downloaded from GenBank: *Allocentrotus fragilis* (KC89820); *Arbacia lixula* (NC_001770); *Echinocardium cordatum* (FN562581); *Hemicentrotus pulcherrimus* (NC_023771); *Mesocentrotus franciscanus* (NC_024177); *Mesocentrotus nudus* (NC_020771); *Pseudocentrotus depressus* (NC_023773); *Strongylocentrotus droebachensis* (NC_009940); *Strongylocentrotus intermedius* (NC_023772); *Strongylocentrotus pallidus* (NC_009941); *Strongylocentrotus purpuratus* (NC_001453); and *Paracentrotus lividus* (NC_001572). The values are reported as percentage of node label. Data on nucleotides are reported in Table S1.

BLAST top hit species distribution of matches for all the scaffolds with known sequences indicated that the majority of *P. lividus* scaffolds show the highest homology with *S. purpuratus* (BLAST Hits = 52,000) (Figure 3).

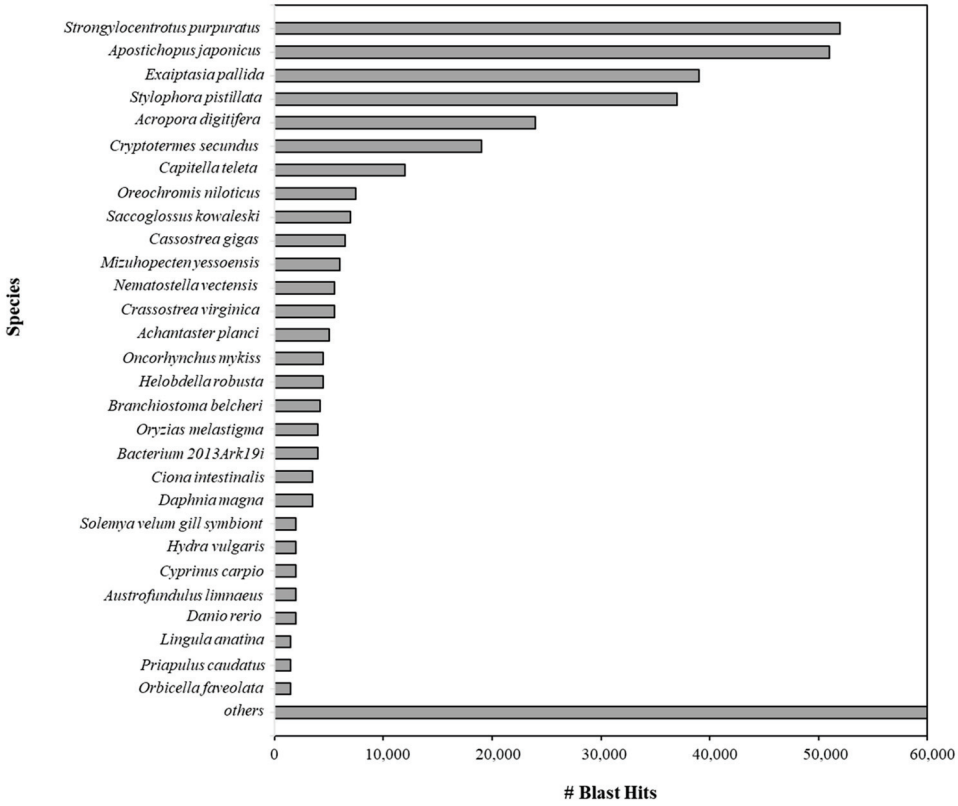


Figure 3. BLAST top hit species distribution (reported as number) of matches with known sequences aligned during the BLAST step using the NCBI database. In total, 3142 species were obtained with at least one BLAST hit, but in the histogram only the top thirty hits were reported; the remaining species are listed in "Others".

The most-represented species included *Apostichopus japonicus* (sea cucumber, BLAST hits: 51,000); *Exaiptasia pallida* (sea anemone, BLAST: hits 39,000); and *Stylophora pistillata* (coral, BLAST hits: 37,000). All alignments were carried out by setting the E-value thresholds as $\leq 1 \times 10^{-5}$. By using ABySS (version 2.0), 252,999 contigs and 252,952 scaffolds were obtained (Table 2). ABySS represents a resource-efficient assembly of large genomes using a Bloom filter [22].

Table 2. Statistical analysis of *P. lividus* genome size.

Parameter	Quast	ABYSS	BUSCO
Assembly	scaffolds (min. Length = 500 bp)	scaffolds (min. Length = 500 bp)	
contigs (≥0 bp)	252,952	252,952	252,952
contigs (≥1000 bp)	280	_	
contigs (≥5000 bp)	5	_	
Total_length (≥0 bp)	42,528,692	_	42,528,692
Total_length (≥1000 bp)	515,753	_	
Total_length (≥5000 bp)	28,242	_	
Total_length (≥10,000 bp)	0	_	
contigs	1757	1757	

Table 2. *Cont.*

Parameter	Quast	ABYSS	BUSCO
Largest_contig	6806	6805	
Total_length	1,488,145	1,486,080	
GC(%)	34.77%	31.88%	
N50	792	791	153

The maximum scaffold length for the genome was 6805 nucleotides with an N50 of 792 (min length 500 bp). The total length of the scaffolds, considering contigs $\geq$ 500 bp, was 1,486,080 nucleotides and the GC content corresponded to 31.9%. Using Geneious, 337,545 ORF sequences were obtained, belonging to 140,726 unique scaffold sequences. The annotation of ORF sequences (8508 were unique) was performed using Blast2GO, obtaining 33,529 ORF. Of these sequences, 13,523 and 21,380 were identified with InterPro and GO IDs, respectively, as shown in Table S2.

A total of 48 GO terms were enriched, including 16 in "Biological Process" followed by 16 in "Molecular Function" and 16 in "Cellular Component" ($p < 0.05$) (Figure 4).

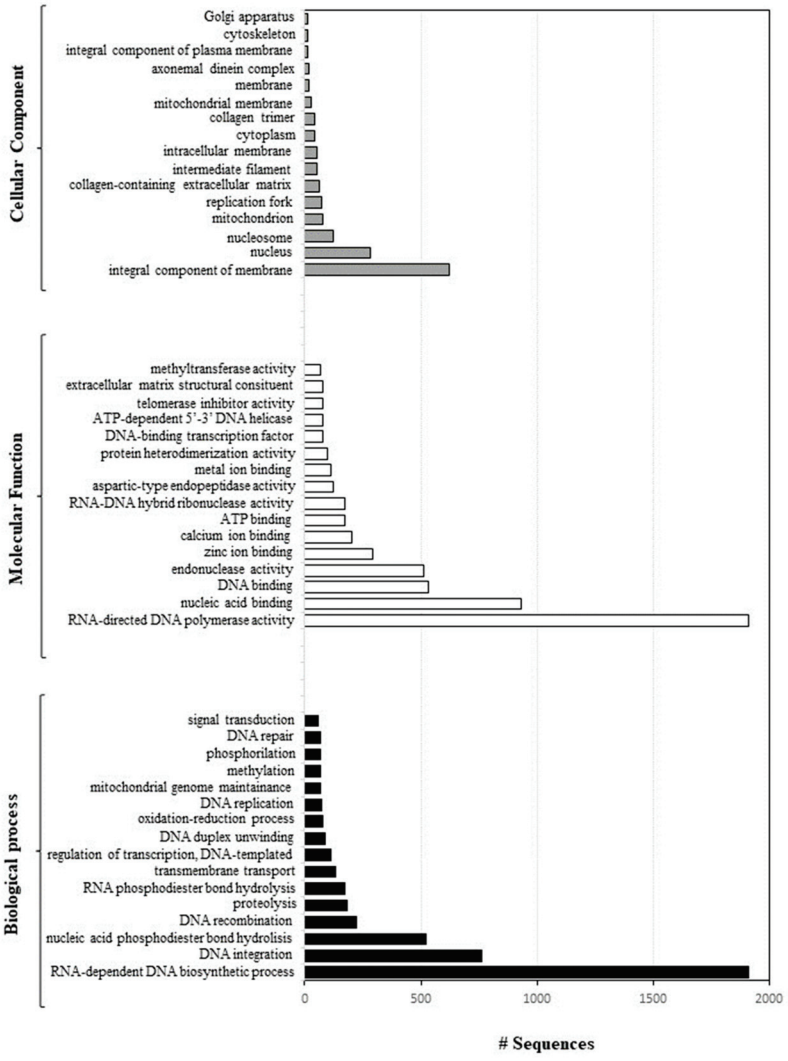


Figure 4. Direct GO count distribution, showing Molecular Function (a chart for the Molecular Function GO category, which shows the most frequent GO terms within a data-set without taking into account the GO hierarchy: white bars), Biological Process (same as above but for Biological Process: black bars), and Cellular Component (same as above but for Cellular Component: grey bars).

Over-represented GO categories included RNA-dependent DNA biosynthetic process; DNA integration; nucleic acid phosphodiester bond hydrolysis; DNA recombination; RNA-directed DNA polymerase activity; nucleic acid binding; DNA binding endonuclease activity; zinc ion binding; calcium ion binding; and DNA binding (see also Figure 5).

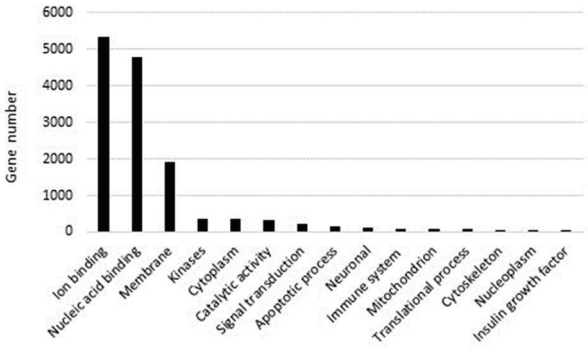


Figure 5. Analysis of *P. lividus* genes reporting the gene number involved in different cellular processes.

Moreover, these genes are an integral component of the cell membrane and were mainly localised in the nucleus, nucleosome, mitochondrion, and replication fork.

The draft genome of *P. lividus* reported here provided additional information to those reported in Marlétaz et al. [14], allowing for a comprehensive survey of the main gene pathways available for further developmental investigations. Evidently, the peculiar phylogenetic position of echinoderms offers the possibility to perform comparisons between protostomes and deuterostomes, and between invertebrate and vertebrate deuterostomes. Many aspects of development and cell–cell interactions will provide new perspectives on those genes that evolved to control important developmental processes. Sea urchins are evolutionarily closer to not only other deuterostomes, including vertebrates, but also to protostomes, such as *Drosophila* and *Caenorhabditis elegans*. From this perspective, they may provide a clearer view into the evolution of vertebrates, including their developmental signalling and evolution. Evidently, sea urchin embryos are a key evolutionary link to vertebrate development, although the extent of molecular commonality has only now become measurable with the genomic sequence data.

2.2. Key Findings and Genes

2.2.1. Complex Innate Immune Responses

Among the classes of innate receptors, toll-like receptor (TLR) genes from 48 proteins, TLR-1, TLR-2, TLR-3, TLR-4, were identified (see Table 3). TLRs represent an important part of innate immunity, playing key roles in the defence against pathogen invasion [23].

Table 3. Summary of all genes/proteins identified in the genome of *P. lividus*.

	Genes/Proteins
Immune response	Toll-like receptor 1
	Toll-like receptor 2
	Toll-like receptor 3
	Toll-like receptor 4
	Toll-like receptor 5
	Toll-like receptor 6
	Toll-like receptor 7
	Toll-like receptor 8
	Toll-like receptor 9
	Toll-like receptor 10
	Toll-like receptor 11
	Toll-like receptor 12
	Toll-like receptor 13
	E3 ubiquitin-protein ligase pellino homolog 1

Table 3. *Cont.*

	Genes/Proteins
Signal transduction	Ras
	Rab
	Ral
	Arf
	Rhodopsin
	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-4
	Nuclear factor kappa B
	Allatostatin-A receptor-like
	Calcium-independent protein kinase C
Membrane receptors	suREJ1
	suREJ2
	suREJ3
	Ankyrin-containing gene specific for Apical Tuft
	Fibrillin A
	Rhodopsin
Neuronal genes	Calcineurin
	Neurexin
	Neurocan
	Neuroendocrine convertase 1 gene
	Neuron navigator 3-like
	Neuronal acetylcholine receptor subunit alpha-5-like
	Beta-adrenergic receptor kinase 2
Kinome	Adenosine kinase
	A-kinase anchor protein 17A
	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase isoform X1
	Cell division cycle 7-related protein kinase
	Cyclin-dependent kinase 2-like
	Dolichol kinase
	Dual specificity mitogen-activated protein kinase kinase 7 isoform X2
	Inositol hexakisphosphate
	Diphosphoinositol-pentakisphosphate kinase 1 isoform X1
	L-fucose kinase
	MAP kinase
	Maternal embryonic leucine zipper kinase isoform X2
	Membrane-associated guanylate kinase
	WW
	PDZ domain-containing protein 2-like
	Receptor tyrosine-protein kinase erbB-4-like
	Receptor-like protein kinase feronia, serine/threonine-protein kinase PAK 2
	Tyrosine-protein kinase receptor Tie-1-like
	Wall-associated receptor kinase and wee1-like protein kinase 1-A
Oxidative metabolism	CYP 1-like
	CYP 2-like
	CYP 3-like
	CYP 4-like
	CYP 6-like
	CYP 20-like
	CYP 26-like
	CYP 27-like
	CYP 46-like
	CYP 51-like
	CYP 120-like

Cell surface TLRs include TLR1; TLR2; TLR4; TLR5; TLR6; and TLR10, whereas intracellular TLRs include TLR3; TLR7; TLR8; TLR9; TLR11; TLR12; and TLR13, which are localised in the endosome [24]. Even if TLRs were first described as important recognition receptors in mice and humans, they were also extensively studied in several animal species, including invertebrates such as the sea urchin, where a great expansion of TLR genes occurred. Gene expression analysis reported TLR receptors specifically expressed at the two-cell cleavage stage [14].

Despite the apparent simplicity of their body organization, echinoderms exhibit an immune system able to perform complex innate immune responses, a phenomenon which is, to date, quite far from being completely understood [25]. In fact, the *S. purpuratus* genome represents a window on the functions of the innate immune system complexity and sensing capacity, revealing a close genetic relationship between sea urchins and humans and thus reinforcing the relevance of these model organisms. In addition, echinoderms are advanced invertebrates, representing a bridge with primitive chordates, because they possess numerous receptors and effectors used to obtain a rapid immune response. After an infection, their cellular immune response triggers a network formed by membrane and endosomal receptors, which in turn triggers an immune response by stimulating consecutive intracellular events [26,27].

The genome of this sea urchin holds a vast set of at least 222 TLR genes, accompanied by a moderate expansion of downstream adaptors, different from that of chordates [28–30]. The abundance of TLRs in sea urchins suggests that this class of receptors plays an important role in the innate immune defence, possibly the case in lower animals as well. The vast majority of sea urchin TLR genes are more similar to each other than to those of other animals, suggesting a gene expansion specific to the sea urchin lineage [28]. The recognition of non-self molecules by specific membrane receptors triggers the immune response, stimulating consecutive intracellular events [30]. An E3 ubiquitin-protein ligase pellino homolog 1 gene was identified, as it has been for *S. purpuratus* (LOC577851). This gene, located in the cytosol, enables ubiquitin protein ligase activity and is involved in several processes, including the negative regulation of the necroptotic process; it also participates in protein poly-ubiquitination.

2.2.2. Molecular Switches in Signal Transduction

Our data revealed for the first time the presence in the *Paracentrotus lividus* genome of four genes belonging to the small guanosine triphosphatase (GTPases) families: Ras, Rab, Ral, ARF and Rho. Comparing these data with those reported in the case of the genome of *S. purpuratus*, four families of RAS GTPases are in common with humans: Ras, Rho, Rab, and ARF. The genes of this family are usually expressed during embryogenesis [18,31]. They represent a large family of hydrolase enzymes able to bind to the nucleotide guanosine triphosphate (GTP), and hydrolyse it to guanosine diphosphate (GDP), which in turn functions as a molecular switch in signal transduction, nuclear import and export, lipid metabolism, and vesicle docking [7]. Vertebrate GTPase families were expanded after their divergence from echinoderms, thanks to whole genome duplications [19,20]. This is different from the sea urchin genome, which did not undergo a whole-genome duplication. In the case of four Ras GTPase families, (Ras, Rho, Rab, and Arf), a local gene duplication occurred, resulting in a comparable number of monomeric GTPases in the genomes of humans and sea urchins, and signalling complexity mediated by GTPases.

Another gene involved in signal transduction processes is 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-4 (PLCB4), also identified in the genome of *P. lividus*. Phosphatidylinositol-specific phospholipase C enzymes mediate the production of the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3); in humans this form has a role in retina signal transduction [32]. The protein encoded by this gene is a phosphodiesterase, which catalyses the hydrolysis of phosphatidylinositol 4,5-bisphosphate to the second messengers, inositol 1,4,5-trisphosphate (IP3) and diacylglycerol. The encoded protein is activated by G proteins and is involved in the signal transduction pathway of the type-2 taste receptor. In addition, the nuclear factor kappa B (which is also identified in *P. lividus* genome) can regulate the transcription of this gene, whose protein product can also act as an important regulator of platelet responses.

Another interesting gene here found (orthologous, also found in *S. purputarus*), is the allatostatin-A receptor-like gene, which belongs to the type A allatostatins (AST-As), a family of insect peptides with a conserved C-terminal FGL-amide motif [33]. The insect allatostatin-A receptors (AST-ARs), considered orthologues of galanin receptors (GALR) in

vertebrates [34], are activated by AST-A peptides. Vertebrate GALR receptors have a close relationship, from an evolutionary point of view, with kisspeptin receptors (KISSR), and are in turn activated by galanin (GAL) and spexin (SPX) peptides, which are unrelated to insect AST-As.

In contrast, the ankyrin (ANK) repeat protein family is largely distributed across plants and has been found to participate in multiple processes such as plant growth and development, hormone response, and response to biotic and abiotic stresses [35].

Finally, the calcium-independent protein kinase C (PKC) has a key role in signal transduction mechanisms. In particular, this gene is involved in the initiation and maintenance of motility in the spermatozoa of the sea urchin *Lytechinus pictus*. White et al. [32] demonstrated the existence of a correlation between motility and the level of phospho-PKC substrates, so PKC activation and phosphorylation of its target proteins represent a significant requirement for the maintenance of motility in the spermatozoa of intact sea urchins. In *S. purpuratus*, the levels of PKC substrates contribute to the production of immotile and motile spermatozoa, and non-competitive PKC inhibitors are involved in diminishing the circular velocity of spermatozoa [36].

2.2.3. Genes Regulating the Membrane Receptors

The alpha-like subunit of the acetylcholine receptor (AChR) binds acetylcholine, and just after is subject to an extensive change in its conformation, thereby affecting all subunits and the opening of an ion-conducting channel present in the plasma membrane. ATPases comprise a superfamily of proteins involved in several cellular processes essential for physiology, (control of proteins; homeostasis; DNA replication; recombination; chromatin re-modelling; ribosomal RNA processing; molecular targeting; organelle biogenesis; and membrane fusion); they are often associated with diverse cellular activities (AAA+) [37–40]. In fact, the members of this superfamily are defined by the presence of the AAA+ domain, containing the canonical Walker A and B motifs necessary for ATP binding and its hydrolysis [37].

The receptor for egg jelly precursor is an integral component of the membrane, with an important role in calcium channel activities and ion binding. Three of these receptors, suREJ1, suREJ2, and suREJ3, were previously described in *S. purpuratus* testis [41]. In particular, suREJ1 is composed of one transmembrane segment, able to bind to the fucose sulphate polymer of egg jelly, thus inducing the sperm acrosome reaction. On the other hand, suREJ3 consists of 11 putative transmembrane segments localised in the plasma membrane over the acrosomal vesicle. In contrast, suREJ2 is an intracellular plasma membrane protein with no extracellular projection from the plasma membrane, and two transmembrane segments; it is present in the entire spermatozoa plasma membrane, mainly concentrated over the spermatozoan mitochondrion. REJ is a common module present in all three sea urchin sperm REJ proteins, and is shared by the human autosomal dominant polycystic kidney disease protein, polycystin-1, and PKDREJ—a testis-specific protein found in mammals [42]. Ankirin is another integral protein component of the membrane, involved in protein heterodimerization. The ankyrin (ANK) repeat domain was identified for the first time in some yeast cell-cycle regulators, and in the *Drosophila melanogaster* signalling protein Notch3 [43,44]. It is considered the most common conserved protein domain, being distributed in organisms ranging from viruses to humans [35,45]. Yaguchi et al. [46] isolated the gene ankAT-1, the Ankyrin-containing gene specific for apical tuft, the expression of which is normally found in the animal pole region of the very early blastula stages of sea urchin embryos. This gene is involved in the regulation of the length of apical tuft cilia, mediating apical tuft formation in the sea urchin embryo, as demonstrated by experimental knock-down of this gene, resulting in much shorter embryos and with decreased motility with respect to the motile cilia in other ectodermal cells[46]. Although the specific functions of the ANK domain proteins are still not known, they were identified in several proteins with diverse functions, highlighting their roles as mediators of protein–protein interactions and acting as molecular chaperones [42,47–49].

Lamin B is a structural constituent of the extracellular matrix, with a key role in calcium ion binding and cell-matrix adhesion. In the sea urchin *L. pictus*, it is imported into the nucleus from a soluble pool at a later stage of pronuclear formation; the resulting incorporation is necessary for pronuclear swelling and growth of the nuclear envelope [50].

The genome of *Paracentrotus lividus* possesses a putative precursor of fibrillin-2, a matrix protein involved in protein kinase activity, ATP binding, and protein phosphorylation. In the *S. purpuratus* genome, two genes encoding fibrillin homologs were identified, suggesting an expansion of this family in deuterostomes, clustering with vertebrate, honeybee and ascidian fibrillins [51]. Fibrillin A is expressed during cleavage and by primary mesenchyme cells, with a role in the fibrillar components of the blastocoel extracellular matrix [52]. There are three fibrillin genes in mammals (FBN-1, FBN-2, and FBN-3), and only one each in *C. elegans* (fbn-1) and *Drosophila* (CG31999). Mutations in human fibrillins are responsible for Marfan syndrome and the related disease, contractural arachnodactyly.

Rhodopsin is a light-sensitive receptor protein belonging to G-protein-coupled receptors (GPCRs), involved in photo-transduction and contributing to the majority of sensory receptors in vertebrates. In the sequenced sea urchin genome, they represent the largest GPCR family with 979 members constituting more than 3% of all predicted genes [53]. Moreover, four greatly expanded subfamilies of rhodopsin-type GPCRs were identified, which rapidly expand the lineages of GPCRs (surreal-GPCRs). This group is mostly expressed in different classes of pedicellariae and in the tube feet of adult sea urchins, harbouring sensory neurons involved in the reaction to chemical stimuli in echinoderms. In addition, these structures also express different opsins, indicating that sea urchins possess a complex system for sensing their environment. These genes may have arisen by rapid duplication in the echinoid lineage, acting as chemosensory receptors.

2.2.4. Nervous System and Neuronal Genes

The nervous system of echinoderms is dispersed both in larvae and adults, and thus differs from both vertebrates and hemichordates, but it is not a simple nerve net. In fact, vertebrates do not have a dispersed nervous system, and hemichordates have nerve nets [54]. Adult sea urchins have thousands of appendages with sensory neurons, ganglia, and motor neurons arranged in local reflex arcs. These peripheral appendages are connected to each other and to radial nerves for overall control and coordination.

The genome of *Paracentrotus lividus* showed the presence of a calcineurin gene. Since the 1970s, a calmodulin-binding protein was found only in the brain of sea urchins [55–57]. This protein, named calcineurin, is localised in neurons and is associated with post synapsis and dendrite microtubules [58]. Stewart et al. [59] found that protein phosphatase 2B, involved in glycogen metabolism in skeletal muscle, is similar to calcineurin, which in turn has similar activity.

In *S. purpuratus* there is a single predicted gene for a neurexin (a synaptic adhesion component [51]), also found in the genome of *Paracentrotus lividus*. There are several predictions for neuroligins, the postsynaptic receptors for neurexins, also known from other invertebrates, where β-neurexin binds neuroligin and is clustered to recruit presynaptic components [60].

In contrast, neurocan is a member of the lectican/chondroitin sulphate proteoglycan protein families; it consists of neurocan core protein (identified in the *Paracentrotus lividus* genome) and chondroitin sulphate. It is involved in the modulation of cell adhesion and migration. The neuroendocrine convertase 1 gene isolated in *Paracentrotus lividus* has some orthologs in *S. purpuratus*, such as furin and subtilisin, which may process TGF-β precursors [61]. The neurogenic locus Notch protein has a neural ortholog in the sea urchin *Lytechinus variegatus*. The Notch intercellular signalling pathway mediates the specification of numerous cell fates in both invertebrate and vertebrate development [62]. The navigator 3-like neuron belongs to the neuron navigator family and is expressed predominantly in the nervous system. Genes related to neurotransmitter receptors were also found, such as the neuronal acetylcholine receptor subunit alpha-5-like gene. These genes also include

beta-adrenergic receptor kinase 2 and are confirmed to be localised to a subset of ectoderm, consistent with a neural population [63].

2.2.5. The Kinome of *P. lividus* Resembles That of Drosophila and Human

Several kinases were identified in the genome of *Paracentrotus lividus*: 52 kDa repressor of the inhibitor of the kinase-like protein; adenosine kinase; A-kinase anchor protein 17A; bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase isoform X1; cell division cycle 7-related protein kinase; 2-like cyclin-dependent kinase; dolichol kinase; dual specificity mitogen-activated protein kinase 7 isoform X2; inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase 1 isoform X1; L-fucose kinase; MAP kinase-activating death domain protein isoform X1; MAPK/MAK/MRK overlapping kinase; maternal embryonic leucine zipper kinase isoform X2; membrane-associated guanylate kinase; WW and PDZ domain-containing protein 2-like; erbB-4-like receptor tyrosine-protein kinase; receptor-like protein kinase Feronia; serine/threonine-protein kinase PAK 2; Tie-1-like tyrosine-protein kinase receptor; and wall-associated receptor kinase and wee1-like protein kinase 1-A. All these genes define the kinome of *Paracentrotus lividus*, representing the complete set of protein kinases encoded in the genome [64]. The genomic sequence of *S. purpuratus* and the predicted gene models were used to identify the predicted protein kinases in this genome, according to both function and kinase domain taxonomy. The results showed that the sea urchin kinome: i, consists of 353 protein kinases, and ii, is closer to the *Drosophila* kinome (239) than the human kinome (518), according to the total kinase number. However, it has been established that the diversity of sea urchin kinases is surprisingly similar to those of humans. In fact, the urchin kinome misses only 4 of 186 human subfamilies, while *Drosophila* lacks 24, thus combining the simplicity of a non-duplicated genome with the diversity of function and signalling, which was previously considered to be very specific to vertebrates [64]. More than half of the sea urchin kinases are involved in signal transduction, and approximately 88% of the signalling kinases are expressed in the developing embryo. An example is the dolichol kinase, for which Rossignol et al. [65] reported that both *de novo* synthesis of dolichol and its phosphorylation may play an important role in the observed increase of glycoprotein synthesis in early embryonic development, prior to gastrulation. Protein kinases play crucial roles in the regulation of signalling pathways, in coordination with protein phosphatases, in the genome of *S. purpuratus* [66]. In fact, high expression of kinases was detected in endomesoderm formation.

2.2.6. Homologies with Human Oxidative Metabolism

About 30 CYP-related genes were identified in the *Paracentrotus lividus* genome, mostly belonging to CYP 1, 2, and 3. Several CYP 2 genes (28) were found, while nine CYP1-like genes and only three CYP 3 genes (CYP3-like) were detected. The genome of *P. lividus* contains homologs of proteins involved in metabolism (CYP 27), and three other families: the CYP 6, CYP 20, and CYP 120 genes. These enzymes, belonging to the CYP1, CYP2, CYP3, and CYP4 families, play key roles in the oxidative biotransformation of chemicals to more hydrophilic products. The sea urchin *S. purpuratus* contains 120 CYP genes, of which 80% of the total are related to CYP gene families 1 to 4, suggesting a selective pressure to expand functionality in these gene families.

Eleven CYP1-like genes are present in the genome of *Paracentrotus lividus*, which represent more than twice the number of these genes among chordates. A greater amount of CYP2-like and CYP3-like genes are also observed in comparison to other deuterostomes. In addition to the CYPs in families 1 to 4, the sea urchin genome contains homologs of proteins involved in developmental patterning (CYP26), cholesterol synthesis (CYP51), and metabolism (CYP27 and CYP46). Homologs of some CYPs with endogenous functions in vertebrates are not found. These CYP genes, in concert with additional expanded defensive gene families, represent a large diversification of defence gene families by the sea urchin relative to mammals.

3. Materials and Methods

3.1. Ethics Statement

Paracentrotus lividus (Lamarck) were collected from a site in the Bay of Naples that is not privately owned or protected in any way, according to Italian legislation (DPR 1639/68, 19 September 1980, confirmed on 10 January 2000). Field studies did not include endangered or protected species. All experimental procedures on animals were in compliance with the guidelines of the European Union (Directive 609/86).

3.2. Sample Collection and DNA Extraction

Adult sea urchins were collected during the breeding season during scuba-diving in the Gulf of Naples, transported in a thermic box to the laboratory within 1 h after collection and maintained in tanks with circulating sea water until testing [67]. Sea urchins were injected with 2 M KCl through the peribuccal membrane to obtain the emission of gametes. Concentrated spermatozoans were collected and immediately used for DNA extraction. Genomic DNA was extracted from single male sperm using 1× TEN buffer (50 mM Tris pH 7.6, 10 mM EDTA, 100 mM NaCl) plus 1% sodium dodecyl sulfate. Digestion with proteinase K (100 mg/mL) was performed at 55 °C overnight, and contaminating RNA was degraded by treating RNase (10 mg/mL) at 37 °C for one hour. Extraction with phenol–chloroform–isoamyl alcohol (25:24:1) was then followed by precipitation via the addition of 3M sodium acetate/95% ethanol. The amount of total DNA extracted was estimated by the absorbance at 260 nm.

3.3. De Novo Genome Assembly

The pipeline followed for the genome assembly and annotation was as follows:

1. Genome sequencing: the next generation sequencing experiment and bioinformatics analysis were performed using Genomix4life S.R.L. (Baronissi, Salerno, Italy). DNA concentration was assayed with a ND-1000 spectrophotometer (NanoDrop, ND-1000 UV-Vis Spectrophotometer; NanoDrop Technologies, Wilmington, DE, USA), and its quality assessed with an Agilent 4200 Tapestation (Agilent Technologies, Santa Clara, CA, USA; according manufacturer instructions). An indexed library was prepared from 1 μg of purified DNA with a Truseq DNA Nano Library Prep Kit according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The library was quantified using the Tape Station 4200 (Agilent Technologies, Santa Clara, CA, USA) and a Qubit fluorometer (Invitrogen Co., Carlsbad, CA, USA), and diluted with a final concentration of 2 nM. The sample was subject to cluster generation and sequencing using an Illumina NextSeq 500 System (Illumina) in a 2 × 150 paired-end format, according NextSeq 500 System Documentation.
2. Sequencing outputs, quality control and cleaning: the most common metric was used to assess the accuracy of a sequencing platform (base calling accuracy, measured by the Phred quality score (Q score). The first step was a quality check of the raw Illumina sequencing data to remove adapter sequences and low-quality reads, using ad hoc script. The FastQC tool (available on http://www.bioinformatics.babraham.ac.uk/projects/fastqc; 1 February 2021) was used to check the quality of raw data sequencing.
3. Genome assembly: to perform the *de novo* assembly, a KmerGenie (version 1.7044) tool was necessary to estimate the best k-mer length 66. In this case, the best k-mer predicted was 121. ABySS 2.0, an implementation of ABySS 1.0, was used to perform the *de novo* assembly on fastq files. The bloom filter of ABySS 2.0 was applied to avoid duplicate sequences.
4. Genome assembly stats and validation: the integrity assembly was also evaluated, using several statistical tools, such as QUAST, Abyss, BBMAP, and BUSCO (Table S3).
5. Genome annotation and functional analysis: Geneious software 69 was used to identify all the ORF sequences, and Blast2GO was applied to perform a blast alignment of all ORF sequences identified and to annotate everything in the Gene Ontology database.

The assembly has been deposited in the SRA database (submission ID: SUB6921168; BioProject ID: PRJNA604684; BioSample: Processed Successfully loaded SAMN13978365: PARLIV_1.0; TaxID: 7656). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAWLRT000000000. The version described in this paper is version JAWLRT010000000.

3.4. Phylogenetic Tree

The phylogenetic tree was based on complete nucleotide mitochondrial genomes available for 12 sea urchin species: *Allocentrotus fragilis* (KC89820); *Arbacia lixula* (NC_001770); *Echinocardium cordatum* (FN562581); *Hemicentrotus pulcherrimus* (NC_023771); *Mesocentrotus franciscanus* (NC_024177); *Mesocentrotus nudus* (NC_020771); *Pseudocentrotus depressus* (NC_023773); *Strongylocentrotus droebachensis* (NC_009940); *Strongylocentrotus intermedius* (NC_023772), *Strongylocentrotus pallidus* (NC_009941); *Strongylocentrotus purpuratus* (NC_001453); and *Paracentrotus lividus* (NC_001572). The sequences were aligned with SeaView [68], a software permitting the performance of a complete phylogenetic analysis of a set of homologous DNA or protein sequences, from network-based sequence extraction from public databases to tree building and display, using up-to-date alignment and a maximum-likelihood tree-building algorithm [69]. The phylogenetic tree was produced using RAxML [69] and figures with FigTree V.1.4.3 [70]. In detail, a set with the function "G blocks" was created, permitting smaller final blocks, gap positions within the final blocks, and less strict flanking positions. It reports bootstrap resampling for branch-support estimation with a substitution matrix (GTR) and substitution rates (Gamma), offering nucleotide sequence alignments that were evolved with non-stationary (NS) and non-reversible (NR) substitution models.

4. Conclusions

This *de novo* assembly greatly expands on the previous analysis of the *Paracentrotus lividus* sea urchin genome reported very recently by Marlétaz et al. [14], highlighting genomic and regulatory evolution in deuterostomes. In fact, our data added new information on several classes of genes which have not been previously identified. Several genes involved in complex innate immune responses, oxidative metabolism, signal transduction, and kinome, as well as genes regulating the membrane receptors, were identified analysing the content of the assembly. This represents a very significant step in understanding the evolution of this key species, not only by the deuterostome phylogeny implying information on vertebrate biology, but also with respect to the vertebrates themselves, which we are planning to expand by adding long-read analysis. The analyses of *P. lividus* genes could also shed light on biological processes, and on differences/similarities among species or genera of sea urchins.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25031685/s1.

Author Contributions: Conceptualization: M.C. and V.Z.; Methodology: A.C., G.M.V., N.R., R.E. and D.C.; Investigation: R.E., N.R., M.C. and V.Z.; Supervision: M.C. and V.Z.; Writing—original draft: M.C. and V.Z.; Writing—review & editing: M.C., R.E., N.R., D.C., A.C., G.M.V. and V.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The assembly has been deposited in the SRA database (submission ID: SUB6921168; BioProject ID: PRJNA604684; BioSample: Processed; Successfully loaded SAMN13978365: PARLIV_1.0 (TaxID: 7656). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAWLRT000000000. The version described in this paper is version JAWLRT010000000.

Acknowledgments: The research was partially supported by the Flagship Research Project ModRes of the Stazione Zoologica Anton Dohrn. We also thank WG Nelson for critical reading of the English manuscript.

Conflicts of Interest: A.C. and G.M.V. have been involved as a consultant and expert witness in Genomix4life Company. Other authors declare no conflicts of interest.

References

1. Werner, T. Next generation sequencing in functional genomics. *Brief. Bioinform.* **2010**, *2*, 499–511. [CrossRef]
2. Kulski, J.K. *Next-Generation Sequencing—An Overview of the History, Tools, and "Omic" Applications*; Spinger: Berlin/Heidelberg, Germany, 2016; Volume 10, p. 61964. ISBN 0000957720.
3. Giani, A.M.; Gallo, G.R.; Gianfranceschi, L.; Formenti, G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 9–19. [CrossRef]
4. Hurd, P.J.; Nelson, C.J. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief. Funct. Genom. Proteom.* **2009**, *8*, 174–183. [CrossRef] [PubMed]
5. Ekblom, R.; Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **2011**, *107*, 1–15. [CrossRef] [PubMed]
6. Dupont, S.; Ortega-Martínez, O.; Thorndyke, M. Impact of near-future ocean acidification on echinoderms. *Ecotoxicology* **2010**, *19*, 449–462. [CrossRef] [PubMed]
7. Sodergren, E.; Weinstock, G.M.; Davidson, E.H.; Cameron, R.A.; Gibbs, R.A.; Angerer, R.C.; Angerer, L.M.; Arnone, M.I.; Burgess, D.R.; Burke, R.D.; et al. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **2006**, *314*, 941–952. [CrossRef] [PubMed]
8. Pawson, D.L. *Narcissia ahearnae*, a new species of sea star from the Western Atlantic (Echinodermata: Asteroidea: Valvatida). *Zootaxa* **2007**, *58*, 53–58. [CrossRef]
9. Boudouresque, C.F.; Bernard, G.; Bonhomme, P.; Charbonnel, E.; Le Diréach, L.; Ruitton, S. Monitoring methods for *Posidonia oceanica* seagrass meadows in Provence and the French Riviera. *Sci. Rep.* **2007**, *22*, 17–38.
10. Guidetti, P.; Terlizzi, A.; Boero, F. Effects of the edible sea urchin, *Paracentrotus lividus*, fishery along the Apulian rocky coast (SE Italy, Mediterranean Sea). *Fish. Res.* **2004**, *66*, 287–297. [CrossRef]
11. Furesi, R.; Madau, F.; Palomba, A.; Pulina, P. Stated preferences for consumption of sea urchin: A choice experiment in Sardinia (Italy). *Int. J. Food Syst. Dyn.* **2014**, *5*, 111–119.
12. Boudouresque, C.F.; Verlaque, M. Ecology of *Paracentrotus lividus*. In *Edible Sea Urchins: Biology and Ecology*; Developments in Aquaculture and Fisheries Science; Elsevier: Amsterdam, The Netherlands, 2001; pp. 177–216.
13. Kober, K.M.; Bernardi, G. Phylogenomics of strongylocentrotid sea urchins. *BMC Evol. Biol.* **2013**, *13*, 88. [CrossRef]
14. Marlétaz, F.; Couloux, A.; Poulain, J.; Labadie, K.; Da Silva, C.; Mangenot, S.; Noel, B.; Poustka, A.J.; Dru, P.; Pegueroles, C.; et al. Analysis of the *P. lividus* sea urchin genome highlights contrasting trends of genomic and regulatory evolution in deuterostomes. *Cell Genom.* **2023**, *3*, 100295. [CrossRef]
15. Materna, S.C.; Berney, K.; Cameron, R.A. The *S. purpuratus* genome: A comparative perspective. *Dev. Biol.* **2006**, *300*, 485–495. [CrossRef]
16. Cameron, R.A.; Davidson, E.H. A basal deuterostome genome viewed as a natural experiment. *Gene* **2007**, *406*, 1–7. [CrossRef]
17. Lawrence, J.M. Edible sea urchins: Use and life-history strategies. In *Developments in Aquaculture and Fisheries Science*; Elsevier: Amsterdam, The Netherlands, 2007; pp. 1–9.
18. Beane, W.S.; Gross, J.M.; McClay, D.R. RhoA regulates initiation of invagination, but not convergent extension, during sea urchin gastrulation. *Dev. Biol.* **2006**, *292*, 213–225. [CrossRef] [PubMed]
19. Dehal, P.; Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **2005**, *3*, e314. [CrossRef] [PubMed]
20. Gu, X.; Wang, Y.; Gu, J. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.* **2002**, *31*, 205–209. [CrossRef]
21. Lee, T.H.; Naitoh, H.; Yamazaki, F. Chromosome studies on the mitten crabs *Eriocheir japonica* and *E. sinensis*. *Fish. Sci.* **2004**, *70*, 211–214. [CrossRef]
22. Jackman, S.D.; Vandervalk, B.P.; Mohamadi, H.; Chu, J.; Yeo, S.; Hammond, S.A.; Jahesh, G.; Khan, H.; Coombe, L.; Warren, R.L.; et al. ABySS 2.0: Resource-efficient assembly of large genomes using a bloom filter effect of bloom filter false positive rate. *Genome Res.* **2017**, *27*, 768–777. [CrossRef] [PubMed]
23. Fink, I.R.; Pietretti, D.; Voogdt, C.G.P.; Westphal, A.H.; Savelkoul, H.F.J.; Forlenza, M.; Wiegertjes, G.F. Molecular and functional characterization of Toll-like receptor (Tlr)1 and Tlr2 in common carp (*Cyprinus carpio*). *Fish Shellfish Immunol.* **2016**, *56*, 70–83. [CrossRef] [PubMed]
24. Kawasaki, T.; Kawai, T. Toll-like receptor signaling pathways. *Front. Immunol.* **2014**, *5*, 461. [CrossRef] [PubMed]
25. Pinsino, A.; Alijagic, A. Sea urchin *Paracentrotus lividus* immune cells in culture: Formulation of the appropriate harvesting and culture media and maintenance conditions. *Biol. Open* **2019**, *8*, bio039289. [CrossRef] [PubMed]
26. Chiaramonte, M.; Russo, R. The echinoderm innate humoral immune response. *Ital. J. Zool.* **2015**, *82*, 300–308. [CrossRef]

27. Smith, L.C.; Arriza, V.; Barela Hudgell, M.A.; Barone, G.; Bodnar, A.G.; Buckley, K.M.; Cunsolo, V.; Dheilly, N.M.; Franchi, N.; Fugmann, S.D.; et al. Echinodermata: The complex immune system in echinoderms. In *Advances in Comparative Immunology*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 773–791. ISBN 9783319767680.
28. Hibino, T.; Loza-Coll, M.; Messier, C.; Majeske, A.J.; Cohen, A.H.; Terwilliger, D.P.; Buckley, K.M.; Brockton, V.; Nair, S.V.; Berney, K.; et al. The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* **2006**, *300*, 349–365. [CrossRef] [PubMed]
29. Materna, S.C.; Cameron, R.A. The sea urchin genome as a window on function. *Biol. Bull.* **2008**, *214*, 266–273. [CrossRef] [PubMed]
30. Chiaramonte, M.; Arizza, V.; Russo, R. Evolutionary conserved pathway of the innate immune response after a viral insult in *Paracentrotus lividus* sea urchin. *Int. J. Immunogenet.* **2019**, *46*, 192–202. [CrossRef]
31. Adonin, L.; Drozdov, A.; Barlev, N.A. Sea urchin as a universal model for studies of gene networks. *Front. Genet.* **2021**, *11*, 1739. [CrossRef] [PubMed]
32. White, D.; De Lamirande, E.; Gagnon, C. Protein kinase C is an important signaling mediator associated with motility of intact sea urchin spermatozoa. *J. Exp. Biol.* **2007**, *210*, 4053–4064. [CrossRef]
33. Felix, R.C.; Trindade, M.; Pires, I.R.P.; Fonseca, V.G.; Martins, R.S.; Silveira, H.; Power, D.M.; Cardoso, J.C.R. Unravelling the evolution of the allatostatin-type A, KISS and galanin peptide-receptor gene families in bilaterians: Insights from *Anopheles mosquitoes*. *PLoS ONE* **2015**, *10*, e0130347. [CrossRef]
34. Mirabeau, O.; Jean-Stéphane, J. Molecular evolution of peptidergic signaling systems in bilaterians. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E2028–E2037. [CrossRef]
35. Lopez-Ortiz, C.; Peña-Garcia, Y.; Natarajan, P.; Bhandari, M.; Abburi, V.; Dutta, S.K.; Yadav, L.; Stommel, J.; Nimmakayala, P.; Reddy, U.K. The ankyrin repeat gene family in *Capsicum* spp.: Genome-wide survey, characterization and gene expression profile. *Sci. Rep.* **2020**, *10*, 4044. [CrossRef] [PubMed]
36. Loza-Huerta, A.; Pacheco-Castillo, H.; Darszon, A.; Beltrán, C. Crosstalk between protein kinases A and C regulates sea urchin sperm motility. *Zygote* **2022**, *30*, 398–409. [CrossRef] [PubMed]
37. Hanson, P.I.; Whiteheart, S.W. AAA+ proteins: Have engine will work. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 519–529. [CrossRef] [PubMed]
38. Erzberger, J.P.; Berger, J.M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 93–114. [CrossRef]
39. Snider, J.; Thibault, G.; Houry, W.A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **2008**, *9*, 216. [CrossRef]
40. Shorter, J.; Houry, W.A. Editorial: The role of AAA+ proteins in protein repair and degradation. *Front. Mol. Biosci.* **2018**, *5*, 85. [CrossRef]
41. Galindo, B.E.; Moy, G.W.; Vacquier, V.D. A third sea urchin sperm receptor for egg jelly module protein, suREJ2, concentrates in the plasma membrane over the sperm mitochondrion. *Dev. Growth Differ.* **2004**, *46*, 53–60. [CrossRef] [PubMed]
42. Michaely, P.; Bennett, V. The ANK repeat: A ubiquitous motif involved in macromolecular recognition. *Trends Cell Biol.* **1992**, *2*, 127–129. [CrossRef]
43. Breeden, L.; Nasmyth, K. Cell cycle control of the yeast HO gene: Cis- and Trans-acting regulators. *Cell* **1987**, *48*, 389–397. [CrossRef]
44. Breeden, L.; Nasmyth, K. Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of Drosophila. *Nature* **1987**, *329*, 651–654. [CrossRef]
45. Sedgwick, S.G.; Smerdon, S.J. The ankyrin repeat: A diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **1999**, *24*, 311–316. [CrossRef] [PubMed]
46. Yaguchi, S.; Yaguchi, J.; Wei, Z.; Shiba, K.; Angerer, L.M.; Inaba, K. ankAT-1 is a novel gene mediating the apical tuft formation in the sea urchin embryo. *Dev. Biol.* **2010**, *348*, 67–75. [CrossRef] [PubMed]
47. Mosavi, L.K.; Cammett, T.J.; Desrosiers, D.C.; Peng, Z.Y. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **2004**, *13*, 1435–1448. [CrossRef] [PubMed]
48. Li, J.; Mahajan, A.; Tsai, M.D. Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* **2006**, *45*, 15168–15178. [CrossRef] [PubMed]
49. Shen, G.; Kuppu, S.; Venkataramani, S.; Wang, J.; Yan, J.; Qiu, X.; Zhang, H. ANKYRIN REPEAT-CONTAINING PROTEIN 2A Is an essential molecular chaperone for peroxisomal membrane-bound ASCORBATE PEROXIDASE3 in Arabidopsis. *Plant Cell* **2010**, *22*, 811–831. [CrossRef] [PubMed]
50. Collas, P.; Courvalin, J.C.; Poccia, D. Targeting of membranes to sea urchin sperm chromatin is mediated by a lamin B receptor-like integral membrane protein. *J. Cell Biol.* **1996**, *135*, 1715–1725. [CrossRef]
51. Whittaker, C.A.; Bergeron, K.F.; Whittle, J.; Brandhorst, B.P.; Burke, R.D.; Hynes, R.O. The echinoderm adhesome. *Dev. Biol.* **2006**, *300*, 252–266. [CrossRef]
52. Hardin, J. 4 The Cellular Basis of Sea Urchin Gastrulation. *Curr. Top. Dev. Biol.* **1996**, *33*, 159–262.
53. Raible, F.; Tessmar-Raible, K.; Arboleda, E.; Kaller, T.; Bork, P.; Arendt, D.; Arnone, M.I. Opsins and clusters of sensory G-protein-coupled receptors in the sea urchin genome. *Dev. Biol.* **2006**, *300*, 461–475. [CrossRef]
54. Xie, W.; Evans, R. Orphan nuclear receptors: The exotics of xenobiotics. *J. Biol. Chem.* **2001**, *276*, 37739–37742. [CrossRef]

55. Klee, C.; Krinks, M. Purification of cyclic 3′5′-nucleotide phosphodiesterase inhibitory protein by affinity chromatography on activator protein coupled to sepharose. *Biochemistry* **1978**, *17*, 120–126. [CrossRef] [PubMed]
56. Wallace, R.; Lynch, T.; Tallant, E.; Cheung, W. Purification and characterization of an inhibitor protein of brain adenylate cyclase and cyclic nucleotide phosphodiesterase. *J. Biol. Chem.* **1979**, *254*, 377–382. [CrossRef] [PubMed]
57. Iwasa, F.; Ishiguro, K. Calmodulin-binding protein (55K+17K) of sea urchin eggs has a Ca2+ and calmodulin-dependent phosphoprotein phosphatase activity. *J. Biochem.* **1986**, *99*, 1353–1358. [CrossRef]
58. Wood, J.G.; Wallace, R.W.; Whitaker, J.N.; Cheung, W.Y. Immunocytochemical localization of calmodulin and a heat-labile calmodulin-binding protein (CaM-BP80) in basal ganglia of mouse brain. *J. Cell Biol.* **1980**, *84*, 66–76. [CrossRef] [PubMed]
59. Stewart, A.; Ingebritsen, T.; Manalan, A.; Klee, C.; Cohen, P. Discovery of A Ca2+-and calmodulin-dependent protein phosphatase. *FEBS Lett.* **1982**, *137*, 80–84. [CrossRef] [PubMed]
60. Washbourne, P.; Dityatev, A.; Scheiffele, P.; Biederer, T.; Weiner, J.; Christopherson, K.; El-Husseini, A. Cell adhesion molecules in synapse formation. *J. Neurosci.* **2004**, *24*, 9244–9249. [CrossRef] [PubMed]
61. Lapraz, F.; Röttinger, E.; Duboc, V.; Range, R.; Duloquin, L.; Walton, K.; Wu, S.Y.; Bradham, C.; Loza, M.A.; Hibino, T.; et al. RTK and TGF-β signaling pathways genes in the sea urchin genome. *Dev. Biol.* **2006**, *300*, 132–152. [CrossRef] [PubMed]
62. Sherwood, D.R.; Mcclay, D.R. Identification and localization of a sea urchin Notch homologue: Insights into vegetal plate regionalization and Notch receptor regulation. *Development* **1997**, *124*, 3363–3374. [CrossRef]
63. Meyer, A.; Ku, C.; Hatleberg, W.; Telmer, C.A.; Hinman, V. New hypotheses of cell type diversity and novelty from comparative single cell and nuclei transcriptomics in echinoderms. *bioRxiv* **2022**. [CrossRef]
64. Bradham, C.A.; Foltz, K.R.; Beane, W.S.; Arnone, M.I.; Rizzo, F.; Coffman, J.A.; Mushegian, A.; Goel, M.; Morales, J.; Geneviere, A.M.; et al. The sea urchin kinome: A first look. *Dev. Biol.* **2006**, *300*, 180–193. [CrossRef]
65. Rossignol, D.P.; Lennarz, W.J.; Waechter, C.J. Induction of phosphorylation of dolichol during embryonic development of the sea urchin. *J. Biol. Chem.* **1981**, *256*, 10538–10542. [CrossRef]
66. Byrum, C.A.; Walton, K.D.; Robertson, A.J.; Carbonneau, S.; Thomason, R.T.; Coffman, J.A.; McClay, D.R. Protein tyrosine and serine-threonine phosphatases in the sea urchin, *Strongylocentrotus purpuratus*: Identification and potential functions. *Dev. Biol.* **2006**, *300*, 194–218. [CrossRef]
67. Ruocco, N.; Fedele, A.M.; Costantini, S.; Romano, G.; Ianora, A.; Costantini, M. New inter-correlated genes targeted by diatom-derived polyunsatured aldehydes in the sea urchin *Paracentrotus lividus*. *Ecotoxicol. Environ. Saf.* **2017**, *142*, 355–362. [CrossRef] [PubMed]
68. Galtier, N.; Gouy, M.; Gautier, C. Seaview and phylo_ win: Two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* **1996**, *12*, 543–548. [CrossRef] [PubMed]
69. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef] [PubMed]
70. Rambaut, A. FigTree 1.4.3. Available online: https://github.com/rambaut/figtree/releases (accessed on 1 December 2023).

International Journal of
Molecular Sciences

MDPI

Article

Genome-Wide Analysis of MYB Genes in *Primulina eburnea* (Hance) and Identification of Members in Response to Drought Stress

Jie Zhang [1,†], Yi Zhang [1,2,†] and Chen Feng [1,*]

[1] Jiangxi Provincial Key Laboratory of Ex Situ Plant Conservation and Utilization, Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang 332900, China; zhangjie@lsbg.cn (J.Z.); zhangyi@lsbg.cn (Y.Z.)
[2] School of Life Sciences, Nanchang University, Nanchang 330031, China
* Correspondence: fengc@lsbg.cn
† These authors contributed equally to this work.

Abstract: Due to periodic water deficiency in karst environments, *Primulina eburnea* experiences sporadic drought stress in its habitat. Despite being one of the largest gene families and functionally diverse in terms of plant growth and development, MYB transcription factors in *P. eburnea* have not been studied. Here, a total of 230 MYB genes were identified in *P. eburnea*, including 67 1R-MYB, 155 R2R3-MYB, six 3R-MYB, and two 4R-MYB genes. The R2R3-type *PebMYB* genes could be classified into 16 subgroups, while the remaining *PebMYB* genes (1R-MYB, 3R-MYB, and 4R-MYB genes) were divided into 10 subgroups. Notably, the results of the phylogenetic analysis were further supported by the motif and gene structure analysis, which showed that individuals in the same subgroup had comparable motif and structure organization. Additionally, gene duplication and synteny analyses were performed to better understand the evolution of *PebMYB* genes, and 291 pairs of segmental duplicated genes were found. Moreover, RNA-seq analysis revealed that the *PebMYB* genes could be divided into five groups based on their expression characteristics. Furthermore, 11 *PebMYB* genes that may be involved in drought stress response were identified through comparative analysis with *Arabidopsis thaliana*. Notably, seven of these genes (*PebMYB3*, *PebMYB13*, *PebMYB17*, *PebMYB51*, *PebMYB142*, *PebMYB69*, and *PebMYB95*) exhibited significant differences in expression between the control and drought stress treatments, suggesting that they may play important roles in drought stress response. These findings clarified the characteristics of the MYB gene family in *P. eburnea*, augmenting our comprehension of their potential roles in drought stress adaptation.

Keywords: *Primulina eburnea*; MYB transcription factor; gene family; drought stress

Citation: Zhang, J.; Zhang, Y.; Feng, C. Genome-Wide Analysis of MYB Genes in *Primulina eburnea* (Hance) and Identification of Members in Response to Drought Stress. *Int. J. Mol. Sci.* **2024**, *25*, 465. https://doi.org/10.3390/ijms25010465

Academic Editors: Wajid Zaman and Setsuko Komatsu

Received: 14 November 2023
Revised: 24 December 2023
Accepted: 28 December 2023
Published: 29 December 2023

1. Introduction

Abiotic stresses primarily comprise factors such as drought stress, which can negatively impact plant growth and development and even lead to plant mortality [1,2]. Drought stress affects physiological, cellular, and molecular processes that are driven by a complex regulatory network in which transcription factors (TFs), kinases, and abscisic acid (ABA) play key roles in signal transduction [3,4]. Analysis of the genes involved in the drought stress response improves understanding of their function in molecular pathways, which ultimately paves the way for genetically modifying stress tolerance in crops such as rice and maize [5,6].

The MYB family is one of the largest gene families in plants and plays a crucial role in a variety of physiological functions, including signal transduction, primary and secondary metabolism, and stress responses [7–9]. MYB TFs in plants can be divided into four primary categories based on the arrangement and number of repetitions: 1R-MYB, 2R-MYB, 3R-MYB, and 4R-MYB with one to four repeats, respectively [10,11]. Each repeat forms a motif fold of the helix-turn-helix (HTH) structure with approximately 50 amino acids, and the

regular interval of tryptophan plays a key role in maintaining the configuration of the HTH structure [12,13]. Furthermore, R2R3-MYB (2R-MYB) are predominantly present and widely investigated in plants such as *Arabidopsis thaliana*, rice, maize, and grape [14–16]. In addition, numerous research studies on model plants such as *Arabidopsis*, rice, and maize demonstrated that MYB genes are implicated in abiotic stress response [17–19].

Primulina eburnea (Hance) Yin Z. Wang is an evergreen perennial herb belonging to the family Gesneriaceae that is extensively distributed in southern China among karst landforms [20–22]. This contrasts sharply with the endemic distribution of other species in the genus *Primulina*. The characteristic habitat of *P. eburnea* in the karst landscape has high calcium content, thin soil, poor water holding capacity, and a shortage of nutrients, as well as a distribution of terrestrial islands [23–25]. Notably, *P. eburnea* stands out as one of the top genetic resources within the genus *Primulina* due to its excellent adaptability, wide distribution, and horticultural potential [26–28]. As a plant resource with broad development potential, *P. eburnea* was adapted to unpredictable drought stress in the karst landform. The abiotic stress response has been studied in many species. However, the mechanism of the drought stress response in *P. eburnea* has not been investigated.

In this study, we focused on the genome-wide identification of MYB genes in *P. eburnea* and investigated members related to drought stress response. A total of 230 genes were identified through a genome-wide survey, and phylogenetic analysis of the *PebMYB* genes was performed. To gain further insight, a combination of high-throughput expression analysis of genes in different tissues was performed, and the expression patterns of selected genes under drought stress treatments were measured using real-time quantitative PCR (qRT-PCR). Seven *PebMYB* genes involved in the drought stress response in *P. eburnea* were finally identified. The aims of this study are to lay a solid foundation for understanding the regulatory mechanism of MYB genes under stress conditions and to provide fresh perspectives on plant conservation in karst landforms.

2. Results

2.1. Identification and Characterization of MYB Genes in P. eburnea

A total of 230 *P. eburnea* MYB sequences were identified in *P. eburnea*, and the genes were labeled with the prefix *PebMYB* for clear designation (Supplemental Table S1, Figure S1). The *PebMYB* genes were unevenly distributed in the chromosome (Figure S1). Notably, 10 *PebMYB* genes (*PebMYB22*, *PebMYB32*, *PebMYB70*, *PebMYB78*, *PebMYB89*, *PebMYB153*, *PebMYB154*, *PebMYB155*, *PebMYB166*, and *PebMYB167*) were not located on any of the 18 assembled chromosomes (Supplemental Table S1). The average gene density across the entire genome of *P. eburnea* was approximately 39.14 genes per megabase (Mb), whereas the density of the *PebMYB* gene per Mb was notably lower at 0.28.

The R2R3-PebMYB proteins had an average length of 312.4 amino acid residues, ranging from 162 (PebMYB48) to 868 (PebMYB120). In contrast, the remaining PebMYB proteins displayed a broader span from 75 (PebMYB204, PebMYB218) to 1056 (PebMYB4R2), with an average length of 386. This considerable variation in amino acid length highlights the variation in the PebMYB proteins (Supplemental Table S1). A corresponding variance in relative molecular weight (Mw) was also observed in relation to amino acid length. For instance, the Mw of R2R3-PebMYB proteins ranged from 18,644.97 (PebMYB48) to 99,015.85 (PebMYB139), while that of the other PebMYB proteins ranged from 8645.75 (PebMYB218) to 117,733.6 (PebMYB4R2). Moreover, the isoelectric point (pI) exhibited a range from 4.92 (PebMYB98) to 10.2 (PebMYB145) in R2R3-PebMYB proteins and from 4.18 (PebMYB192) to 10.14 (PebMYB178) for the remaining types, highlighting the differences in the charge properties of the PebMYB proteins.

2.2. Phylogenetic Analysis and Classification of PebMYB Genes

To investigate the phylogenetic relationship of the *PebMYB* and *AtMYB* genes, ML phylogenetic trees were constructed. The members of the phylogenetic tree were classified

into several subgroups based on the topology of the tree and classifications in *A. thaliana* (Figures 1 and 2).

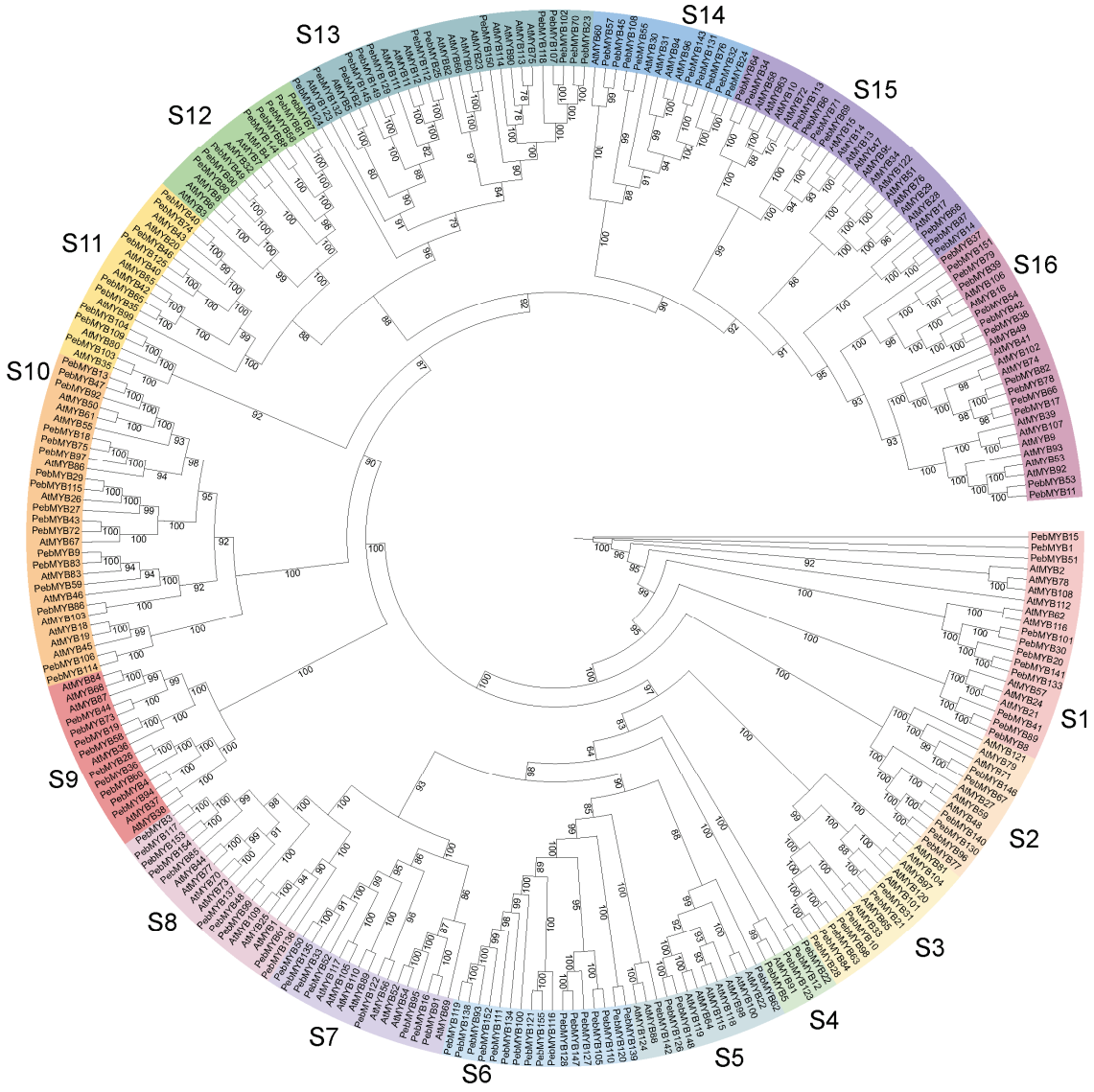


Figure 1. The phylogenetic relationship among R2R3-MYB genes in *Arabidopsis thaliana* and *Primulina eburnea*. The MYB genes were classified into 16 subgroups named S1 to S16 with different colors.

In the phylogenetic tree of R2R3-MYB genes, 126 *A. thaliana* protein sequences and 155 *P. eburnea* sequences were classified into 16 subgroups and designated S1 to S16 (Figure 1). Notably, subgroups S4 and S5 were the smallest group and contained only four *PebMYB* genes. In contrast, subgroup S6 consisted of 17 *PebMYB* genes, making it the largest group. Interestingly, all the members in S6 belong to the *PebMYB* genes (Figure 1). Furthermore, the majority of subgroups contained *AtMYB* genes with known functions, which will be useful for studying the function of *PebMYB* genes. For example, members in S13 (*AtMYB66*, *AtMYB114*, *AtMYB90*, etc.) have been shown to play important roles in the biosynthesis of anthocyanin and flavonol, as well as seed germination in *A. thaliana*, suggesting that *PebMYB150* in S13 may also play a crucial role in anthocyanin biosynthesis [29–31].

The phylogenetic tree of the other MYB genes (1R-MYB, 3R-MYB, and 4R-MYB) contained 72 protein sequences from *A. thaliana* (64 1R-MYB, 5 3R-MYB, and 3 4R-MYB genes, respectively) and 75 protein sequences from *P. eburnea* (67 1R-MYB, 6 3R-MYB, and two 4R-MYB genes, respectively) (Figure 2). The members were classified into 10 categories

(C1 to C10). C9, which was the smallest group, contained only two *PebMYB* genes, whereas the largest group, C4, included 20 members. Most *AtMYB* genes in C6 belonged to *CCA1-like* (*Circadian Clock Associated 1*), suggesting the likelihood that *PebMYB* genes in C6 may also function in circadian regulation. C2 contained *AtMYB* genes associated with *CPC-like* (*CAPRICE*), indicating that members of *PebMYB* genes in C2 could be involved in trichome formation and root development. C4 contained *A. thaliana* members of *TBP-like* (*telomeric DNA-binding protein*), which are known to play important roles in enhancing gene expression and cell development. Finally, C9 contained *R-R-type* MYB-like genes, while C10 contained *I-Box-Binding-like* genes from *A. thaliana* [16].

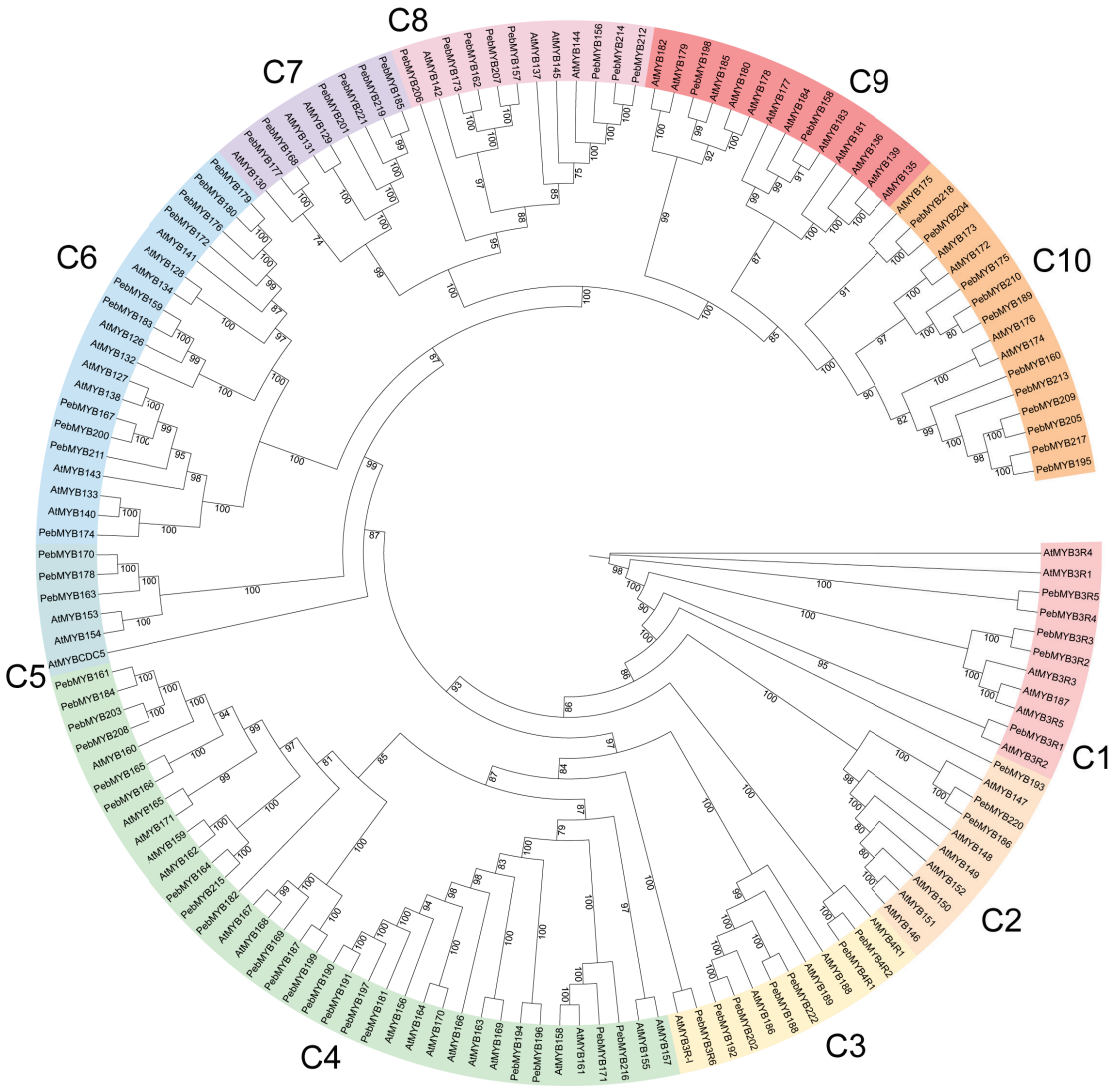


Figure 2. The phylogenetic tree of 1R-MYB, 3R-MYB, and 4R-MYB genes in *A. thaliana* and *P. eburnea*. The MYB genes were classified into 10 subgroups named C1 to C10 with different colors.

2.3. Motif Analysis and Gene Structure

ML phylogenetic trees of 155 *R2R3-PebMYB* and 75 remaining *PebMYB* genes were constructed (Figures 3 and 4). The topology of the phylogenetic tree was similar to the grouping pattern observed in the phylogenetic analysis in Figures 1 and 2, confirming the reliability of the phylogenetic analysis and classification of *PebMYB* genes. Furthermore, the presence of similar motif compositions within the same subgroup further supported the grouping results.

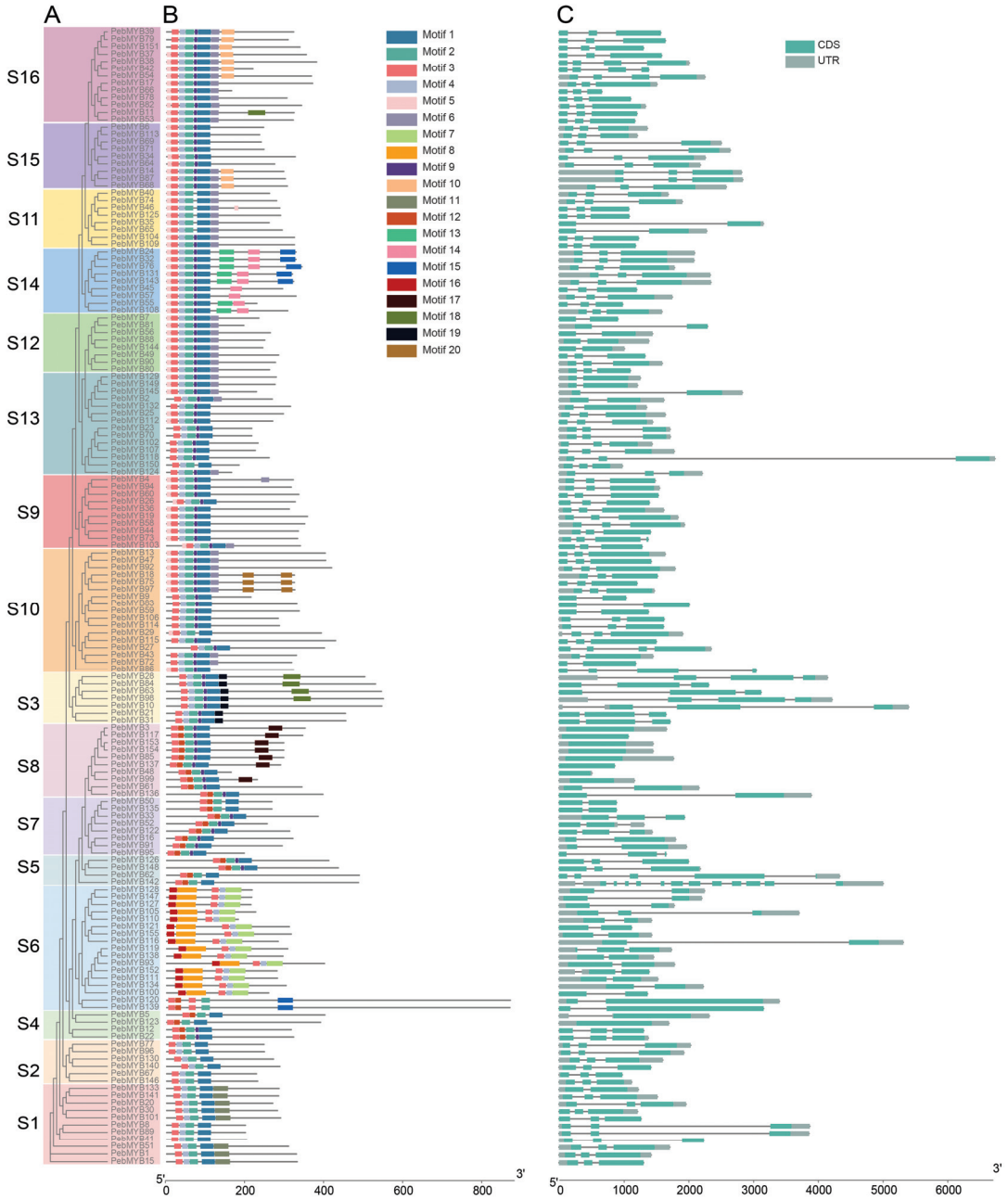


Figure 3. The phylogenetic tree, conserved motifs, and gene structure of R2R3-MYB genes in *P. eburnea*. (**A**) The ML phylogenetic tree of *PebMYB* genes grouped into 16 subgroups designated S1 to S16. (**B**) The conserved motif structure of *PebMYB* genes. (**C**) The gene structure of *PebMYB* genes. The grey bar indicates the untranslated region (UTR), whereas the cyan bar indicates the CDS. The x-axis in (**B**,**C**) indicates the lengths of proteins and genes, respectively.

To gain a comprehensive understanding of the conserved domains in *PebMYB* genes, motif analysis was performed by the MEME program with a parameter of up to 20 motifs found (Figures 3B and 4B). The most common conserved motifs identified in the motif analysis of R2R3-MYB genes were Motif 1, Motif 2, Motif 3, and Motif 4, which consisted of 34, 22, 18, and 15 amino acids, respectively. The R2 repeat was composed of Motif 3, Motif 4, and Motif 2, while Motif 1 corresponded to the R3 repeat (Supplementary Figure S2A, Table S2). The HTH (helix-turn-helix) structure was formed by regularly spaced tryptophan residues (arrows in Figure S2A), which was consistent with previous studies in other species [32,33].

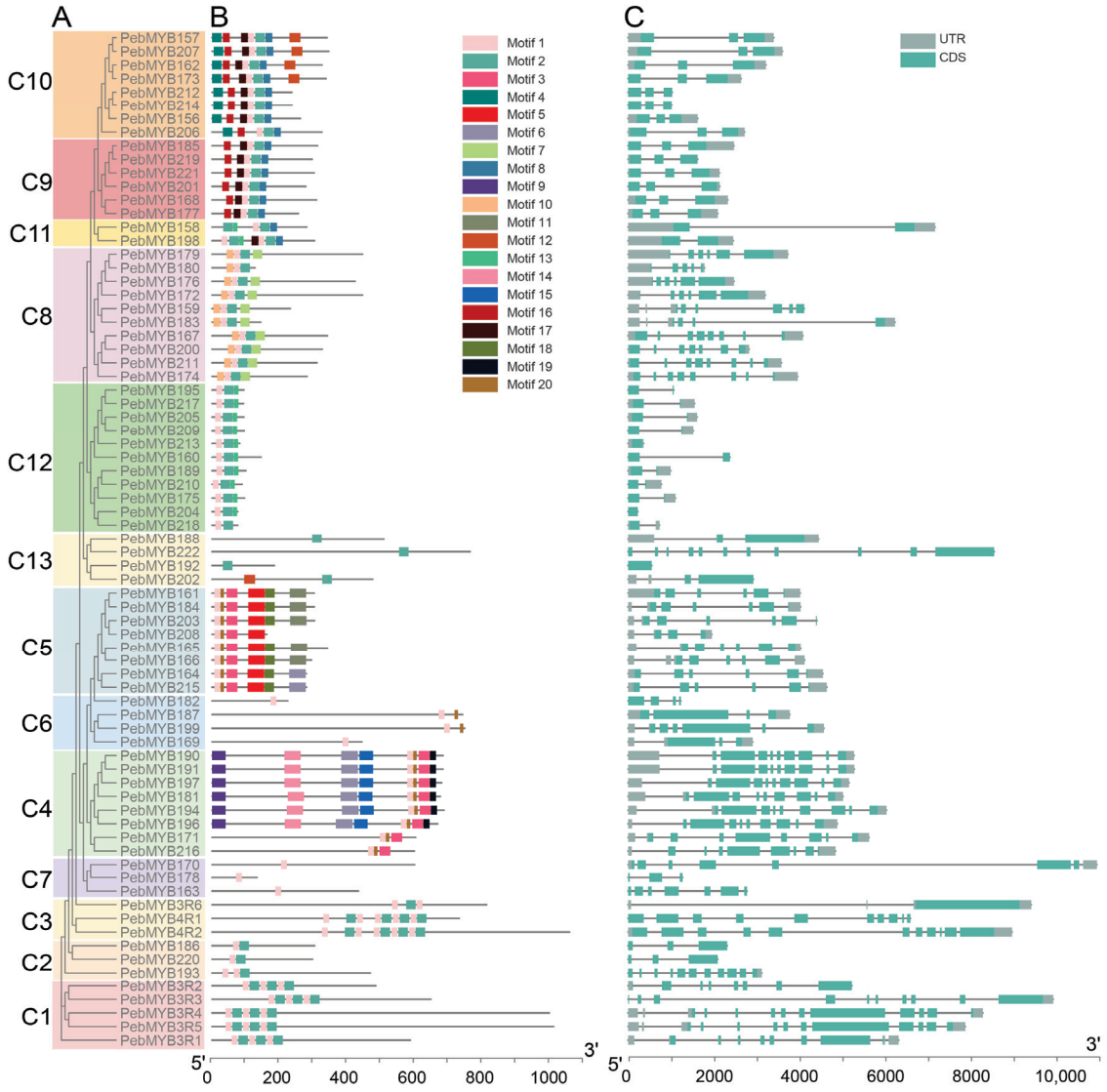


Figure 4. The phylogenetic tree, conserved protein motifs, and gene structure of 1R-MYB, 3R-MYB, and 4R-MYB genes of *P. eburnea*. (**A**) The ML phylogenetic tree of *PebMYB* genes grouped into 13 subgroups designated C1 to C13. (**B**,**C**) Conserved motifs and gene structure of *PebMYB* genes.

With only a few exceptions in S5, S6, and S7, the length of the R2R3 domain within the majority of *PebMYB* genes ranged from 10 to 150 amino acids (Figure 3B). Notably, Motif 1, Motif 2, and Motif 4 were either partially or completely absent in members of S4 and S6. This pattern was particularly evident in all 17 members of S6, which lacked either Motif 1 or Motif 2, separating them from other subgroups. These findings were similar to those found in Chinese jujube and upland cotton, indicating that similar motif compositions are features of the same subgroup and implying potential functional commonalities among members of the same subgroup [34,35]. The analysis of gene structures further revealed variations among subgroups (Figure 3C). For example, the majority of the R2R3-MYB genes in *P. eburnea* had two (40/155) or three (96/155) exons, only one gene (*PebMYB*62) had five exons, and one (*PebMYB*142) had 12 exons. Interestingly, these two genes belonged to S5. In addition, most members of S8 contained only one exon.

In the motif analysis of 1R-MYB, 3R-MYB, and 4R-MYB genes, two predominant motifs (Motif 1 and Motif 2) were identified. Motif 1 consisted of 18 amino acids and was composed of R2 MYB repeats, whereas Motif 2 contained 29 amino acids and was composed of both R1 and R3 repeats (Supplementary Figure S2B, Table S3). Motif 1 existed in most of the 1R-MYB, 3R-MYB, and 4R-MYB genes except for the members in C13, which

distinguished them from the other groups. Regarding gene structure, approximately half of the genes (36 out of 75 genes) contained one to three exons, while 19 out of 75 genes exhibited five to seven exons. Notably, five genes contained as many as 11 exons. These findings highlighted a greater variation in gene structure and motif composition among 1R-MYB, 3R-MYB, and 4R-MYB genes compared to R2R3-MYB genes.

2.4. Gene Duplications and Synteny Analysis of PebMYB Genes

Gene duplication events play a crucial role in the diversity and evolution of gene families. To further understand the evolution and expansion of the *PebMYB* genes, a duplication analysis was performed. The analysis revealed a total of 291 pairs of segmental duplicated genes, including 176 *PebMYB* genes (Supplemental Table S4). The nonsynonymous and synonymous substitution ratios (Ka/Ks) values of all segmental duplicated gene pairs were significantly less than 1, suggesting that purifying selection may have played an important role in the evolution of the *PebMYB* genes. Interestingly, only 3 pairs of tandem duplicated *PebMYB* genes (*PebMYB147*, *PebMYB195*, and *PebMYB209*) were found.

Most *PebMYB* genes were located at the ends of chromosomes and were distributed unevenly across these regions (Figure 5 and Figure S1). Additionally, gene duplications were found in all chromosomes (Figure 5).

Synteny analysis of MYB genes between *P. eburnea* and two species (*A. thaliana* and *Oryza sativa*) revealed 234 orthologous gene pairs between *P. eburnea* and *A. thaliana*, whereas 35 orthologous gene pairs were found between *P. eburnea* and *O. sativa* (Figure S3). The synteny analysis results showed 26 collinear gene pairs across the three species, indicating that 107 collinear pairs between *A. thaliana* and *P. eburnea* no longer existed between *P. eburnea* and *O. sativa* (Supplemental Table S5).

2.5. Expression Profiles of PebMYB Genes in Different Tissues

The expression patterns of *PebMYB* genes in stems, roots, buds, and leaves of *P. eburnea* were studied using RNA-seq data. Twelve *PebMYB* genes that showed no expression in any of the four tissues were excluded from the expression analysis. The remaining genes could be divided into five groups (I to V) based on their expression profiles. In group I, 30 *PebMYB* genes were highly expressed in leaf tissue, while 55 genes in group II were predominantly expressed in buds. The members of groups III and V (22 and 77 genes, respectively) showed high expression levels in the stem and root, respectively. Furthermore, the members of group IV (34 genes) showed high expression in both root and stem tissues (Figure 6A, Supplemental Table S6).

Furthermore, the *cis*-elements present in the promoters of *PebMYB* genes were analyzed to gain insights into their regulatory mechanisms. The *cis*-elements were primarily divided into four groups based on their functions: light response, stress response, hormone response, and others (Figure 6B, Supplemental Table S7). Notably, the hormone response and light response *cis*-elements were particularly abundant among the analyzed genes.

To verify the expression profiles from the RNA-seq data, several genes with high expression levels (fragments per kilobase of exon model per million mapped fragments, FPKM > 2) in at least one tissue were randomly selected, and the expression levels were determined by qRT-PCR. The relative expression levels of the chosen genes (11 genes) determined by qRT-PCR were consistent with the trends observed in the RNA-seq data, further supporting the reliability and accuracy of the RNA-seq results (Supplementary Figure S4, Table S6).

2.6. Expression Analysis of PebMYB in Drought Stress

MYB genes play a variety of roles, including anthocyanin biosynthesis, drought stress response, and many other functions. This notion was supported by the results of the promoter analysis, which highlighted numerous *PebMYB* genes associated with the stress response (Figure 6B). To investigate the role of *PebMYB* genes in the drought stress response, an experiment involving drought stress treatment was carried out. We

conducted a homology search with the genome of *A. thaliana* using the protein sequences of 230 *PebMYB* genes. Based on functional annotation of the homologous genes in *A. thaliana* and their phylogenetic relationships, 11 *PebMYB* genes were selected for further investigation. The relative expression levels of the selected genes were determined by qRT-PCR at different developmental stages (ST0, ST1, and ST2).

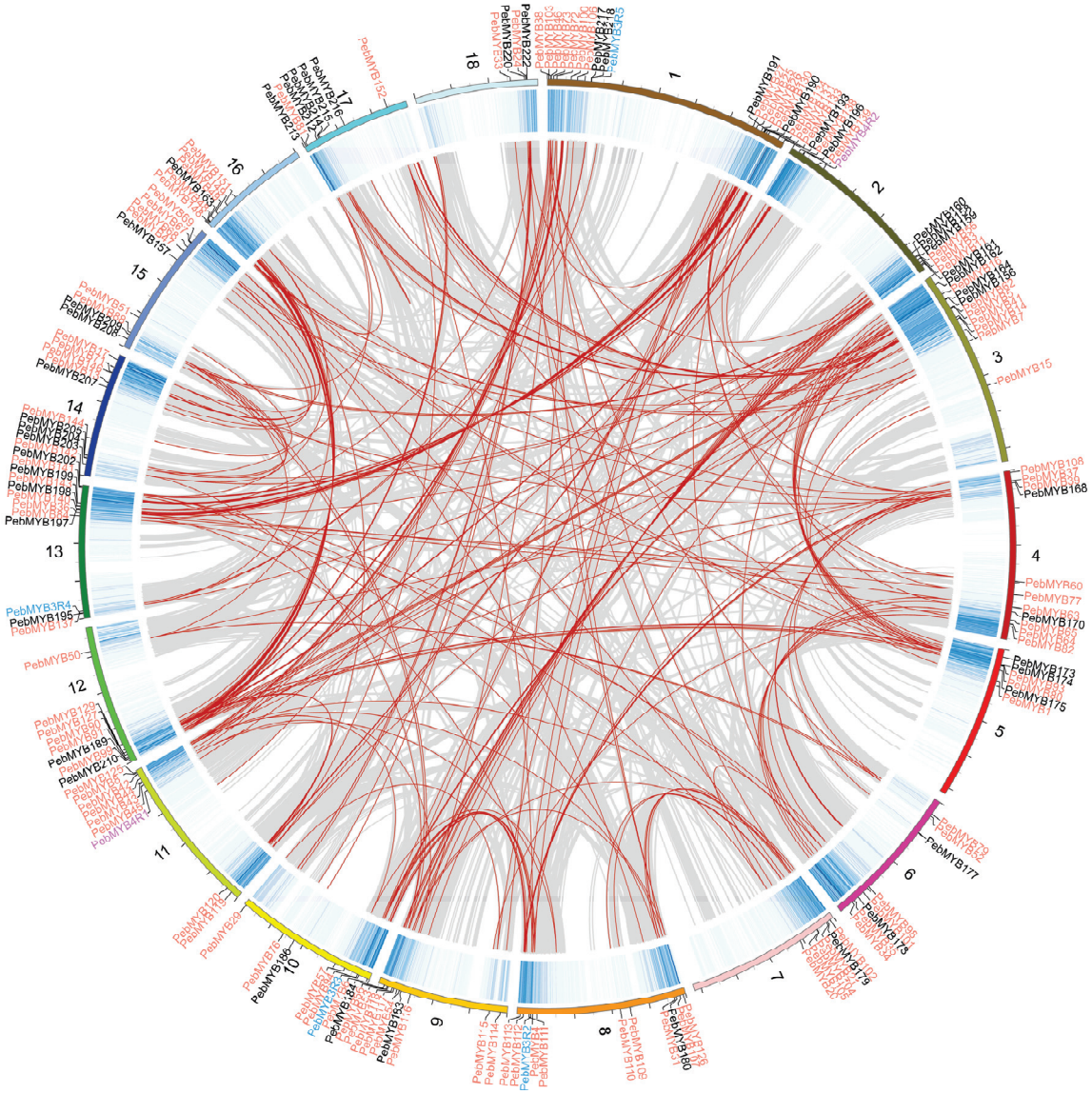


Figure 5. Duplication analysis of *PebMYB* genes. The heatmap indicates gene density. Grey lines indicate duplications of all genes, while duplications of *PebMYB* genes are highlighted by red lines. The black and grey scales on chromosomes indicate 10 Mb and 5 Mb, respectively. The 1R-MYB, 2R-MYB, 3R-MYB, and 4R-MYB genes are listed on chromosomes with black, salmon, blue, and purple colors.

Among these selected genes, *PebMYB3* exhibited homology to *AtMYB44* (*AT5G67300*) and *AtMYB77* (*AT3G50060*), which is recognized for their involvement in mediating crosstalk between different signaling pathways in response to drought stress. Notably, these pathways included the ABA, auxin, salicylic acid (SA), and methyl jasmonate (MeJA) regulatory pathways, ultimately activating genes to prevent reactive oxygen species (ROS) accumulation [36–38]. Differential expression was observed between the control and treatment groups after treatment for 2 months (ST2 in Figure 7D). *PebMYB13*, along with its homologous gene *AT1G09540* (*AtMYB61*), was selected due to its role in processes such as

stoma movement, root development, and seed germination. Interestingly, we noticed that the expression levels were significantly decreased in the treatment group compared to the control group in ST1 and ST2 (Figure 7E). In *A. thaliana*, a high expression level of *AtMYB61* could reduce stomatal aperture and stomatal conductance [39]. Likewise, *PebMYB17*, a homologous gene of *AT4G28110* (*AtMYB41*, function in suberin synthesis and assembly), was expressed significantly differential expression in ST2 (Figure 7F) [40,41]. *AT2G47190* (*AtMYB2*) and *AT3G06490* (*AtMYB108*) are homologous genes of *PebMYB51* that play crucial roles in response to drought and salt stress and regulate filament elongation and anther dehiscence through the jasmonic acid (JA) and gibberellic acid (GA) regulatory pathways [42,43]. The expression of *PebMYB51* displayed a rapid decline in the treatment group (Figure 7G). As the homologous gene of *PebMYB142*, *AT2G02820* (*AtMYB88*) is responsible for limiting cell division in the stomatal lineage and promoting stomatal closure in response to abiotic stress through the ABA regulatory pathway [44]. Furthermore, *AT3G23250* (*AtMYB15*, homologous to *PebMYB69*) was involved in responding to various stresses via ABA biosynthesis and signaling in *A. thaliana*. *PebMYB95* was homologous to *AT1G17950* (*AtMYB52*), which played crucial roles in cell wall architecture formation and affected ABA biosynthesis and response in *A. thaliana* [45]. In *P. eburnea*, the relative expression of *PebMYB142*, *PebMYB69*, and *PebMYB95* significantly differed between the treatment and control groups in ST1 and ST2 (Figure 7H–J and Figure S5).

PebMYB45 and *PebMYB57* shared homology with *AT1G08810* (*AtMYB60*), a gene known for its involvement in the drought stress response, and were specifically expressed in guard cells and promoted stomatal opening when highly expressed [46,47]. However, the relative expression levels of *PebMYB45* and *PebMYB57* showed no significant difference between the control and treatment groups in *P. eburnea* (Figure 7M,N). Similarly, no significant differences were observed for *PebMYB156*, which is a homologous gene of *AT5G56840* with functions related to dehydration stress memory and sugar metabolism (Figure 7L) [48,49]. The homologous genes of *PebMYB24*, *AT3G47600* (*AtMYB94*), and *AT5G62470* (*AtMYB96*) played important roles in cuticular wax biosynthesis and accumulation in response to drought stress. However, no significant differences in expression were found between the control and treatment groups (Figure 7K) [50,51].

These results suggest that seven *PebMYB* genes (*PebMYB3*, *PebMYB13*, *PebMYB17*, *PebMYB51*, *PebMYB142*, *PebMYB69*, and *PebMYB95*) likely play important roles in the drought stress response in *P. eburnea*. However, it appears that *PebMYB45*, *PebMYB57*, *PebMYB24*, and *PebMYB156* do not respond to drought stress, as observed in *A. thaliana*. Further studies are needed to validate these findings and to elucidate the specific functions of these genes in the molecular mechanisms underlying drought stress adaptation in *P. eburnea*.

2.7. The Potential Co-Expression Network between PebMYBs and Other TFs

MYB genes could cooperate with other TFs to regulate their expression. To investigate the cooperative interactions, we calculated the PCC between *PebMYB* genes and other TFs across four tissues (Supplementary Table S8). A total of 2359 TFs from 48 families were predicted in this study using a cutoff PCC value greater than 0.95. These TFs were used to construct an interaction network (Figure S6). Among the interacting TFs, members from the AP2/ERF, C2H2, and WRKY families displayed prominent TFs that interact with *PebMYB* genes (Figure S6A). Further investigation revealed interesting patterns of interaction clustering in the network. For example, a notable cluster comprised more than 30 R2R3-type *PebMYB* genes, forming a complex network (Figure S6B). This observation suggested that developmental processes may be regulated by an intriguing network of multiple interacting partners. The interaction network of *PebMYB* genes and other TFs provided useful information for further investigation of the interactions between *PebMYB* genes and other TFs.

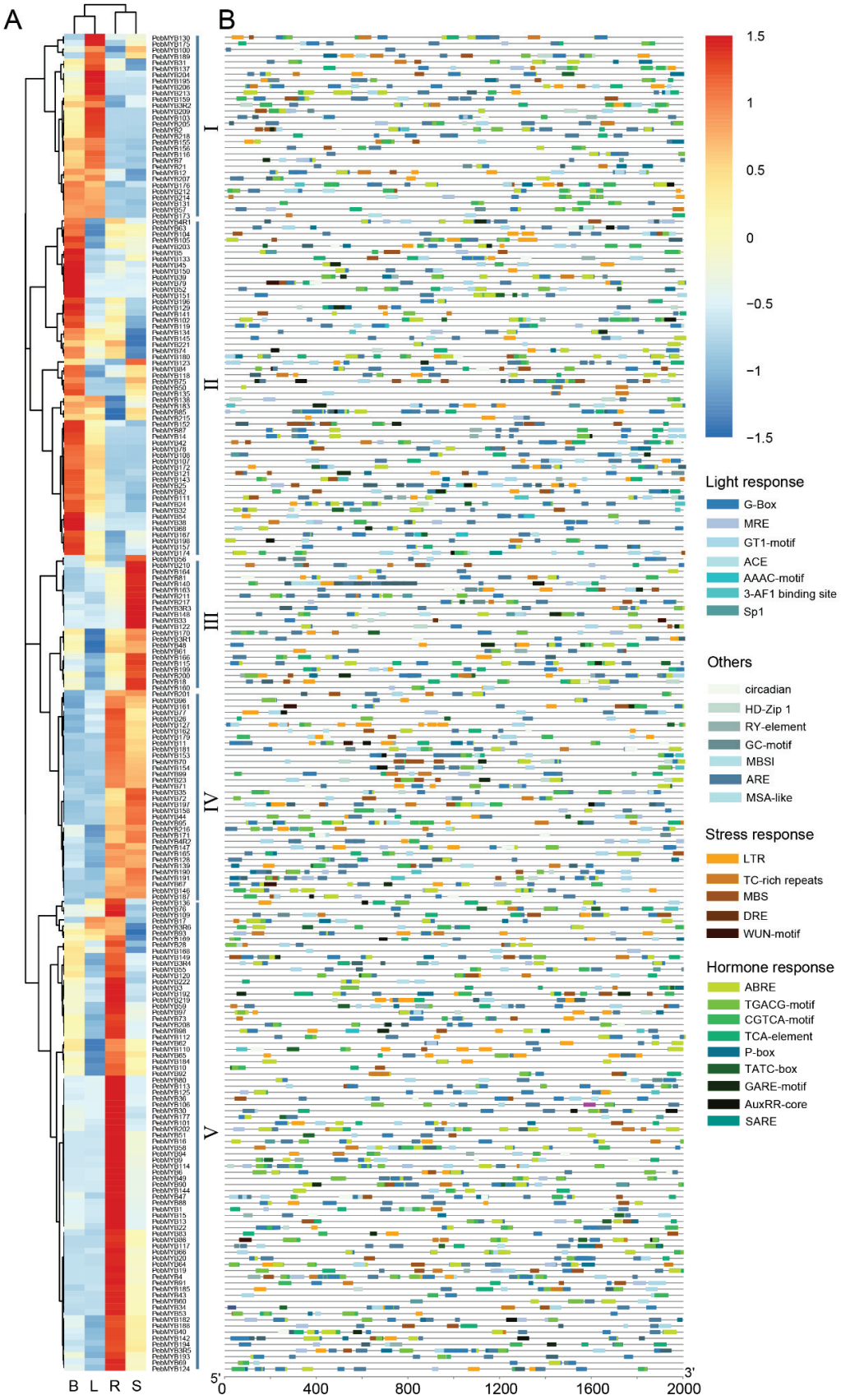


Figure 6. Expression pattern and promoter analysis of *PebMYB* genes. (**A**) The expression pattern of *PebMYB* genes in four different tissues (B, L, R, and S indicate bud, leaf, root, and stem, respectively) based on the RNA-seq data. (**B**) *Cis*-elements in the 2000 bp promoter region of *PebMYB* genes.

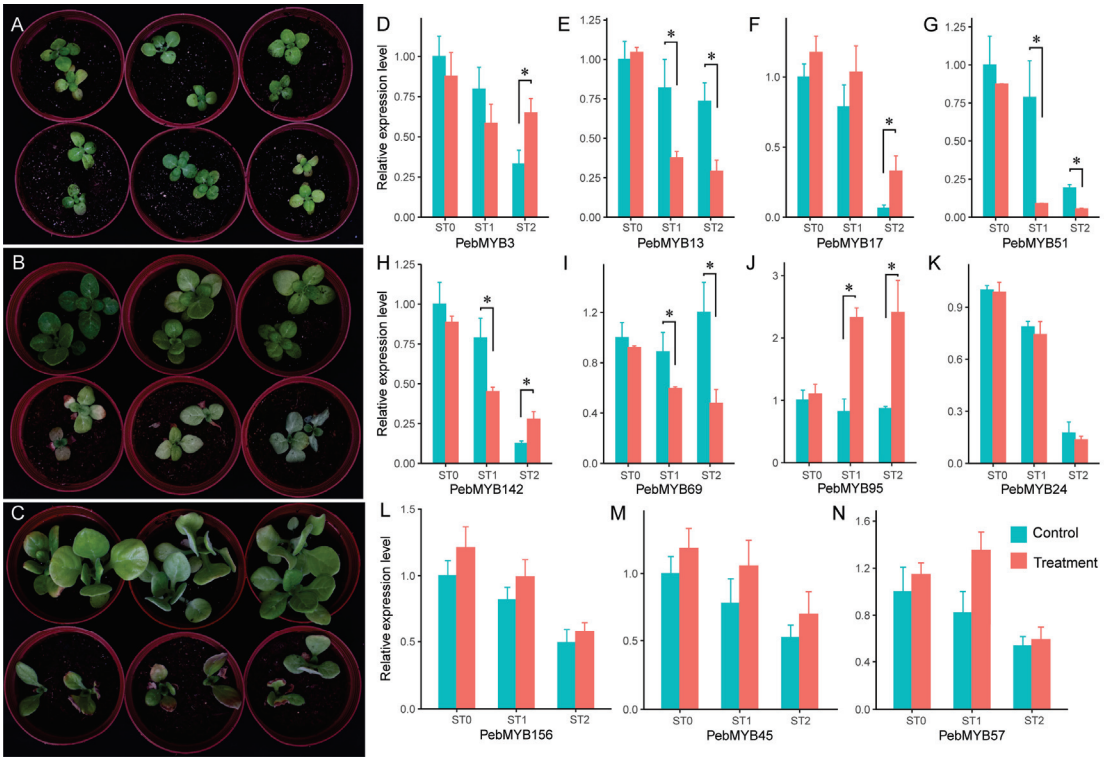


Figure 7. Expression analysis of selected *PebMYB* genes between the control and drought stress treatments. (**A**–**C**) The top three pots in (**A**–**C**) represent the control group, while the bottom three pots are the treatment group. The seedlings of *P. eburnea* before (**A**) and after drought stress treatment for one (**B**) and two months (**C**), respectively. (**D**–**N**) The relative expression levels of 11 selected *PebMYB* genes. ST0, ST1, and ST2 on the *x*-axis represent the seedling stages in (**A**–**C**), respectively. Asterisks above the columns show statistically significant differences at $p < 0.05$.

3. Discussion

P. eburnea is a promising candidate for development into ornamental plants, primarily due to its unique flower shape and colors [27,45]. Unlike the other species in the genus *Primulina*, which have limited endemic distributions, *P. eburnea* is widespread across several provinces in southern China, including diverse landscapes such as the Danxia and karst landforms. In addition, the natural soil conditions of its habitat do not retain water for extended periods, resulting in aperiodic drought stress. However, little is known about how *P. eburnea* adapts and responds to unpredictable drought stress. Unraveling the mechanisms by which *P. eburnea* adapts to the challenges of drought stress has important implications. Not only does it enrich our knowledge of how this species responds to drought conditions, but it also lays a foundation for advancing its conservation.

The MYB gene family is involved in regulating various biological activities, such as signal transduction, anthocyanin biosynthesis, and abiotic stress response [9,11,52]. However, comprehensive studies on the MYB gene family in *P. eburnea* have yet to be conducted. Here, we identified 230 *PebMYB* genes in *P. eburnea*. This number was greater than those found in *A. thaliana* (196) and *Populus trichocarpa* (197) but less than those in *Primulina swinglei* (264) and *Brassica rapa* (293) [9,53]. The reason may be the differential rates of gene family contraction or expansion among different lineages [54].

In the majority of subgroups, the number of *PebMYB* genes closely followed that of *AtMYB* genes, with a few exceptions. For instance, there was only one *AtMYB* gene in subgroup S4 while harboring four *PebMYB* genes. Similarly, S10 exhibited a comparable pattern with 11 *AtMYB* genes and 17 *PebMYB* genes. This gene expansion may be responsible for the gene duplication and differentiation of gene function [55]. For instance, *AtMYB91* in S4 is known to be involved in the specification of the leaf proximodistal axis [56]. Similar

observations were found in S10, where members (*AtMYB103*, *AtMYB26*, etc.) contribute to lignin biosynthesis and stamen development [57,58]. Conversely, gene contraction was also found in certain groups. For instance, nine *AtMYB* genes in S5 coexisted with four *PebMYB* genes, while S15 consisted of 16 *AtMYB* genes and nine *PebMYB* genes (Figure 1). The majority of members in S5 (*AtMYB115*, *AtMYB119*, etc.) were involved in the regulation of glucosinolate biosynthesis. These natural chemicals likely enhance plant defenses against pests and confer the characteristic bitter flavor property in cruciferous vegetables [59]. Members of S15 (*AtMYB15*, *AtMYB17*, *AtMYB28*, etc.) play roles in lignin biosynthesis and stress resistance [60–62]. The expansion or contraction of *PebMYB* genes may result from asymmetric gene duplication events in different subgroups [63].

Furthermore, certain *PebMYB* genes (*PebMYB45*, *PebMYB57*, *PebMYB47*, *PebMYB13*, etc.) may also be involved in the response to drought stress based on gene homology analysis and functional characterization. This finding was also confirmed by qRT-PCR analysis, showing that seven *PebMYB* genes were associated with the drought stress response (Figure 7D–J). Notably, the similar functions of the *PebMYB* genes and their homologous genes in *A. thaliana* were found in other species as well. For example, *AtMYB60* and its ortholog in grape (*Vitis vinifera*), *VvMYB60*, also respond to drought stress, salt stress, and ABA treatment [64]. The homologous gene of *AtMYB15* in *Chrysanthemum morifolium* (*CmMYB15*) was associated with biotic stress resistance [65].

Gene duplication events, such as whole-genome duplication, tandem duplication, and segmental duplication, are major sources of new gene formation and functional diversity, playing a crucial role in evolution [66]. In this study, we screened duplication events and found that 176 *PebMYB* genes were derived from segmental duplication. Remarkably, these genes accounted for 76.5% of the total *PebMYB* genes and a mere 1.03% of the total duplicated genes (17,026 genes). This finding suggested that segmental duplications likely contributed to the diversification and expansion of *PebMYB* genes under purifying selection, which was also observed in the study of pearl millet [67]. Interestingly, we found that only three *PebMYB* genes were derived from tandem duplication, suggesting a relatively limited role of tandem duplication in the expansion of MYB genes in *P. eburnea* that was also observed in soybean and tobacco [68,69].

4. Materials and Methods

4.1. Identification of MYB Members

The whole genome sequences and protein sequences of *P. eburnea* were retrieved from a previous study [70]. A total of 198 sequences of *Arabidopsis* MYB genes (including 126 2R-MYB, 64 1R-MYB, five 3R-MYB, and three 4R-MYB) were downloaded from TAIR (www.arabidopsis.org, accessed on 25 May 2023) according to a previous study [16]. The *Arabidopsis* protein sequences were used as queries in BLAST v2.13.0 to identify *P. eburnea* MYB candidates with an e-value $\leq 1 \times 10^{-5}$ and a bit score ≥ 100 [10,71]. The conserved domains of each candidate were confirmed in the InterPro database (www.ebi.ac.uk/interpro, accessed on 5 June 2023), while members lacking MYB domains were eliminated [72]. The Mw and pI of the proteins were determined by the TBtools program and the Expasy online program (https://web.expasy.org/compute_pi, accessed on 15 June 2023) [73,74]. The locations of *PebMYB* genes were obtained from the annotation file and visualized in TBtools.

4.2. Sequence Alignment and Phylogenetic Analysis

The protein sequences of *PebMYB* genes and 198 *AtMYB* genes were divided into two distinct categories based on their types (R2R3-MYB genes and the remaining types) and then used for multiple sequence alignment by MAFFT v7.505 software with default parameters [75,76]. To reveal the relationship between the *PebMYB* and *AtMYB* genes, phylogenetic analysis using the maximum-likelihood (ML) method was performed by IQ-TREE v1.6.12 software with parameters of 1000 ultrafast bootstraps and automatic

model selection [77]. The phylogenetic results were then visualized in the iTOL (http://itol.embl.de, accessed on 10 July 2023) online program [78].

4.3. Analysis of Gene Structure, Motif and Cis-Acting Elements

The online MEME program (https://meme-suite.org/meme/tools/meme, accessed on 25 July 2023) was used to investigate the conserved motifs in the MYB genes of *P. eburnea* with the following parameters: maximum number of motifs detected = 20 and number of repeats = any [79]. The annotation file of the *P. eburnea* genome was used to define the gene structure, and the *PebMYB* protein sequences were used to construct an ML tree with the aforementioned method.

The promoter sequences of *PebMYB* genes (2000 bp upstream of the start codon) were extracted from the genome sequences according to the annotation file. The sequences were used to search for potential *cis*-acting elements in the PlantCARE online program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on 5 September 2023). These elements were grouped into four groups based on their functions.

4.4. Gene Duplication and Synteny Analysis of PebMYB Genes

A BLASTp search was performed using 230 *PebMYB* protein sequences as queries against all protein sequences of *P. eburnea* with an e-value cutoff of 1×10^{-5}. The collinearity relationship of the genome was calculated using MCScanX software according to the BLAST results and the genome annotation results [80]. The segmentally duplicated genes, gene density, and location information in the given annotated file were then visualized in Circos v0.69.9, while genes situated on unanchored scaffolds were omitted [81]. The Ka/Ks were calculated in TBtools based on the collinearity results of MCScanX analysis. The collinearity results between *P. eburnea* and other species were analyzed in MCScanX and visualized with TBtools.

4.5. RNA-seq Data Analysis and Network Construction

The RNA-seq data of four tissues (bud leaf, mature leaf, root, and stem) in *P. eburnea* were retrieved from a previous study [70]. After mapping reads to the genome of *P. eburnea* in HISAT2, the FPKM values of each gene were generated [82]. The FPKM values were normalized with the Z score method and visualized by the pheatmap package of R v4.2.2 to generate a heatmap.

For interacting network analysis, all the TFs with an FPKM greater than 1 in any of the tissues were selected for calculating the Pearson correlation coefficient (PCC). Only the absolute value of PCC greater than 0.95 was considered a potential interaction. The interacting network was constructed by Cytoscape v3.10.0 [83].

4.6. Drought Stress Treatment and qRT-PCR

For drought stress treatment, the seeds of *P. eburnea* were germinated in plastic pots with soil in a chamber at 25 °C and relative humidity at 70%. To ensure the uniformity of seedlings, every two uniform seedlings were transplanted to a pot filled with the same amount of soil and grown at Lushan Botanical Garden, Chinese Academy of Sciences, Nanchang, Jiangxi Province, China (115.8382° E; 28.9112° N). Six pots were divided into two groups (control and treatment). All plants were irrigated equally before treatment until they reached the 4-leaf stage. Drought stress was induced by ceasing irrigation for the treatment group, while seedlings in the control group were irrigated normally. Seedling leaves were sampled at three specific time points: before the initiation of drought stress treatment (stage 0, ST0) and after one month (stage 1, ST1) and two months (stage 2, ST2) of exposure to drought stress treatment.

For qRT-PCR analysis, total RNA was extracted from samples using an *EASYspin* Plus Complex Plant RNA kit (Aidlab Biotech, Beijing, China) according to the manufacturer's instructions. First-strand cDNA was synthesized from 1 μg RNA with One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen, Beijing, China). qRTC-PCR

was performed with PerfectStart Green qPCR SuperMix (TransGen, Beijing, China) on a CFX Connect Real-Time System (Bio-Rad, Hercules, CA, USA). The expression levels of the selected *PebMYB* genes were determined according to the comparative CT ($2^{-\Delta\Delta Ct}$) method [84]. The *P. eburnea* glyceraldehyde-3-phosphate dehydrogenase (*PebGAPDH1*) was used as the internal reference gene. All qRT-PCRs were performed with three biological and three technical replicates. All primer sequences were designed by Primer Premier 5.0 and are listed in Supplementary Table S9.

5. Conclusions

In this study, a total of 230 *PebMYB* genes were identified in *P. eburnea*, including 67 1R-MYB, 155 R2R3-MYB, six 3R-MYB, and two 4R-MYB genes. A comprehensive analysis was performed to investigate phylogenetic relationships, motifs, gene structure, and synteny in *P. eburnea*. The prevalent purifying selection observed in the evolution of the *PebMYB* genes highlights their functional conservation based on the Ka/Ks results. Further analysis of the expression of the *PebMYB* genes in four different tissues revealed that they could be divided into five groups based on their expression characteristics. The expression profiles were verified by qRT-PCR analysis. Importantly, the identification of *PebMYB* genes that were homologous to drought-responsive genes in *A. thaliana*, as well as the characterization of their expression under drought stress, broadened our understanding of their potential roles in drought adaptation. The seven genes (*PebMYB3*, *PebMYB13*, *PebMYB17*, *PebMYB51*, *PebMYB142*, *PebMYB69*, and *PebMYB95*) showed significant differences between the control and drought treatments, suggesting that they are likely involved in the drought stress response in *P. eburnea*.

Our findings shed a basis on understanding MYB genes in *P. eburnea* and provide valuable insights for future studies investigating the molecular mechanisms underlying drought stress. Moreover, these results serve as a valuable resource for future genetic engineering and breeding programs aimed at improving the drought tolerance of *P. eburnea* and related species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25010465/s1.

Author Contributions: C.F. designed and conceived the study. J.Z. and Y.Z. performed the bioinformatic analysis and wet experiments. J.Z. wrote the paper with contributions from all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Biological Resources Program, Chinese Academy of Sciences (KFJ-BRP-007-013), Innovation Leading Talent Program in Jiangxi Province (JXSQ2023101107), and the Plant Special Project of Lushan Botanical Garden, Chinese Academic of Science (2022ZWZX05).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The transcriptomic sequencing data can be downloaded in the NCBI Sequence Read Archive under accession number PRJNA934730.

Acknowledgments: We would like to thank Nanchang Botanical Garden for providing the greenhouse for seedling cultivation.

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

References

1. Farooq, M.; Wahid, A.; Kobayashi, N.S.M.A.; Fujita, D.B.S.M.A.; Basra, S.M.A. Plant drought stress: Effects, mechanisms and management. *Agron. Sustain. Dev.* **2009**, *29*, 185–212. [CrossRef]
2. Zhang, H.; Zhu, J.; Gong, Z.; Zhu, J.K. Abiotic stress responses in plants. *Nat. Rev. Genet.* **2022**, *23*, 104–119. [CrossRef]
3. Shinozaki, K.; Yamaguchi-Shinozaki, K. Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* **2006**, *58*, 221–227. [CrossRef] [PubMed]
4. Zhu, J.K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **2002**, *53*, 247–273. [CrossRef] [PubMed]

5. Xiong, H.; Yu, J.; Miao, J.; Li, J.; Zhang, H.; Wang, X.; Liu, P.; Zhao, Y.; Jiang, C.; Yin, Z.; et al. Natural Variation in *OsLG3* Increases Drought Tolerance in Rice by Inducing ROS Scavenging. *Plant Physiol.* **2018**, *178*, 451–467. [CrossRef] [PubMed]
6. 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]
7. Zhang, L.; Zhao, G.; Xia, C.; Jia, J.; Liu, X.; Kong, X. A wheat R2R3-MYB gene, *TaMYB30-B*, improves drought stress tolerance in transgenic *Arabidopsis*. *J. Exp. Bot.* **2012**, *63*, 5873–5885. [CrossRef]
8. Du, H.; Zhang, L.; Liu, L.; Tang, X.-F.; Yang, W.-J.; Wu, Y.-M.; Huang, Y.-B.; Tang, Y.-X. Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry* **2009**, *74*, 1–11. [CrossRef]
9. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2010**, *15*, 573–581. [CrossRef]
10. Stracke, R.; Werber, M.; Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **2001**, *4*, 447–456. [CrossRef]
11. Jin, H.; Martin, C. Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* **1999**, *41*, 577–585. [CrossRef] [PubMed]
12. Zhou, L.; Yarra, R.; Jin, L.; Cao, H. Genome-wide identification and expression analysis of MYB gene family in oil palm (*Elaeis guineensis* Jacq.) under abiotic stress conditions. *Environ. Exp. Bot.* **2020**, *180*, 104245. [CrossRef]
13. Li, Y.; Liang, J.; Zeng, X.; Guo, H.; Luo, Y.; Kear, P.; Zhang, S.; Zhu, G. Genome-wide Analysis of MYB Gene Family in Potato Provides Insights into Tissue-specific Regulation of Anthocyanin Biosynthesis. *Hortic. Plant J.* **2021**, *7*, 129–141. [CrossRef]
14. Du, H.; Feng, B.-R.; Yang, S.-S.; Huang, Y.-B.; Tang, Y.-X. The R2R3-MYB transcription factor gene family in maize. *PLoS ONE* **2012**, *7*, e37463. [CrossRef] [PubMed]
15. Matus, J.T.; Aquea, F.; Arce-Johnson, P. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. *BMC Plant Biol.* **2008**, *8*, 83. [CrossRef] [PubMed]
16. Chen, Y.; Yang, X.; He, K.; Liu, M.; Li, J.; Gao, Z.; Lin, Z.; Zhang, Y.; Wang, X.; Qiu, X.; et al. The MYB transcription factor superfamily of *Arabidopsis*: Expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* **2006**, *60*, 107–124. [CrossRef]
17. Baldoni, E.; Genga, A.; Cominelli, E. Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. *Int. J. Mol. Sci.* **2015**, *16*, 15811–15851. [CrossRef]
18. Park, M.Y.; Kang, J.-y.; Kim, S.Y. Overexpression of *AtMYB52* confers ABA hypersensitivity and drought tolerance. *Mol. Cells* **2011**, *31*, 447–454. [CrossRef]
19. Wu, J.; Jiang, Y.; Liang, Y.; Chen, L.; Chen, W.; Cheng, B. Expression of the maize MYB transcription factor *ZmMYB3R* enhances drought and salt stress tolerance in transgenic plants. *Plant Physiol. Biochem.* **2019**, *137*, 179–188. [CrossRef]
20. Li, S.; Xin, Z.-B.; Chou, W.-C.; Huang, Y.; Pan, B.; Maciejewski, S.; Wen, F. Five new species of the genus *Primulina* (Gesneriaceae) from Limestone Areas of Guangxi Zhuangzu Autonomous Region, China. *PhytoKeys* **2019**, *127*, 77–91. [CrossRef]
21. Weber, A.; Middleton, D.J.; Forrest, A.; Kiew, R.; Lim, C.L.; Rafidah, A.; Sontag, S.; Triboun, P.; Wei, Y.-G.; Yao, T.L. Molecular systematics and remodelling of Chirita and associated genera (Gesneriaceae). *Taxon* **2011**, *60*, 767–790. [CrossRef]
22. Wang, Y.; MAO, R.; Liu, Y.; LI, J.M.; Dong, Y.; LI, Z.Y.; Smith, J.F. Phylogenetic reconstruction of Chirita and allies (Gesneriaceae) with taxonomic treatments. *J. Syst. Evol.* **2011**, *49*, 50–64. [CrossRef]
23. Hao, Z.; Kuang, Y.; Kang, M.; Niu, S. Untangling the influence of phylogeny, soil and climate on leaf element concentrations in a biodiversity hotspot. *Funct. Ecol.* **2015**, *29*, 165–176. [CrossRef]
24. Valente, L.; Phillimore, A.B.; Melo, M.; Warren, B.H.; Clegg, S.M.; Havenstein, K.; Tiedemann, R.; Illera, J.C.; Thébaud, C.; Aschenbach, T.; et al. A simple dynamic model explains the diversity of island birds worldwide. *Nature* **2020**, *579*, 92–96. [CrossRef]
25. Martín-Queller, E.; Albert, C.; Dumas, P.-J.; Saatkamp, A. Islands, mainland, and terrestrial fragments: How isolation shapes plant diversity. *Ecol. Evol.* **2017**, *7*, 6904–6917. [CrossRef] [PubMed]
26. Wang, Z.; Huang, S.; Hong, X.; Wen, F. Pollination biology of *Primulina eburnea*. *Guihaia* **2021**, *41*, 671–683.
27. Feng, C.; Feng, C.; Kang, M. The first genetic linkage map of *Primulina eburnea* (Gesneriaceae) based on EST-derived SNP markers. *J. Genet.* **2016**, *95*, 377–382. [CrossRef] [PubMed]
28. Yue, A.W.; Fang, W.; Qi-xiang, Z. Evaluation on Ornamental Characterstics and Selection for Promising Species and Varieties of Chirita and Chiritopsis Plants in Guangxi. *Acta Hortic. Sin.* **2008**, *35*, 239–250.
29. Wang, W.; Ryu, K.H.; Bruex, A.; Barron, C.; Schiefelbein, J. Molecular Basis for a Cell Fate Switch in Response to Impaired Ribosome Biogenesis in the *Arabidopsis* Root Epidermis. *Plant Cell* **2020**, *32*, 2402–2423. [CrossRef]
30. Liang, T.; Shi, C.; Peng, Y.; Tan, H.; Xin, P.; Yang, Y.; Wang, F.; Li, X.; Chu, J.; Huang, J.; et al. Brassinosteroid-Activated *BRI1-EMS-SUPPRESSOR 1* Inhibits Flavonoid Biosynthesis and Coordinates Growth and UV-B Stress Responses in Plants. *Plant Cell* **2020**, *32*, 3224–3239. [CrossRef]
31. Maier, A.; Schrader, A.; Kokkelink, L.; Falke, C.; Welter, B.; Iniesto, E.; Rubio, V.; Uhrig, J.F.; Hülskamp, M.; Hoecker, U. Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in Arabidopsis. *Plant J.* **2013**, *74*, 638–651. [CrossRef] [PubMed]
32. Wilkins, O.; Nahal, H.; Foong, J.; Provart, N.J.; Campbell, M.M. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* **2009**, *149*, 981–993. [CrossRef] [PubMed]

33. Zhang, Z.; Zhang, L.; Liu, Y.; Shang, X.; Fang, S. Identification and Expression Analysis of R2R3-MYB Family Genes Associated with Salt Tolerance in *Cyclocarya paliurus*. *Int. J. Mol. Sci.* **2022**, *23*, 3429. [CrossRef] [PubMed]
34. Qing, J.; Dawei, W.; Jun, Z.; Yulan, X.; Bingqi, S.; Fan, Z. Genome-wide characterization and expression analyses of the MYB superfamily genes during developmental stages in Chinese jujube. *PeerJ* **2019**, *7*, e6353. [CrossRef] [PubMed]
35. Jia, T.; Ge, Q.; Zhang, S.; Zhang, Z.; Liu, A.; Fan, S.; Jiang, X.; Feng, Y.; Zhang, L.; Niu, D.; et al. UDP-Glucose Dehydrogenases: Identification, Expression, and Function Analyses in Upland Cotton (*Gossypium hirsutum*). *Front. Genet.* **2021**, *11*, 597890. [CrossRef] [PubMed]
36. Persak, H.; Pitzschke, A. Dominant repression by *Arabidopsis* transcription factor MYB44 causes oxidative damage and hypersensitivity to abiotic stress. *Int. J. Mol. Sci.* **2014**, *15*, 2517–2537. [CrossRef] [PubMed]
37. Lü, B.B.; Li, X.J.; Sun, W.W.; Li, L.; Gao, R.; Zhu, Q.; Tian, S.M.; Fu, M.Q.; Yu, H.L.; Tang, X.M.; et al. *AtMYB44* regulates resistance to the green peach aphid and diamondback moth by activating EIN2-affected defences in Arabidopsis. *Plant Biol.* **2013**, *15*, 841–850. [CrossRef]
38. Shin, R.; Burch, A.Y.; Huppert, K.A.; Tiwari, S.B.; Murphy, A.S.; Guilfoyle, T.J.; Schachtman, D.P. The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *Plant Cell* **2007**, *19*, 2440–2453. [CrossRef]
39. Liang, Y.-K.; Dubos, C.; Dodd, I.C.; Holroyd, G.H.; Hetherington, A.M.; Campbell, M.M. *At*MYB61, an R2R3-MYB Transcription Factor Controlling Stomatal Aperture in Arabidopsis thaliana. *Curr. Biol.* **2005**, *15*, 1201–1206. [CrossRef]
40. Kosma, D.K.; Murmu, J.; Razeq, F.M.; Santos, P.; Bourgault, R.; Molina, I.; Rowland, O. *AtMYB41* activates ectopic suberin synthesis and assembly in multiple plant species and cell types. *Plant J.* **2014**, *80*, 216–229. [CrossRef]
41. Cominelli, E.; Sala, T.; Calvi, D.; Gusmaroli, G.; Tonelli, C. Over-expression of the Arabidopsis *AtMYB41* gene alters cell expansion and leaf surface permeability. *Plant J.* **2008**, *53*, 53–64. [CrossRef] [PubMed]
42. Mandaokar, A.; Thines, B.; Shin, B.; Lange, B.M.; Choi, G.; Koo, Y.J.; Yoo, Y.J.; Choi, Y.D.; Choi, G.; Browse, J. Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant J.* **2006**, *46*, 984–1008. [CrossRef] [PubMed]
43. Urao, T.; Yamaguchi-Shinozaki, K.; Urao, S.; Shinozaki, K. An *Arabidopsis* myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* **1993**, *5*, 1529–1539. [CrossRef] [PubMed]
44. Xie, Z.; Li, D.; Wang, L.; Sack, F.D.; Grotewold, E. Role of the stomatal development regulators *FLP/MYB88* in abiotic stress responses. *Plant J.* **2010**, *64*, 731–739. [CrossRef] [PubMed]
45. Zhong, R.; Ye, Z.H. Secondary cell walls: Biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* **2015**, *56*, 195–214. [CrossRef] [PubMed]
46. Rusconi, F.; Simeoni, F.; Francia, P.; Cominelli, E.; Conti, L.; Riboni, M.; Simoni, L.; Martin, C.R.; Tonelli, C.; Galbiati, M. The *Arabidopsis thaliana MYB60* promoter provides a tool for the spatio-temporal control of gene expression in stomatal guard cells. *J. Exp. Bot.* **2013**, *64*, 3361–3371. [CrossRef] [PubMed]
47. Negi, J.; Moriwaki, K.; Konishi, M.; Yokoyama, R.; Nakano, T.; Kusumi, K.; Hashimoto-Sugimoto, M.; Schroeder, J.I.; Nishitani, K.; Yanagisawa, S.; et al. A Dof Transcription Factor, SCAP1, Is Essential for the Development of Functional Stomata in *Arabidopsis*. *Curr. Biol.* **2013**, *23*, 479–484. [CrossRef] [PubMed]
48. Ding, Y.; Liu, N.; Virlouvet, L.; Riethoven, J.-J.; Fromm, M.; Avramova, Z. Four distinct types of dehydration stress memory genes in *Arabidopsis thaliana*. *BMC Plant Biol.* **2013**, *13*, 229. [CrossRef]
49. Muller, R.; Morant, M.; Jarmer, H.; Nilsson, L.; Nielsen, T.H. Genome-Wide Analysis of the Arabidopsis Leaf Transcriptome Reveals Interaction of Phosphate and Sugar Metabolism. *Plant Physiol.* **2006**, *143*, 156–171. [CrossRef]
50. Lee, S.B.; Suh, M.C. Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in *Arabidopsis*. *Plant Cell Physiol.* **2015**, *56*, 48–60. [CrossRef]
51. Seo, P.J.; Lee, S.B.; Suh, M.C.; Park, M.J.; Go, Y.S.; Park, C.M. The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in *Arabidopsis*. *Plant Cell* **2011**, *23*, 1138–1152. [CrossRef] [PubMed]
52. Zhang, T.; Zhao, Y.; Wang, Y.; Liu, Z.; Gao, C. Comprehensive Analysis of MYB Gene Family and Their Expressions Under Abiotic Stresses and Hormone Treatments in *Tamarix hispida*. *Front. Plant Sci.* **2018**, *9*, 1303. [CrossRef] [PubMed]
53. Jin, J.; Tian, F.; Yang, D.-C.; Meng, Y.-Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2016**, *45*, D1040–D1045. [CrossRef] [PubMed]
54. Guo, Y.-L. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. *Plant J.* **2013**, *73*, 941–951. [CrossRef] [PubMed]
55. Chen, Q.; Zhang, X.; Fang, Y.; Wang, B.; Xu, S.; Zhao, K.; Zhang, J.; Fang, J. Genome-Wide Identification and Expression Analysis of the R2R3-MYB Transcription Factor Family Revealed Their Potential Roles in the Flowering Process in Longan (*Dimocarpus longan*). *Front. Plant Sci.* **2022**, *13*, 820439. [CrossRef] [PubMed]
56. Luo, L.; Ando, S.; Sakamoto, Y.; Suzuki, T.; Takahashi, H.; Ishibashi, N.; Kojima, S.; Kurihara, D.; Higashiyama, T.; Yamamoto, K.T.; et al. The formation of perinucleolar bodies is important for normal leaf development and requires the zinc-finger DNA-binding motif in Arabidopsis ASYMMETRIC LEAVES2. *Plant J.* **2020**, *101*, 1118–1134. [CrossRef]
57. Öhman, D.; Demedts, B.; Kumar, M.; Gerber, L.; Gorzsás, A.; Goeminne, G.; Hedenström, M.; Ellis, B.; Boerjan, W.; Sundberg, B. MYB103 is required for *FERULATE-5-HYDROXYLASE* expression and syringyl lignin biosynthesis in Arabidopsis stems. *Plant J.* **2013**, *73*, 63–76. [CrossRef]

58. Nelson, M.R.; Band, L.R.; Dyson, R.J.; Lessinnes, T.; Wells, D.M.; Yang, C.; Everitt, N.M.; Jensen, O.E.; Wilson, Z.A. A biomechanical model of anther opening reveals the roles of dehydration and secondary thickening. *New Phytol.* **2012**, *196*, 1030–1037. [CrossRef]
59. Zhang, Y.; Li, B.; Huai, D.; Zhou, Y.; Kliebenstein, D.J. The conserved transcription factors, MYB115 and MYB118, control expression of the newly evolved benzoyloxy glucosinolate pathway in *Arabidopsis thaliana*. *Front. Plant Sci.* **2015**, *6*, 343. [CrossRef]
60. Ma, K.W.; Niu, Y.; Jia, Y.; Ordon, J.; Copeland, C.; Emonet, A.; Geldner, N.; Guan, R.; Stolze, S.C.; Nakagami, H.; et al. Coordination of microbe-host homeostasis by crosstalk with plant innate immunity. *Nat. Plants* **2021**, *7*, 814–825. [CrossRef]
61. Coleto, I.; Bejarano, I.; Marín-Peña, A.J.; Medina, J.; Rioja, C.; Burow, M.; Marino, D. Arabidopsis thaliana transcription factors MYB28 and MYB29 shape ammonium stress responses by regulating Fe homeostasis. *New Phytol.* **2021**, *229*, 1021–1035. [CrossRef] [PubMed]
62. Curci, P.L.; Zhang, J.; Mähler, N.; Seyfferth, C.; Mannapperuma, C.; Diels, T.; Van Hautegem, T.; Jonsen, D.; Street, N.; Hvidsten, T.R.; et al. Identification of growth regulators using cross-species network analysis in plants. *Plant Physiol.* **2022**, *190*, 2350–2365. [CrossRef] [PubMed]
63. Jiang, C.K.; Rao, G.Y. Insights into the Diversification and Evolution of R2R3-MYB Transcription Factors in Plants. *Plant Physiol.* **2020**, *183*, 637–655. [CrossRef] [PubMed]
64. Galbiati, M.; Matus, J.T.; Francia, P.; Rusconi, F.; Cañón, P.; Medina, C.; Conti, L.; Cominelli, E.; Tonelli, C.; Arce-Johnson, P. The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress. *BMC Plant Biol.* **2011**, *11*, 142. [CrossRef] [PubMed]
65. An, C.; Sheng, L.; Du, X.; Wang, Y.; Zhang, Y.; Song, A.; Jiang, J.; Guan, Z.; Fang, W.; Chen, F.; et al. Overexpression of *CmMYB15* provides chrysanthemum resistance to aphids by regulating the biosynthesis of lignin. *Hortic. Res.* **2019**, *6*, 84. [CrossRef] [PubMed]
66. Zhang, X.; Wang, H.; Chen, Y.; Huang, M.; Zhu, S. Comprehensive Genome-Wide Analyses of Poplar R2R3-MYB Transcription Factors and Tissue-Specific Expression Patterns under Drought Stress. *Int. J. Mol. Sci.* **2023**, *24*, 5389. [CrossRef] [PubMed]
67. Lin, M.; Dong, Z.; Zhou, H.; Wu, G.; Xu, L.; Ying, S.; Chen, M. Genome-Wide Identification and Transcriptional Analysis of the MYB Gene Family in Pearl Millet (*Pennisetum glaucum*). *Int. J. Mol. Sci.* **2023**, *24*, 2484. [CrossRef]
68. Zhu, Y.; Wu, N.; Song, W.; Yin, G.; Qin, Y.; Yan, Y.; Hu, Y. Soybean (*Glycine max*) expansin gene superfamily origins: Segmental and tandem duplication events followed by divergent selection among subfamilies. *BMC Plant Biol.* **2014**, *14*, 93. [CrossRef]
69. Yang, J.; Zhang, B.; Gu, G.; Yuan, J.; Shen, S.; Jin, L.; Lin, Z.; Lin, J.; Xie, X. Genome-wide identification and expression analysis of the R2R3-MYB gene family in tobacco (*Nicotiana tabacum* L.). *BMC Genom.* **2022**, *23*, 432. [CrossRef]
70. Zhang, Y.; Zhang, J.; Zou, S.; Liu, Z.; Huang, H.; Feng, C. Genome-wide analysis of the cellulose toolbox of *Primulina eburnea*, a calcium-rich vegetable. *BMC Plant Biol.* **2023**, *23*, 259. [CrossRef]
71. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef] [PubMed]
72. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; et al. InterPro in 2022. *Nucleic Acids Res.* **2022**, *51*, D418–D427. [CrossRef] [PubMed]
73. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
74. Bjellqvist, B.; Basse, B.; Olsen, E.; Celis, J.E. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **1994**, *15*, 529–539. [CrossRef] [PubMed]
75. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [CrossRef] [PubMed]
76. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
77. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
78. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
79. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 28–36.
80. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]
81. Krzywinski, M.I.; Schein, J.E.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef] [PubMed]
82. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef] [PubMed]

83. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
84. 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]

International Journal of
Molecular Sciences

MDPI

Article

Multi-Omics Analysis Reveals Intricate Gene Networks Involved in Female Development in Melon

Zhongyuan Wang, Vivek Yadav, Xiaoyao Chen, Siyu Zhang, Xinhao Yuan, Hao Li, Jianxiang Ma, Yong Zhang, Jianqiang Yang, Xian Zhang * and Chunhua Wei *

State Key Laboratory of Crop Stress Biology in Arid Areas, College of Horticulture, Northwest A&F University, Xianyang 712100, China; zydx@nwafu.edu.cn (Z.W.); vivekyadav@nwafu.edu.cn (V.Y.); xiaoyaochen7186@163.com (X.C.); zhangsiyu@nwafu.edu.cn (S.Z.); yuanxinhao@nwafu.edu.cn (X.Y.); yuanyilihao123@163.com (H.L.); majianxiang@126.com (J.M.); zhangyong123@nwsuaf.edu.cn (Y.Z.); yangjq1208@126.com (J.Y.)

* Correspondence: zhangxian@nwafu.edu.cn (X.Z.); xjwend020405@nwafu.edu.cn (C.W.); Tel.: +86-029-8708-2613 (C.W.)

Abstract: Sexual differentiation is an important developmental phenomenon in cucurbits that directly affects fruit yield. The natural existence of multiple flower types in melon offers an inclusive structure for studying the molecular basis of sexual differentiation. The current study aimed to identify and characterize the molecular network involved in sex determination and female development in melon. Male and female pools separated by the F_2 segregated generation were used for sequencing. The comparative multi-omics data revealed 551 DAPs and 594 DEGs involved in multiple pathways of melon growth and development, and based on functional annotation and enrichment analysis, we summarized four biological process modules, including ethylene biosynthesis, flower organ development, plant hormone signaling, and ubiquitinated protein metabolism, that are related to female development. Furthermore, the detailed analysis of the female developmental regulatory pathway model of ethylene biosynthesis, signal transduction, and target gene regulation identified some important candidates that might have a crucial role in female development. Two CMTs ((cytosine-5)-methyltransferase), one AdoHS (adenosylhomocysteinase), four ACSs (1-aminocyclopropane-1-carboxylic acid synthase), three ACOs (ACC oxidase), two ARFs (auxin response factor), four ARPs (auxin-responsive protein), and six ERFs (Ethylene responsive factor) were identified based on various female developmental regulatory models. Our data offer new and valuable insights into female development and hold the potential to offer a deeper comprehension of sex differentiation mechanisms in melon.

Keywords: cucurbits; ethylene; omics; sex differentiation; female development

Citation: Wang, Z.; Yadav, V.; Chen, X.; Zhang, S.; Yuan, X.; Li, H.; Ma, J.; Zhang, Y.; Yang, J.; Zhang, X.; et al. Multi-Omics Analysis Reveals Intricate Gene Networks Involved in Female Development in Melon. *Int. J. Mol. Sci.* **2023**, *24*, 16905. https://doi.org/10.3390/ijms242316905

Academic Editor: Wajid Zaman

Received: 11 October 2023
Revised: 21 November 2023
Accepted: 24 November 2023
Published: 29 November 2023

1. Introduction

Fruit development and seed formation in plants are directly linked with flowering patterns. Many plant species naturally exist with perfect flowers. Our current understanding of natural sex differentiation in many plants is not extensive. The mechanism of sex differentiation and development is limited because many model plants are not dioecious in nature [1,2]. Sexual differentiation in plants is crucial for reproduction. In practice, the economic values often differ between the existence of male and female flowers. Female flowers can produce fruits and seeds, while male flowers are the only source of pollen to fertilize female flowers [3]. Understanding sexual differentiation is essential in plant breeding programs, enabling the development of cultivars with desirable traits [3,4]. The first transcriptome-based study for sex-related genes was carried out in *Silene latifolia*, where dosage compensation in the sex chromosome was identified [5]. Transcriptome-based studies were carried out in many plants, including some cucurbits, to reveal the genetic mechanisms involved in sex differentiation and flower development. For instance,

transcriptome profiling of the flower buds of *Coccinia grandis* (L.) Voigt was performed by Mohanty [6]. The results revealed that many hormone-responsive genes, including ADP-ribosylation factor 6 (ARF6), 1-Aminocyclopropane-1-carboxylic acid synthase (ACC/ACS), sucrose non-fermenting-1-related protein kinase 2 (SNRK2), and BRI1-associated receptor kinase 1 (BAK1), were involved in flower differentiation [6].

The Cucurbitaceae family is composed of 120 genera and 960 species, which include many economically important cultivated species [4]. The Cucurbitaceae family holds significant agricultural and nutritional importance due to its diverse members, including cucumbers, pumpkins, and watermelons [1,7–9]. These plants are valued for their economic contributions as staple food crops and sources of essential nutrients. In melon, some studies have shown important information about floral sex differentiation and development. For example, the mechanisms regulating the bisexual flowers in melon by transcriptome showed that the ethylene downstream gene *CmERF1* significantly contributes to bisexual bud formation [10]. In another study, bulk segregant populations were utilized for transcriptome analysis, which revealed ethylene, abscisic acid (ABA), auxin/indole-3-acetic acid (IAA), and aminocyclopropanecarboxylate (ACC) oxidase were involved in different sex type expressions, suggesting that they might have important roles in melon sex determination [11]. The transcriptome studies provided some reliable findings regarding in-depth information about sex forms in many crops. However, there are some limitations in transcriptome-based studies, such as that gene expression abundance may not be reliable and not always precisely reflect the proteins coding the phenotypic character or biological activities [12,13]. Therefore, post-transcriptional modification should be validated through proteomics studies [14,15].

The most common sex forms in cultivated melon plants are andromonoecious, where bisexual flowers are only found on the leaf axil of lateral branches and male flowers on the main stem and lateral branches [10,16]. The fruit development is from bisexual flowers in plants [17]. Melon flowers serve as a model system for studying physiological and molecular aspects of sex determination in plants due to the availability of different types of stable sex forms in melon species. Moreover, the flowering pattern in melons also provides an ideal system of choice for researchers to study the biological process from flowering to fruit development [18]. Based on the distribution of flowers of different sexes on the same plant, melons can be classified as andromonoecy, androecy, gynoecy, hermaphrodite, and monoecy [19]. In melon, all four whorls of floral organ primordia emerge at the preliminary stage of flower development. However, due to the function of the sex-determining gene, one of the sex organ primordia (stamen or carpels) stagnates at a certain stage, resulting in abortion and loss of function in this sex organ and leading to bisexual, female, and male flowers in melon [20]. The floral primordia of melon flowers are initially bisexual, and the differentiation in sex forms occurs at lateral stages of development [21]. The restricted development of the carpel or the stamen results in a female flower or a unisexual male flower. The research reports to date suggest that sex differentiation in melon is mainly regulated by a set of genes that includes *CmACS7*, *CmWIP1*, and *CmACS11* [21–23]. It was evidenced that the expression of *CmACS7* is limited to the carpel of bisexual and female flowers, which makes it an important gene that regulates sex differentiation in melon [22]. Further studies revealed restricted expression of *CmACS7* in the male flower, and it inhibits stamen development in the female flower of melon. Some reports also revealed that mutations in *CmACS7* can result in the formation of bisexual or male flowers [22]. On the other hand, *CmWIP1* acts as an inhibitor that suppresses female organ development through a complex network where it has a strong interaction with *CmbZIP48* [24]. In *CmWIP1* mutant plants, only bisexual or female flowers developed. Some research results revealed that *CmWIP1* can suppress the expression of *CmACS7* because it works as an upstream suppressor on male flowers [20,24]. Similarly, *CmACS11* has an important role at an upstream level that restricts the expression of *CmWIP1*. The expression of the *CmACS11* gene was not detected in male flowers; rather, the expression was detected in vascular bundles of female flowers or bisexual flowers. Therefore, it was assumed that *CmACS11*

plays a critical role in female sex organ development in melon flowers [21]. The mutant of *CmACS11* showed a generation of androecy sex forms in flowers. In addition to the genetic network controlling sex expression in cucurbits, phytohormones, including gibberellin (GA) and ethylene (ET), profoundly affect flower expression in cucurbits [25]. External factors such as temperature, photoperiod, and ethylene-controlling agents also regulate flowering patterns in cucurbits [26]. The ET-based pathways are thoroughly studied in many cucurbits because they significantly regulate sex expression [8,27]. For instance, the development of female flowers required more ethylene than what is required for male flowers in watermelon [1]. Ethephon and IAA applications in pumpkins significantly increased the number of female flowers [28]. Similarly, the application of an ethylene inhibitor resulted in a higher number of male flowers in cucumber [29]. In the case of melon, the external application of ethephon treatment boosts the number of bisexual and female flowers [16].

However, it is quite possible that other genes, complex genetic networks, and non-genetic factors could contribute to sex expression [11]. The application of multi-omics can lead to some concrete conclusions and contribute to more detailed information on sex expression research in cucurbit crops. The main mechanisms and some candidate pathways that regulate sex differentiation in melon are controlled by several mechanisms, and some important mechanisms are still elusive and require further research [30]. Therefore, the current study was designed to decipher the gene networks that may contribute to female flower development. We mapped the comparative transcriptome and total proteome profiling of male and female dominant flower bud pools generated from crossing in melon plants. In the current study, stable '0426' (monoecious) and 'Y101' (hermaphrodite line) were used as parent material because of their distinct flowering characteristics [26]. With the advancement of genomic techniques, we have more opportunities to link the available information with previously identified results. The current research provides important data about molecular mechanisms involved in female flower development in plants. Transcriptome data supported by proteome analysis of male and female dominant pools evidenced an intricate gene network that contributes to female flower development in melon.

2. Results

2.1. Proteome and Transcriptome Raw Data Analysis and Protein/Gene Identification

Melon sex determination primarily occurs during the early stage of flower bud development and cannot be distinguished by the naked eye. The female phenotype 'Y101' and the male phenotype '0426' differ in the mechanism of sex differentiation. In the F_2 population, there are two female phenotypes (gynoicous and hermaphrodite) and two male phenotypes (monoecious and andromonoecious). To eliminate the influence of individual sex type differences, two types of mixed pools (F pool and M pool) were selected for comparative analysis from the F_2 isolated population. Due to the specific temporal and spatial nature of the sex-determining gene, the apex tissue, which is rich in flower bud primordia, was selected for proteome and transcriptome analysis.

The original data analysis of the proteome showed that a total of 41,998 peptides were obtained from the female pool (F) and the male pool (M). Among these, 7130 proteins were identified. After comparison with the melon reference gene proteome (DHL92, V4.0), a total of 39,812 (94.79%) peptide matches were found. Finally, 6870 (96.35%) melon-specific proteins were identified (Figure 1). Among the specific proteins, more than 76.27% are composed of more than two peptides (Figure 1C). The identified proteins (6870) have a wide range of molecular weight distribution, with the majority (78.05%) concentrated in the range of 10–70 kDa. Among these, proteins of size 30–40 kDa are the most abundant, with 1131 instances.

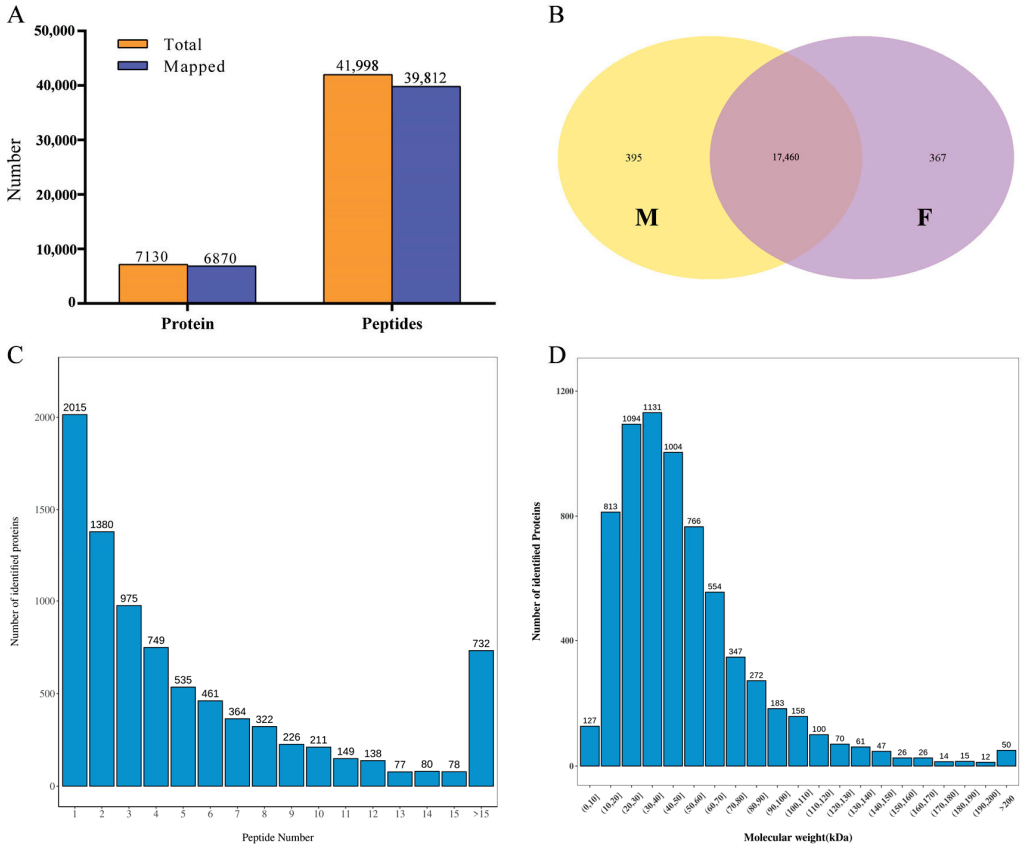


Figure 1. Overview of the multi-omics data. (**A**) Protein and peptide identification information; (**B**) Venn diagram of transcriptome gene identification; (**C**) Statistics of protein peptides; (**D**) Protein molecular mass distribution map (kDa).

The transcriptome raw data analysis revealed that 45,179,104 and 44,568,262 clean reads were obtained from the F pool and M pool samples, respectively. The matching rate after comparison with the reference genome (DHL92, V3.6.1) was above 95%. Approximately 94% of the total matching reads were uniquely mapped to a single position in the genome (Unique map), while around 1% of the reads matched multiple positions (Multimap) in the genome. Both positive and negative matches accounted for approximately 47% each (Table 1). From the matched reads, a total of 18,222 melon genes were identified. Among these genes, 17,460 were shared between the female phenotype (F) and the male phenotype (M), while 367 genes were specific to the female phenotype and 395 genes were specific to the male phenotype (Figure 1B).

Table 1. Overview of the transcriptome data.

Pool	Total	Total Map	Unique Map	Multimap	Positive	Negative
F	45,179,104	43,344,617 (95.94%)	42,839,380 (94.82%)	505,237 (1.12%)	21,413,961 (47.4%)	21,425,419 (47.42%)
M	44,568,262	42,604,523 (95.59%)	42,122,518 (94.51%)	482,005 (1.08%)	2,1056,584 (47.25%)	2,1065,934 (47.27%)

2.2. Identification and Analysis of Differential Proteins and Differential Genes

Among the identified proteins, the expression value of the F pool protein was higher than the expression value of the M pool protein, with a fold change value. The significance of the difference between F and M was determined using a *t*-test, resulting in a *p*-value. To screen for differentially abundant proteins (DAPs), the following criteria were applied: fold change ≥ 1.5 or fold change ≤ 0.67, and *p*-value ≤ 0.05. As a result, a total of 551 DAPs

were screened from the identified proteins, out of which 434 DAPs were up-regulated and 117 DAPs were down-regulated (Figure 2A,B and Tables S1 and S2).

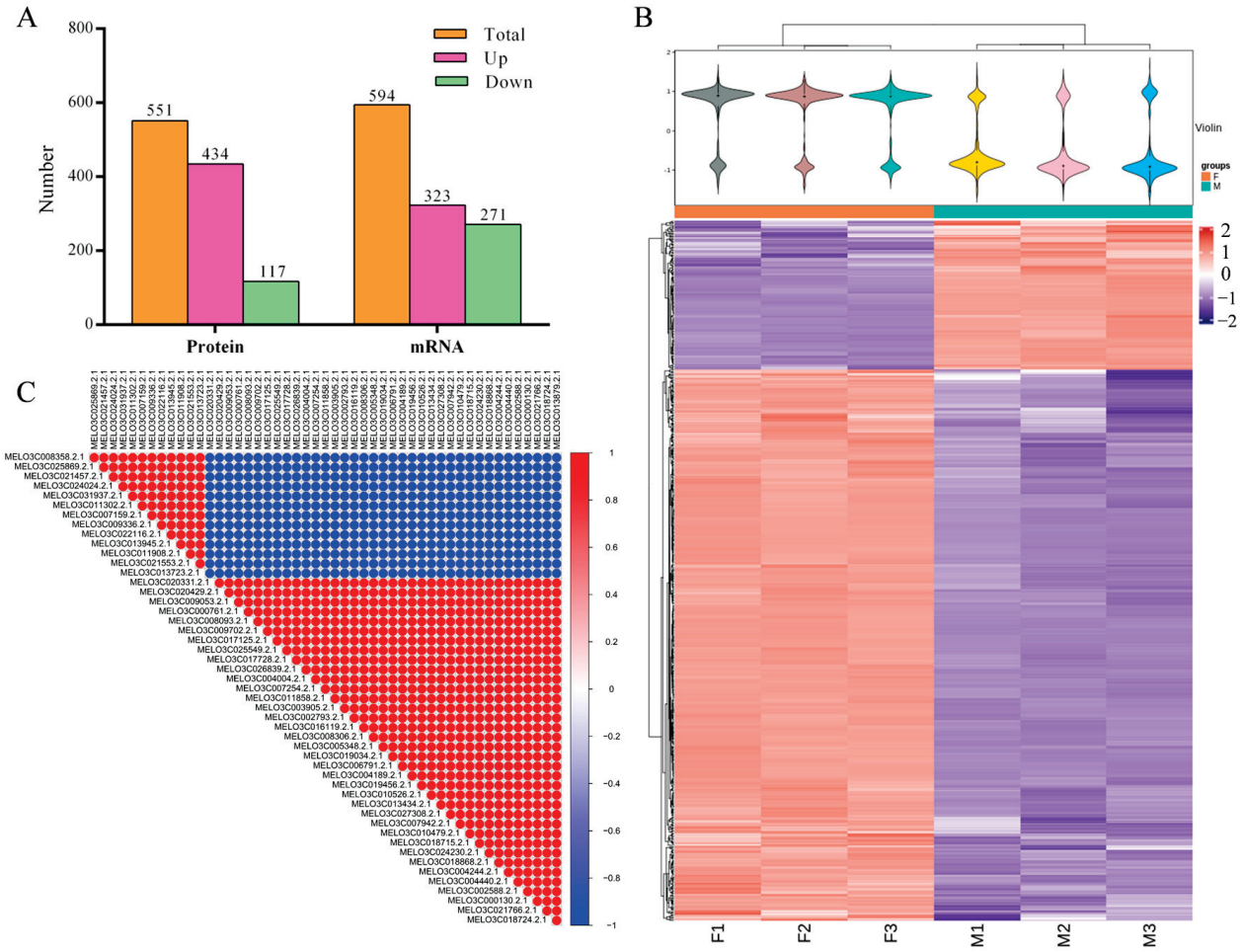


Figure 2. Differential protein/gene identification and analysis. (**A**) Differential protein and differential gene statistics; (**B**) Heat map of differential protein clustering of female (F1, F2, and F3) and male (M1, M2, and M3) pools; (**C**) Correlation analysis of the expression of the top 50 DAPs, with the color bar representing the correlation coefficient.

In this study, correlation analysis was conducted on the top 50 differentially significant proteins (ranked by *p*-value). It was found that 13 DAPs exhibited a highly negative correlation with the remaining 47 DAPs, showing opposite expression patterns. This indicates the presence of opposite mechanisms of action between these genes in melon female regulation (Figure 2C). Regarding the 18,222 genes identified by RNA-seq, the expression value of the F pool gene was compared to that of the M pool gene using fold change values, and the significance of the difference was determined using a *t*-test and a *p*-value. The screening condition used was $|\log2 \text{ (Fold Change)}| > 0.4$ and $p\text{-value} < 0.05$, resulting in a total of 594 differentially expressed genes (DEGs) identified from the gene set. Among these, 323 DEGs were up-regulated and 271 DEGs were down-regulated (Figure 2A and Tables S3 and S4).

2.3. Omics Data Validation Test

Out of the total identified differently expressed proteins, 17 up-regulated proteins and 3 down-regulated proteins were selected from a pool of 551 differential proteins based on descriptions in previous studies, and their roles in various pathways were studied and compared to identify candidates. Among these proteins, 7 are hormone-related proteins, including genes related to auxin and gibberellin. Additionally, there are 5 ethylene

synthesis-related proteins, 2 transcription factors associated with growth and development, 2 proteins associated with flower organ development, 6 signal transduction-related proteins, and 2 membrane proteins. Figure 3 illustrates that the majority of the protein-coding genes, specifically 19 in total, display concordant patterns of transcriptional activity that correspond with their protein expression profiles. Furthermore, the female phenotypic parent 'Y101' and the male phenotypic parent '0426' show a strong correlation with the F pool and M pool, respectively. However, it is worth noting that the gene expression level of one ABA response protein (MELO3C010850.2.1) contradicts the observed protein expression trend (Figure 3N). This finding suggests that the transcription level of a few protein-coding genes may not be fully representative of their protein abundance (Figure 3).

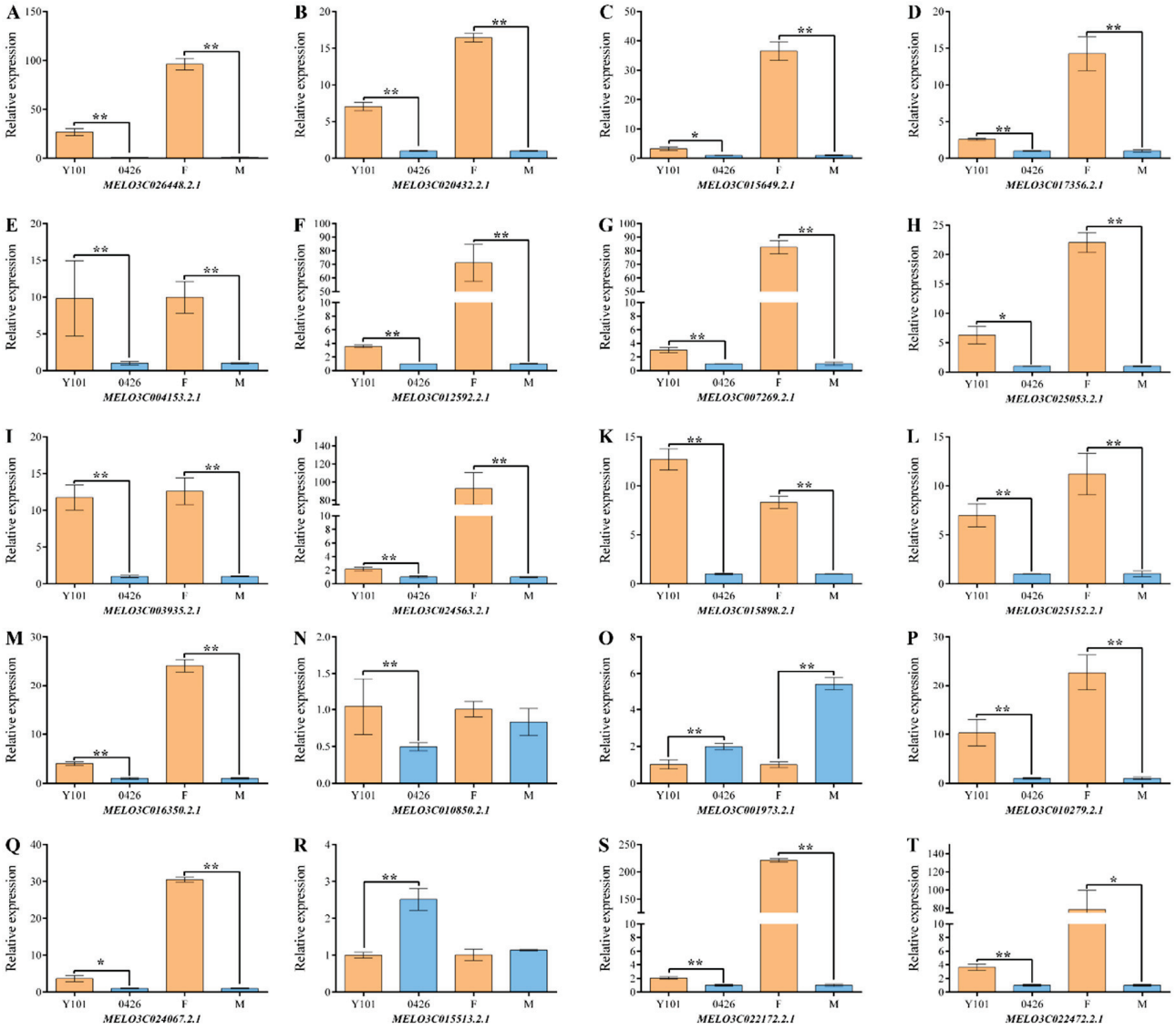


Figure 3. Identification of transcription levels of differential proteins. The DAPs that are down-regulated include (**N**,**O**,**R**), whereas those that are up-regulated include the remaining set. (**A**–**E**) ethylene synthesis-related protein; (**F**) gibberellin-associated protein; (**G**) auxin-associated protein; (**H**,**I**) proteins associated with flower organ development; (**J**–**O**) phytohormone signaling and transduction; (**P**) sexual cycle protein; (**Q**,**R**) transcription factors that regulate growth and development; (**S**,**T**) membrane proteins. The relative expression (*Y*-axis) was calculated using the $2^{\Delta\Delta Ct}$ method, and melon varieties ('Y101' and '0426') with F and M pools are shown on the *X*-axis. Bar graphs are plotted with the mean ± SD of three replications. * and ** represent significant differences in expression levels at $p < 0.05$ and $p < 0.01$, respectively (Student's *t*-test).

Similarly, 15 up-regulated genes and 5 down-regulated genes were selected from a pool of 594 differentially expressed genes based on GO, KEGG, and functional annotation analysis of identified DEGs. These include 5 genes associated with flower organ development, 2 members of the ACS family of ethylene synthetic rate-limiting enzymes, 2 genes involved in hormone response, 2 zinc finger protein-coding genes, and one transmembrane protein-coding gene. Figure 4 shows that the expression trends of all the genes in the female (F) pool and male (M) pool are consistent with the transcriptome data. Moreover, the female phenotypic parent 'Y101' and the male phenotypic parent '0426' exhibit the same expression trend as the F pool and M pool (Figure 4).

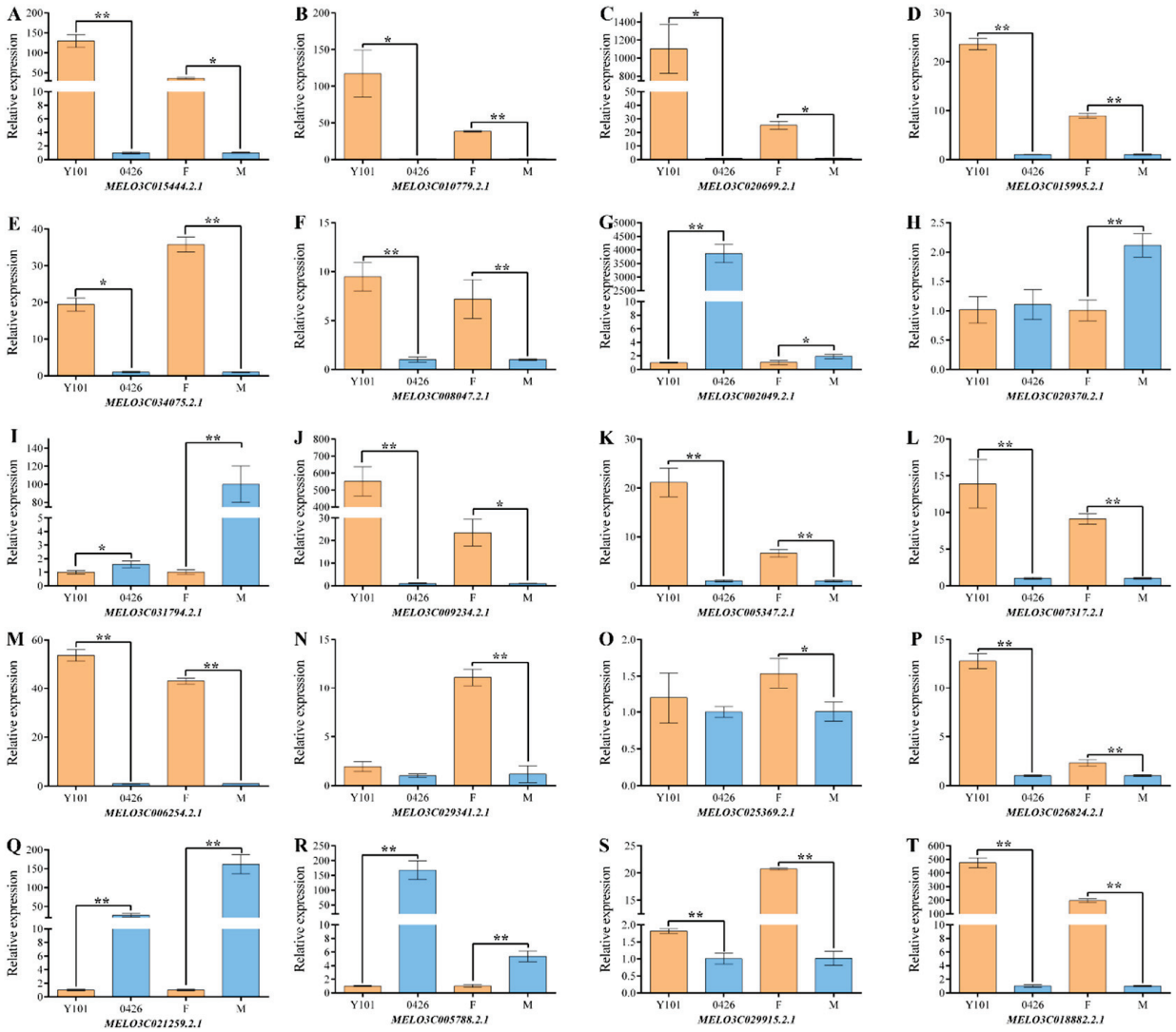


Figure 4. qRT-PCR verification of differential genes. The DEGs that are down-regulated include (**G**–**I**,**Q**,**R**), whereas those that are up-regulated include the remaining set. (**A**,**B**) ethylene synthesis-related proteins; (**C**) gibberellin-associated protein; (**D**) auxin-associated protein; (**E**–**I**) proteins associated with flower organ development; (**J**,**K**) zinc finger proteins; (**L**) transmembrane protein; (**M**) membrane proteins; (**N**–**R**) enzymes; (**S**,**T**) unknown proteins. The relative expression (*Y*-axis) was calculated using the $2^{\Delta\Delta Ct}$ method, and melon varieties ('Y101' and '0426') with F and M pools are shown on the *X*-axis. Bar graphs are plotted with the mean ± SD of three replications. * and ** represent significant differences in expression levels at $p < 0.05$ and $p < 0.01$, respectively (Student's *t*-test).

In summary, the proteome and transcriptome data obtained in this study exhibit high confidence and accurately reflect protein/gene expression levels, thereby providing

a reliable foundation for further analytical studies. Consequently, researchers can now proceed with confidence and utilize the data from both sources.

2.4. Differential Protein and Differential Gene Function Enrichment Analysis

GO and KEGG enrichment analyses were conducted to verify the authenticity of the data and explore the potential regulatory network and related pathways of female traits in melon. A total of 551 differential proteins and 594 differential genes were subjected to analysis. The 551 differential proteins were categorized into three groups based on their GO functions: 380 DAPs were annotated for molecular function, 297 DAPs for biological processes, and 279 DAPs for cellular components (Figure 5A). The top 30 entries in the GO annotation revealed that the largest proportion of DAPs in terms of molecular functions belonged to "ATP binding" (GO:0005524), with 92 DAPs (90 up-regulated and 2 down-regulated). This was followed by "Metal binding" (GO:0046872) and "DNA binding" (GO:0003677), with 45 (41 up-regulated, 4 down-regulated) and 31 (27 up-regulated, 4 down-regulated) DAPs, respectively. Additionally, "Zinc ion binding" (GO:0008270) showed a high proportion of 15 DAPs, all of which were up-regulated. In terms of cellular components, the differential proteins were widely distributed in various organelles, such as the nucleus, cytoplasm, mitochondria, and plasma membrane. The nucleus (GO:0005634) accounted for the largest proportion, with 59 DAPs (46 up-regulated, 13 down-regulated). Regarding biological processes, the 297 DAPs were primarily involved in protein transport (GO:0006886 and GO:0006606), energy metabolism (GO:0005975), transcription and transcriptional regulation (GO:0006351 and GO:0006355), and ubiquitinated protein metabolism (GO:0006511) (Figure 5A).

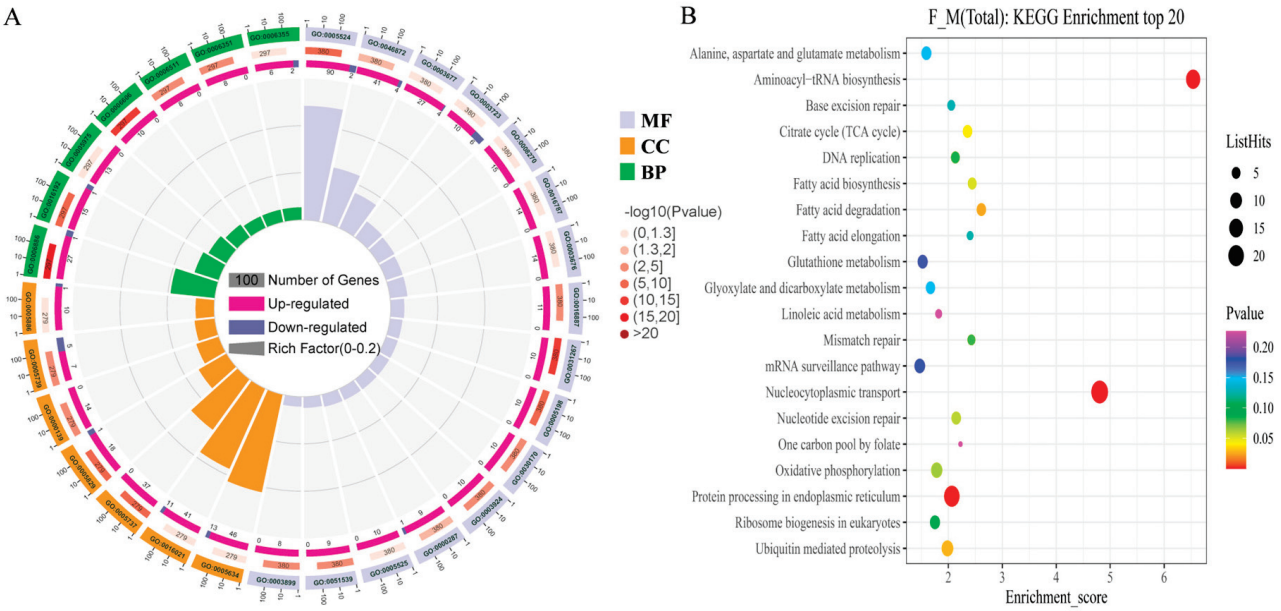


Figure 5. GO and KEGG enrichment analyses of differential proteins. (**A**) TOP30 information under GO item (GO entries with the number of corresponding differential proteins greater than 1 in the three categories were screened, and 10 entries were sorted from largest to smallest according to the −log10pvalue corresponding to each entry). MF: molecular function; CC: component; BP: biological process. (**B**) TOP20 KEEG enrichment information (each entry is sorted from largest to smallest by its −log10pvalue). The dot size represents the number of DAPs under this enriched item.

The KEGG enrichment analysis identified a total of 132 DAPs, distributed among the top 20 enrichment categories. Notably, the category of "Nucleocytoplasmic transport" (cmo03013), which encompasses the transport of molecules between the nucleus and cytoplasm, exhibited the highest representation with 24 DAPs (23 up-regulated and 1 down-regulated). Following this, the category of "Protein processing in endoplasmic reticulum"

(cmo:04141), involved in the synthesis and folding of proteins, contained 20 DAPs (10 up-regulated and 10 down-regulated). Another significant category was "Ubiquitin-mediated proteolysis" (CMO:04120), which is associated with the degradation of proteins by the ubiquitin system. This category is composed of 10 DAPs, all of which were up-regulated. Furthermore, two additional categories, "Alanine, aspartate, and glutamate metabolism" (cmo:00250) and "Plant hormone signal transduction" (cmo:04075), related to the ethylene synthesis pathway, exhibited notable proportions with 7 DAPs each. Within the "Alanine, aspartate, and glutamate metabolism" category, all 7 DAPs were up-regulated. However, the distribution of up-regulated and down-regulated DAPs within the "Plant hormone signal transduction" category was not specified (Figure 5B).

GO function analysis annotated 594 differential genes, out of which 62 were included in the top 30 entries, divided into three categories: molecular function, cell composition, and biological process. From a molecular function perspective, the largest proportion is associated with catalytic activity, specifically "Binding" and "Catalytic activity," which contain 38 and 20 differential genes, respectively. The main functional locations of the differentially expressed genes are distributed in the organelle ("Organelle") and the membrane system ("Membrane"). These differential genes are primarily involved in the biological processes of the "Cellular process", "Metabolic process", and "Developmental process," which include 43, 39, and 23 differential genes, respectively (Figure 6A). KEGG enrichment analysis revealed 51 differential genes among the top 20 enriched entries. Among them, particular attention was given to "Plant hormone signal transduction" (CMO:04075), which contains 6 differential genes, and "Cysteine methionine metabolism" (CMO:00270), which contains 2 differential genes. Based on the functional enrichment analysis of differential proteins and genes, it can be inferred that transcriptional regulation occurring in the nucleus plays a significant role in the female regulatory pathway of melon, with transcription factors containing zinc ions being relatively important. Additionally, plant hormone synthesis and signal transduction processes are closely related to female differentiation, particularly the significance of ethylene. Ubiquitination-mediated metabolism is also important (Figure 6B).

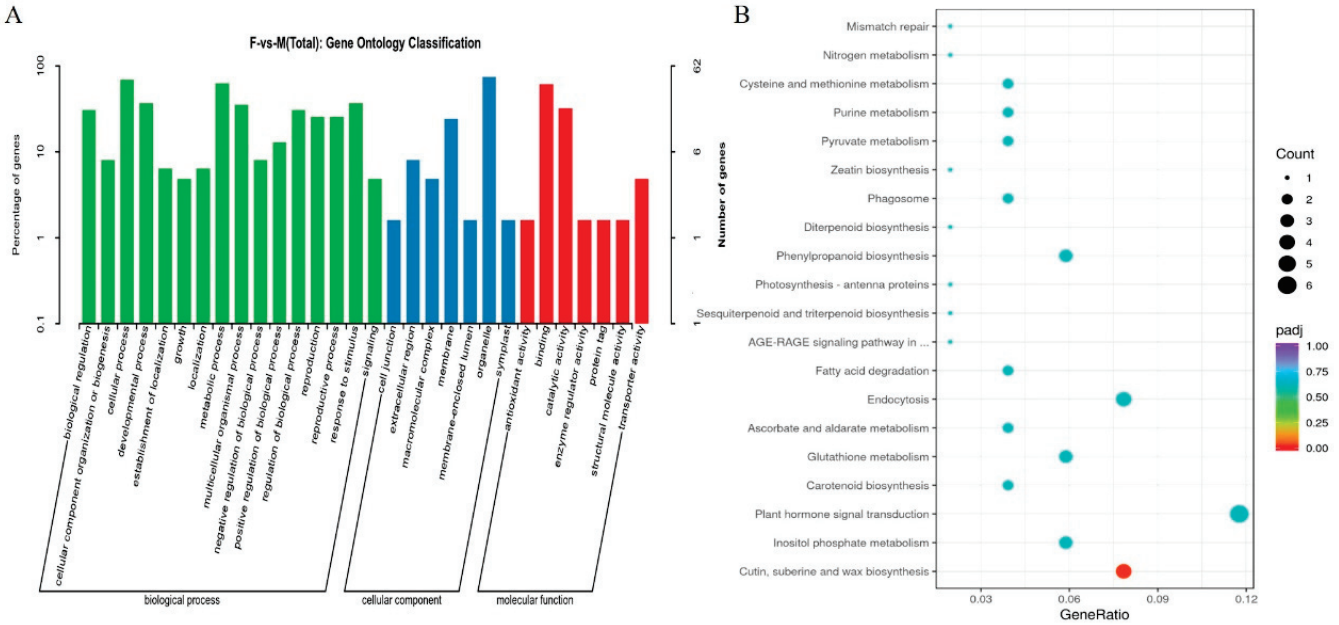


Figure 6. GO and KEGG enrichment analyses of differential genes. (**A**) TOP30 information under the GO item (GO entries with the number of corresponding differential proteins greater than 1 in the three categories were screened, and 10 entries were sorted from largest to smallest according to the −log10pvalue corresponding to each entry). (**B**) TOP20 KEEG enrichment information (each entry is sorted from largest to smallest by its −log10pvalue). The dot size represents the number of DEGs under the enriched item.

2.5. Differential Protein and Differential Gene Interaction Network Regulates Female Differentiation in Melon

Through functional enrichment analysis of 551 differentially abundant proteins and 594 differentially expressed genes, we have made preliminary inferences about the relevant pathways involved in the regulation of female development. The interactions between these differentially expressed proteins and genes within these pathways are believed to play a role in the regulation of female development. In this study, we performed an interaction network analysis of the differentially expressed proteins and genes potentially involved in sex regulation to investigate their interaction relationships.

In this study, 38 DAPs and 43 DEGs were screened to identify their potential involvement in the sex-determining pathway (Table S5). After performing a functional enrichment analysis, it was found that these 81 DAPs/DEGs were primarily associated with ethylene biosynthesis, regulation of flower organ development, plant hormone signaling, and ubiquitinated protein metabolism. These four biological process modules formed a sex regulatory network (Figure 7A). Ethylene, which plays a crucial role in female differentiation, was found to be influenced by the up-regulation of eight DAPs and seven DEGs involved in the ethylene synthesis pathway (Figure 7B,C). Notably, SAH (MELO3C020432.2.1), CMT3 (MELO3C015649.2.1), and MET4 (MELO3C026448.2.1) were found to interact with multiple DAPs and DEGs. Additionally, two ACS genes (ACS7 and ACS11) and two ACO genes (ACO1 and ACO2) exhibited high connectivity in the network (Figure 7A). The interaction network revealed the significant roles of the ethylene receptor ETR (MELO3C003906.2.1) and MAPK protein kinase (MELO3C005705.2.1) in plant hormone signaling. These proteins were crucial intermediaries between ethylene synthesis and flower organ development (Figure 7A). They directly or indirectly promoted the expression of DL (MELO3C034075.2.1) and HEC3 (MELO3C008047.2.1), which are associated with carpel development. Notably, DL exhibited high expression levels among all the network DEGs. Moreover, ETR and MAPK indirectly inhibited the expression of another development-related gene, TGA9.1 (MELO3C021427.2.1) and TGA9.2 (MELO3C004728.2.1), through flower allotypes PMADS2 (MELO3C010515.2.1) and DEFA (MELO3C003778.2.1). Among all the network DEGs, TGA9.2 displayed lower expression levels (Figure 7B,C). Furthermore, it was observed that almost all proteins related to plant hormone signaling were directly or indirectly associated with the ubiquitinated protein degradation pathway. This finding suggests that the ubiquitinated protein degradation pathway plays a crucial role in hormone signaling and the melon sex-determining pathway.

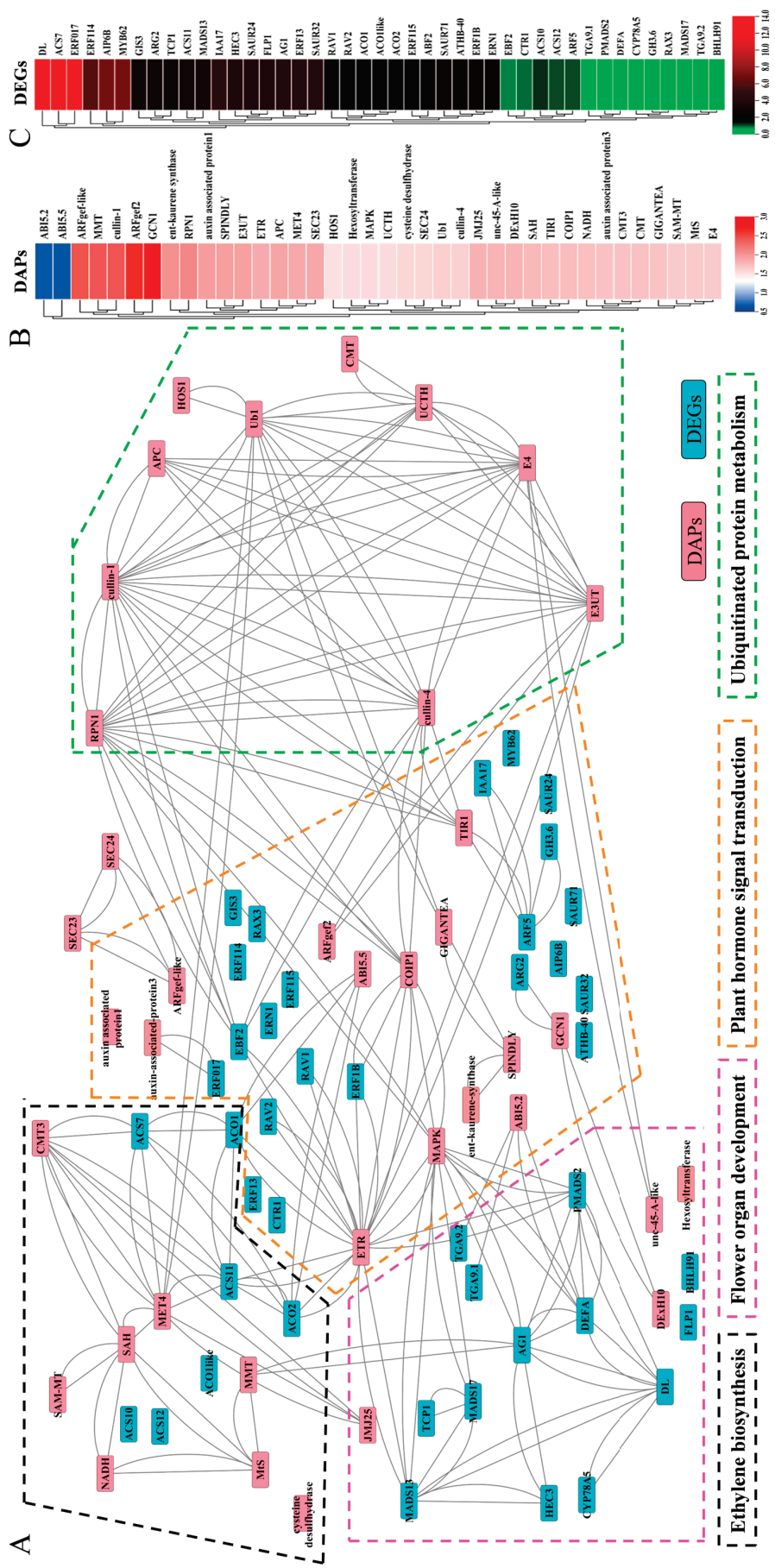


Figure 7. Integrated network analysis of key DAPs and DEGs. (**A**) Interaction network of key DAPs and DEGs involved in ethylene synthesis, floral organ development, plant hormone signaling, and ubiquitinated protein metabolism. The nodes represent DAPs or DEGs, the solid lines represent interaction relationships, and the dashed lines in different colors represent biological process modules. (**B**,**C**) Heat maps of DAP and DEG expression in the network. The color bars are fold-change values.

2.6. Ethylene-Mediated Study of Female Differentiation Regulatory Pathways in Melon

According to omics analysis and previous research reports, this study proves that the ethylene-mediated sex determination pathway plays a key regulatory role in female development. A regulatory pathway model for melon female development is proposed based on functional enrichment analysis and interaction network analysis of differential proteins and differential genes.

As shown in Figure 8, ethylene is synthesized de novo with methionine (L-Methionine) as a substrate. S-adenosylmethionine synthionine synthase catalyzes the formation of S-adenosylmethionine (S-AdoMet). Part of S-AdoMet continues to enter the ethylene synthesis pathway, while another part enters the methionine cycle to continue synthesizing methionine. In the methionine cycle, two key enzymes catalyze the process: (cytosine-5)-methyltransferase (CMT; MELO3C026448.2.1, and MELO3C015649.2.1) and adenosine homocysteinase (AdoHS; MELO3C020432.2.1). Both enzymes have higher expression levels of protein (FvsM) and therefore contribute to more ethylene synthesis substrate (L-Methionine) in the female phenotype (F). S-adenosylmethionine is then catalyzed by ACS to form ACC (an ethylene precursor). There are four ACS-coding genes (MELO3C015444.2.1, MELO3C010779.2.1, MELO3C005597.2.1, and MELO3C006840.2.1) that are highly expressed (FvsM) to promote the production of more ethylene precursors. ACC is subsequently decomposed by ACO (ACC oxidase) to produce ethylene molecules. At this stage, three ACO-coding genes (MELO3C006437.2.1, MELO3C019735.2, and MELO3C014437.2) exhibit up-regulated expression (FvsM), catalyzing the production of more ethylene molecules. Ethylene acts as a small gas molecule for signaling by binding to ethylene receptors in the membrane system. The female phenotype has more ethylene receptors (MELO3C003906.2.1) capable of binding to C_2H_4 molecules (Figure 8). Ethylene signaling is mediated by MPK6, EIN2, and EIN3. The expression level of the MPK6 (MELO3C005705.2.1) protein is high in the female phenotype, and EIN3 is inhibited by EBF1/2 in this pathway. At this stage, the auxin signaling mechanism plays an important role. A series of auxin response proteins (ARPs) (AUX/IAA, GH3, and SAUR) can inhibit EBF1/2 [31]. In females, four auxin response protein-coding genes (MELO3C011268.2.1, MELO3C002378.2.1, MELO3C020765.2.1, and MELO3C013403.2) show high transcription levels, weakening the inhibitory effect of EBF1/2 on EIN3 in female manifestations. Auxin response factor (ARF) plays a crucial signaling role by producing ARPs that not only inhibit EBF1/2 but also directly up-regulate ACS expression [32]. In the female phenotype, two ARFs (MELO3C006181.2.1 and MELO3C023883.2.1) have extremely high protein abundance levels (Figure 8). Ethylene undergoes a signaling process that activates the expression of downstream ethylene response factor (ERF), which can promote the auxin signaling process. In addition, the ERF ultimately promotes melon female development through the regulation of downstream target genes. Six highly expressed ERFs (MELO3C015543.2.1, MELO3C026158.2.1, MELO3C005940.2.1, MELO3C006430.2.1, MELO3C006149.2.1, and MELO3C022358.2.1) may have important regulatory roles in this process.

Based on the above results, we finally proposed a pathway model for female regulation in melon. The model starts with the de novo synthesis of ethylene, which undergoes signal transduction and polyhormonal co-regulation of downstream target genes, ultimately leading to female sexual differentiation (Figure 8).

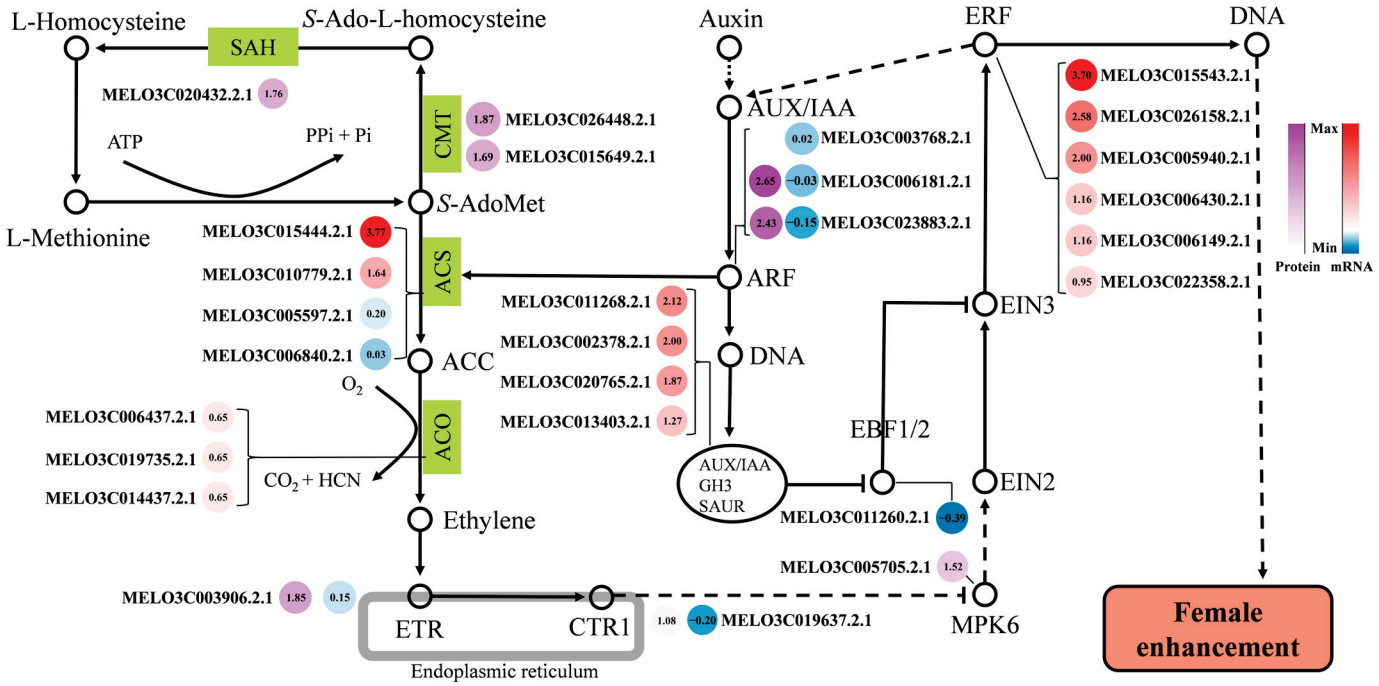


Figure 8. Ethylene-mediated regulation of female development in melon. Hollow dots represent molecules or products, green squares represent catalytic enzymes at that stage, sharp lines represent synthesis or promotion, T-shaped lines represent inhibition, dashed lines represent the presence of other processes, different colored origin points represent DAPs or DEGs, and protein color bars represent fold-change values of DAPs. mRNA color bars represent Log_2 fold-change values of DEGs.

3. Discussion

3.1. Candidate Genes for Melon Female Development Revealed by Omics

Mining key genes involved in sex determination has always been a top priority in melon sexual differentiation research. For many years, researchers have been searching for sex-determining genes in melons using forward and reverse genetic approaches. While genetic analysis and genetic mapping have the advantages of high accuracy and strong relevance, they are often limited by the presence of mutants in the population. This is where the advantages of functional omics can be better utilized. Functional omics examines the differences between different sexes comprehensively and explores sex-determining genes and proteins through comparative studies involving sex-determining pathways. In this study, the differences between female and male phenotypes of melons were systematically analyzed at the multi-omics level, including the proteome and transcriptome. Through integrated comparative analysis, several important genes related to female development were discovered.

The study identified 551 DAPs and 594 DEGs involved in multiple pathways of melon growth and development. Based on functional annotation and enrichment analysis, four biological process modules closely related to female development were summarized: ethylene biosynthesis, flower organ development, plant hormone signaling, and ubiquitinated protein metabolism. These modules included 38 DAPs and 43 DEGs. Ethylene, a gaseous plant hormone that regulates sex differentiation in melon [3,33,34], plays a crucial role in female flower development. Therefore, two genes coding for cytosine-specific methyltransferase (CMT) (*MELO3C026448.2.1* and *MELO3C015649.2.1*), one gene coding for Adenosylhomocysteinase (AdoHS) (*MELO3C020432.2.1*), four genes coding for ACS (*MELO3C015444.2.1*, *MELO3C010779.2.1*, *MELO3C005597.2.1*, and *MELO3C006840.2.1*), and three genes coding for 1-aminocyclopropane-1-carboxylate oxidase homolog 1-like (ACO) (*MELO3C006437.2.1*, *MELO3C019735.2*, and *MELO3C014437.2*) are considered important candidate genes. Among these, two ACS genes (*MELO3C015444.2.1* and *MELO3C010779.2.1*) have been identified as

melon sex-determining genes [21,22]. Additionally, an ACO-coding gene (*MELO3C007425.2.1*) has been reported to be associated with melon female development. Hence, the importance of the two ACS genes and three ACO-coding genes in this study is further highlighted.

Ethylene promotes female development through hormone signaling pathways, making the high-expression protein MPK6 coding gene (*MELO3C005705.2.1*) a significant candidate for female regulation in this process. Interactive protein network analysis has shown crosstalk clues between different phytohormones (Figure 7). Among these, there is a high degree of connection between ethylene and auxin, gibberellin (GA), methyl jasmonate (MeJA), and other signaling pathways. Many studies have indicated a mutual promotional relationship between ethylene and auxin, particularly during signal conduction. ARF can directly and positively regulate the expression of ACS, while ARP can indirectly inhibit EBF1/2, thereby promoting ethylene signaling [11,35,36]. Therefore, in this process, two ARF-coding genes (*MELO3C006181.2.1* and *MELO3C023883.2.1*) and four ARP-coding genes (*MELO3C011268.2.1, MELO3C002378.2.1, MELO3C020765.2.1,* and *MELO3C013403.2.1*) are also important candidates. Among them, the positive regulation of the two ARFs aligns more closely with the mechanism of 'Y101′ dominant strong female regulation. As a gaseous phytohormone, ethylene plays a crucial role in regulating the development of flower organs and acts as a significant mediator in the ethylene response factor (ERF) regulation mechanism. Previous studies have demonstrated that *CsERF110* in cucumbers can directly bind to *CsACS11*, activating its expression to promote female development. This finding has also been confirmed in melons [34]. Additionally, *CsERF31* has been shown to enhance cucumber female differentiation [37]. Based on these findings, we hypothesize that the six highly expressed ERFs identified in this study, namely *MELO3C015543.2.1*, *MELO3C026158.2.1*, *MELO3C005940.2.1*, *MELO3C006430.2.1*, *MELO3C006149.2.1*, and *MELO3C022358.2.1*, may have a regulatory effect on melon female differentiation.

3.2. A Regulatory Network Model for the Female Development of Melon

The hermaphroditic flower type material 'Y101' is a typical female phenotype and a wild-type thin-skinned melon. We conducted comparative omics analysis of the female pool (F) and male pool (M) constructed by the melon material F_2 population by 'Y101' and the male phenotype '0426'. This approach helped us eliminate background errors between species.

A series of candidate genes that may have a role in female flower development were revealed by functional enrichment analysis and network interaction analysis. Results indicated that ethylene pathway genes play an essential role in determining female flower development. Through omics analysis, the study speculates that the enhanced expression of *CmACS7* and *CmACS11* may also be regulated by ARF. Additionally, the study shows that 'Y101' exhibits a high expression of (cytosine-5)-methyltransferase (CMT) and adenosine homocysteine (SAH), leading to an abundance of ethylene synthesis substrates. With the high expression of the ethylene precursor ACC oxidase (ACO), the flower buds of 'Y101' contain more ethylene during development. Sufficient ethylene activates a series of ERF expressions through the signaling process, ultimately promoting female development by regulating downstream target genes. Network interaction analysis suggests that these downstream target genes of ERF may be related to the development of female organ-related genes, including *CRABS CLAW* (MELO3C034075.2.1), *HECATE3* (*MELO3C008047.2.1*), and *MADS13* (*MELO3C033521.2.1*). An interesting fact identified in previous studies is that the majority of the genes associated with sex differentiation are ethylene synthase genes [3]. Our study proposes a putative model (see Figure 8) that further supports the hypothesis that the ethylene synthase network plays a crucial role in melon, regulating sex differentiation. A comparison of female (F) and male (M) pools in our study has uncovered several candidate genes within the ethylene signaling pathway.

As illustrated in Figure 8, four catalytic enzymes, including ACO, ACS, CMT, and SAH-related genes and proteins, exhibited a multifold increase. Additionally, proteins

encoding AUX/IAA were up-regulated, suggesting the involvement of more genes in female enhancement. It is noteworthy that only selective genes were studied in previous experiments, and Li et al. (2019) recommended, in their review on gene interactions regulating sex determination, conducting expression regulation studies for each gene to obtain clearer evidence of sex differentiation in cucurbits [3]. Interestingly, an integrated network analysis of key DAPs and DESs from our comparative groups highlighted four interaction networks, including those involving plant hormones, transduction, and ethylene biosynthesis genes. Our proposed model indicates that auxin-related genes, such as ARPs and ARFs, also exhibited higher expression, suggesting their potential role in sex differentiation.

It is well established in previous literature that auxins promote ethylene biosynthesis by up-regulating ACS, the first regulatory step in ethylene biosynthesis, and our findings support these hypotheses. Similarly, several auxin-responsive factors were differentially expressed in our study. The proteome analysis also revealed that two key ARFs (ME-LO3C023883.2.1 and MELO3C006181.2.1) showed a two-fold increase in the female dominant melon. The female regulatory candidate genes and regulatory networks identified through omics analysis require further verification. Nonetheless, the current research results hold important reference value and have significant implications for in-depth research. It is believed that with the improvement of the melon transformation system and further research, the regulatory mechanism of sex determination in melons can be better understood in the future. The current study encompasses various aspects, including transcriptomics, proteomics, and the generation of sex type-specific pools for research. A detailed explanation and analysis of all crucial pathways in terms of female development provide essential information for researchers. However, the omics data presented in the study are complex and high-dimensional, posing challenges for analysis and interpretation. Furthermore, the study unveiled new candidates that require further investigation to validate their true function in melon female development. The candidate genes and proteins identified in the current study can be utilized in other cucurbits as a hypothesis to identify the sex differentiation mechanism.

4. Materials and Methods

4.1. Plant Material and Flower Bud Collection

The plant monoecious line '0426' (ssp. *melo*) and hermaphrodite line 'Y101' (ssp. *agrestis*) were grown in the cucurbit research and experimental station, College of Horticulture, Northwest A & F University, Yangling, Shaanxi, China [26]. Plants of '0426' were used as the female parent and 'Y101' as the male parent to obtain the F_1 (0426 $\times$ Y101) genetic population. F_2 plants were obtained by selfing F_1 plants in the greenhouse. The isolated population of F_2 (0426 $\times$ Y101) contained four sex types of individuals: gynoecy, hermaphrodite, monoecy, and andromonoecy [26]. The gynoecy and hermaphrodite plants were considered female sexual types and kept under the F-pool [38,39]. The flower buds collected from monoecy and andromonoecy plants are considered male-dominant plants and grouped under the M-pool [38,39]. At the 15-node plant growth stage, when sex type judgment is reliable, 50 gynoecious plants and 50 hermaphrodite plants were selected from the isolated group, and the apex tissue was collected to construct the F (female) pool (Figure 9). Similarly, 50 each of monoecious plants and andromonoecious plants were selected, and the apex tissue was sampled to construct the M (male) pool. Samples were quickly placed in liquid nitrogen after collection, and whole proteins and RNA were extracted for proteome and transcriptome sequencing.

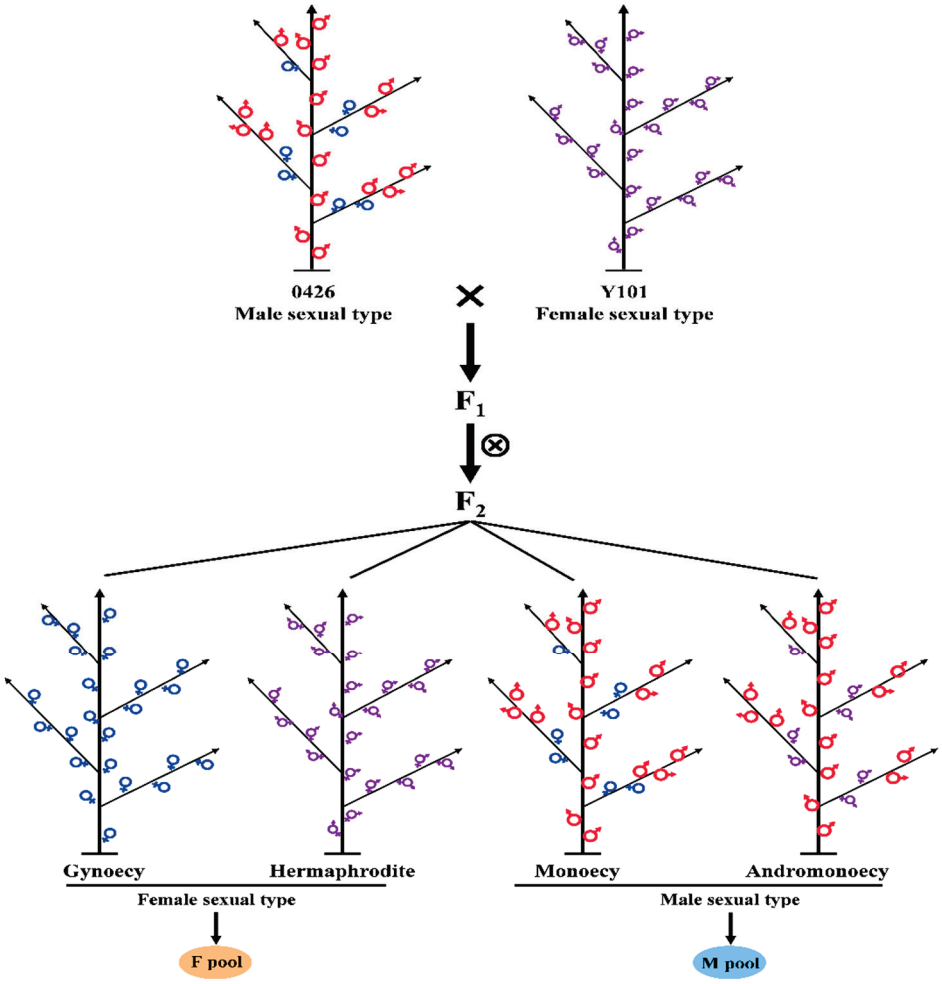


Figure 9. Representation of sexual phenotypes and female (F) and male (M) pools of plant materials used in the current study. Monoecious line '0426' (ssp. *melo*) and hermaphrodite line 'Y101' (ssp. *agrestis*) were used as parents, and the F/M pool was constructed using the F_2 population.

4.2. TMT Labeling Quantitative Proteome Sequencing and Analysis

The samples were collected and then mixed in a single pool before being sent to Shanghai Luming Biotechnology Co., Ltd. (Shanghai, China) for protein detection and comparative analysis of the F and M-pool samples [40,41]. The main experimental process involved extracting the total protein from the F pool and the M pool, which was a mixture of collected apex tissue. A small amount of the extracted total protein was taken for polyacrylamide gel electrophoresis (SDS-PAGE) to detect protein quality and integrity. Once the protein bands were successfully detected on the gel, trypsin was used to enzymatically hydrolyze the total protein, breaking it down into peptides. The peptides were then labeled using TMT reagents. TMT (6 standard) labeled peptides were separated using reversed-phase chromatography on an Agilent 1100 HPLC liquid chromatograph (column: Agilent Zorbax Extend—C18 narrow bore column, 2.1 × 150 mm, 5 μm). Finally, LC-MS analysis was performed on the separated samples to obtain the raw data. The data were compared with the melon reference proteome data (http://cucurbitgenomics.org/v2/ftp/genome/melon/DHL92/v4.0/, accessed on 16 September 2023) for qualitative and quantitative analysis. Quality control was conducted on the searched data, followed by expression level analysis and functional analysis. The false discovery rate (FDR) was adjusted to <1%, and the minimum score for peptides was set at >40. Differentially abundant proteins (DAPs) were identified by performing a Student's *t*-test with a fold change >1.5 (or <0.67) and a *p*-value <0.05, comparing them to the healthy control [42,43]. Additionally, Fisher's exact test was used to test DAPs against the background of the identified proteins.

4.3. F/M Mixed Pool RNA-Seq and Parametric Transcriptome Analysis

The bud samples were collected and sent to Shanghai Luming Biotechnology Co., Ltd. for RNA-seq and comparative analysis of the F pool and M pool samples. The main steps followed for comparative transcriptome analysis included extracting total RNA from the apex tissue mixed in the F pool and M pool and constructing the cDNA library after assessing the purity, integrity, and concentration of the RNA sample. In this study, total RNA extraction was carried out utilizing the TIANGEN RNAprep Pure kit (DP432) from Tiangen Biotech Co., Ltd. (Beijing, China). Following the manufacturer's instructions, apex tissue of melon varieties was collected, and any specified pre-processing steps, such as homogenization or cell lysis, were performed accordingly. The RNA extraction protocol provided by the kit was rigorously followed, incorporating specific incubation times and temperatures. Purification was executed according to the provided guidelines, utilizing recommended columns or membranes. The elution step involved the use of RNase-free water or elution buffer, following the manufacturer's instructions for optimal recovery. To assess the quality of the extracted RNA, measurements of purity and concentration were performed using NanoDrop (Agilent Technologies, Santa Clara, USA), providing key parameters such as the A260/A280 ratio. Furthermore, the integrity of the RNA was evaluated using agarose gel electrophoresis, ensuring the suitability of the extracted RNA for downstream applications. The Illumina HiSeq platform was used to bilaterally sequence the library and obtain the raw data. After performing quality control on the original sequencing data (reads), the final clean reads were summarized. The obtained sequence data were then compared with the melon reference genome (DHL92, V3.6.1, http://cucurbitgenomics.org/v2/organism/23, accessed on 16 September 2023) to obtain information about the reads' alignment to the reference genome. Reads with fewer than 10 comparisons, reads that aligned to multiple regions, and reads without alignment to the reference were subsequently removed. Finally, the quantitative expression level of each gene was obtained. This was followed by an analysis of the expression levels and functional annotations.

4.4. Functional Enrichment of DEGs and DAPs

The GO enrichment analysis of the differentially expressed genes (DEGs) and differentially abundant proteins (DAPs) was performed using the GOseq R package, which is based on the Wallenius non-central hypergeometric distribution [44]. This approach allows for the adjustment of gene length bias in DEGs and DAPs. The GO annotation proteome used in the analysis was obtained from the UniProt-GOA database (http://www.ebi.ac.uk/GOA/, accessed on 16 September 2023). To test the statistical enrichment of DEGs and DAPs in KEGG pathways, we utilized the KOBAS V3.0 software (http://bioinfo.org/kobas/download/, accessed on 16 September 2023) [45].

4.5. Protein Interaction Network Analysis

According to the analysis results of the proteome and transcriptome, relevant proteins and genes that may be involved in the sex-determining pathway were screened from the KEGG-enriched DAPs and DEGs. To analyze the interactions among these proteins, an online protein interaction prediction and analysis software, STRING (https://cn.string-db.org/, accessed on 16 September 2023), was employed. Furthermore, Cytoscape V3.7.2 software was utilized to map the protein–protein interaction (PPI) networks.

4.6. qRT-PCR Analysis of Omics Data and Statistical Analysis

To estimate the transcript levels of selected differentially enriched proteins and genes, qRT-PCR was conducted using terminal bud tissue samples that were previously used for proteomics and RNA-seq. Twenty genes, both up-regulated and down-regulated, were randomly chosen from the proteome and transcriptome results to validate the reliability of the sequencing data. The cDNA synthesis was carried out using two micrograms of total RNA with SuperScript IV reverse transcriptase (Invitrogen) and an oligo(dT) primer.

The qRT-PCR reaction (25 μL) included 1 μL of the cDNA template, 0.5 μL of each primer, 12.5 μL of 2 × SYBR qPCR Mix, and 11 μL ddH_2O to a final volume of 25 μL. Three technical replicates per sample set were performed. qRT-PCR amplification conditions were set to 94 °C for 3 min, followed by 40 cycles of 94 °C for 10 s, and 60 °C for 30 s. The qRT-PCR was performed on a BIO-RAD CFX96 machine using gene-specific forward and reverse primers (Table S6). We also examined the transcript levels of selected DEGs and DAPs in the female type parent (Y101) and male type parent (0426) to additionally analyze and confirm the dependability of the sequencing data and the potential functionalities in sex regulation.

The data were analyzed using the $2^{-\Delta\Delta CT}$ method, and *CmActin* was selected as the reference gene for normalization. Significance analysis was performed using IBM SPSS 19.0 software for ANOVA, and the Student's *t*-test was employed for further analysis.

5. Conclusions

In the present study, we performed transcriptome analysis followed by proteome analysis of two different sex type pools, including flower buds of male and female characteristics. A high number of DEGs and DAPs were identified by proteome and transcriptome analysis of the female pool (F) and male pool (M) identified by the segregated population in the F_2 generation. Through functional enrichment analysis, 38 DAPs and 43 DEGs were identified, which were involved in the regulatory network of female flower development. Through systematic analysis, the female developmental regulatory pathway model of ethylene biosynthesis, signal transduction, and target gene regulation was proposed. Two CMTs (*MELO3C026448.2.1* and *MELO3C015649.2.1*), one AdoHS, four ACSs (*MELO3C015444.2.1*, *MELO3C010779.2.1*, *MELO3C005597.2.1*, and *MELO3C006840.2.1*), three ACOs (*MELO3C006437.2.1*, *MELO3C019735.2*, and *MELO3C014437.2*), two ARFs (*MELO3C006181.2.1* and *MELO3C023883.2.1*), four ARPs (*MELO3C011268.2.1*, *MELO3C002378.2.1*, *MELO3C020765.2.1*, and *MELO3C013403.2.1*), and six ERFs (*MELO3C015543.2.1*, *MELO3C026158.2.1*, *MELO3C005940.2.1*, *MELO3C006430.2.1*, *MELO3C006149.2.1*, and *MELO3C022358.2.1*) were found to have an important role in pathways associated with female flower development. Future investigations into the regulatory mechanisms of some newly identified candidate genes and proteins may enable researchers to manipulate specific sexual expressions in melon. The functional annotation of high-throughput transcriptomic and proteomics data generated in this study may serve as genetic resources for the development of crop improvement strategies for melon and various other cucurbits.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms242316905/s1.

Author Contributions: Conceptualization, data analysis, experimentation, and manuscript preparation, Z.W. and V.Y.; supervision, X.Z.; plant growth management, X.C., S.Z. and X.Y.; data analysis and trait identification, H.L., J.M. and Y.Z.; collection and identification of germplasm J.Y.; project implementation and manuscript development, C.W. and X.Z. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Seed Innovation Project of Northwest A&F University (2452022116), the National Natural Science Foundation of Shaanxi Province, China [No. 2023-JC-YB-199], the Key Research and Development Project of Yangling Seed Industry Innovation Center (Ylzy-sc-01), and the High-quality Development and Ecological Protection Science and Technology Innovation Project of Ningxia Academy of Agriculture and Forestry Sciences (NGSB-2021-7).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and supplementary materials.

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

References

1. Chibalina, M.V.; Filatov, D.A. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* **2011**, *21*, 1475–1479. [CrossRef] [PubMed]
2. Devani, R.S.; Chirmade, T.; Sinha, S.; Bendahmane, A.; Dholakia, B.B.; Banerjee, A.K.; Banerjee, J. Flower bud proteome reveals modulation of sex-biased proteins potentially associated with sex expression and modification in dioecious *Coccinia grandis*. *BMC Plant Biol.* **2019**, *19*, 330. [CrossRef] [PubMed]
3. Li, D.; Sheng, Y.; Niu, H.; Li, Z. Gene interactions regulating sex determination in Cucurbits. *Front. Plant Sci.* **2019**, *10*, 1231. [CrossRef] [PubMed]
4. Bhowmick, B.K.; Jha, S. Dynamics of sex expression and chromosome diversity in Cucurbitaceae: A story in the making. *J. Genet.* **2015**, *94*, 793–808. [CrossRef]
5. Law, T.F.; Lebel-Hardenack, S.; Grant, S.R. Silver enhances stamen development in female white campion (*Silene latifolia* [Caryophyllaceae]). *Am. J. Bot.* **2002**, *89*, 1014–1020. [CrossRef]
6. Mohanty, J.N.; Nayak, S.; Jha, S.; Joshi, R.K. Transcriptome profiling of the floral buds and discovery of genes related to sex-differentiation in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. *Gene* **2017**, *626*, 395–406. [CrossRef]
7. Zhang, Y.; Zhao, G.; Li, Y.; Mo, N.; Zhang, J.; Liang, Y. Transcriptomic Analysis Implies That GA Regulates Sex Expression via Ethylene-Dependent and Ethylene-Independent Pathways in Cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* **2017**, *8*, 10. [CrossRef]
8. Lai, Y.S.; Shen, D.; Zhang, W.; Zhang, X.; Qiu, Y.; Wang, H.; Dou, X.; Li, S.; Wu, Y.; Song, J.; et al. Temperature and photoperiod changes affect cucumber sex expression by different epigenetic regulations. *BMC Plant Biol.* **2018**, *18*, 268. [CrossRef]
9. Pawełkowicz, M.; Pryszcz, L.; Skarzyńska, A.; Wóycicki, R.K.; Posyniak, K.; Rymuszka, J.; Przybecki, Z.; Pląder, W. Comparative transcriptome analysis reveals new molecular pathways for cucumber genes related to sex determination. *Plant Reprod.* **2019**, *32*, 193–216. [CrossRef]
10. Ge, C.; Zhao, W.; Nie, L.; Niu, S.; Fang, S.; Duan, Y.; Zhao, J.; Guo, K.; Zhang, Q. Transcriptome profiling reveals the occurrence mechanism of bisexual flowers in melon (*Cucumis melo* L.). *Plant Sci.* **2020**, *301*, 110694. [CrossRef]
11. Gao, P.; Sheng, Y.; Luan, F.; Ma, H.; Liu, S. RNA-Seq Transcriptome profiling reveals differentially expressed genes involved in sex expression in melon. *Crop Sci.* **2015**, *55*, 1686–1695. [CrossRef]
12. Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720–1730. [CrossRef] [PubMed]
13. Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M.C.; Zimmerman, L.J.; Shaddox, K.F.; Kim, S.; et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, *513*, 382–387. [CrossRef] [PubMed]
14. Yang, L.; Gong, F.; Xiong, E.; Wang, W. Proteomics: A promising tool for research on sex-related differences in dioecious plants. *Front. Plant Sci.* **2015**, *6*, 954. [CrossRef] [PubMed]
15. Misra, B.B.; Langefeld, C.D.; Olivier, M.; Cox, L.A. Integrated Omics: Tools, Advances, and Future Approaches. *J. Mol. Endocrinol.* **2018**, *62*, R21–R45. [CrossRef] [PubMed]
16. Girek, Z.; Prodanovic, S.; Zdravković, J.; Zivanovic, T.; Ugrinović, M.; Zdravkovic, M. The effect of growth regulators on sex expression in melon (*Cucumis melo* L.). *Crop Breed. Appl. Biotechnol.* **2013**, *165*, 165–171. [CrossRef]
17. Garcia-Mas, J.; Benjak, A.; Sanseverino, W.; Bourgeois, M.; Mir, G.; Gonzalez, V.M.; Henaff, E.; Camara, F.; Cozzuto, L.; Lowy, E.; et al. The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11872–11877. [CrossRef]
18. Ezura, H.; Owino, W.O. Melon, an alternative model plant for elucidating fruit ripening. *Plant Sci.* **2008**, *175*, 121–129. [CrossRef]
19. Zhang, S.Q.; Tan, F.Q.; Chung, C.H.; Slavkovic, F.; Devani, R.S.; Troadec, C.; Marcel, F.; Morin, H.; Camps, C.; Gomez Roldan, M.V.; et al. The control of carpel determinacy pathway leads to sex determination in cucurbits. *Science* **2022**, *378*, 543–549. [CrossRef]
20. Martínez, C.; Jamilena, M. To be a male or a female flower, a question of ethylene in cucurbits. *Curr. Opin. Plant Biol.* **2021**, *59*, 101981. [CrossRef]
21. Boualem, A.; Troadec, C.; Camps, C.; Lemhemdi, A.; Morin, H.; Sari, M.A.; Fraenkel-Zagouri, R.; Kovalski, I.; Dogimont, C.; Perl-Treves, R. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* **2015**, *350*, 688–691. [CrossRef]
22. Boualem, A.; Fergany, M.; Fernandez, R.; Troadec, C.; Bendahmane, A. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **2008**, *321*, 836–838. [CrossRef]
23. Martin, A.; Troadec, C.; Boualem, A.; Rajab, M.; Fernandez, R.; Morin, H.; Pitrat, M.; Dogimont, C.; Bendahmane, A. A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **2009**, *461*, 1135–1138. [CrossRef]
24. Eleblu, J.S.Y.; Haraghi, A.; Mania, B.; Camps, C.; Bendahmane, A.J.S.R. The gynoecious *CmWIP1* transcription factor interacts with *CmbZIP48* to inhibit carpel development. *Sci. Rep.* **2019**, *9*, 15443. [CrossRef]
25. Zhou, Y.; Ahammed, G.J.; Wang, Q.; Wu, C.; Wan, C.; Yang, Y. Transcriptomic insights into the blue light-induced female floral sex expression in cucumber (*Cucumis sativus* L.). *Sci. Rep.* **2018**, *8*, 14261. [CrossRef]
26. Wang, Z.; Zhang, S.; Yang, Y.; Li, Z.; Li, H.; Yu, R.; Luan, F.; Zhang, X.; Wei, C. Novel bisexual flower control gene regulates sex differentiation in melon (*Cucumis melo* L.). *J. Agric. Food Chem.* **2022**, *70*, 15401–15414. [CrossRef]
27. Peñaranda, A.; Payan, M.C.; Garrido, D.; Gómez, P.; Jamilena, M. Production of fruits with attached flowers in zucchini squash is correlated with the arrest of maturation of female flowers. *J. Hortic. Sci. Biotechnol.* **2007**, *82*, 579–584. [CrossRef]

28. Wang, Y.L.; Yan, C.D.; Zou, B.X.; Wang, C.J.; Xu, W.L.; Cui, C.S.; Qu, S.P. Morphological, transcriptomic and hormonal characterization of trimonoecious and subandroecious pumpkin (*Cucurbita maxima*) suggests important roles of ethylene in sex expression. *Int. J. Mol. Sci.* **2019**, *20*, 3185. [CrossRef]
29. Yamasaki, S.; Fujii, N.; Matsuura, S.; Mizusawa, H.; Takahashi, H. The M locus and ethylene-controlled sex determination in andromonoecious cucumber plants. *Plant Cell Physiol.* **2001**, *42*, 608–619. [CrossRef]
30. Rashid, D.; Devani, R.S.; Rodriguez-Granados, N.Y.; Abou-Choucha, F.; Troadec, C.; Morin, H.; Tan, F.Q.; Marcel, F.; Huang, H.Y.; Hanique, M.; et al. Ethylene produced in carpel primordia controls *CmHB40* expression to inhibit stamen development. *Nat. Plants* **2023**, *9*, 1675–1687. [CrossRef] [PubMed]
31. He, W.; Brumos, J.; Li, H.; Ji, Y.; Ke, M.; Gong, X.; Zeng, Q.; Li, W.; Zhang, X.; An, F.; et al. A small-molecule screen identifies l-kynurenine as a competitive inhibitor of TAA1/TAR activity in ethylene-directed auxin biosynthesis and root growth in *Arabidopsis*. *Plant Cell* **2011**, *23*, 3944–3960. [CrossRef]
32. Robles, L.; Stepanova, A.; Alonso, J. Molecular mechanisms of ethylene-auxin interaction. *Mol. Plant* **2013**, *6*, 1734–1737. [CrossRef]
33. Forlani, S.; Masiero, S.; Mizzotti, C. Fruit ripening: The role of hormones, cell wall modifications, and their relationship with pathogens. *J. Exp. Bot.* **2019**, *70*, 2993–3006. [CrossRef]
34. Tao, Q.; Niu, H.; Wang, Z.; Zhang, W.; Wang, H.; Wang, S.; Zhang, X.; Li, Z. Ethylene responsive factor ERF110 mediates ethylene-regulated transcription of a sex determination-related orthologous gene in two *Cucumis* species. *J. Exp. Bot.* **2018**, *69*, 2953–2965. [CrossRef]
35. Jing, H.; Korasick, D.A.; Emenecker, R.J.; Morffy, N.; Wilkinson, E.G.; Powers, S.K.; Strader, L.C. Regulation of auxin response factor condensation and nucleo-cytoplasmic partitioning. *Nat. Commun.* **2022**, *13*, 4015. [CrossRef]
36. Li, S.B.; Xie, Z.Z.; Hu, C.-G.; Zhang, J.Z. A Review of auxin response factors (ARFs) in plants. *Front. Plant Sci.* **2016**, *7*, 47. [CrossRef]
37. Pan, J.; Wen, H.; Chen, G.; Lin, W.-H.; Du, H.; Chen, Y.; Zhang, L.; Lian, H.; Wang, G.; Cai, R.; et al. A positive feedback loop mediated by *CsERF31* initiates female cucumber flower development. *Plant Physiol.* **2021**, *186*, 1088–1100. [CrossRef]
38. Rosa, J.T. The inheritance of flower types in *Cucumins* and *Citrullus*. *Hilgardia* **1928**, *3*, 233–250. [CrossRef]
39. Poole, C.F.; Grimball, P.C. Inheritance of new sex forms in *Cucumis melo* L. *J. Hered.* **1939**, *30*, 21–25. [CrossRef]
40. Jiang, A.L.; Liu, Y.N.; Liu, R.; Ren, A.; Ma, H.Y.; Shu, L.B.; Shi, L.; Zhu, J.; Zhao, M.W. Integrated proteomics and metabolomics analysis provides insights into ganoderic acid biosynthesis in response to methyl jasmonate in *Ganoderma lucidum*. *Int. J. Mol. Sci.* **2019**, *20*, 6116. [CrossRef] [PubMed]
41. Xu, D.; Zhu, X.; Ren, J.; Huang, S.; Xiao, Z.; Jiang, H.; Tan, Y. Quantitative proteomic analysis of cervical cancer based on TMT-labeled quantitative proteomics. *J. Proteom.* **2022**, *252*, 104453. [CrossRef]
42. Bo, C.; Geng, X.; Zhang, J.; Sai, L.; Zhang, Y.; Yu, G.; Zhang, Z.; Liu, K.; Du, Z.; Peng, C.; et al. Comparative proteomic analysis of silica-induced pulmonary fibrosis in rats based on tandem mass tag (TMT) quantitation technology. *PLoS ONE* **2020**, *15*, e0241310. [CrossRef]
43. Du, Y.; Wang, Y.; Xu, Q.; Zhu, J.; Lin, Y. TMT-based quantitative proteomics analysis reveals the key proteins related with the differentiation process of goat intramuscular adipocytes. *BMC Genom.* **2021**, *22*, 417. [CrossRef]
44. Zhang, W.; Huang, L.; Li, J.; Li, Y.; Wei, S.; Cai, L.; Hua, W. Transcriptomic analysis of differentially expressed genes in the oviduct of *Rhacophorus omeimontis* provides insights into foam nest construction. *BMC Genom.* **2019**, *20*, 562. [CrossRef]
45. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Communication

Physical Pretreatments Applied in Three Commercial Kits for the Extraction of High-Quality DNA from Activated Sewage Sludge

Claudio Vásquez [1,2], Benjamín Leyton-Carcaman [1,2], Fernanda P. Cid-Alda [2,*], Iñaky Segovia [3], Fernanda Pinto [4] and Michel Abanto [2,*]

[1] Doctorado en Ciencias Mención Biología Celular y Molecular Aplicada, Universidad de La Frontera, Temuco 4811230, Chile; claudio.vasquez@ufrontera.cl (C.V.); benjamin.leyton@ufrontera.cl (B.L.-C.)
[2] Scientific and Technological Bioresource Nucleus (BIOREN), Universidad de La Frontera, Temuco 4811230, Chile
[3] Carrera de Tecnología Médica, Facultad de Medicina, Universidad de La Frontera, Temuco 4811230, Chile; i.segovia01@ufromail.cl
[4] Departamento de Procesos Industriales, Facultad de Ingeniería, Universidad Católica de Temuco, Casilla 15-D, Temuco 4780000, Chile; fpinto@uct.cl
* Correspondence: fernanda.cid@ufrontera.cl (F.P.C.-A.); mfabanto@gmail.com or michel.abanto@ufrontera.cl (M.A.)

Abstract: Obtaining sufficient and high-quality genomic DNA from sludge samples is a fundamental issue of feasibility and comparability in genomic studies of microbial diversity. Commercial kits for soil are often used for the extraction of gDNA from sludge samples due to the lack of specific kits. However, the evaluation of the performance of commercial kits for sludge DNA extraction is scarce and optimization of these methods to obtain a high quantity and quality of DNA is necessary, especially for downstream genomic sequencing. Sequential batch reactors (SBRs) loaded with lignocellulosic biomass are used for the synthesis of renewable resources such as levulinic acid (LA), adipic acid (AA), and polyhydroxyalkanoates (PHAs), and the biochemical synthesis of these compounds is conducted through the inoculation of microbes present in the residual activated sludge (AS) obtained from a municipal wastewater treatment plant. To characterize these microbes, the extraction of DNA from residual sewage sludge was conducted with three different commercial kits: Nucleospin® Soil from Macherey-Nagel, DNEasy® PowerSoil® from Qiagen, and E.Z.N.A.® Plant DNA Kit from Omega BIO-TEK. Nevertheless, to obtain the highest load and quality of DNA for next-generation sequencing (NGS) analysis, different pretreatments and different combinations of these pretreatments were used. The pretreatments considered were an ultrasonic bath and a temperature of 80 °C, together and separately with different incubation time periods of 30, 60, and 90 min. The results obtained suggest a significant improvement in the efficiency and quality of DNA extraction with the three commercial extraction kits when used together with the ultrasonic bath and 80 °C for 60 min. Here, we were able to prove that physical pretreatments are a viable alternative to chemical lysis for DNA extraction from complex samples such as sludge.

Keywords: DNA extraction; complex samples; next-generation sequencing; metagenomics; microbiome; bioreactor

Citation: Vásquez, C.; Leyton-Carcaman, B.; Cid-Alda, F.P.; Segovia, I.; Pinto, F.; Abanto, M. Physical Pretreatments Applied in Three Commercial Kits for the Extraction of High-Quality DNA from Activated Sewage Sludge. *Int. J. Mol. Sci.* **2023**, *24*, 15243. https://doi.org/10.3390/ijms242015243

Academic Editor: Wajid Zaman

Received: 24 August 2023
Revised: 27 September 2023
Accepted: 27 September 2023
Published: 17 October 2023

1. Introduction

Feedstocks such as lignocellulosic biomass could be used to produce high-added value compounds such as levulinic acid (LA), adipic acid (AA), and polyhydroxyalkanoates (PHAs) moving from an economy based on fossil fuels, to an economy based on renewable resources [1,2]. For example, LAs as raw material, could be applied as fuel additives, antifreeze, textiles, animal feed, herbicides, pharmaceuticals, and flavor substances [3,4]. On the other hand, AA is a straight-chain dicarboxylic acid with many uses

in chemical, pharmaceutical, and lubricant manufacturing, mainly for resins and nylon 6–6 fibers production [5,6]. Instead, PHAs are used as a bioplastic with physicochemical and mechanical properties like those in petroleum-based plastics and could also be used in medical, pharmaceutical, and food applications [7]. Sequential batch reactors (SBRs) are used for the conversion of lignocellulosic waste to LA, AA, and PHAs, and microbes from activated sludge (AS) from a municipal sludge wastewater treatment plant were used for the biochemical synthesis of these high-added value compounds [8,9]. Microbes from the AS develop mixed microbial cultures responsible for the conversion of different carbon sources from lignocellulosic waste such as synthetic, molasses, sugarcane bagasse, and grape pomace hydrolysate [10]. Nevertheless, these microbes are likely an unknown black box, and this is why it is important to determine the identity of these microbes.

Nowadays, next-generation sequencing (NGS) studies are conducted to identify these microorganisms usually described as a taxonomically and metabolically diverse microbial community with complex relationships among their members [11]. The main problem, however, is that the disruption of this activated sludge from SBRs is difficult to achieve and time-consuming due to a heavy encapsulation of microorganisms in flocs or aerobic granules [12–14]. These microbial populations are usually aggregated or embedded in a polymeric matrix or gel, referred to as extracellular polymeric substances (EPSs), commonly hard to disaggregate [13].

Commercial DNA extraction kits provide the protocols for the extraction of nucleic acids from standard samples, but the extraction of genomic DNA (gDNA) from SBR samples is challenging, especially in terms of achieving sufficient concentration for NGS analysis and identification of these microbes. For Illumina sequencing, a load above 5 $ng \cdot \mu L^{-1}$ is required as described by the protocol "16S Metagenomic Sequencing Library Preparation" (Illumina, Inc., San Diego, CA, USA) [15]. Whereas for third-generation sequencing such as Pacific Bioscience (PacBio) and Oxford Nanopore sequencing (ONS), related to long-length reads, even more than 10 times this amount of DNA is necessary and has been described as the main bottleneck for genome sequencing [16]. A high quantity and quality of DNA, beside the amplification of a certain marker gene, such as *16S rRNA gene* or internal transcribed spacer (ITS), allows the application of this gDNA extraction for whole genome sequencing approaches (WGS) and obtains a more comprehensive and major resolution of taxonomic and functional profiles.

In addition, a high quality of DNA free of contamination is necessary with 260/280 absorbances in the range of 1.7 and 2.0 [17]. Thus, commercial kits require additional steps to obtain a sufficient concentration of gDNA free of contamination. Vortexing with glass beads, for example, has been used for mechanical sludge lysis to allow better hydration with the extraction buffers [14], whereas CTAB was added to improve chemical lysis and help to precipitate polysaccharides and humic substances [18].

Here, we aimed to explore the gDNA extraction efficiency obtained with three commercial extraction kits combined with different physical pretreatments tested to determine their efficiency. The commercial kits used were Genomic DNA from Soil (Nucleospin® Soil from Macherey-Nagel), DNEasy® PowerSoil® from Qiagen, and E.Z.N.A.® Plant DNA Kit from Omega BIO-TEK, and the pretreatments tested here were an ultrasonic bath (UB) for 90 min, a temperature of 80 °C for 90 min and the combination of UB + 80 °C for different periods of 30, 60, and 90 min.

2. Results

The DNA concentrations acquired using the three available commercial kits as per the provider's standard protocol were either extremely low or indetectable (<0.1 $ng \cdot \mu L^{-1}$) (Figure 1, Tables S1–S3). Notably, the application of 80 °C for 90 min and the combination of an ultrasonic bath (UB) with a temperature of 80 °C for 60 and 90 min prolonged periods resulted in increased DNA concentrations exceeding the mandatory 5 $ng \cdot \mu L^{-1}$ limit required for NGS across all three kits (as depicted in Figure 1). In the case of Nucleospin® Soil in Figure 1A, it is possible to observe that the application of a temperature of 80 °C

for 90 min is the main variable affecting DNA obtention, compared to the UB applied for 90 min with no significant differences from the default extraction. Similarly, the combined effect of UB + 80 °C for 30 min did not present significant differences from the default kit extraction protocol. The combined effect of 80 °C + UB for longer periods of 60 and 90 min were significantly higher than the default extraction protocol and surpasses the threshold concentration of 5 ng·µL^{-1} of DNA. In the case of the DNEasy® PowerSoil® kit, a similar trend to the Nucleospin® Soil kit was observed (Figure 1A,B). The pretreatment of 80 °C for 90 min and the combined effect of UB + 80 °C temperature exposure for longer periods of 60 and 90 min were crucial to obtaining significant differences from the standard extraction protocol and a higher concentration above the threshold of 5 ng·µL^{-1} of gDNA.

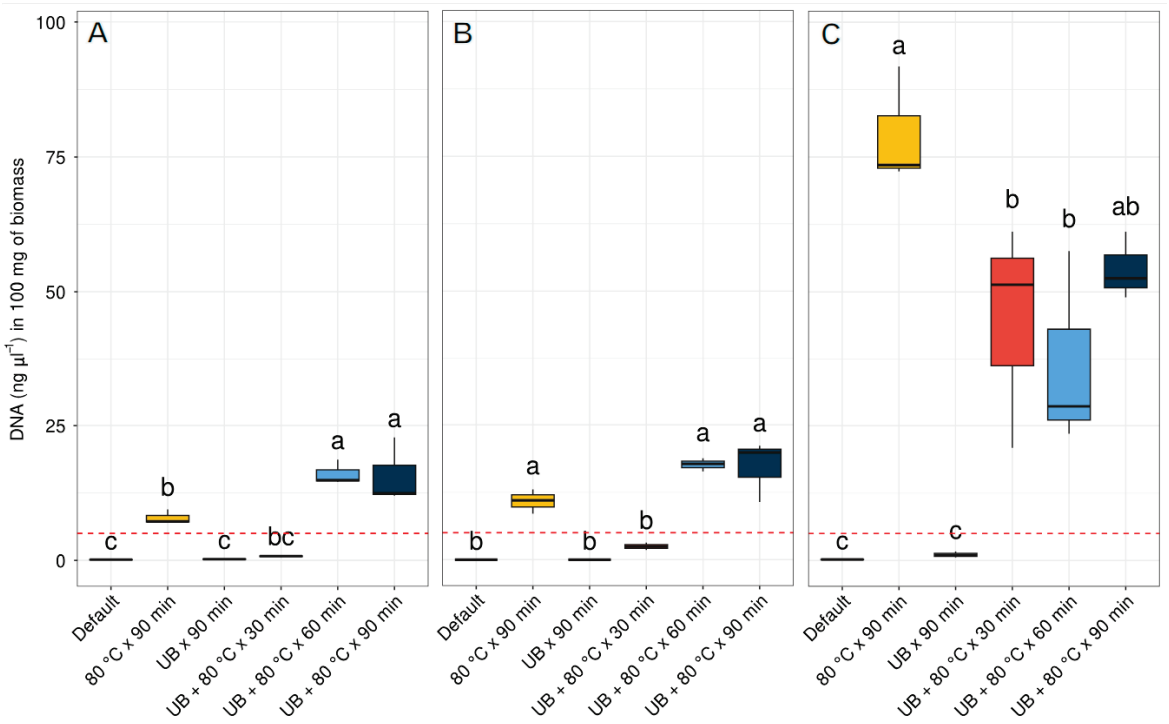


Figure 1. Concentration of DNA with the three commercial kits used. (**A**) Nucleospin® Soil, (**B**) DNEasy® PowerSoil®, and (**C**) E.Z.N.A.® Plant DNA Kit. Different letters indicate statistical differences according to the results of ANOVA followed by Tukey's post hoc test. UB = ultrasonic bath. The dashed red line indicates the optimal DNA concentration threshold for next-generation sequencing studies of 5 ng·µL^{-1}.

Regarding the E.Z.N.A.® Plant DNA Kit, as shown in Figure 1C, the application of 80 °C temperature was the main factor that significantly improved the DNA concentration, compared to the UB applied for 90 min, which did not show significant differences from the standard extraction protocol. As for the combined use of 80 °C + UB for the three periods of 30, 60, and 90 min, they resulted in significantly higher amounts of gDNA obtained with the standard kit extraction protocol. As can be seen in this figure, the differences in the concentration of gDNA obtained with the 80 °C applied for 90 min and the combination of 80 °C + UB for 30, 60, and 90 min were up to four times higher than the differences seen with the other two kits: Nucleospin® Soil and the DNEasy® PowerSoil®.

For the quality of the DNA, the 260/280 absorbance ratio was measured for the default extraction and all pretreatments tested. As shown in Figure 2A for the Nucleospin® Soil kit, the application of 80 °C for 90 min, the UB applied for 90 min and the combination of UB + 80 °C for 60 min reached values in the ranges of 1.7 and 2.0 in the 260/280 absorbance ratio and were significantly higher than the default extraction protocol. The other pretreatments such as the combination of UB + 80 °C for 30 min, presented a significant difference from the default extraction; nevertheless, the values did not reach the quality ranges of

1.7 and 2.0 in the 260/280 absorbance ratio. Whereas with the pretreatment combination of UB + 80 °C for 90 min, no statistical differences were seen from the default extraction protocol and presented an average value of 1550, much lower than the ranges of 1.7 and 2.0 accepted for the 260/280 absorbance ratio (Table S1). Regarding the DNEasy® PowerSoil® kit, as seen in Figure 2B, qualities with significant differences from the default extraction protocol were seen when a temperature of 80 °C was applied for 90 min, and the combination of UB + 80 °C for 60 and 90 min reached the threshold for quality between 1.7 and 2.0 in the 260/280 absorbance ratio. Using this kit, the combination of UB + 80 °C for 30 min presented a significant difference with the default extraction protocol; however, the average did not reach the quality threshold of 1.7 and 2.0 in the 260/280 absorbance ratio. Regarding the application of UB for 90 min alone, no statistical difference was obtained compared to the default extraction protocol. Finally, for the E.Z.N.A.® Plant DNA Kit, the application of 80 °C for 90 min and the combination of UB + 80 °C temperature no matter the time periods used presented a significant difference from the default extraction protocol, surpassing the quality threshold ranges of 1.7 and 2.0 in the 260/280 absorbance ratio. The pretreatment of UB for 90 min instead did not present any significant difference from the default extraction protocol and did not reach the threshold ranges for the 260/280 absorbance ratio.

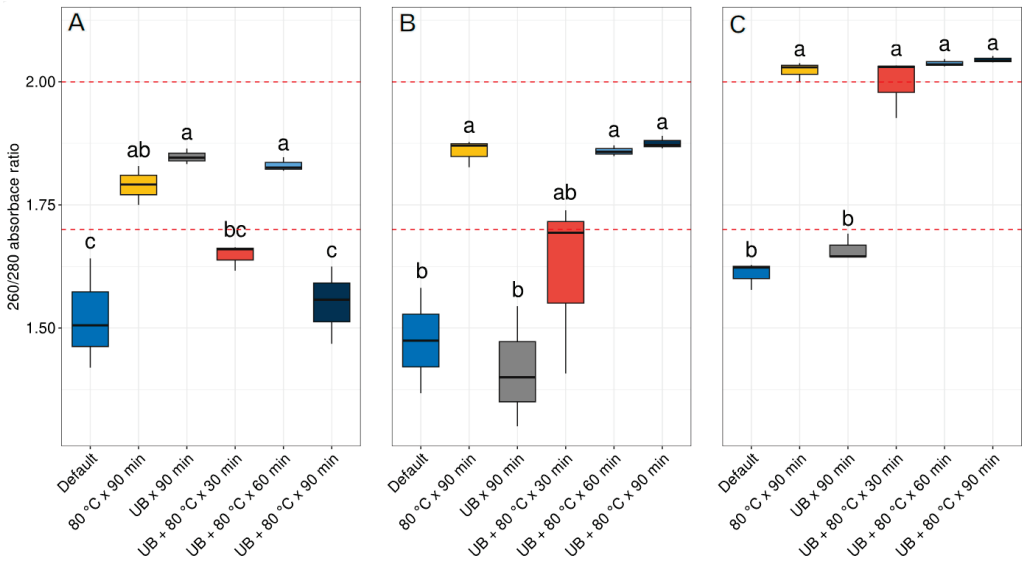


Figure 2. Quality of DNA extracted using three commercial kits. (**A**) Nucleospin® Soil, (**B**) DNEasy® PowerSoil® and (**C**) E.Z.N.A.® Plant DNA Kit, and the pretreatments applied before the extractions with the respective kits. Different letters indicate statistical differences according to the results of ANOVA followed by Tukey's post hoc test. UB = ultrasonic bath. The dashed red line indicates the DNA quality threshold for the 260/280 ratio considered adequate in the ranges 1.7 and 2.0.

Regarding the time required for each protocol, the Nucleospin® Soil has a standard extraction time of approximately 25 min per sample. The DNEasy® PowerSoil® extraction protocol requires approximately 30 min per sample and the E.Z.N.A.® Plant DNA Kit requires approximately 40 min per sample. For the pretreatments tested, applications of 80 °C for 90 min improved the quantity and quality of DNA, but the combination of UB + 80 °C helped to reduce the time from 1.5 h to 1 h and even half an hour in the case of the E.Z.N.A.® Plant DNA Kit.

3. Discussion

As shown in the results, we could observe here that the pretreatments tested were essential to improve the quantity and quality of gDNA extractions from SBR samples for NGS and effectively favored the amounts of gDNA obtained with the three commercial kits used. Although the three kits responded adequately to the pretreatments tested, the

kit that obtained the highest concentrations of gDNA was the E.Z.N.A.® Plant DNA Kit with the pretreatments of 80 °C for 90 min and the combination of UB + 80 °C for 30, 60, and 90 min. This was followed by the two other soil extraction kits, Nucleospin® Soil and DNEasy® PowerSoil®, which showed a similar response to the pretreatments with higher DNA concentrations and quality with the pre-treatments of 80 °C for 90 min and the combination of UB + 80 °C for 60 min.

One of the reasons that could explain a better performance of the commercial kit for plants, is that it includes reagents to break down the rigid polysaccharide cell wall and other compounds such as secondary metabolites like phenols, carbohydrates, and waxes that protect plant cells and hamper the extraction of DNA [14,19]. These E.Z.N.A.® Plant DNA Kit reagents, in the presence of sonication and high temperature, probably helped in the disruption of the EPS that hampered the obtention of DNA from the SBRs.

Regarding the 260/280 absorbance ratio, the quality obtained with the E.Z.N.A.® Plant DNA Kit was on average above the threshold of 1.7 and 2.0 for the 260/280 absorbance ratio established by Chen et al. (2010) [17]. In the same line, other studies established that the accepted absorbances could vary between 1.6 and 1.9 [20,21]. In general, the peak of UV absorption occurs at 260 nm for DNA and at 280 nm for proteins, and a 260/280 absorbance ratio of 1.8 indicates that the extraction is free from protein contamination [14,17,22]. Nevertheless, higher values slightly above 2.0 also indicate that the sample contains more DNA than proteins in the extraction obtained. Indeed, other studies, such as that of Vilanova et al., 2020 [23], accepted values slightly above 2.0 due to the high concentration of DNA and the effective results obtained by enzymatic digestion.

It is important to note that other measurements not considered in this study, such as the 260/230 absorbance ratio [19,22,23] or absorbance at 340 nm [24,25], help to determine the presence of contaminants such as humic acids in the extracted gDNA. Therefore, it is important to consider this kind of measurement with similar samples in future studies. The main reason for considering this measurement is that humic acids from organic matter have similar size and charge of DNA resulting in their co-extraction and influencing the efficiency and purity of the extraction process [26,27]. Some of the mechanisms proposed in the literature to mitigate the effect of humic acids when present in gDNA extraction procedures include the use of aluminum sulphate and powdered activated charcoal [25,26]. Although the 260/230 absorbance ratio measurement was not considered in this study, the improvements in gDNA extraction obtained here were sufficient for the amplification of the *16S rRNA gene* and internal transcribed spacer (ITS) fragments for Illumina sequencing, which will be part of another manuscript in progress.

As mentioned above, the other two extraction kits responded similarly, with the pretreatments tested reaching the threshold concentration of 5 $ng \cdot \mu L^{-1}$ required for NGS when exposed to a temperature of 80 °C for 90 min and the combination of UB + 80 °C for 60 and 90 min. Regarding the quality obtained, it is important to mention that the only protocol that exhibited a high quantity of gDNA but a low quality, was the combination of UB + 80 °C for 90 min pretreatment with the kit Nucleospin® Soil (Figure 2A). The main reason for the low quality obtained is probably that the longer time-lapsed exposure of 90 min favored not only DNA release but also a higher proportion of proteins from the EPSs.

As seen in our study, the mechanical effect of the UB for 90 min pretreatment was not sufficient to obtain significant differences with the standard extraction protocol with the three commercial kits tested. Similar results were obtained by Bourrain et al. (1999) [28] in which the sample dispersion was performed with an ultrasonic bath and by stirring in a cation exchange resin, with better results with the latter. Nevertheless, the use of a UB combined with a temperature of 80 °C helped to reduce the time from 90 min to 1 h (Figures 1 and 2). In other studies, such as that of Shan et al. (2008) [22], the use of SDS, mechanical milling, and thermal shock were the most effective ways to obtain better DNA yield and quality.

The results obtained here show that the main variable that significantly improved gDNA extraction from SBR samples was the application of a temperature of 80 °C for 90 min alone and the combination of a temperature of 80 °C and a UB for 60 min. Previous studies have reported the use of a combination of mechanical and chemical lysis. For example, glass bead vortexing was used for mechanical sludge lysis to allow better hydration with the extraction buffers [14], while CTAB was added to improve chemical lysis and helped to precipitate polysaccharides and humic substances [18].

4. Materials and Methods

A residual sewage sludge sample was first homogenized with vortex for 2 min and aliquoted in 54 tubes of 1 mL to consider the three extraction kits, the six treatments of temperature and sonication added, and the three replicates per sample analyzed ($3 \times 6 \times 3 = 54$). The tubes were centrifuged for 10 min at 10,000$\times$ g, and the supernatant was removed. The resulting biomass was weighed, and biomasses were equalized to 100 mg.

Three commercial DNA extraction kits were used: Genomic DNA from Soil (Nucleospin® Soil) Macherey-Nagel, Düren, Germany; DNEasy® PowerSoil® from Qiagen, Hilden, Germany; and E.Z.N.A.® Plant DNA Kit from Omega BIO-TEK, Norcross, GA, USA. To maintain the initial conditions used with the kits Genomic DNA from Soil and the DNEasy® PowerSoil® which contain 2 mm steel beads to homogenize samples, this step was also added for the E.Z.N.A.® Plant DNA Kit to homogenize samples. For the use of the three commercial kits, samples were homogenized twice in FastPrep-24™ MP Biomedicals™ equipment, at 6.5 m s^{-1} per minute.

To improve the efficiency and quality of the DNA extractions, different pretreatments were tested: (1) the default extraction protocol described in each kit, (2) the application of a temperature of 80 °C in a dry bath incubator (Allsheng, model MK 200-1, Hangzhou, China) for 90 min, (3) a pretreatment with only an ultrasonic bath (UB) (Elmasonic E30H, Singen, Germany) for 90 min, (4) a UB combined with 80 °C for 30 min, (5) a UB combined with 80 °C for 60 min, and (6) a UB combined with 80 °C for 90 min. For each pretreatment, the extractions were conducted in triplicate and the median of concentrations measured were used for graph constructions and the concentrations were normalized to 100 mg of biomass. After the application of the different pretreatments tested, the extraction protocols were followed as described by the corresponding provider protocols.

The concentration of DNA from the samples extracted was found with equipment Qubit 4.0 (Invitrogen, Woodlands, Singapore) with a 1X dsDNA HS Assay Kit, following the manufacturer's instructions. To determine the quality of the DNA obtained, the relation 260/280 absorbance was obtained with the spectrophotometer Synergy H1 Hybrid Reader (Biotek, Santa Clara, CA, USA), with an adaptor for the quantification of nucleic acids Take3 and each sample was read in duplicate following the manufacturer's instructions. A load above 5 ng·μL^{-1} was established for Illumina sequencing as described by the protocol "16S Metagenomic Sequencing Library Preparation" (Illumina, Inc., San Diego, CA, USA) [16]. While the ranges of quality were established between 1.7 and 2.0 for the absorbance ratio of 260/280 as described by Chen et al. (2010) [17].

The visualization and statistical analysis were conducted In R language using the libraries from *ggpubr*, *ggplot2*, *car*, *agricolae*, and *dplyr* [29–32]. To evaluate the normal distribution of data, the Levene and Shapiro–Wilk test were first conducted, and to compare the pretreatments tested, a one-way ANOVA was conducted followed by a Tukey's post hoc test to determine the significance of the results obtained.

5. Conclusions

As seen in this study, the pretreatments tested provided substantial improvement to the quantity and quality of genomic DNA (gDNA) from samples from sequential batch reactors' (SBRs) residues extracted with commercial kits based on silica matrices where gDNA binds. For the three commercial kits tested, these pretreatments increased the concentration of

high molecular weight gDNA and its quality, especially when a temperature of 80 °C was applied for 90 min, and in the treatments including a combination of 80 °C temperature and an ultrasonic bath (UB) for 60 min with all commercial kits used. The pretreatments tested in this communication were found to be adequate for next-generation sequencing (NGS) analysis, which will be part of an incoming original paper under construction.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms242015243/s1.

Author Contributions: Conceptualization, C.V., F.P.C.-A. and M.A.; methodology, C.V., I.S., F.P.C.-A. and B.L.-C.; formal analysis, B.L.-C. and C.V.; writing—original draft preparation, C.V., F.P.C.-A. and B.L.-C.; writing—review and editing, F.P.C.-A. and F.P.; supervision, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: F.P.C.-A. was supported by ANID Postdoctoral fellowship 3200728, and while writing, F.P.C.-A. was supported by PIA21-0002 from Dirección de Investigación y Postgrado, Universidad de La Frontera. F.P. wishes to acknowledge the financial support provided by ANID FONDECYT POSTDOCTORAL project No. 3210626. M.A. was supported by FONDEF ID19I10299. C.V. received the support of the Doctoral scholarship 21210454 from ANID.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the SmartC-BIOREN (Service Management Analytical Research and Training Center), ANID Project No. CCSS210005. This work was partially funded by the Dirección de Investigación, Universidad de La Frontera. This research was partially supported by the Soroban supercomputing infrastructure (SATREPS MACH Project) at the Centro de Modelación y Computación Científica of the Universidad de La Frontera (Chile). Two anonymous reviewers provided useful criticism that improved an earlier version of this communication.

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 the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Jeong, H.; Park, S.Y.; Ryu, G.H.; Choi, J.H.; Kim, J.H.; Choi, W.S.; Lee, S.M.; Choi, J.W.; Choi, I.G. Catalytic conversion of hemicellulosic sugars derived from biomass to levulinic acid. *Catal. Commun.* **2018**, *117*, 19–25. [CrossRef]
2. Dietrich, K.; Dumont, M.J.; Del Rio, L.F.; Orsat, V. Sustainable PHA production in integrated lignocellulose biorefineries. *New Biotechnol.* **2019**, *49*, 161–168. [CrossRef] [PubMed]
3. Ren, H.; Zhou, Y.; Liu, L. Selective conversion of cellulose to levulinic acid via microwave-assisted synthesis in ionic liquids. *Bioresour. Technol.* **2013**, *129*, 616–619. [CrossRef] [PubMed]
4. Lappalainen, K.; Dong, Y. Simultaneous production of furfural and levulinic acid from pine sawdust via acid-catalysed mechanical depolymerization and microwave irradiation. *Biomass Bioenergy* **2019**, *123*, 159–165. [CrossRef]
5. Kruyer, N.S.; Wauldron, N.; Bommarius, A.S.; Peralta-Yahya, P. Fully biological production of adipic acid analogs from branched catechols. *Sci. Rep.* **2020**, *10*, 13367. [CrossRef]
6. Wu, M.; Di, J.; Gong, L.; He, Y.C.; Ma, C.; Deng, Y. Enhanced adipic acid production from sugarcane bagasse by a rapid room temperature pretreatment. *Chem. Eng. J.* **2023**, *452*, 139320. [CrossRef]
7. Cassuriaga, A.P.A.; Freitas, B.C.B.; Morais, M.G.; Costa, J.A.V. Innovative polyhydroxybutyrate production by Chlorella fusca grown with pentoses. *Bioresour. Technol.* **2018**, *265*, 456–463. [CrossRef]
8. Pinto-Ibieta, F.; Cea, M.; Cabrera, F.; Abanto, M.; Felissia, F.E.; Area, M.C.; Ciudad, G. Strategy for biological co-production of levulinic acid and polyhydroxyalkanoates by using mixed microbial cultures fed with synthetic hemicellulose hydrolysate. *Bioresour. Technol.* **2020**, *309*, 123323. [CrossRef]
9. Pinto-Ibieta, F.; Cea, M.; Serrano, A.; Felissia, F.E.; Area, M.C.; Cabrera, F.; Ciudad, G. Toward the use of mixed microbial cultures for the biological production of adipic and levulinic acid. *Front. Microbiol.* **2023**, *14*, 1224543. [CrossRef]
10. Pinto-Ibieta, F.; Serrano, A.; Cea, M.; Ciudad, G.; Fermoso, F.G. Beyond PHA: Stimulating intracellular accumulation of added-value compounds in mixed microbial cultures. *Bioresour. Technol.* **2021**, *337*, 125381. [CrossRef]

11. Begmatov, S.; Dorofeev, A.G.; Kadnikov, V.V.; Beletsky, A.V.; Pimenov, N.V.; Ravin, N.V.; Mardanov, A.V. The structure of microbial communities of activated sludge of large-scale wastewater treatment plants in the city of Moscow. *Sci. Rep.* **2022**, *12*, 3458. [CrossRef] [PubMed]
12. Hulshoff, L.W.; De Castro Lopes, S.I.; Lettinga, G.; Lens, P.N.L. Anaerobic sludge granulation. *Water Res.* **2004**, *38*, 1376–1389. [CrossRef] [PubMed]
13. Seviour, R.; Nielsen, P.H. *Microbial Ecology of Activated Sludge*; McIlroy, S., Porter, K., Eds.; DNA and RNA extraction; IWA Publishing: London, UK, 2010; ISBN 9781441980731.
14. Singka, D.; Kumdhitiahutsawakul, L.; Rekkriangkrai, P.; Pathom-Aree, W. A simple method for DNA extraction from activated sludge. *Chiang Mai J. Sci.* **2012**, *39*, 111–118.
15. Yarimizu, K.; Fujiyoshi, S.; Kawai, M.; Norambuena-Subiabre, L.; Cascales, E.K.; Rilling, J.I.; Vilugrón, J.; Cameron, H.; Vergara, K.; Morón-López, J.; et al. Protocols for monitoring harmful algal blooms for sustainable aquaculture and coastal fisheries in chile. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7642. [CrossRef]
16. Vaillancourt, B.; Buell, C.R. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv* **2019**. [CrossRef]
17. Chen, H.; Rangasamy, M.; Tan, S.Y.; Wang, H.; Siegfried, B.D. Evaluation of five methods for total DNA extraction from western corn rootworm beetles. *PLoS ONE* **2010**, *5*, e11963. [CrossRef] [PubMed]
18. Roose-Amsaleg, C.; Garnier-Sillam, E.; Harry, M. Extraction and purification of microbial DNA from soil and sediment samples. *Appl. Soil Ecol.* **2001**, *18*, 47–60. [CrossRef]
19. Rana, M.M.; Aycan, M.; Takamatsu, T.; Kaneko, K.; Mitsui, T.; Itoh, K. Optimized nuclear pellet method for extracting next-generation sequencing quality genomic DNA from fresh leaf tissue. *Methods Protoc.* **2019**, *2*, 54. [CrossRef] [PubMed]
20. Martínez-González, C.R.; Ramírez-Mendoza, R.; Jiménez-Ramírez, J.; Gallegos-Vázquez, C.; Luna-Vega, I. Improved method for genomic DNA extraction for *Opuntia* Mill. (Cactaceae). *Plant Methods* **2017**, *13*, 82. [CrossRef]
21. Cerveira de Souza, D.; Teixeira, T.A. A simple and effective method to obtain high DNA quality and quantity from Cerrado plant species. *Mol. Biol. Rep.* **2019**, *46*, 4611–4615. [CrossRef]
22. Shan, G.; Jin, W.; Lam, E.K.; Xing, X. Purification of total DNA extracted from activated sludge. *J. Environ. Sci.* **2008**, *20*, 80–87. [CrossRef] [PubMed]
23. Vilanova, S.; Alonso, D.; Gramazio, P.; Plazas, M.; García-Fortea, E.; Ferrante, P.; Schmidt, M.; Díez, M.J.; Usadel, B.; Giuliano, G.; et al. SILEX: A fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods* **2020**, *16*, 110. [CrossRef] [PubMed]
24. Sharma, S.; Sharma, K.K.; Kuhad, R.C. An efficient and economical method for extraction of DNA amenable to biotechnological manipulations, from diverse soils and sediments. *J. Appl. Microbiol.* **2014**, *116*, 923–933. [CrossRef] [PubMed]
25. Devi, S.G.; Fathima, A.A.; Radha, S.; Arunraj, R.; Curtis, W.R.; Ramya, M. A rapid and economical method for efficient DNA extraction from diverse soils suitable for metagenomic applications. *PLoS ONE* **2015**, *10*, e0132441. [CrossRef] [PubMed]
26. Dong, D.; Yan, A.; Liu, H.; Zhang, X.; Xu, Y. Removal of humic substances from soil DNA using aluminium sulfate. *J. Microbiol. Methods* **2006**, *66*, 217–222. [CrossRef]
27. Wnuk, E.; Waśko, A.; Walkiewicz, A.; Bartmiński, P.; Bejger, R.; Mielnik, L.; Bieganowski, A. The effects of humic substances on DNA isolation from soils. *PeerJ* **2020**, *8*, e9378. [CrossRef]
28. Bourrain, M.; Achouak, W.; Urbain, V.; Heulin, T. DNA extraction from activated sludges. *Curr. Microbiol.* **1999**, *38*, 315–319. [CrossRef]
29. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag, Ed.; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
30. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 3rd ed.; Sage Publications: Thousand Oaks, CA, USA, 2019; ISBN 9781544336473.
31. Wickham, H.; Francois, R.; Müller, K.; Vaughan, D. Dplyr: A Grammar of Data Manipulation. 2023. Available online: https://dplyr.tidyverse.org (accessed on 4 May 2023).
32. Mendiburu, F. Agricolae: Statistical Procedures for Agricultural Research. *R Packag. Version* **2010**, *1*, 1–8.

International Journal of
Molecular Sciences

MDPI

Article

Crystal Structure of DNA Replication Protein SsbA Complexed with the Anticancer Drug 5-Fluorouracil

Hsin-Hui Su [1], Yen-Hua Huang [2], Yi Lien [3], Po-Chun Yang [2] and Cheng-Yang Huang [2,4,*]

[1] Department of Pharmacy, Chia Nan University of Pharmacy and Science, Tainan City 717, Taiwan
[2] Department of Biomedical Sciences, Chung Shan Medical University, Taichung City 402, Taiwan
[3] Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
[4] Department of Medical Research, Chung Shan Medical University Hospital, Taichung City 402, Taiwan
* Correspondence: cyhuang@csmu.edu.tw

Abstract: Single-stranded DNA-binding proteins (SSBs) play a crucial role in DNA metabolism by binding and stabilizing single-stranded DNA (ssDNA) intermediates. Through their multifaceted roles in DNA replication, recombination, repair, replication restart, and other cellular processes, SSB emerges as a central player in maintaining genomic integrity. These attributes collectively position SSBs as essential guardians of genomic integrity, establishing interactions with an array of distinct proteins. Unlike *Escherichia coli*, which contains only one type of SSB, some bacteria have two paralogous SSBs, referred to as SsbA and SsbB. In this study, we identified *Staphylococcus aureus* SsbA (SaSsbA) as a fresh addition to the roster of the anticancer drug 5-fluorouracil (5-FU) binding proteins, thereby expanding the ambit of the 5-FU interactome to encompass this DNA replication protein. To investigate the binding mode, we solved the complexed crystal structure with 5-FU at 2.3 Å (PDB ID 7YM1). The structure of glycerol-bound SaSsbA was also determined at 1.8 Å (PDB ID 8GW5). The interaction between 5-FU and SaSsbA was found to involve R18, P21, V52, F54, Q78, R80, E94, and V96. Based on the collective results from mutational and structural analyses, it became evident that SaSsbA's mode of binding with 5-FU diverges from that of SaSsbB. This complexed structure also holds the potential to furnish valuable comprehension regarding how 5-FU might bind to and impede analogous proteins in humans, particularly within cancer-related signaling pathways. Leveraging the information furnished by the glycerol and 5-FU binding sites, the complexed structures of SaSsbA bring to the forefront the potential viability of several interactive residues as potential targets for therapeutic interventions aimed at curtailing SaSsbA activity. Acknowledging the capacity of microbiota to influence the host's response to 5-FU, there emerges a pressing need for further research to revisit the roles that bacterial and human SSBs play in the realm of anticancer therapy.

Keywords: 5-fluorouracil; SSB; anticancer drug; single-stranded DNA-binding protein; SsbA; crystal structure; docking; conformational change

Citation: Su, H.-H.; Huang, Y.-H.; Lien, Y.; Yang, P.-C.; Huang, C.-Y. Crystal Structure of DNA Replication Protein SsbA Complexed with the Anticancer Drug 5-Fluorouracil. *Int. J. Mol. Sci.* **2023**, *24*, 14899. https://doi.org/10.3390/ijms241914899

Academic Editor: Wajid Zaman

Received: 31 August 2023
Revised: 27 September 2023
Accepted: 2 October 2023
Published: 4 October 2023

1. Introduction

Nucleobases play a pivotal role as fundamental constituents within nucleic acids, orchestrating the replication of genetic information across all biological systems [1]. The accurate synthesis of nucleotides stands as a critical linchpin for the survival and proliferation of both eukaryotic and prokaryotic cells [2]. The structural alteration of nucleobases holds the potential to exert substantial impacts, displaying potent biological effects [3]. Throughout an expansive range encompassing anticancer [4], antiviral [5], antibacterial [6], anti-inflammatory [7], and antitumor activities [8], numerous derivatives of uracil have garnered longstanding employment. One standout exemplar in this domain is the FDA-approved anticancer agent, 5-fluorouracil (5-FU) [4]. In 5-FU, the hydrogen situated at the C5 position of uracil is supplanted by a fluorine atom, culminating in a fluoropyrimidine

configuration. This modification empowers 5-FU to effectively target the enzyme thymidylate synthase (TSase) for anticancer chemotherapy [9]. The cytotoxic influences of 5-FU emerge through its adept ability to impede the operation of TSase, induce RNA miscoding, and activate apoptosis. In the dynamic landscape of drug development, although numerous novel agents have been conceived, 5-FU persists as a cornerstone within the arsenal of chemotherapeutic modalities. It prominently features in systemic treatments for an array of cancers spanning the gastrointestinal tract, breast, head, and neck [9]. Notably, 5-FU manifests diverse interactions with over a dozen distinct proteins, including dihydropyrimidinase (DHPase) [10], an enzyme involved in the pyrimidine degradation [11]. Strikingly, 5-FU-induced toxicity has been found in asymptomatic patients with DHPase deficiency. Such patients, undergoing anticancer therapy with 5-FU, confront severe toxicity manifestations, including instances of mortality [12]. Beyond the confines of human genetics [12,13] and the associations with human gene products [14], the intricate web of host–microbiota interactions ushers in additional dimensions to the 5-FU narrative [15–17]. Evidently, the active gut microbiota, equipped with the capacity to synthesize bromovinyluracil, can exert profound regulatory influences on systemic 5-FU concentrations [17,18]. This unforeseen modulation results in an adverse outcome, as demonstrated by a tragic occurrence where 5-FU exposure triggered the demise of 16 patients in Japan [17,18]. Given the intricate tapestry of interactions that envelop 5-FU, the imperative emerges to comprehensively construct its interactome. Such a feat is essential for in-depth analyses of clinical pharmacokinetics and toxicity profiles [19–22]. The holistic elucidation of the multifaceted relationships woven by 5-FU holds the promise of enhancing our grasp of its intricate dynamics, paving the way for refined therapeutic strategies and personalized medicine. To attain this objective, the initial stride involves the identification of additional proteins that interact with 5-FU.

DNA is subject to constant challenges, such as replication errors, DNA damage, and the formation of secondary structures [23]. To ensure its integrity and stability, cells have evolved an intricate network of proteins involved in DNA metabolism [24–26]. Single-stranded DNA (ssDNA)-binding protein (SSB) is one of central players in these processes, safeguarding and manipulating ssDNA during various cellular events [27]. Through its multifaceted roles in DNA replication, recombination, repair, replication restart, and other cellular processes, SSB emerges as a critical player in maintaining genomic integrity [28–30]. SSB binds specifically to ssDNA with high affinity, preventing its re-annealing, protecting it from nucleases, and promoting its accessibility to other DNA-binding proteins [31]. SSBs are conserved across organisms, from bacteria [32] to humans [33,34], highlighting their fundamental importance. SSBs have undergone extensive investigation in eubacteria, with a notable focus on the *Escherichia coli* SSB (EcSSB) [31]. The majority of SSBs adopt homotetrameric configurations for activity [35]. This arrangement involves four oligonucleotide/oligosaccharide-binding folds (OB folds) coalescing to constitute a DNA-binding domain [36,37]. Beyond their DNA-binding functions, SSBs also establish interactions with a multitude of DNA-metabolism proteins, collectively forming the SSB interactome [38]. EcSSB comprises two primary domains: an N-terminal ssDNA-binding/oligomerization domain (SSBn) and a flexible C-terminal domain for protein–protein interactions (SSBc). Within EcSSBc, a further division into two sub-domains emerges, namely an intrinsically disordered linker (IDL) and an acidic tip [39]. Bacterial SSBs from distinct sources share moderate sequence homology, especially within the SSBn domain, encompassing roughly the initial 110 residues. The ubiquity of this homology designates SSBn as a prospective common target for devising inhibitors against various SSBs [40–43]. Diverging from *E. coli*, which harbors a solitary SSB variant (EcSSB), certain bacteria like *Staphylococcus aureus* exhibit a more complex scenario with the presence of three paralogous SSBs, specifically referred to as SaSsbA [44,45], SaSsbB [46], and SaSsbC [47]. Of these, SaSsbA may be an EcSSB counterpart due to its possession of an acidic tip and sequence resemblance to EcSSBn. Given this alignment, the ongoing pursuit of molecules that bind to and inhibit SaSsbA holds notable promise for future applications

in combating pathogens [44,48]. Although the identification of SaSsbB as a protein that binds to 5-FU is established [49], it holds significance to investigate whether 5-FU can also interact with SaSsbA, and delving into the mode of this interaction is a valuable pursuit.

Although 5-FU has been under investigation for over six decades [50,51], its interactions with proteins remain challenging to predict. In this study, we identified SaSsbA as a fresh addition to the roster of 5-FU binding proteins, and, thus, the 5-FU interactome was extended to include this essential DNA replication protein. In our pursuit of comprehending the precise interaction sites between 5-FU and SaSsbA, we solved the complexed crystal structure at a resolution of 2.3 Å (PDB ID 7YM1). The structure of SaSsbA complexed with glycerol was also determined at 1.8 Å (PDB ID 8GW5). The binding sites within the structure of this OB-fold protein emerge as prime contenders for potential drug design efforts. To further characterize the binding affinity and confirm the interacting sites of 5-FU with SaSsbA, fluorescence quenching and mutational analysis were conducted. In addition, we also compared the 5-FU binding modes between SaSsbA and SaSsbB. The interaction of SaSsbB with 5-FU relies on specific residues T12, K13, T30, F48, and N50 [49], and these residues remain conserved within SaSsbA [45,46]. Given that SaSsbA exhibits structural parallels with SaSsbB, one might initially infer that 5-FU's binding capabilities would align, and the 5-FU binding mode of SaSsbA would echo that of SaSsbB. However, our investigation unveiled a disparity between their 5-FU binding sites. Accordingly, this complexed crystal structure unfurls a molecular insight into the distinct manner in which the anticancer drug 5-FU binds to a cognate protein, even when the structural scaffold appears analogous yet divergent.

2. Results

2.1. Sequence Analysis of SaSsbA

According to the nucleotide sequence available on NCBI, the projected monomeric SaSsbA protein spans 167 amino acid residues (aa), with a calculated molecular mass of 19 kDa. The alignment consensus of sequenced SSB homologs, facilitated by ConSurf analysis, delineated the extent of variability at distinct positions along the sequence (Figure 1). It became apparent that the aa 107–148 segment (teal) in SaSsbA lacks conservation within the SSB homologs. Drawing from insights garnered from the EcSSB-ssDNA complex [52], it emerges that four crucial aromatic residues, namely W40, W54, F60, and W88, universally preserved within most SSB families as F/Y/W, engage in ssDNA binding through stacking interactions. Correspondingly, in SaSsbA, the corresponding residues manifest as F37, F48, F54, and Y82. A notable absence in SaSsbA is the W residue. Moreover, the EcSSB's significant C-terminal tail DDDIPF, which plays a role in protein–protein interaction, undergoes a modification to DDDLPF in SaSsbA. The GGRQ motif postulated as a regulatory switch for ssDNA binding [53], which in SaSsbA could be replaced by the GGQR motif (aa 112–115). In the context of the PXXP motifs within EcSSB, positioned at aa 139 (PQQP), 156 (PQQS), and 161 (PAAP), and acknowledged for their role in mediating protein–protein interactions [39], their presence is notably absent in SaSsbA (Figure 1). While SaSsbA is presumed to be a primary SSB homolog similar to EcSSB in structure and function, divergences arise in their gene locations [45]. In the genetic map of *S. aureus*, the *ssbA* gene resides flanked by *rpsF* (encoding the ribosomal protein S6) and *rpsR* (encoding ribosomal protein S18) genes, all encompassed within one operon regulated by the SOS response [54]. The scenario deviates in *E. coli*, where the *ssb* gene is found adjacent to the *uvrA* gene, positioned distant from that *S. aureus* operon. Instead, the *prib* gene (coding for PriB, an ssDNA-binding protein) [55] is flanked by the *rpsF* and *rpsR* genes in the genetic map of *E. coli*. The rationale behind the necessity for the evolution of distinct SSBs in particular species remains a subject that warrants elucidation.

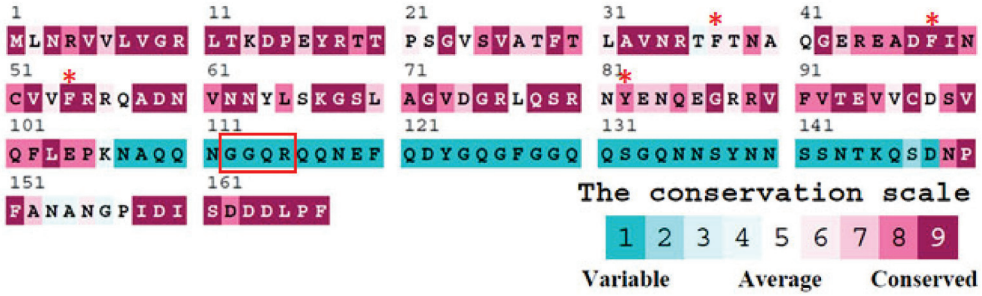


Figure 1. Sequence analysis of SaSsbA reveals an alignment consensus of sequenced SSB homologs through ConSurf analysis, which effectively showcases the extent of variability exhibited at each position along the sequence. In this depiction, amino acid residues that span a wide range of variability are depicted in teal, while those that remain highly conserved are marked in burgundy. Residues F37, F48, F54, and Y82, which potentially participate in ssDNA binding through stacking interactions, are denoted by asterisks. Of significance is the corresponding GGRQ motif found in SaSsbA, which is highlighted within a red box. The PXXP motifs seen in EcSSB are notably absent in SaSsbA. Another notable contrast arises in the critical C-terminal tail. While the DDDIPF sequence of EcSSB plays a vital role in protein–protein interaction, this sequence takes the form of DDDLPF in SaSsbA.

2.2. Crystallization of the Glycerol-Bound SaSsbA and the SaSsbA-5-FU Complex

SaSsbA with a His tag was overexpressed in *E. coli* through heterologous expression and subsequently isolated from the soluble supernatant using Ni^{2+}-affinity chromatography. The purified SaSsbA was concentrated to a concentration of 20 mg/mL, with the addition of the cryoprotectant glycerol to attain a final concentration of 25%. This glycerol-enhanced formulation allowed for storage at −20 °C. Previously, we observed that crystals of apo-SaSsbA could be cultivated at room temperature using the hanging drop vapor diffusion method in a solution composed of 22% PEG 4000, 100 mM HEPES, and 100 mM sodium acetate at pH 7.5 [45]. Efforts were initially directed towards soaking and cocrystallization of SaSsbA (20 mg/mL) with 5-FU (200 µM) under identical crystallization conditions to those for apo-SaSsbA. The goal was to obtain crystals of the SaSsbA–5-FU complex, yet these attempts proved unfruitful. Subsequent rescreening was undertaken utilizing commercial crystallization kits. Under these new conditions (JBScreen Classic 2, Jena Bioscience, Jena, Germany), crystals of the SaSsbA–5-FU complex emerged at room temperature within a mixture containing 16% PEG 4000, 100 mM Tris-HCl, and 200 mM $MgCl_2$ at pH 8.5. For glycerol-bound SaSsbA, crystallization occurred in a solution containing 30% PEG 4000, 100 mM HEPES, and 200 mM $CaCl_2$ at pH 7.5 (Table 1).

Table 1. Data collection and refinement statistics.

Data Collection		
Crystal	SaSsbA–5-FU complex	Glycerol-bound SaSsbA
Wavelength (Å)	1	1
Resolution (Å)	27.8–2.36	27.9–1.80
Space group	$P4_12_12$	$P4_12_12$
Cell dimension		
a, *b*, *c* (Å)	88.09, 88.09, 57.78	88.22, 88.22, 58.00
β (°)	90.00	90.00
Redundancy	7.8 (7.9)	9.4 (9.4)
Completeness (%)	99.9 (100.0)	100.0 (100.0)
<I/σI>	31.6 (8.2)	35.7 (4.4)
$CC_{1/2}$	0.997 (0.974)	0.989 (0.955)

Table 1. *Cont.*

Data Collection		
Refinement		
No. of reflections	9797	21,823
R_{work}/R_{free}	0.203/0.244	0.199/0.224
No. of atoms		
Protein	199	200
Water	52	151
5-FU	1	0
Glycerol	1	2
r.m.s deviations		
Bond lengths (Å)	0.008	0.007
Bond angles (°)	1.06	1.05
Ramachandran plot		
Favored (%)	98.43	97.40
Allowed (%)	1.57	2.60
Outliers (%)	0	0
PDB entry	7YM1	8GW5

Values in parentheses are for the highest resolution shell. $CC_{1/2}$ is the percentage of correlation between intensities of random half-data sets.

2.3. Crystal Structure of Glycerol-Bound SaSsbA

The crystal structure of glycerol-bound SaSsbA was successfully determined at a resolution of 1.8 Å (Table 1). The crystals of glycerol-bound SaSsbA belonged to the $P4_12_12$ space group, featuring cell dimensions with a = 88.22 Å, b = 88.22 Å, and c = 58.00 Å. This structure of SaSsbA with glycerol (PDB ID 8GW5) was elucidated via molecular replacement, utilizing the apo-SaSsbA as a model (PDB ID 5XGT). Examination of a Ramachandran plot revealed no presence of unallowed regions (outliers) within this structure. While SaSsbA is typically active as tetramers [45], this glycerol-bound structure contained only two monomers of SaSsbA within each asymmetric unit (Figure 2A). In contrast to the crystal structure of apo-SaSsbA [45], this structure displayed two additional amino acid residues, namely P105 and K106, as indicated by the blue mesh (Figure 2A), solely in the subunit A of SaSsbA. The C-terminal region in SaSsbA, comprising aa 107–167 in the subunit A and aa 105–167 in the subunit B, was not observed. This suggests that the C-terminal region of SaSsbA exhibits dynamic behavior, a feature reminiscent of EcSSB [56]. For this SaSsbA dimer, the electron density mostly exhibited satisfactory quality. Nonetheless, some sections remained disordered and unobserved, including aa 38–44 (loop L_{23}) in subunit A and aa 40–42 (loop L_{23}) in subunit B within the ternary structure of this glycerol-bound SaSsbA. In congruence with the apo form, the overall architecture of this glycerol-bound SaSsbA monomer remained characteristic of an OB-fold structure, marked by a β-barrel (composed of 5 β-strands) crowned with an α-helix. It is noteworthy that SaSsbA did not encompass the β6 strand, a component found in numerous other SSBs such as those from *E. coli* [52], *Salmonella enterica* [41], *Klebsiella pneumoniae* [53], and *Pseudomonas aeruginosa* [40,42,46,57,58]. In tetrameric SSBs, the β6 strand has been proposed to be involved in mediating diverse protein–DNA and protein–protein interaction specificities among distinct SSBs [59]. Within this structure of an SaSsbA dimer, two glycerol molecules (designated as Glycerol 1 and Glycerol 2) were present. However, these glycerol molecules exhibited distinct binding patterns to SaSsbA (see below).

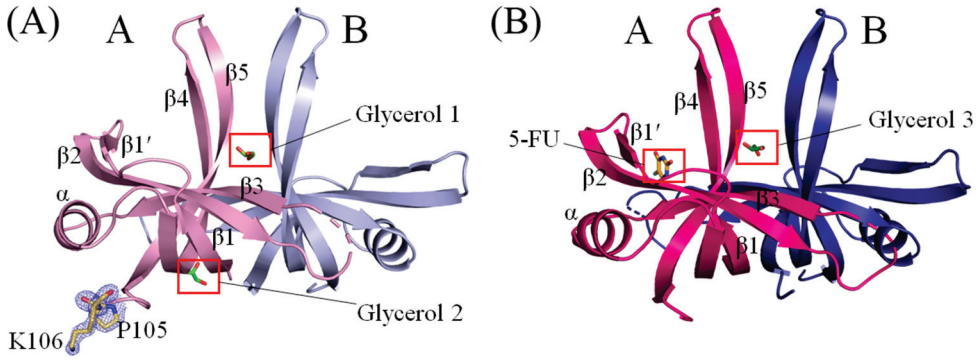


Figure 2. Crystal Structures of SaSsbA. (**A**) The crystal structure of SaSsbA featuring two glycerol molecules is displayed. The SaSsbA monomer adopts an OB-fold structure, composed of a β-barrel comprising 5 β-strands, capped with an α-helix. These two glycerol molecules are designated as Glycerol 1 (splitpea) and Glycerol 2 (green). Different SaSsbA monomers are colored in light pink and light blue. In contrast to the crystal structure of apo-SaSsbA, this complexed structure displayed two additional amino acid residues, namely P105 and K106, as indicated by the blue mesh. (**B**) The complexed crystal structure of SaSsbA with one 5-FU molecule and one glycerol molecule is presented. Different SaSsbA monomers are colored in hot pink and deep blue. Notably, 5-FU was localized solely in monomer A, with no presence in monomer B. The glycerol molecule is labeled as Glycerol 3 (forest).

2.4. Crystal Structure of SaSsbA Complexed with 5-FU

The complexed crystal structure of SaSsbA with 5-FU (PDB ID 7YM1) was successfully determined at a resolution of 2.3 Å (Table 1) with molecular replacement employing the apo-SaSsbA as a model (PDB ID 5XGT). The crystals of the SaSsbA–5-FU complex were categorized under the $P4_12_12$ space group, showcasing cell dimensions of a = 88.09 Å, b = 88.09 Å, and c = 57.78 Å. The completeness exceeded 99%. Upon scrutinizing a Ramachandran plot, there were no regions featuring unallowed conformations (outliers) within this structure. The electron density mostly exhibited satisfactory quality. However, aa 39–44 (loop L_{23}) in subunit A and aa 39–41 (loop L_{23}) in subunit B within the ternary structure of the SaSsbA–5-FU complex were disordered and unobserved. In both subunits, the range of aa 105–167 was also not detected. The individual monomer within this complexed SaSsbA revealed the characteristic OB-fold structure, featuring a β-barrel comprising 5 β-strands capped with an α-helix. Within this complex structure of a SaSsbA dimer, a single glycerol molecule and one 5-FU molecule were encapsulated (Figure 2B). The positioning of this glycerol molecule, designated as Glycerol 3, within the SaSsbA–5-FU complex closely mirrored that of Glycerol 1 within the glycerol-bound SaSsbA.

2.5. Glycerol 1 Binding Mode of SaSsbA

The presence of the cryoprotectant glycerol within the protein solution led to its binding to SaSsbA. In this study, three distinct glycerol binding sites within our two structures were found. Notably, the binding sites for Glycerol 1 and Glycerol 3 exhibited similarity. As revealed by the crystal structure (PDB ID 8GW5), Glycerol 1 is sandwiched by SaSsbA monomers A and B (Figures 2A and 3A). A comprehensive analysis of the interactions transpiring between Glycerol 1 and SaSsbA was undertaken, resulting in the identification of multiple residues that came within contact distance (<4 Å) with the glycerol molecule. Among these interacting residues were F48 (Subunit B), N50 (Subunit B), S79 (Subunit A), R80 (Subunit A), F91 (Subunit A), V92 (Subunit A), and T93 (Subunit A). Delving into the nature of these interactions, hydrogen bonds formed between the ligand and SaSsbA were meticulously examined by leveraging PLIP (the protein–ligand interaction profiler) [60]. In light of the interactions identified through PLIP, it was established that the main chains of R80 and V92, along with the side chains of S79 and T93, partook in hydrogen bonding with Glycerol 1 (Figure 3B).

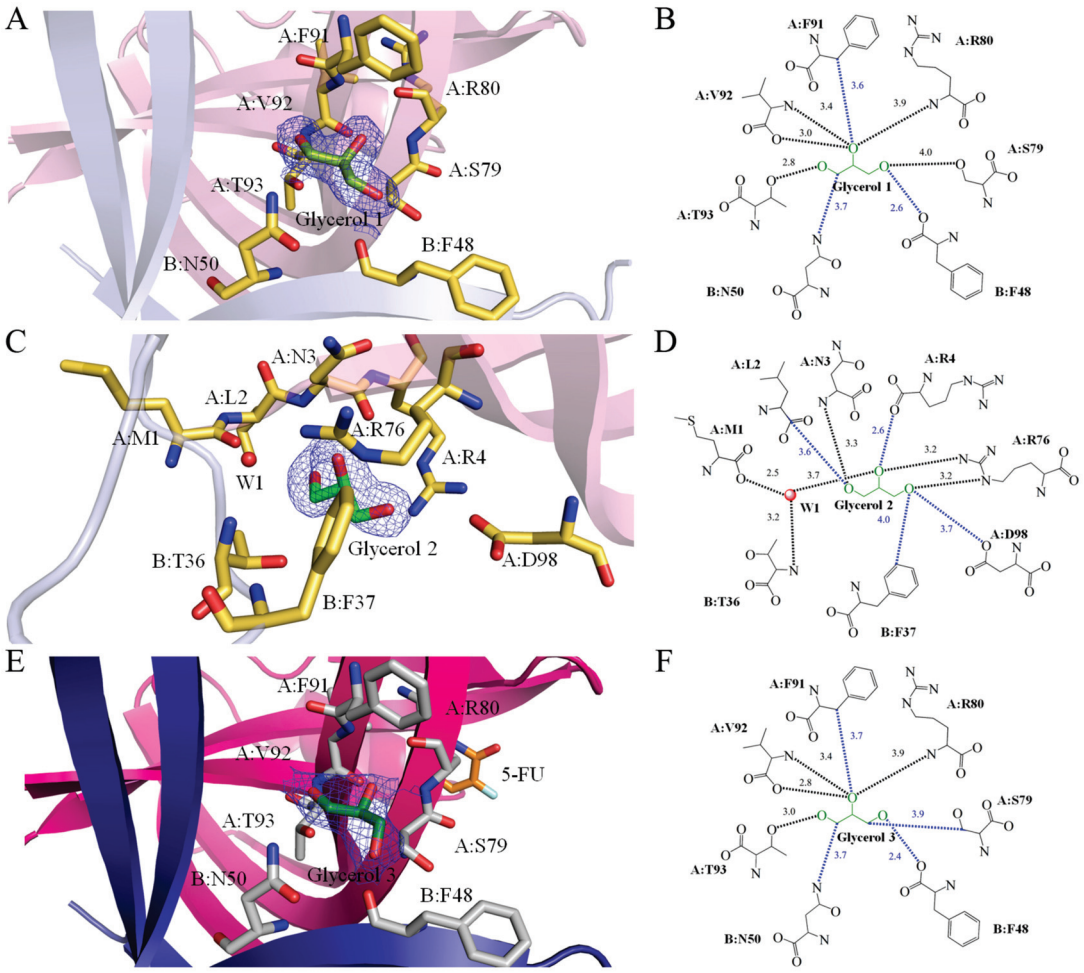


Figure 3. Glycerol binding modes. (**A**) The binding site of Glycerol 1 within SaSsbA was unveiled through the structure of glycerol-bound SaSsbA (PDB ID 8GW5). Glycerol 1 (splitpea) was positioned between SaSsbA monomers A (light pink) and B (light blue). Residues engaging with Glycerol 1 are colored in yellow. The composite omit map (presented as blue mesh, contoured at 1 σ) indicated the presence of Glycerol 1 within a cavity formed at the SaSsbA monomers A and B interface. (**B**) Depiction of the binding mode for Glycerol 1. Residues F48 (Subunit B), N50 (Subunit B), S79 (Subunit A), R80 (Subunit A), F91 (Subunit A), V92 (Subunit A), and T93 (Subunit A), situated within a contact distance of <4 Å, were instrumental in binding Glycerol 1. The interactive distances are also shown in Å. Based on interactions detected by PLIP, hydrogen bonds were formed between the main chains of R80 and V92, as well as the side chains of S79 and T93, and Glycerol 1 (indicated in black). (**C**) The binding site of Glycerol 2 within SaSsbA was unveiled by the structure of glycerol-bound SaSsbA (PDB ID 8GW5). Similar to Glycerol 1, Glycerol 2 (green) interacted with both SaSsbA monomers A (light pink) and B (light blue). However, the binding poses and locations between Glycerol 1 and Glycerol 2 exhibited distinctions. (**D**) Depiction of the binding mode for Glycerol 2. Within a contact distance of <4 Å, M1 (Subunit A), L2 (Subunit A), N3 (Subunit A), R4 (Subunit A), T36 (Subunit B), F37 (Subunit B), R76 (Subunit A), and D98 (Subunit A) were engaged in binding Glycerol 2. Structural analysis via PLIP revealed hydrogen bonds formed between the main chains of M1, N3, and T36, as well as the side chain of R76, and Glycerol 2. (**E**) The binding site of Glycerol 3 within SaSsbA was revealed by the SaSsbA–5-FU complex structure (PDB ID 7YM1). Glycerol 3 (forest) was nestled between SaSsbA monomers A (hot pink) and B (deep blue). The binding mode for Glycerol 3 exhibited similarities to that of Glycerol 1. 5-FU (orange) was also showcased within this structure. (**F**) Depiction of the binding mode for Glycerol 3. Through interactions detected using PLIP, it was determined that hydrogen bonds were formed between the main chains of R80 and V92, along with the side chain of T93, and Glycerol 3.

2.6. Glycerol 2 Binding Mode of SaSsbA

Similar to Glycerol 1, which finds itself ensconced between SaSsbA monomers A and B, Glycerol 2 also forms interactions with both monomers (Figure 3C). Nonetheless, it is important to note that their binding poses and spatial arrangements between Glycerol 1 and 2 differed significantly. The interacting residues associated with Glycerol 1 and Glycerol 2 displayed complete variation (PDB ID 8GW5). Specifically, M1 (Subunit A), L2 (Subunit A), N3 (Subunit A), R4 (Subunit A), T36 (Subunit B), F37 (Subunit B), R76 (Subunit A), and D98 (Subunit A) were observed to participate within contact distance (<4 Å) in binding interactions with Glycerol 2 (Figure 3D). A notable aspect of Glycerol 2 binding was the inclusion of a water molecule in the interaction. This water molecule, in conjunction with M1 (Subunit A) and T36 (Subunit B), also contributed to interactions with 5-FU, facilitated through a hydrogen bonding network. The structural evaluation conducted via PLIP underscored that the main chains of M1, N3, and T36, as well as the side chain of R76, were pivotal components in the hydrogen bonding network, fostering the binding of Glycerol 2 (Figure 3D).

2.7. Glycerol 3 Binding Mode of SaSsbA

The binding configuration of Glycerol 3 (PDB ID 7YM1) exhibited similarities to that of Glycerol 1 (Figure 3E). In a manner akin to Glycerol 1, Glycerol 3 was positioned between SaSsbA monomers A and B. The interaction between Glycerol 3 and SaSsbA (<4 Å) involved F48 (Subunit B), N50 (Subunit B), S79 (Subunit A), R80 (Subunit A), F91 (Subunit A), V92 (Subunit A), and T93 (Subunit A). The structural evaluation conducted via PLIP revealed that the main chains of R80 and V92, along with the side chain of T93, partook in hydrogen bonding with Glycerol 3 (Figure 3F).

2.8. 5-FU Binding Mode of SaSsbA

SsbA, a crucial DNA replication protein, plays multifaceted roles in nucleic acid metabolism [45,54,61,62]. Prior to this study, it remained uncertain whether the FDA-approved clinical drug 5-FU [4], renowned as a prominent pyrimidine derivative in anticancer therapy, could indeed interact with SsbA. Consequently, the complexed crystal structure of SaSsbA with 5-FU was meticulously established to pinpoint the binding site and delve into the binding mechanism (Figure 4A). The electron density corresponding to 5-FU exhibited a well-defined clarity (Figure 4A). The arrangement of 5-FU was discernible, notably due to the positioning of its substituent (Figure 4A). A comprehensive scrutiny was carried out to decipher the interactions between 5-FU and SaSsbA (Figure 4B). Residues R18, P21, V52, F54, Q78, R80, E94, and V96, positioned within a contact distance of <4 Å, were instrumental in the binding of 5-FU. Through analysis conducted using PLIP [60], it was revealed that a water molecule also participated in the binding of 5-FU, facilitated by E94 in SaSsbA, which engaged in water-molecule-mediated hydrogen bonding (Figure 4B). Based on interactions discerned via PLIP, the four side chains of R18, Q78, R80, and E94 were observed to form hydrogen bonds with 5-FU (Figure 3B). The electrostatic potential surface of the SaSsbA complexed with 5-FU unveiled that 5-FU effectively occupied the groove within SaSsbA (Figure 4C), a site significant for single-stranded DNA binding (Figure 4D). The positive (blue) and negative (red) charge distributions underscored that several critical basic residues on the SaSsbA surface, which are exposed to the solvent and collectively form a binding pathway conducive for accommodating ssDNA binding ssDNA (gold). This complex structure insightfully revealed that the presence of 5-FU in the groove potentially influences the wrapping of ssDNA by SaSsbA.

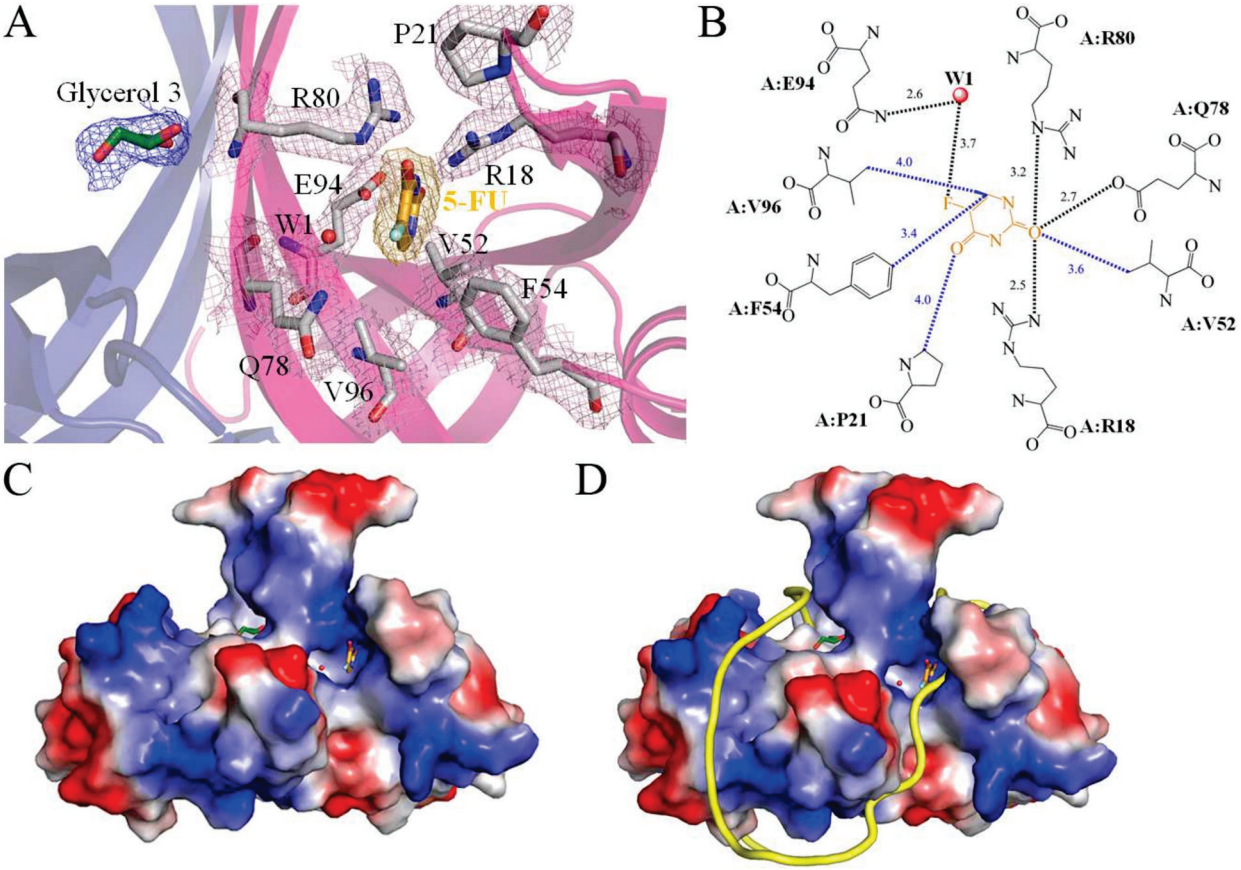


Figure 4. 5-FU interaction mode. (**A**) The binding site for 5-FU within SaSsbA was unveiled through the SaSsbA–5-FU complex structure (PDB ID 7YM1). This complexed structure of an SaSsbA dimer contained one glycerol molecule (Glycerol 3; forest) and one 5-FU molecule (orange). The residues engaged in interactions with 5-FU are depicted in gray. An orange mesh, contoured at 1 σ, illustrates the presence of 5-FU within the groove of SaSsbA monomer A (hot pink). The electron density for these interactive residues is also distinctly visible (light pink mesh, contoured at 1 σ). (**B**) Depiction of the binding mode for 5-FU. Residues R18, P21, V52, F54, Q78, R80, E94, and V96, situated within a contact distance of <4 Å, played pivotal roles in binding with 5-FU. The corresponding interactive distances are also indicated (Å). Based on the interactions identified via PLIP, the side chains of R18, Q78, R80, and E94 engaged in hydrogen bonding with 5-FU (highlighted in black). (**C**) The electrostatic potential surface portrayal of the SaSsbA complexed with 5-FU elucidates the distribution of positive (blue) and negative (red) charges. Notably, 5-FU (orange) occupies a groove within SaSsbA, potentially pertinent to ssDNA binding. (**D**) The superimposed structures of the SaSsbA–5-FU complex and the EcSSB–ssDNA complex (PDB ID 1EYG). The crystal structures of SaSsbA and EcSSB exhibit similarity. For clarity, the EcSSB structure is omitted. The distribution of positive (blue) and negative (red) charges showcases a collection of fundamental basic residues on the surface of SaSsbA, which are exposed to the solvent and collectively form a binding pathway conducive for accommodating ssDNA binding (gold). Our complex structure highlights that the 5-FU presence within the groove may potentially affect and regulate the ssDNA wrapping phenomenon by SaSsbA.

2.9. Comparative Analysis of 5-FU Binding Sites in Different Binding States of SaSsbA

In this study, we identified that residues R18, P21, V52, F54, Q78, R80, E94, and V96 within SaSsbA engage in interactions with 5-FU. Furthermore, upon comparing the complexed structures of SaSsbA monomers A and B (PDB ID 7YM1), distinct binding site configurations emerged between the 5-FU-bound state (monomer A) and the Glycerol 3-bound state (monomer B). This comparison revealed noteworthy conformational

changes (Figure 5A). When accommodating 5-FU (Figure 5B), the positions of R18 and P21 experienced shifts of approximately 5.2 and 7.3 Å, respectively. Moreover, V52, F54, Q78, and E94 exhibited angular shifts of 180, 22, 23, and 52°, respectively, due to 5-FU binding. In addition, the sizes of the binding groove differed between monomer A (the 5-FU-bound state) and monomer B (the Glycerol 3-bound state) (Figure 5C). While the OB folds exhibited a similar appearance, it was observed that their sizes were divergent. Structurally, the angles between strands β1′ and β4 in monomer A and B were measured at 41.2° and 65.9°, respectively (Figure 5C).

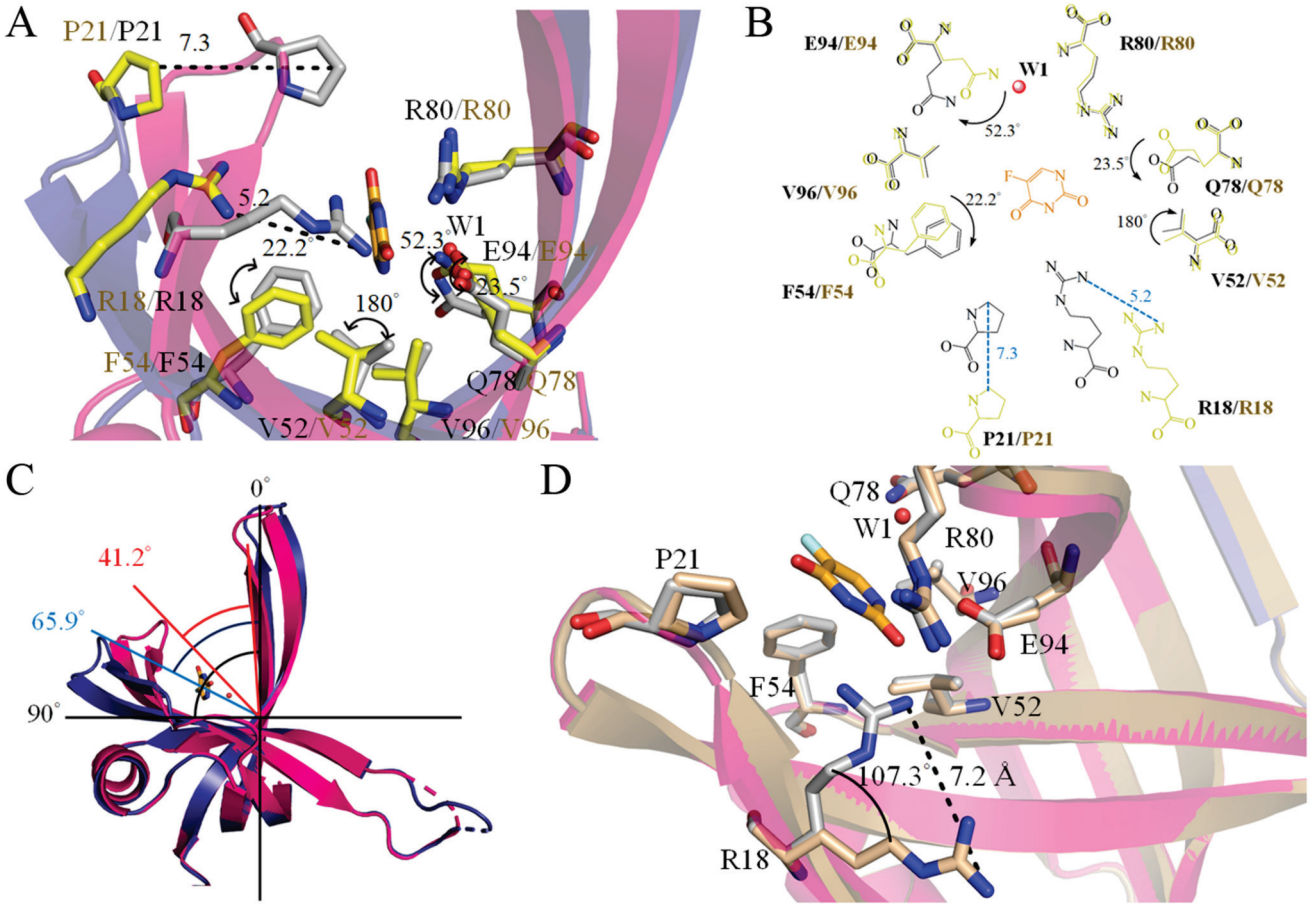


Figure 5. Comparative analysis of 5-FU binding sites in different states of SaSsbA. (**A**) Superposition of monomer A (hot pink) and monomer B (deep blue) within the 5-FU complexed SaSsbA structure (PDB ID 7YM1). Residues in monomer A and B are depicted in gray and yellow, respectively, with the presence of a 5-FU molecule shown in orange. (**B**) Comparison of the 5-FU binding sites between the 5-FU-bound and glycerol 3-bound states of SaSsbA. Binding of 5-FU led to spatial shifts of R18 and P21 by distances of 5.2 and 7.3 Å, respectively. Additionally, V52, F54, Q78, and E94 underwent angular shifts of 180, 22, 23, and 52°, respectively, upon 5-FU binding. (**C**) Evaluation of the sizes of the binding groove in monomer A and B. Despite the comparable appearance of their OB folds, variations in their sizes were noted. Notably, the structural angles between strands β1′ and β4 in monomer A and B were found to be 41.2 and 65.9°, respectively. (**D**) Superposition of the 5-FU-bound (hot pink; PDB ID 7YM1) and unbound (pale yellow; PDB ID 5XGT) states of SaSsbA. Residues within the 5-FU-bound and unbound states are presented in gray and wheat hues, respectively. A 5-FU molecule is visualized in orange. Remarkably, the side chain of R18 experienced a considerable shift, spanning a distance of 7.2 Å and an angular shift of 107.3° upon 5-FU binding.

2.10. Comparative Structural Analysis of 5-FU Binding Sites in the 5-FU-Bound and Unbound States of SaSsbA

Having previously determined the crystal structure of the apo-form of SaSsbA [45], we have a basis for comparing the structural aspects of the 5-FU binding sites between the 5-FU-bound state (PDB ID 7YM1) and the unbound state (PDB ID 5XGT) of SaSsbA. Through superimposition of these structures, R18 in the apo-form SaSsbA was significantly shifted by a distance of 7.2 Å and an angle of 107.3° upon binding of 5-FU (Figure 5D). Accordingly, the side chain of R18 is likely a pivotal element in facilitating 5-FU binding.

2.11. Structure-Based Mutational Analysis

Our complex structure of SaSsbA with 5-FU has elucidated the binding mode and identified the interactive residues. Notably, a substantial conformational change was observed in R18 upon 5-FU binding, resulting in a shift of its side chain position by 7.2 Å and an angular alteration of 77.4°. This suggests that R18 might play a role in initiating or mediating the 5-FU binding process. Since R18 is highly conserved among SSB homologs (Figure 1), we generated an alanine substitution mutant (Table 2), which was subsequently purified and analyzed to investigate its contribution to binding (Figure 6). The strength of interaction between the R18A mutant and 5-FU was assessed through fluorescence quenching and compared to that of the wild-type SaSsbA (WT). The quenching phenomenon involves the formation of a complex that diminishes the protein's fluorescence intensity. SaSsbA exhibited prominent intrinsic fluorescence, with a peak wavelength at 339 nm upon excitation at 279 nm (Figure 6A). As concentrations of 5-FU were incrementally added to the SaSsbA solution, the intrinsic fluorescence underwent gradual quenching (Figure 6A). Upon introducing 200 μM of 5-FU, the intrinsic fluorescence of SaSsbA was reduced by 83.8% (Table 3). The binding of 5-FU induced a red shift of the SaSsbA emission wavelength from 339 nm to 347 nm (about 8 nm), as indicated by the change in λ_{max} (Table 3). These observations collectively confirm the formation of a stable complex between SaSsbA and 5-FU. Comparable to SaSsbA, the R18A mutant also exhibited strong intrinsic fluorescence, with a peak wavelength at 339 nm upon excitation at 279 nm (Figure 6B). However, the addition of 200 μM 5-FU resulted in only a 34.9% reduction in the intrinsic fluorescence of R18A (Table 3). Furthermore, the λ_{max} of R18A shifted only minimally from 339 nm to 340 nm upon exposure to 200 μM 5-FU. Analysis of the titration curves (Figure 6C) facilitated the determination of K_d values of 497.6 ± 13.5 μM for R18A and 55.9 ± 0.7 μM for WT (Table 3). These experimental findings observed from structural and functional investigations collectively underscore the significance of R18 in SaSsbA as a crucial residue for 5-FU binding.

Table 2. Primers used for construction of the plasmid.

Oligonucleotide	Primer
SaSsbA-R18A-N	GAAAGATCCGGAATACGCAACCACTCCCTC
SaSsbA-R18A-C	ACACCTGAGGGAGTGGTTGCGTATTCCGGA

Underlined nucleotides indicate the designated site for mutation site.

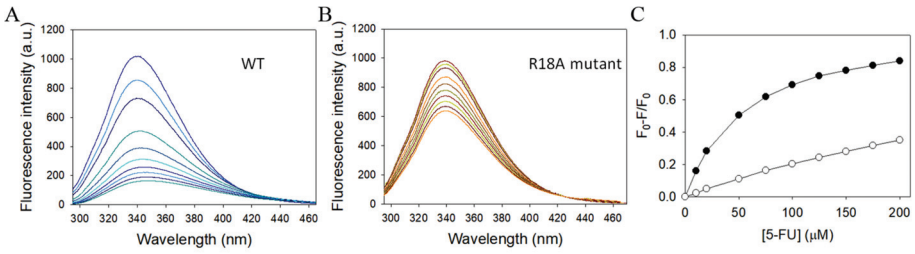


Figure 6. Fluorescence titration of SaSsbA with 5-FU. (**A**) The fluorescence emission spectra of WT with 5-FU of different concentrations (0–200 μM; 0, 10, 20, 50, 75, 100, 125, 150, 175, and 200 μM). The

decrease in intrinsic fluorescence of protein was measured at 339 nm upon excitation at 279 nm with a spectrofluorometer. The fluorescence intensity emission spectra of SaSsbA were significantly quenched by 5-FU. (**B**) The fluorescence emission spectra of the mutant R18A with 5-FU of different concentrations (0–200 μM). The strength of interaction between the R18A mutant and 5-FU was assessed and compared to that of WT. (**C**) The titration curves for determining the K_d values of 5-FU for WT (●) and R18A (○). The K_d was obtained by the equation: $\Delta F = \Delta F_{max} - K_d(\Delta F/[5\text{-FU}])$. Data points are an average of 2–3 determinations within a 10% error.

Table 3. Binding parameters of SaSsbA WT and the mutant R18A.

SaSsbA	λ_{max} (nm)	λ_{em} Shift (nm)	Quenching (%)	K_d Value (μM)
SaSsbA WT	339	347	83.81	55.9 ± 0.7
SaSsbA-R18A	339	340	34.99	497.6 ± 13.5

The decrease in the intrinsic fluorescence of SaSsbA was measured with a spectrofluorometer (Hitachi F-2700; Hitachi High-Technologies, Tokyo, Japan). The K_d was obtained using the following equation: $\Delta F = \Delta F_{max} - K_d(\Delta F/[5\text{-FU}])$.

2.12. Distinct 5-FU Binding Modes in SaSsbA and SaSsbB

Unlike the case of *E. coli*, which possesses a singular type of SSB (EcSSB), certain bacteria, particularly some Gram-positive bacteria [54], harbor two paralogous SSBs, namely SsbA and SsbB. Recently, we elucidated the crystal structure of SaSsbB [46], as well as its complex with 5-FU [49]. Consequently, the complex structure of SaSsbB is available for a comparative analysis of the 5-FU binding mode in relation to SaSsbA (PDB ID 7YM1) and SaSsbB (PDB ID 7D8J). Given the structural similarity between SaSsbA and SaSsbB, one might naturally assume a congruent 5-FU binding mode. However, intriguingly, our complexed structure revealed distinct 5-FU binding configurations for SaSsbA and SaSsbB (Figure 7). In the complexed structure, specific residues including T12, K13, T30, F48, and N50 of SaSsbB were identified to interact with 5-FU, forming an integral part of the binding site (Figure 7A) [49]. Particularly noteworthy is the essential stacking interaction between the pyrimidine ring of 5-FU and the aromatic ring of F48 in SaSsbB, which underpins the drug–protein interaction. While these residues are entirely conserved in SaSsbA, they did not engage in interactions with 5-FU (Figure 7A). Upon superposition analysis, a considerable distance of 18.8 Å was evident between these divergent 5-FU binding sites (Figure 7B). Furthermore, the residues in SaSsbA that interact with 5-FU are similarly conserved in SaSsbB; however, they do not collectively form a 5-FU binding site in SaSsbB (Figure 7C). Given the perfect conservation of these 5-FU binding residues in both SaSsbA and SaSsbB, the substantial dissimilarity observed warrants further investigation to ascertain whether other inherent species-specific differences contribute to this phenomenon. Additional biophysical studies are warranted to comprehensively explore these disparities.

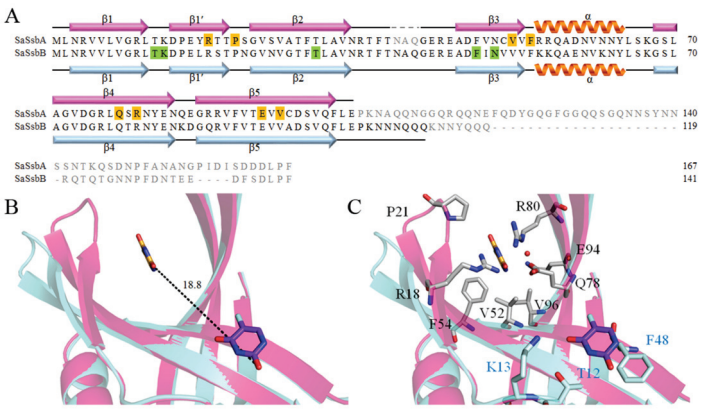


Figure 7. Distinct 5-FU binding modes in SaSsbA and SaSsbB. (**A**) Alignment of the sequences of SaSsbA and SaSsbB, with secondary structural elements indicated. Unobserved amino acids in these

crystal structures are colored in gray. The residues responsible for 5-FU binding in SaSsbA and SaSsbB are shaded in orange and green, respectively. (**B**) Superposition of the 5-FU-bound structures of SaSsbA (hot pink) and SaSsbB (aquamarine). The 5-FU molecules in SaSsbA (PDB ID 7YM1) and SaSsbB (PDB ID 7D8J) are represented in orange and purple-blue, respectively. The distance between these distinct 5-FU binding sites measures 18.8 Å. (**C**) Disparate 5-FU binding sites. Despite the perfect conservation of these 5-FU binding residues in SaSsbA (gray) and SaSsbB (aquamarine), each protein exhibits a preferred binding site for 5-FU.

3. Discussion

Metabolic reprogramming is the strategy adopted by cancer cells to expedite their proliferation, resist the effects of chemotherapy, invade tissues, metastasize, and endure within nutrient-scarce microenvironments [2]. Various uracil derivatives have long been harnessed as pyrimidine-based antimetabolites in the battle against cancer [63]. Chief among these agents is 5-FU [4], a prominent fluoropyrimidine drug esteemed for its role in targeting TSase during anticancer chemotherapy [9]. Over the past 60 years, chemotherapeutic agents designed to thwart thymidylate biosynthesis have emerged as stalwarts in cancer treatment. In addition, the synergistic administration of 5-FU alongside other chemotherapeutic agents amplifies treatment efficacy and overall survival rates, particularly in cancers involving the head, breast, and neck [64]. However, some mechanistic details, including signaling pathways, remain unexplained [65–67]. It is important to recognize that the purview of 5-FU's interactions goes beyond merely engaging human TSase. For example, other human proteins, including dihydroorotase, PARP (procyclic acidic repetitive protein), VEGFR1 (vascular endothelial growth factor receptor 1), and CASP-3 (caspase-3 protein), are also known to interact with 5-FU [14,68]. Moreover, the intricate interplay between microbiota and chemotherapeutic drugs, such as 5-FU, holds the potential to influence host responses, further adding to the complexity of the landscape [17]. Consequently, a comprehensive elucidation of the complete 5-FU interactome is imperative, serving as the foundation for exhaustive clinical pharmacokinetic assessments and toxicity analyses [21,22]. The holistic elucidation of the multifaceted relationships woven by 5-FU holds the promise of enhancing our grasp of its intricate dynamics, paving the way for refined therapeutic strategies and personalized medicine.

In this study, our findings have unveiled SaSsbA's capacity to engage in interaction with the anticancer drug 5-FU (Figure 6). In comparison with SaSsbB, a paralogous protein of SaSsbA in *S. aureus*, the K_d values for 5-FU binding to SaSsbA and SaSsbB are 55.9 ± 0.7 μM (Table 3) and 152.8 ± 2.5 μM [49], respectively. Based on the fact that the K_d value of human dihydroorotase bound to 5-FU is 91.2 ± 1.7 μM [14], the hierarchy of binding affinities for 5-FU can be delineated as follows: SaSsbA > human dihydroorotase > SaSsbB. This outcome may imply that, in scenarios where 5-FU enters the human system, it exhibits a preference for binding to the bacterial DNA replication protein SaSsbA within bacterial cells, as opposed to its interaction with the human enzyme dihydroorotase. However, it is important to underscore that this supposition necessitates comprehensive validation through a thorough investigation spanning biochemical and cellular dimensions. Considering the potential diversity of the gut microbiome across individuals, it remains imperative to ascertain the binding affinities of 5-FU to any feasible proteins within the human body, encompassing locales like the gastrointestinal tract and bloodstream. Such investigations are crucial to facilitate comprehensive comparisons and subsequent clinical analyses.

For the investigation of the binding mode, we solved the complexed crystal structure of SaSsbA with 5-FU at a resolution of 2.3 Å (Table 1). The interaction between 5-FU and SaSsbA was found to involve R18, P21, V52, F54, Q78, R80, E94, and V96 (Figure 4). Unexpectedly, this pattern of interactive residues deviated entirely from those identified in the 5-FU binding sites of SaSsbB [49]. In contrast to SaSsbA, where the interaction with 5-FU relies on a distinct set of residues, SaSsbB's 5-FU interaction hinges on T12, K13, T30,

F48, and N50 (Figure 7). Notably, the stacking interaction between the aromatic ring of F48 and the pyrimidine ring of 5-FU assumes a pivotal role in the drug–protein interaction within SaSsbB [49]. Intriguingly, despite the presence of these identical residues within both SaSsbA and SaSsbB, none of them (T12, K13, T30, F48, and N50) participate in 5-FU binding within SaSsbA. Biochemically, reconciling the divergence in 5-FU binding sites, despite the residue conservation between SaSsbA and SaSsbB, presents a challenge. Although the OB folds share a striking visual resemblance, we noted a subtle difference in the size of the binding groove between SaSsbA and SaSsbB. This structural divergence is underscored by the angles between strands β1′ and β4 in their monomers A, which measure 41.2° and 65.9°, respectively (Figure 5). This variance in binding groove width may potentially influence the mechanisms governing 5-FU binding. A noteworthy observation pertains to the substantial conformational alteration that accompanies 5-FU binding (Figure 5). This observation led us to propose a hypothesis wherein these seemingly identical residues result in divergent 5-FU binding modes. Within various contexts, OB folds can exhibit broad ligand-binding capabilities, targeting both single-stranded DNA and proteins [35]. This is evident in cases such as the tumor suppressor BRCA2, where two OB folds bind to ssDNA while a third participates in protein–protein interactions rather than ssDNA binding [69]. ssDNA bound by *Pseudomonas aeruginosa* SSB (PaSSB) only occupies half of the binding sites of two OB folds rather than four OB folds through the ssDNA-binding mode $(SSB)_{3:1}$ [57,58]. Similarly, in RPA, two different binding modes involve two and four OB folds, respectively [70]. Insights gleaned from the SaSsbA-glycerol complex structure (Figure 3) indicate that Glycerol 1 and 2 do not necessarily need to occupy corresponding sites within monomers A and B. Previous experimental observations through single molecule experiments [71,72] also suggest that the unoccupied OB fold within SSB could adopt an open conformation to facilitate sliding, and, therefore, the ligand-binding groove within its OB-fold structure could be regulated to accommodate the requirements of dynamic binding processes (Figure 7). Consequently, it becomes plausible to consider that even when similar sites exist, 5-FU may bind to different locations due to the influence of these adaptable binding grooves.

Differing from some enzymes characterized by a single active site, the binding behavior of the DNA replication protein SaSsbA, which engages with diverse ssDNA and proteins, can introduce unpredictability into its ligand binding site(s). A multitude of solvent-exposed surfaces on SaSsbA functions as binding sites for both ssDNA and partner proteins, further complicating the task of foreseeing which specific pocket serves as the binding site. The prospect of utilizing docking tools, like MOE Dock [73], to anticipate a protein's ligand binding site presents itself as a viable avenue to explore. The analysis conducted through MOE Dock highlighted five preferred binding modes (Figure 8). However, none of these predicted sites (Table 4) aligned precisely with the binding site evident in the complexed crystal structure of SaSsbA with 5-FU (Figure 2). Therefore, it becomes evident that the generation of additional complexed crystal structures remains an imperative in facilitating more robust binding analyses and supporting the development of structure-based approaches to drug design.

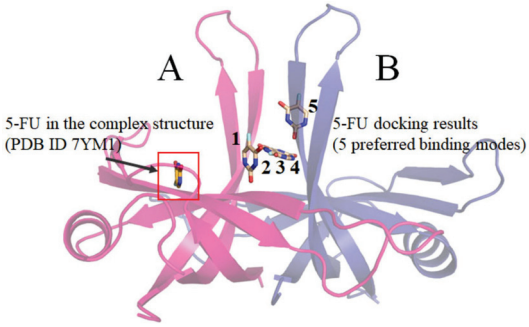


Figure 8. Molecular docking of 5-FU with SaSsbA. Comparison between 5-FU molecules extracted from the complexed SaSsbA structure and the outcomes of docking simulations. SaSsbA monomers

A and B are colored in light pink and light blue. Molecular docking of 5-FU was performed using the crystal structure of the apo-form SaSsbA (PDB ID 5XGT). The binding affinities of SaSsbA with 5-FU, considering all feasible binding orientations, were evaluated using the S score. Utilizing MOE-Dock (version 2019.0102) software, we identified and numbered the five most favorable binding modes for 5-FU. The five preferred binding modes of 5-FU are indicated as 1–5. Additionally, the position of 5-FU in the complexed crystal structure of SaSsbA (PDB ID 7YM1) was included for reference.

Table 4. Molecular docking analysis.

Mode	S Score	Receptor Residue	Interaction	Distance (Å)	E (kcal/mol)
1	−4.2104	Phe 48 (A)	H-donor	2.96	−5.6
		Thr 93 (A)	H-donor	2.98	−1.9
		Asn 50 (A)	H-acceptor	3.08	−1.1
2	−4.1638	Arg 80 (A)	H-donor	3.42	−0.9
		Val 92 (A)	H-donor	3.03	−1.2
		Asn 81 (A)	H-acceptor	3.14	−2.2
3	−3.9845	Arg 80 (A)	H-donor	3.27	−1.0
		Val 92 (A)	H-donor	3.05	−1.2
		Phe 91 (A)	Pi-H	4.03	−0.6
4	−3.8762	Asn 50 (B)	H-donor	3.37	−1.8
		Asn 81 (A)	H-acceptor	3.52	−0.8
5	−3.8744	Phe 91 (A)	H-donor	3.07	−1.6
		Phe 91 (A)	H-acceptor	3.15	−1.2

It is believed that all current cells trace their lineage back to a shared ancestor, suggesting that fundamental principles learned from experiments conducted with one cell type possess broad applicability across diverse cells. This perspective implies that the mechanisms governing essential cellular activities, such as DNA replication, transcription, and translation, should exhibit similarities across various cell types. Nonetheless, in response to challenging environmental circumstances, organisms tend to evolve new enzymes or auxiliary components to enhance their survival prospects and adaptive capabilities throughout evolutionary processes. In contrast to the situation in *E. coli* and many other bacteria, which feature a single SSB, certain microorganisms like *S. aureus* and other Gram-positive bacteria manifest multiple paralogous SSBs, including SsbA [74], SsbB [59], and SsbC [47]. Intriguingly, the positioning of the *ssbA* gene in the *S. aureus* genetic map does not align with the location of the *ssb* gene in *E. coli* and other Gram-negative bacteria [53]. Notably, this corresponding position in *E. coli* is occupied by *priB* [53], another variant of SSB [55,75]. Given the diversity of primosomal proteins with which these SSBs interact [76,77], SaSsbA and EcSSB confront an array of binding partners within their respective cellular contexts. This intricacy suggests that the presence of these distinct SSBs might necessitate their co-evolution with partner proteins, enabling the development of species-specific functions to address survival demands and secure a competitive edge. This co-evolutionary dynamic might elucidate the lack of conservation in PXXP motifs and amino acid residues within the IDL among different SSBs, including SaSsbA (Figure 1) [39]. Furthermore, intriguing disparities come to light, such as myricetin's inhibitory effect on PaSSB but not on SaSsbA [44] or *Klebsiella pneumoniae* SSB [40]. Even within proteins that share structural similarities, as demonstrated by SaSsbA and SaSsbB, their 5-FU binding modes exhibit complete divergence (Figure 7). Consequently, it is conceivable that 5-FU could bind to these distinct SSBs present in both human and microorganisms, subsequently

influencing various cellular signaling pathways. Despite these observations, additional research is indispensable to elucidate the precise mechanisms underpinning the recognition of 5-FU binding sites and the rationale behind the evolution of these diverse SSBs within specific species.

4. Materials and Methods

4.1. Protein Expression and Purification

The expression vector pET21b-SaSsbA [45] was transformed into *E. coli* BL21 (DE3) cells and grown in LB medium at 37 °C. The overexpression was induced by incubating with 1 mM isopropyl thiogalactopyranoside for 9 h. Recombinant SaSsbA was purified from the soluble supernatant by using Ni^{2+}-affinity chromatography. The recombinant protein was eluted with a linear imidazole gradient and dialyzed against a dialysis buffer (20 mM Tris-HCl and 0.1 M NaCl, pH 7.9; Buffer A). Protein concentration was measured using Biorad protein (Bradford) assay. The protein purity remained at >97% as determined using SDS–PAGE.

4.2. Site-Directed Mutagenesis

The SaSsbA mutant was generated according to the QuikChange site-directed mutagenesis kit protocol (Stratagene; LaJolla, CA, USA), by using the wild-type plasmid pET21b–SaSsbA as a template. The presence of the mutation was verified by DNA sequencing. The recombinant mutant proteins were purified using the protocol for the wild-type SaSsbA by Ni^{2+}-affinity chromatography.

4.3. Crystallization Experiments

Purified SaSsbA was concentrated to 20 mg/mL with addition of the cryoprotectant glycerol to a final concentration of 25% for storage at −20 °C. The crystals of the SaSsbA-5-FU complex appeared at room temperature through hanging drop vapor diffusion in 16% PEG 4000, 100 mM Tris-HCl, 200 μM 5-FU, and 200 mM $MgCl_2$ at pH 8.5. For the SaSsbA–glycerol complex, the crystals were grown in 30% PEG 4000, 100 mM HEPES, 200 mM $CaCl_2$ at pH 7.5. These crystals reached full size in 7–13 days and validated in the beamline 15A of the National Synchrotron Radiation Research Center (NSRRC; Hinchu, Taiwan).

4.4. X-ray Diffraction Data and Structure Determination

Data were collected in the beamline 15A using an Rayonix MX300HE CCD Area Detector at NSRRC. Data sets were indexed, integrated, and scaled using HKL-2000 [78] and XDS [79]. The initial phase, density modification, and model building were performed using the AutoSol program in the PHENIX [80]. The iterative model building and structure refinement were performed using Refmac in the CCP4 software suite (version v7.1.008) [81] and phenix.refine in the PHENIX software suite (Phenix1.19.1-4122) [82]. The initial phase of SaSsbA complexed with 5-FU was determined through the molecular replacement software Phaser MR (Phenix1.19.1-4122) [83] by using SaSsbA (PDB ID 5XGT) as a search model. The correctness of the stereochemistry of the models was verified using MolProbity [84].

4.5. Fluorescence Quenching

The K_d value of purified SaSsbA was determined using the fluorescence quenching method previously described for the DHOase [85] and DHPase [86,87]. Briefly, an aliquot of the compound was added into the solution containing SaSsbA (1 μM) and 50 mM HEPES at pH 7.0. SaSsbA displayed strong intrinsic fluorescence with a peak wavelength of 339 nm when excited at 279 nm at 25 °C. The decrease in the intrinsic fluorescence of SaSsbA was measured at 339 nm with a spectrofluorometer (Hitachi F-2700; Hitachi High-Technologies, Japan). The K_d was obtained using the following equation: $\Delta F = \Delta F_{max} - K_d(\Delta F/[5\text{-FU}])$. Data points are an average of 2–3 determinations within a 10% error.

4.6. MOE-Dock Analysis

The binding analysis of 5-FU to SaSsbA was carried out using MOE-Dock (version 2019.0102) [73]. The binding capacity was also calculated using MOE-Dock. The crystal structure of SaSsbA (PDB ID 5XGT) was used [45]. Before docking, any water molecules present in the crystal structure were removed using MOE. To ensure accuracy, a 3D protonation step followed by energy minimization was applied to add hydrogen atoms to the protein structure. The binding modes were generated and predicted through the MOE-Dock tool and visualized by PyMOL.

5. Conclusions

SsbA represents a captivating molecular apparatus that orchestrates a multitude of indispensable processes vital for maintaining DNA integrity [88]. This study has revealed SaSsbA's hitherto unknown capability of binding to the anticancer drug 5-FU, thereby expanding the roster of proteins within the 5-FU interactome to encompass this pivotal DNA replication protein (Figure 6). In light of the results derived from mutational and structural analyses, it became evident that SaSsbA's mode of binding with 5-FU diverges from that of SaSsbB (Figure 7). Given the insights offered by the glycerol and 5-FU binding sites (Figures 3 and 4), our complexed SaSsbA structures underscore the likelihood that several of these interactive residues could be suitable targets for drug interventions aimed at inhibiting SaSsbA activity. This complexed structure also holds the potential to furnish valuable comprehension regarding how 5-FU and its pyrimidine derivatives might bind to and impede analogous OB-fold proteins in humans, particularly within cancer-related signaling pathways [89]. Acknowledging the capacity of microbiota to influence the host's response to 5-FU, there emerges a pressing need for further research to revisit the roles that bacterial and human SSBs play in the realm of anticancer therapy.

Author Contributions: H.-H.S., Y.-H.H., Y.L. and P.-C.Y. performed the experiments; H.-H.S. and Y.-H.H. analyzed the data; H.-H.S. and C.-Y.H. contributed to the study design and manuscript writing. All authors reviewed the results and contributed to the data interpretation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Atomic coordinates and related structure factors were deposited in the PDB with accession codes 7YM1 and 8GW5.

Acknowledgments: We thank the experimental facility and the technical services provided by the Synchrotron Radiation Protein Crystallography Facility of the National Core Facility Program for Biotechnology, National Science and Technology Council, Taiwan.

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

References

1. Sessler, J.L.; Lawrence, C.M.; Jayawickramarajah, J. Molecular recognition via base-pairing. *Chem. Soc. Rev.* **2007**, *36*, 314–325. [CrossRef] [PubMed]
2. Zhu, J.; Thompson, C.B. Metabolic regulation of cell growth and proliferation. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 436–450. [CrossRef] [PubMed]
3. Fung, S.K.; Lok, A.S. Drug insight: Nucleoside and nucleotide analog inhibitors for hepatitis B. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **2004**, *1*, 90–97. [CrossRef] [PubMed]
4. Longley, D.B.; Harkin, D.P.; Johnston, P.G. 5-fluorouracil: Mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **2003**, *3*, 330–338. [CrossRef]
5. Peters, H.L.; Ku, T.C.; Seley-Radtke, K.L. Flexibility as a Strategy in Nucleoside Antiviral Drug Design. *Curr. Med. Chem.* **2015**, *22*, 3910–3921. [CrossRef]
6. Yssel, A.E.J.; Vanderleyden, J.; Steenackers, H.P. Repurposing of nucleoside- and nucleobase-derivative drugs as antibiotics and biofilm inhibitors. *J. Antimicrob Chemother.* **2017**, *72*, 2156–2170. [CrossRef]

7. Ku, S.K.; Baek, M.C.; Bae, J.S. Anti-inflammatory effects of methylthiouracil in vitro and in vivo. *Toxicol. Appl. Pharmacol.* **2015**, *288*, 374–386. [CrossRef]
8. Alvarez, P.; Marchal, J.A.; Boulaiz, H.; Carrillo, E.; Velez, C.; Rodriguez-Serrano, F.; Melguizo, C.; Prados, J.; Madeddu, R.; Aranega, A. 5-Fluorouracil derivatives: A patent review. *Expert Opin. Ther. Pat.* **2012**, *22*, 107–123. [CrossRef]
9. Wilson, P.M.; Danenberg, P.V.; Johnston, P.G.; Lenz, H.J.; Ladner, R.D. Standing the test of time: Targeting thymidylate biosynthesis in cancer therapy. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 282–298. [CrossRef]
10. Huang, Y.H.; Ning, Z.J.; Huang, C.Y. Crystal structure of dihydropyrimidinase in complex with anticancer drug 5-fluorouracil. *Biochem. Biophys. Res. Commun.* **2019**, *519*, 160–165. [CrossRef]
11. Huang, C.Y. Structure, catalytic mechanism, posttranslational lysine carbamylation, and inhibition of dihydropyrimidinases. *Adv. Protein Chem. Struct. Biol.* **2020**, *122*, 63–96. [PubMed]
12. Van Kuilenburg, A.B.; Meinsma, R.; Zonnenberg, B.A.; Zoetekouw, L.; Baas, F.; Matsuda, K.; Tamaki, N.; van Gennip, A.H. Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity. *Clin. Cancer Res.* **2003**, *9*, 4363–4367.
13. Van Kuilenburg, A.B.; Dobritzsch, D.; Meijer, J.; Meinsma, R.; Benoist, J.F.; Assmann, B.; Schubert, S.; Hoffmann, G.F.; Duran, M.; de Vries, M.C.; et al. Dihydropyrimidinase deficiency: Phenotype, genotype and structural consequences in 17 patients. *Biochim. Biophys. Acta* **2010**, *1802*, 639–648. [CrossRef] [PubMed]
14. Lin, E.S.; Huang, Y.H.; Yang, P.C.; Peng, W.F.; Huang, C.Y. Complexed Crystal Structure of the Dihydroorotase Domain of Human CAD Protein with the Anticancer Drug 5-Fluorouracil. *Biomolecules* **2023**, *13*, 149. [CrossRef] [PubMed]
15. Sillo, T.O.; Beggs, A.D.; Middleton, G.; Akingboye, A. The Gut Microbiome, Microsatellite Status and the Response to Immunotherapy in Colorectal Cancer. *Int. J. Mol. Sci.* **2023**, *24*, 5767. [CrossRef]
16. Jaye, K.; Chang, D.; Li, C.G.; Bhuyan, D.J. Gut Metabolites and Breast Cancer: The Continuum of Dysbiosis, Breast Cancer Risk, and Potential Breast Cancer Therapy. *Int. J. Mol. Sci.* **2022**, *23*, 9490. [CrossRef]
17. Alexander, J.L.; Wilson, I.D.; Teare, J.; Marchesi, J.R.; Nicholson, J.K.; Kinross, J.M. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **2017**, *14*, 356–365. [CrossRef]
18. Nakayama, H.; Kinouchi, T.; Kataoka, K.; Akimoto, S.; Matsuda, Y.; Ohnishi, Y. Intestinal anaerobic bacteria hydrolyse sorivudine, producing the high blood concentration of 5-(E)-(2-bromovinyl)uracil that increases the level and toxicity of 5-fluorouracil. *Pharmacogenetics* **1997**, *7*, 35–43. [CrossRef]
19. Narendra, G.; Choudhary, S.; Raju, B.; Verma, H.; Silakari, O. Role of Genetic Polymorphisms in Drug-Metabolizing Enzyme-Mediated Toxicity and Pharmacokinetic Resistance to Anti-Cancer Agents: A Review on the Pharmacogenomics Aspect. *Clin. Pharmacokinet.* **2022**, *61*, 1495–1517. [CrossRef]
20. Schneider, J.J.; Galettis, P.; Martin, J.H. Overcoming barriers to implementing precision dosing with 5-fluorouracil and capecitabine. *Br. J. Clin. Pharmacol.* **2021**, *87*, 317–325. [CrossRef]
21. Gmeiner, W.H. A narrative review of genetic factors affecting fluoropyrimidine toxicity. *Precis. Cancer Med.* **2021**, *4*, 38. [CrossRef] [PubMed]
22. Kobuchi, S.; Ito, Y. Application of Pharmacometrics of 5-Fluorouracil to Personalized Medicine: A Tool for Predicting Pharmacokinetic-Pharmacodynamic/Toxicodynamic Responses. *Anticancer Res.* **2020**, *40*, 6585–6597. [CrossRef] [PubMed]
23. Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 335–346. [CrossRef]
24. Reyes-Lamothe, R.; Sherratt, D.J. The bacterial cell cycle, chromosome inheritance and cell growth. *Nat. Rev. Microbiol.* **2019**, *17*, 467–478. [CrossRef]
25. Papamichos-Chronakis, M.; Peterson, C.L. Chromatin and the genome integrity network. *Nat. Rev. Genet.* **2013**, *14*, 62–75. [CrossRef] [PubMed]
26. Cox, M.M.; Goodman, M.F.; Kreuzer, K.N.; Sherratt, D.J.; Sandler, S.J.; Marians, K.J. The importance of repairing stalled replication forks. *Nature* **2000**, *404*, 37–41. [CrossRef]
27. Meyer, R.R.; Laine, P.S. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev.* **1990**, *54*, 342–380. [CrossRef]
28. Croft, L.V.; Bolderson, E.; Adams, M.N.; El-Kamand, S.; Kariawasam, R.; Cubeddu, L.; Gamsjaeger, R.; Richard, D.J. Human single-stranded DNA binding protein 1 (hSSB1, OBFC2B), a critical component of the DNA damage response. *Semin. Cell Dev. Biol.* **2019**, *86*, 121–128. [CrossRef] [PubMed]
29. Shereda, R.D.; Kozlov, A.G.; Lohman, T.M.; Cox, M.M.; Keck, J.L. SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 289–318. [CrossRef]
30. Richard, D.J.; Bolderson, E.; Cubeddu, L.; Wadsworth, R.I.; Savage, K.; Sharma, G.G.; Nicolette, M.L.; Tsvetanov, S.; McIlwraith, M.J.; Pandita, R.K.; et al. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* **2008**, *453*, 677–681. [CrossRef]
31. Lohman, T.M.; Ferrari, M.E. Escherichia coli single-stranded DNA-binding protein: Multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* **1994**, *63*, 527–570. [CrossRef] [PubMed]
32. Antony, E.; Lohman, T.M. Dynamics of E. coli single stranded DNA binding (SSB) protein-DNA complexes. *Semin. Cell Dev. Biol.* **2019**, *86*, 102–111. [CrossRef] [PubMed]
33. Byrne, B.M.; Oakley, G.G. Replication protein A, the laxative that keeps DNA regular: The importance of RPA phosphorylation in maintaining genome stability. *Semin. Cell Dev. Biol.* **2019**, *86*, 112–120. [CrossRef]
34. Iftode, C.; Daniely, Y.; Borowiec, J.A. Replication protein A (RPA): The eukaryotic SSB. *Crit. Rev. Biochem. Mol. Biol.* **1999**, *34*, 141–180. [CrossRef] [PubMed]

35. Dickey, T.H.; Altschuler, S.E.; Wuttke, D.S. Single-stranded DNA-binding proteins: Multiple domains for multiple functions. *Structure* **2013**, *21*, 1074–1084. [CrossRef] [PubMed]
36. Yang, C.; Curth, U.; Urbanke, C.; Kang, C. Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 A resolution. *Nat. Struct. Biol.* **1997**, *4*, 153–157. [CrossRef]
37. Murzin, A.G. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J.* **1993**, *12*, 861–867. [CrossRef] [PubMed]
38. Bianco, P.R. The mechanism of action of the SSB interactome reveals it is the first OB-fold family of genome guardians in prokaryotes. *Protein Sci.* **2021**, *30*, 1757–1775. [CrossRef]
39. Bianco, P.R. The tale of SSB. *Prog. Biophys. Mol. Biol.* **2017**, *127*, 111–118. [CrossRef]
40. Lin, E.S.; Luo, R.H.; Huang, C.Y. A Complexed Crystal Structure of a Single-Stranded DNA-Binding Protein with Quercetin and the Structural Basis of Flavonol Inhibition Specificity. *Int. J. Mol. Sci.* **2022**, *23*, 588. [CrossRef]
41. Lin, E.S.; Huang, Y.H.; Luo, R.H.; Basharat, Z.; Huang, C.Y. Crystal Structure of an SSB Protein from *Salmonella enterica* and Its Inhibition by Flavanonol Taxifolin. *Int. J. Mol. Sci.* **2022**, *23*, 4399. [CrossRef] [PubMed]
42. Huang, C.Y. Crystal structure of SSB complexed with inhibitor myricetin. *Biochem. Biophys. Res. Commun.* **2018**, *504*, 704–708. [CrossRef]
43. Glanzer, J.G.; Endres, J.L.; Byrne, B.M.; Liu, S.; Bayles, K.W.; Oakley, G.G. Identification of inhibitors for single-stranded DNA-binding proteins in eubacteria. *J. Antimicrob. Chemother.* **2016**, *71*, 3432–3440. [CrossRef]
44. Liu, H.W.; Chiang, W.Y.; Huang, Y.H.; Huang, C.Y. The Inhibitory Effects and Cytotoxic Activities of the Stem Extract of *Sarracenia purpurea* against Melanoma Cells and the SsbA Protein. *Plants* **2022**, *11*, 3164. [CrossRef]
45. Huang, Y.H.; Guan, H.H.; Chen, C.J.; Huang, C.Y. Staphylococcus aureus single-stranded DNA-binding protein SsbA can bind but cannot stimulate PriA helicase. *PLoS ONE* **2017**, *12*, e0182060. [CrossRef]
46. Chen, K.L.; Cheng, J.H.; Lin, C.Y.; Huang, Y.H.; Huang, C.Y. Characterization of single-stranded DNA-binding protein SsbB from *Staphylococcus aureus*: SsbB cannot stimulate PriA helicase. *RSC Adv.* **2018**, *8*, 28367–28375. [CrossRef]
47. Huang, Y.H.; Huang, C.Y. SAAV2152 is a single-stranded DNA binding protein: The third SSB in *Staphylococcus aureus*. *Oncotarget* **2018**, *9*, 20239–20254. [CrossRef] [PubMed]
48. Lin, E.S.; Huang, Y.H.; Chung, J.C.; Su, H.H.; Huang, C.Y. The Inhibitory Effects and Cytotoxic Activities of the Stem Extract of Nepenthes miranda against Single-Stranded DNA-Binding Protein and Oral Carcinoma Cells. *Plants* **2023**, *12*, 2188. [CrossRef] [PubMed]
49. Lin, E.S.; Huang, C.Y. Crystal structure of the single-stranded DNA-binding protein SsbB in complex with the anticancer drug 5-fluorouracil: Extension of the 5-fluorouracil interactome to include the oligonucleotide/oligosaccharide-binding fold protein. *Biochem. Biophys. Res. Commun.* **2021**, *534*, 41–46. [CrossRef]
50. Curreri, A.R.; Ansfield, F.J.; Mc, I.F.; Waisman, H.A.; Heidelberger, C. Clinical studies with 5-fluorouracil. *Cancer Res.* **1958**, *18*, 478–484.
51. Eidinoff, M.L.; Knoll, J.E.; Klein, D. Effect of 5-fluorouracil on the incorporation of precursors into nucleic acid pyrimidines. *Arch. Biochem. Biophys.* **1957**, *71*, 274–275. [CrossRef]
52. Raghunathan, S.; Kozlov, A.G.; Lohman, T.M.; Waksman, G. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat. Struct. Biol.* **2000**, *7*, 648–652. [CrossRef] [PubMed]
53. Lin, E.S.; Huang, Y.H.; Huang, C.Y. Characterization of the Chimeric PriB-SSBc Protein. *Int. J. Mol. Sci.* **2021**, *22*, 10854. [CrossRef] [PubMed]
54. Lindner, C.; Nijland, R.; van Hartskamp, M.; Bron, S.; Hamoen, L.W.; Kuipers, O.P. Differential expression of two paralogous genes of *Bacillus subtilis* encoding single-stranded DNA binding protein. *J. Bacteriol.* **2004**, *186*, 1097–1105. [CrossRef]
55. Huang, C.Y.; Hsu, C.H.; Sun, Y.J.; Wu, H.N.; Hsiao, C.D. Complexed crystal structure of replication restart primosome protein PriB reveals a novel single-stranded DNA-binding mode. *Nucleic Acids Res.* **2006**, *34*, 3878–3886. [CrossRef] [PubMed]
56. Savvides, S.N.; Raghunathan, S.; Futterer, K.; Kozlov, A.G.; Lohman, T.M.; Waksman, G. The C-terminal domain of full-length *E. coli* SSB is disordered even when bound to DNA. *Protein Sci.* **2004**, *13*, 1942–1947. [CrossRef]
57. Huang, Y.H.; Lin, E.S.; Huang, C.Y. Complexed crystal structure of SSB reveals a novel single-stranded DNA binding mode (SSB)3:1: Phe60 is not crucial for defining binding paths. *Biochem. Biophys. Res. Commun.* **2019**, *520*, 353–358. [CrossRef] [PubMed]
58. Huang, Y.H.; Chen, I.C.; Huang, C.Y. Characterization of an SSB-dT25 complex: Structural insights into the S-shaped ssDNA binding conformation. *RSC Adv.* **2019**, *9*, 40388–40396. [CrossRef]
59. Paradzik, T.; Ivic, N.; Filic, Z.; Manjasetty, B.A.; Herron, P.; Luic, M.; Vujaklija, D. Structure-function relationships of two paralogous single-stranded DNA-binding proteins from *Streptomyces coelicolor*: Implication of SsbB in chromosome segregation during sporulation. *Nucleic Acids Res.* **2013**, *41*, 3659–3672. [CrossRef]
60. Adasme, M.F.; Linnemann, K.L.; Bolz, S.N.; Kaiser, F.; Salentin, S.; Haupt, V.J.; Schroeder, M. PLIP 2021: Expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.* **2021**, *49*, W530–W534. [CrossRef]
61. Torres, R.; Carrasco, B.; Alonso, J.C. *Bacillus subtilis* RadA/Sms-Mediated Nascent Lagging-Strand Unwinding at Stalled or Reversed Forks Is a Two-Step Process: RadA/Sms Assists RecA Nucleation, and RecA Loads RadA/Sms. *Int. J. Mol. Sci.* **2023**, *24*, 4536. [CrossRef] [PubMed]
62. Ramos, C.; Hernández-Tamayo, R.; López-Sanz, M.; Carrasco, B.; Serrano, E.; Alonso, J.C.; Graumann, P.L.; Ayora, S. The RecD2 helicase balances RecA activities. *Nucleic Acids Res.* **2022**, *50*, 3432–3444. [CrossRef]

63. Natarajan, R.; Anthoni Samy, H.N.; Sivaperuman, A.; Subramani, A. Structure-Activity Relationships of Pyrimidine Derivatives and their Biological Activity—A Review. *Med. Chem.* **2022**, *19*, 10–30. [PubMed]
64. Mafi, A.; Rezaee, M.; Hedayati, N.; Hogan, S.D.; Reiter, R.J.; Aarabi, M.H.; Asemi, Z. Melatonin and 5-fluorouracil combination chemotherapy: Opportunities and efficacy in cancer therapy. *Cell Commun. Signal* **2023**, *21*, 33. [CrossRef]
65. Huang, Y.H.; Chiang, W.Y.; Chen, P.J.; Lin, E.S.; Huang, C.Y. Anticancer and Antioxidant Activities of the Root Extract of the Carnivorous Pitcher Plant *Sarracenia purpurea*. *Plants* **2022**, *11*, 1668. [CrossRef] [PubMed]
66. Guan, H.H.; Huang, Y.H.; Lin, E.S.; Chen, C.J.; Huang, C.Y. Plumbagin, a Natural Product with Potent Anticancer Activities, Binds to and Inhibits Dihydroorotase, a Key Enzyme in Pyrimidine Biosynthesis. *Int. J. Mol. Sci.* **2021**, *22*, 6861. [CrossRef] [PubMed]
67. Wyatt, M.D.; Wilson, D.M., 3rd. Participation of DNA repair in the response to 5-fluorouracil. *Cell Mol. Life Sci.* **2009**, *66*, 788–799. [CrossRef]
68. Malik, Z.; Parveen, R.; Zahiruddin, S.; Gaurav; Husain, S.A.; Ahmad, S. HPTLC Stability Indicating Analytical Method of Andrographolide and 5-fluorouracil with Network Pharmacology Analysis against Cancer. *Comb. Chem. High Throughput Screen* **2023**. *ahead of print*. [CrossRef]
69. Yang, H.; Jeffrey, P.D.; Miller, J.; Kinnucan, E.; Sun, Y.; Thoma, N.H.; Zheng, N.; Chen, P.L.; Lee, W.H.; Pavletich, N.P. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **2002**, *297*, 1837–1848. [CrossRef]
70. Brosey, C.A.; Yan, C.; Tsutakawa, S.E.; Heller, W.T.; Rambo, R.P.; Tainer, J.A.; Ivanov, I.; Chazin, W.J. A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **2013**, *41*, 2313–2327. [CrossRef]
71. Suksombat, S.; Khafizov, R.; Kozlov, A.G.; Lohman, T.M.; Chemla, Y.R. Structural dynamics of *E. coli* single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *eLife* **2015**, *4*, e08193. [CrossRef] [PubMed]
72. Zhou, R.; Kozlov, A.G.; Roy, R.; Zhang, J.; Korolev, S.; Lohman, T.M.; Ha, T. SSB functions as a sliding platform that migrates on DNA via reptation. *Cell* **2011**, *146*, 222–232. [CrossRef]
73. Scholz, C.; Knorr, S.; Hamacher, K.; Schmidt, B. DOCKTITE-a highly versatile step-by-step workflow for covalent docking and virtual screening in the molecular operating environment. *J. Chem. Inf. Model.* **2015**, *55*, 398–406. [CrossRef] [PubMed]
74. Dubiel, K.; Myers, A.R.; Kozlov, A.G.; Yang, O.; Zhang, J.; Ha, T.; Lohman, T.M.; Keck, J.L. Structural Mechanisms of Cooperative DNA Binding by Bacterial Single-Stranded DNA-Binding Proteins. *J. Mol. Biol.* **2019**, *431*, 178–195. [CrossRef] [PubMed]
75. Huang, Y.H.; Lo, Y.H.; Huang, W.; Huang, C.Y. Crystal structure and DNA-binding mode of *Klebsiella pneumoniae* primosomal PriB protein. *Genes Cells* **2012**, *17*, 837–849. [CrossRef] [PubMed]
76. Windgassen, T.A.; Wessel, S.R.; Bhattacharyya, B.; Keck, J.L. Mechanisms of bacterial DNA replication restart. *Nucleic Acids Res* **2018**, *46*, 504–519. [CrossRef]
77. Bruand, C.; Ehrlich, S.D.; Janniere, L. Primosome assembly site in *Bacillus subtilis*. *EMBO J.* **1995**, *14*, 2642–2650. [CrossRef]
78. Otwinowski, Z.; Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **1997**, *276*, 307–326.
79. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. [CrossRef]
80. Terwilliger, T.C.; Adams, P.D.; Read, R.J.; McCoy, A.J.; Moriarty, N.W.; Grosse-Kunstleve, R.W.; Afonine, P.V.; Zwart, P.H.; Hung, L.W. Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *65*, 582–601. [CrossRef]
81. Lebedev, A.A.; Young, P.; Isupov, M.N.; Moroz, O.V.; Vagin, A.A.; Murshudov, G.N. JLigand: A graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 431–440. [CrossRef]
82. Terwilliger, T.C.; Grosse-Kunstleve, R.W.; Afonine, P.V.; Moriarty, N.W.; Zwart, P.H.; Hung, L.W.; Read, R.J.; Adams, P.D. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64*, 61–69. [CrossRef] [PubMed]
83. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. [CrossRef] [PubMed]
84. Chen, V.B.; Arendall, W.B., 3rd; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12–21. [CrossRef] [PubMed]
85. Peng, W.F.; Huang, C.Y. Allantoinase and dihydroorotase binding and inhibition by flavonols and the substrates of cyclic amidohydrolases. *Biochimie* **2014**, *101*, 113–122. [CrossRef]
86. Huang, Y.H.; Lien, Y.; Chen, J.H.; Lin, E.S.; Huang, C.Y. Identification and characterization of dihydropyrimidinase inhibited by plumbagin isolated from Nepenthes miranda extract. *Biochimie* **2020**, *171–172*, 124–135. [CrossRef] [PubMed]
87. Huang, C.Y. Inhibition of a putative dihydropyrimidinase from *Pseudomonas aeruginosa* PAO1 by flavonoids and substrates of cyclic amidohydrolases. *PLoS ONE* **2015**, *10*, e0127634. [CrossRef]

88. Carrasco, B.; Seco, E.M.; López-Sanz, M.; Alonso, J.C.; Ayora, S. *Bacillus subtilis* RarA modulates replication restart. *Nucleic Acids Res.* **2018**, *46*, 7206–7220. [CrossRef]
89. Par, S.; Vaides, S.; VanderVere-Carozza, P.S.; Pawelczak, K.S.; Stewart, J.; Turchi, J.J. OB-Folds and Genome Maintenance: Targeting Protein-DNA Interactions for Cancer Therapy. *Cancers* **2021**, *13*, 3346. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

Paclitaxel and Caffeine–Taurine, New Colchicine Alternatives for Chromosomes Doubling in Maize Haploid Breeding

Saeed Arshad [1], Mengli Wei [1], Qurban Ali [2], Ghulam Mustafa [2], Zhengqiang Ma [1] and Yuanxin Yan [1,3,*]

[1] State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China; maliksaeedawan@hotmail.com (S.A.); 2019201097@njau.edu.cn (M.W.); zqm2@njau.edu.cn (Z.M.)
[2] Department of Plant Pathology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China; rattarqurban@hotmail.com (Q.A.); 2018201104@njau.edu.cn (G.M.)
[3] Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing 210095, China
* Correspondence: yuanxin.yan@njau.edu.cn

Abstract: The doubled haploid (DH) technology is employed worldwide in various crop-breeding programs, especially maize. Still, restoring tassel fertility is measured as one of the major restrictive factors in producing DH lines. Colchicine, nitrous oxide, oryzalin, and amiprophosmethyl are common chromosome-doubling agents that aid in developing viable diploids (2n) from sterile haploids (n). Although colchicine is the most widely used polyploidy-inducing agent, it is highly toxic to mammals and plants. Therefore, there is a dire need to explore natural, non-toxic, or low-toxic cheaper and accessible substitutes with a higher survival and fertility rate. To the best of our knowledge, the advanced usage of human anticancer drugs "Paclitaxel (PTX)" and "Caffeine–Taurine (CAF–T)" for in vivo maize haploids doubling is being disclosed for the first time. These two antimitotic and antimicrotubular agents (PTX and CAF–T) were assessed under various treatment conditions compared to colchicine. As a result, the maximum actual doubling rates (ADR) for PTX versus colchicine in maize haploid seedlings were 42.1% (400 M, 16 h treatment) versus 31.9% (0.5 mM, 24 h treatment), respectively. In addition, the ADR in maize haploid seeds were CAF–T 20.0% (caffeine 2 g/L + taurine 12 g/L, 16 h), PTX 19.9% (100 μM, 24 h treatment), and colchicine 26.0% (2.0 mM, 8 h treatment). Moreover, the morphological and physiological by-effects in haploid plants by PTX were significantly lower than colchicine. Hence, PTX and CAF–T are better alternatives than the widely used traditional colchicine to improve chromosome-doubling in maize crop.

Keywords: maize haploid; chromosome doubling; paclitaxel; caffeine–taurine; colchicine; anticancer

Citation: Arshad, S.; Wei, M.; Ali, Q.; Mustafa, G.; Ma, Z.; Yan, Y. Paclitaxel and Caffeine–Taurine, New Colchicine Alternatives for Chromosomes Doubling in Maize Haploid Breeding. *Int. J. Mol. Sci.* **2023**, *24*, 14659. https://doi.org/10.3390/ijms241914659

Academic Editor: Lam-Son Phan Tran

Received: 10 July 2023
Revised: 25 August 2023
Accepted: 16 September 2023
Published: 28 September 2023

1. Introduction

During the last century, the development of maize inbred lines relied on recurrent selfing and selection to reach the desired level of homozygosity [1]. It requires six-to-eight generations of selfing after crossing two parent lines. In contrast, the process of double haploid (DH) line production involves generating haploids (by crossing female source parent with an inducer) with only the gametic chromosomes, the identification of haploids from diploids, and the doubling of the chromosomes in the haploids. As a result, the DH line contains two sets of chromosomes that are exact replicas of each other [2,3]. The DH technology facilitates the rapid attainment of genetically homozygous, ~99% pure inbred lines with distinctive characteristics, uniformity, and stability from heterozygous germplasm in a single generation [4], which helps to achieve rapid varietal registration and secure breeders' property [5]. The DH technology has been widely adopted as an effective alternative to conventional recurrent self-breeding. It is being used for upkeeping germplasm, discovering the genetic diversity in maize, and expanding the genetic base of exclusive top germplasm through novel variations [6–8]. During the

development of conventional parental lines, plant breeders have to face different and complex challenges of segregation, linked undesirable traits, uncontrolled morphological variation, unpredicted traits, and an ambiguous selection of recessive unexpressed desired traits. During DH line development, higher heritability and genetic variance facilitate proper selection [5]. The DH technology overcomes obstacles during recurrent self-breeding and provides an efficient, cheaper, timely, and less laborious substitute. Doubled haploids achieved after doubling chromosomes are considered novel, highly uniform genotypes because of homozygosity at all loci [9]. During the last decade, the DH technology opened new avenues of research; for example, genetic engineering [10], marker-trait associated studies [11], genomics and transformation target [12], genome mapping [13], valuation of quantitative trait loci (QTL) × environmental interactions with precision [4], and QTL mapping [14]. The foremost applied approaches used in breeding to develop DH lines include androgenesis, wide hybridization, and gynogenesis [15].

Haploid plants are generally sterile [16], and their fertility can be restored spontaneously or/and by chromosomal doubling agents. Among popular chromosome doubling agents, colchicine is preferred over nitrous oxide (N_2O) and antimitotic herbicides, such as benzamides (pronamide), dinitro–anilines (oryzalin and trifluralin), carbamates (chlorpropham and isopropyl N-3-chlorophenyl carbamate), and phosphor–thioimidates (aminoprophosmethyl/APM) due to its availability and well-developed protocols. These herbicides stop mitosis and cell division, leading to chromosome doubling [17], like colchicine [18,19]. In contrast, higher concentrations of trifluralin, APM, and oryzalin revealed equal results of chromosome doubling by colchicine in wheat during in vitro studies [9]. In maize, flufenacet, oryzalin, and trifluralin were assessed for double haploid production in vitro [20] and in vivo [21], but found unimpressive in terms of the combinations applied and concentrations used [22]. A high concentration of oryzalin significantly decreased callus regeneration; whereas, at a low concentration, oryzalin has been found ineffective regarding chromosome doubling [23]. Oryzalin had a lower survival rate than colchicine and APM but had better chromosome doubling rates than colchicine in maize [24]. Another alternative to colchicine developed by [25] employs N_20, which is considered a comparatively safe and effective alternative [22], but the cost of N_20 pressure chamber is relatively higher [25,26]. N_20 exposure at 0.6 MPa (MegaPascal) on maize at V3–V8 growth stage revealed even potential regarding the doubling of chromosomes, but treatment of adult plants requires big chambers and pots for growth [26].

Colchicine is far more toxic than pronamide, trifluralin, N_2O, APM, and oryzalin [22,24,25], requiring careful handling and disposal [16,21]. "The reported oral LD50 (lethal dose, 50%) values for pronamide range from 5620 to 8350 mg/kg, oryzalin 5000 (mg/kg), trifluralin >10,000 mg/kg, colchicine 5.8 mg/kg and APM 309 mg/kg in rats. However, the LD50 for PTX solid dispersion stated 160 mg/kg, which is ~28 times higher than colchicine." Colchicine expresses low affinity with plant microtubules, which means required in higher concentrations [27,28]. However, it has revealed a high affinity with microtubules of vertebrates [17]. Recently, safer alternatives like N_20 and anti-mitotic herbicides with equal efficacy of doubling chromosomes to toxic colchicine have been acknowledged. However, some basic limitations confine their usages, such as cost, availability, higher doses or concentrations, and unregistered sales, such as pronamide and APM in most African and Asian countries, limit their easy accessibility. The availability of such chemicals in high-quality pure grade is more expensive than colchicine and is limited [5]. No doubt these antimitotic herbicides (Pronamide, APM, trifluralin, and oryzalin) are comparatively less toxic but do not meet the expectations to reduce the cost per DH line. Paclitaxel (PTX) is an effective antineoplastic agent, originally extracted from the bark of *Taxus brevifolia*. Although it was generally considered a particular metabolite of *Taxus* spp., it was recently found in hazel cell cultures [27]. It is an intravenous injection to treat several cancers, including breast, cervical, and ovarian. The World Health Organization (WHO) has enlisted it among essential medicines due to its antitumor activity. According to Tel-Aviv University Safety Unit issue "Standard Operating Procedure for Paclitaxel (Taxol)

in Animals," it is considered a cytotoxic drug with no set safety standards of exposure. According to medical opinion, reducing exposure as much as possible is the best safety approach. Due to its lower solubility in water, a formulation is prepared by dehydrated ethanol (50:50, v/v) and cremophor EL, known as "Taxol" [27]. PTX is an antimitotic drug, which interferes in the normal function of microtubules, defects in chromosome segregation, spindle assembly, and cell division. Chromosomes do not achieve spindle configuration in metaphase, block further progress in mitosis, prolong cell-programmed death, or reverse the quiescent stage (G0) phase without cell division of the cell cycle. PTX with a functionary mechanism primarily inhibits the dynamics of microtubule spindle, deoxyribonucleic acid (DNA) repair, and control cell proliferation [29]. Microtubules consisting of two similar polypeptides (α and β tubulin dimers) are the key component of the cytoskeleton and mitotic apparatus in eukaryotic cells [30]. PTX can selectively bind to the subunit β of tubulin proteins and promote their polymerization and assembly. This polymerization consumes intracellular tubulin, stops the function of the cell, averts the formation of the spindle, causes profound mitotic arrest at G2/M phases, and eventually terminates mitosis [31]. It is also known as a cytoskeletal drug, which targets tubulin. It stabilizes the microtubule polymer and shields it from disassembly, whereas colchicine obstructs the microtubule assembly. It has been found to suppress detachment of microtubules from centrosomes, a normal process triggered during mitosis. In contrast to traditional anticancer drugs, PTX neither affects the synthesis of DNA and ribonucleic acid (RNA) in cancer cells nor damages DNA molecules [32].

Taurine is a non-essential amino acid containing sulfur called 2-amino-ethane-sulfonic acid [33]. It is a beneficial anticancer agent that inhibits reactive oxygen species (ROS) buildup in the tumor cells. It stops the advancement of cancer [34], maintains the concentration of calcium and stability of membranes, and sustains the process of phosphorylation [35,36]. It stimulates tumor suppressors p53 and phosphatase and tensin homolog (PTEN) [37]. Fluorescence studies revealed that taurine has a significant binding affinity with cyclin-dependent kinase 6 (CDK6), which is connected with cyclin partner and initiates a critical role in early phases of cancer development [38]. The muscle toxicities produced during chemotherapy in cancer patients can be reduced by taurine [39]. Taurine has an inhibitory concentration 50 (IC50) value equal to 4.44 μM and, according to an enzyme-inhibition assay, is considered a good inhibitor to treat cancers directed by CDK6 [38]. The supplemental taurine dose helps to protect C2C12 myoblasts against decreasing cell viability and moderately conserve the myotube differentiation capability of cisplatin-impaired myoblasts [39].

Caffeine can increase the early doubling of chromosomes and haploid plants, the androgenic induction, and spontaneously produce fertile plants [40]. It can impact cell division and phragmoplast microtubules during cytokinesis, observed by [41] in suspension-cultured cells, BY-2, of tobacco (*Nicotiana tabacum* "Bright Yellow 2"). Caffeine did not significantly affect the doubling of chromosomes in wheat anther culture. However, several caffeine treatments reported a higher frequency of chromosomes doubled plants [40]. Therefore, we decided to assess the efficacy of caffeine in combination with taurine for a higher chromosome doubling rate in our study.

Continuous efforts proceed to optimize cheaper, innovative, non-toxic, or less toxic chemical agents to improve chromosome-doubling efficiency with higher survival rates than colchicine. To meet this challenge, we designed an advanced study to explore the ability of PTX and caffeine–taurine (CAF–T) to induce chromosome doubling in maize haploids. A systematic approach was adopted, and multiple treatments with different processing times were designed based on replicated field trials to study the chromosomal doubling efficacy of PTX and CAF–T versus colchicine. Moreover, the impact of PTX versus colchicine on the morphology and physiology of haploid plants was assessed and analyzed through seed germination, photosynthetic pigments, plant and ear height, number of silks/ears, and plant weight.

2. Results

2.1. Experiment 1

Field Assessment-Based Comparative Efficiencies in Different Treatments of PTX Application

Under different concentrations and processing periods, the survival rate (SR), doubling rate (DR), and actual doubling rate (ADR) of PTX-treated haploid seeds and seedlings were analyzed. In addition, anthers emergence rate (AER), partial fertility rate (PFR), complete fertility rate (CFR), and total fertility rate (TFR) were also calculated based on four different types of tassels (Figure 1A) usually observed in the DH nursery field. The SRs were decreased with the increasing concentration and the processing time in both treatments of the seed-soaking method (M1) and seedling-immersion method (M2). While comparing the SR of seeds versus seedlings, M1 treatments (Figure 2A) were higher than M2 treatments (Figure 2B), which concluded that both coleoptile and root tips were more sensitive and affected by the chemical treatments. Hence, M1 is a better option than M2 to achieve higher SR.

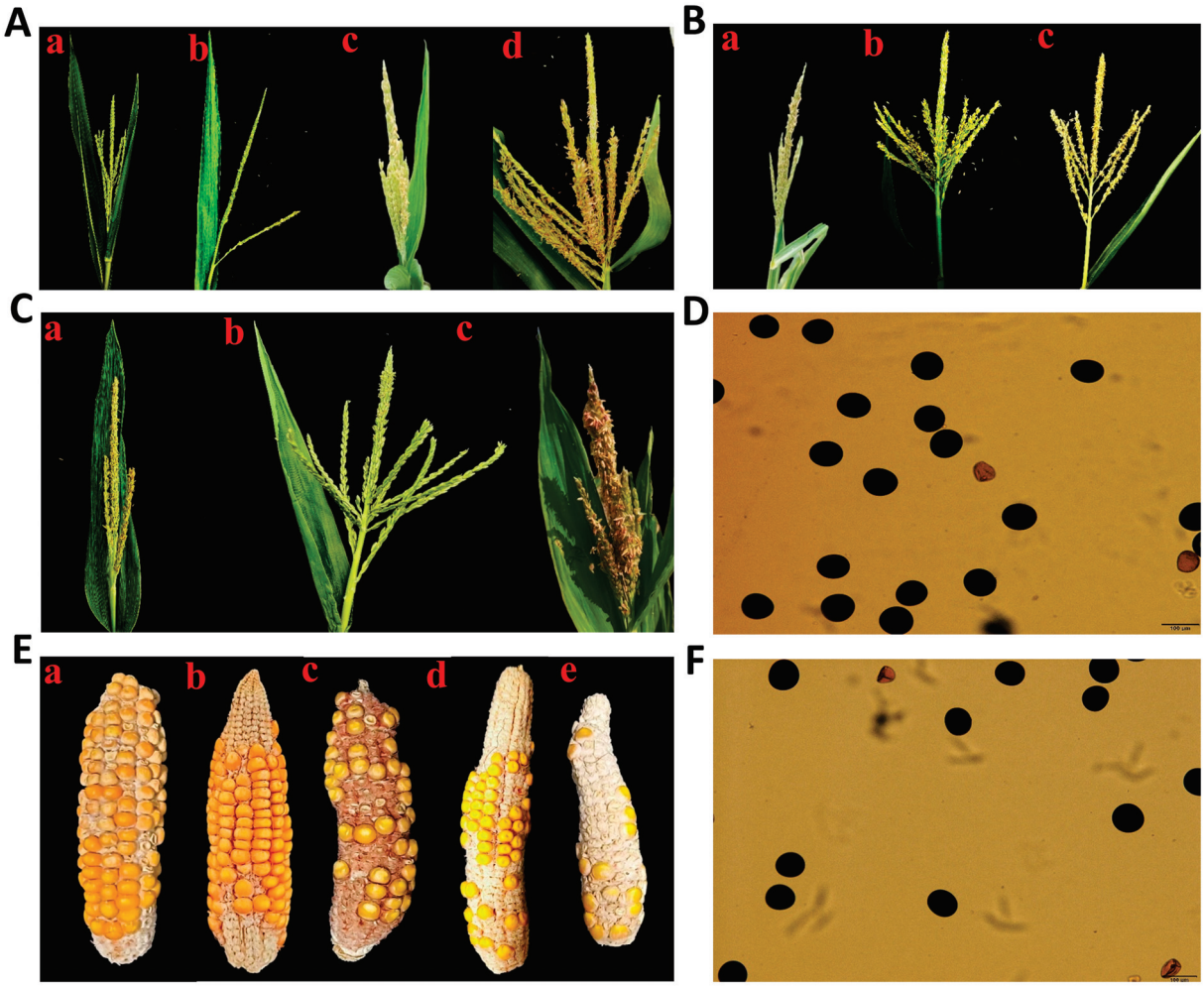


Figure 1. Paclitaxel (PTX) and Caffeine–Taurine (CAF-T) fertility and seed-setting efficacy. (**A**) Four types of maize tassels (**a**) Sterile, (**b**) Pollenless anthers, (**c**) Partially fertile, and (**d**) Complete fertile tassel in the DH field. (**B**) PTX-induced fertility (**a**) Anthers only, (**b**) Partial fertility, (**c**) Complete fertility. (**C**) CAF–T-induced fertility (**a**) Anthers only, (**b**) Partial fertility, (**c**) Complete fertility. (**D**) Effective pollen fertility induced by PTX. (**E**) Maximum DH seed quantity produced by (**a**) PTX seedling inversion method (M2), (**b**) Colchicine M2, (**c**) CAF-T seed soaking method (M1), (**d**) PTX M1, and (**e**) Standard colchicine treatment. (**F**) Effective pollen fertility induced by CAF-T.

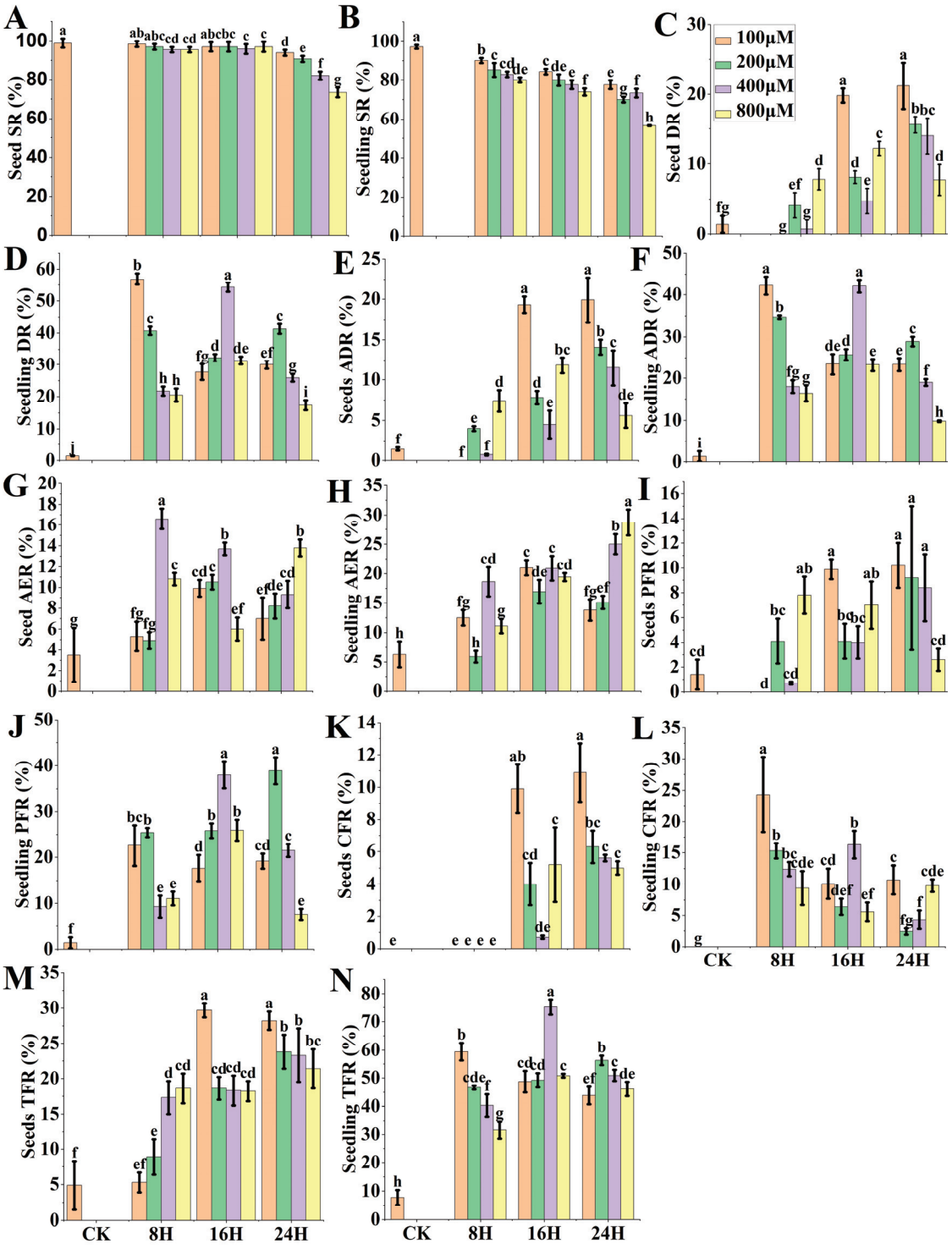


Figure 2. Paclitaxel (PTX) seed-soaking method (M1) versus seedling-immersion method (M2) efficacy. (**A**) Seed SR (survival rate), (**B**) Seedling SR, (**C**) Seed DR (doubling rate), (**D**) Seedling DR, (**E**) Seed ADR (actual doubling rate), (**F**) Seedling ADR, (**G**) Seed AER (anthers emergence rate), (**H**) Seedling AER, (**I**) Seed PFR (partial fertility rate), (**J**) Seedling PFR, (**K**) Seed CFR (complete fertility rate), (**L**) Seedling CFR, (**M**) Seed TFR (total fertility rate), (**N**) Seedling TFR. CK = control; this treatment is identical to all other treatments, but PTX, Di-methyl Sulfoxide (DMSO), and tween-80 solution were replaced by distilled deionized water (ddH_2O). Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard error of the mean ($n = 3$).

In all treatments of M1, the highest DRs induced by PTX were 19.8% and 21.2% in the condition of 100 μM for 16 h and 24 h, respectively. However, both conditions for DR, ADR, PFR, CFR, and TFR were determined to be non-significantly different to each other. The M2 in PTX produced the highest DR of 54.3% and TFR of 75.2% in the condition of 400 μM for 16 h, followed by another treatment that produced a DR of 46.9% and TFR of 59.4% in the condition of 100 μM for 8 h. However, both conditions were statistically non-significant regarding ADR (42.1% and 42.2%, respectively). While comparing both M1 versus M2, M2 is a better line not only to yield higher DR (Figure 2C,D) but also ADR, AER, PFR, CFR, and TFR than M1 (Figure 2E–N). AER, PFR, and CFR induced by PTX were photographed (Figure 1B). The pollen fertility induced by PTX in haploid plants is shown in Figure 1D.

2.2. Experiment 2

2.2.1. Field Assessment-Based Comparative Efficiencies in Different Treatments of CAF-T Application

In Experiment 2, using different treatments and processing times, CAF–T application to maize haploid seeds was assessed through SR, DR, ADR, AER, PFR, CFR, and TFR (Figure 3A–G). The highest DR and TFR produced were 20.9% and 28.1%, respectively, in caffeine 2 g/L + taurine 12 g/L for 16 h (Figure 3B,G). However, the SR revealed by this treatment was 95.8% (Figure 3A), disclosing non-significant difference with control (CK). The SRs decreased non-significantly with the increasing concentration and the processing time in T1–T6 (except T5) of CAF–T versus CK. Hence, CAF–T disclosed its safe use for maize seeds. CAF–T did not show doubling efficiency in treatments of 4 g and 6 g, 16 h and all 24 h treatments (Figure 3B). However, AER was higher than doubling treatments (T1–T6) (Figure 3D), which might reveal the impact of higher concentrations of CAF–T and longer time exposure. AER, PFR, and CFR induced by CAF–T are displayed in Figure 1C. The pollen fertility induced by CAF–T in haploid plants is shown in Figure 1F.

2.2.2. Field Assessment-Based Comparative Efficiencies in Different Treatments of Colchicine Application

Two treatment methods, i.e., M1 and M2, were employed using different colchicine concentrations and processing time to calculate SR, DR, ADR, AER, PFR, CFR, and TFR (Figure 4A–N). In M1 and M2, the survival rates decreased with the increasing concentration and processing time. However, M1 treatments revealed higher SR than M2 treatments (Figure 4A,B). In M1, the highest DR induced by colchicine was 29.6%, with SR 87.7% in the condition of 2.0 mM for 8 h treatment, disclosing ADR 26.0% (Figure 4E).

In M2, the highest DR induced by colchicine was calculated 43.8% with SR 72.9% in the condition of 0.5 mM for 24 h treatment (Figure 4B,D), revealing ADR 31.9% (Figure 4F). Hence, M2 was found to be better than M1. Another treatment of M2 in the condition of 2.0 mM for 8 h revealed the highest AER and TFR 35.9% and 62.1%, respectively (Figure 4H,N) among all treatments. However, DR was only 26.2%, indicating the impact of sectoral diploidization.

2.3. Comparative DH Seed Quantity Produced by PTX, CAF-T versus Colchicine

In addition to fertility rates, the number of seeds on DH ears subsequent from each treatment was widely variable (Supplementary Tables S1 and S2). Mostly plants produced $\leq$ 5 seeds/ear. However, maximum number of seeds produced by PTX (M2), colchicine (M2), and CAF–T (M1) were counted as 115 (800 μM, 16 h), 95 (0.5 mM, 16 h), and 38 (1 g/L, 16 h), respectively (Figure 1E(a–c)). The following treatments of M1 treated by PTX produced 7 seeds (100 μM, 16 h), 13 seeds (100 μM, 24 h), 52 seeds (200 μM, 16 h), and 53 seeds (400 μM, 24 h; Figure 1E(d)) as well. Hence, PTX and CAF–T produced a comparatively higher seed count than standard colchicine treatment M2 (1.5 mM, 8 h) [16, 42], which produced 21 seeds (Figure 1E(e)).

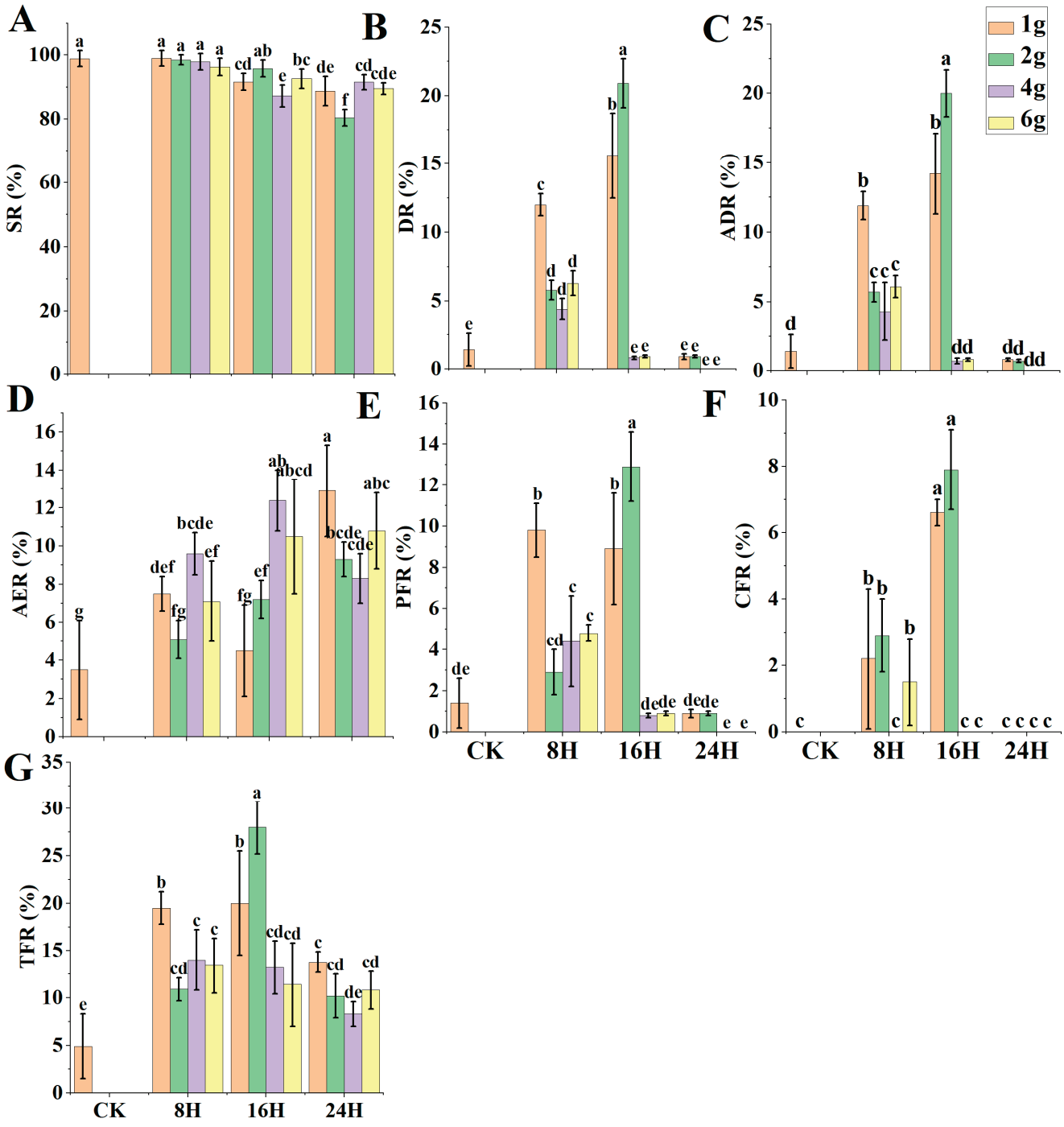


Figure 3. Caffeine–Taurine (CAF–T) seed-soaking method efficacy. (**A**) Seed SR (survival rate), (**B**) Seed DR (doubling rate), (**C**) Seed ADR (actual doubling rate), (**D**) Seed AER (anthers emergence rate), (**E**) Seed PFR (partial fertility rate), (**F**) Seed CFR (complete fertility rate), (**G**) Seed TFR (total fertility rate). CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced CAF–T and Di-methyl Sulfoxide (DMSO) solution. Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard deviation of the mean ($n = 3$).

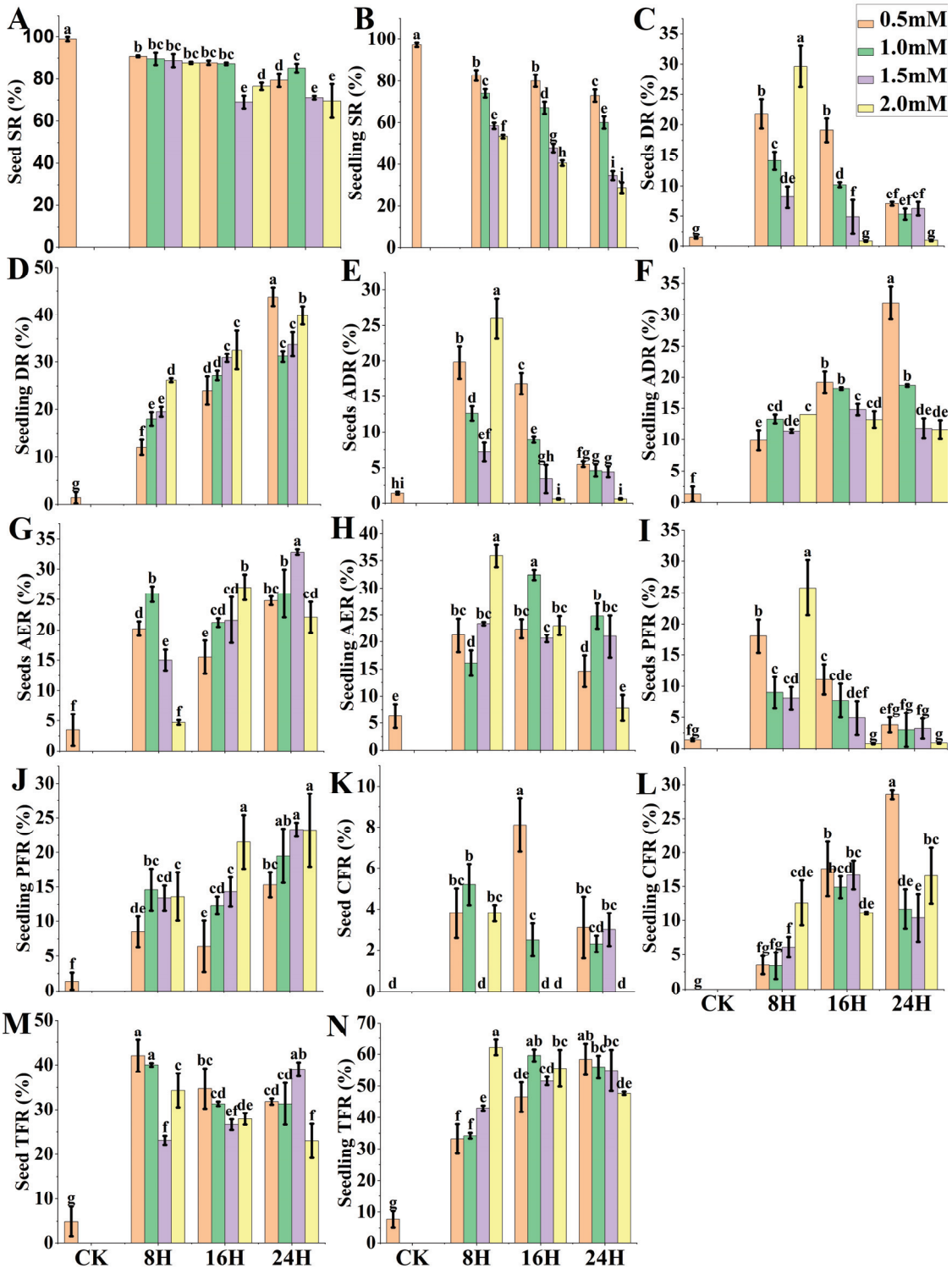


Figure 4. Colchicine seed-soaking method (M1) versus seedling-immersion method (M2) efficacy. (**A**) Seed SR (survival rate), (**B**) Seedling SR, (**C**) Seed DR (doubling rate), (**D**) Seedling DR, (**E**) Seed ADR (actual doubling rate), (**F**) Seedling ADR, (**G**) Seed AER (anthers emergence rate), (**H**) Seedling AER, (**I**) Seed PFR (partial fertility rate), (**J**) Seedling PFR, (**K**) Seed CFR (complete fertility rate), (**L**) Seedling CFR, (**M**) Seed TFR (total fertility rate), (**N**) Seedling TFR. CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced colchicine and Dimethyl Sulfoxide (DMSO) solution. Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard error of the mean ($n = 3$).

2.4. Experiment 4

Large-Scale (LS) Field Efficacy-Based Comparative Studies for Validation of the Best Treatment of PTX for DH Production Pipeline

To establish the efficacy of PTX versus colchicine, 1478 putative haploid seedlings from two populations were used in this experiment. PTX treatment presented consistent and significantly better SR, ADR, and TFR in both genotypes as compared to colchicine treatment (Figure 5A,B). However, DR and CFR were significantly higher in PTX than colchicine, and AER was significantly lower in PTX than colchicine in variety 1 (V1). The CFR in colchicine treatment was significantly higher than PTX, but AER presented by PTX treatment was significantly higher than standard colchicine treatment in variety 2 (V2). However, DR was non-significantly different in both PTX and colchicine in V2.

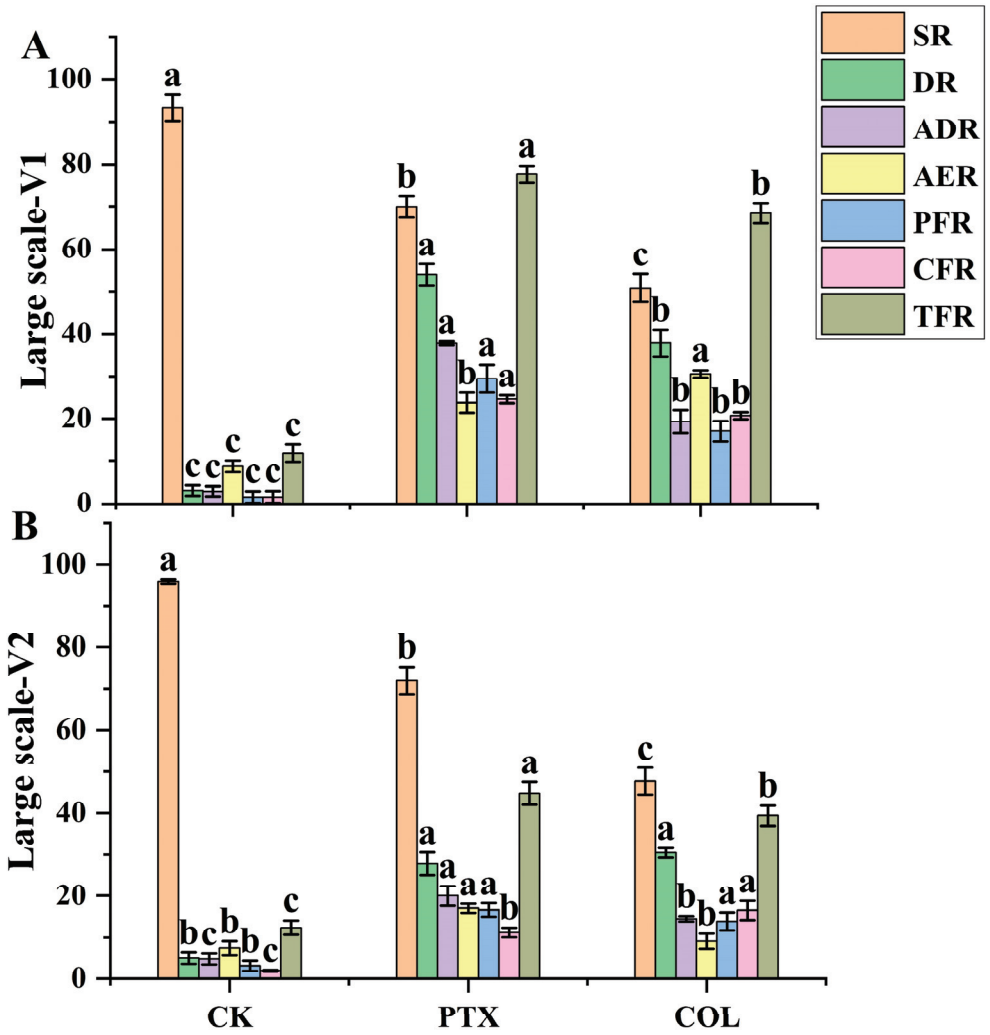


Figure 5. Large-scale efficacy comparison between Paclitaxel (PTX) versus colchicine (COL) of two maize genotypes. (**A**) V1 and (**B**) V2. SR (survival rate), DR (doubling rate), ADR (actual doubling rate), AER (anthers emergence rate), PFR (partial fertility rate), CFR (complete fertility rate), and TFR (total fertility rate). CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced PTX, colchicine, tween-80, and Di-methyl Sulfoxide (DMSO) solution. Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard error of the mean ($n = 3$).

2.5. Experiment 5

2.5.1. Morphological Studies

Comparative Plant and Ear Height of Treated Maize Plants by PTX versus Colchicine and CK

The plant and ear height were significantly lower for both PTX and colchicine than CK in both the genotypes. However, PTX treatment was significantly better than colchicine treatment for both plant and ear height (Figures 6A and 7A,B). Colchicine-treated seedlings exhibited delayed growth compared to PTX.

Comparative Silks Number/Ear Treated by PTX versus Colchicine and CK

The number of silks/ears was significantly lower for both PTX and colchicine than CK treatment in both genotypes. However, PTX treatment was significantly better than colchicine treatment for the number of silk/ear (Figures 6B and 7C), which helps infer that PTX lower toxicity than colchicine will ensure a greater number of D0 seeds as proved above as well (Figure 1E). The count of silks per ear for 20 plants ranged between 220 and 255, 190 and 225, and 252 and 280 for PTX, colchicine, and CK, respectively, for V1. The count of silks per ear for 20 plants ranged between 227 and 270, 175 and 236, and 263 and 320 for PTX, colchicine, and CK, respectively, for V2 (Supplementary Tables S3 and S4).

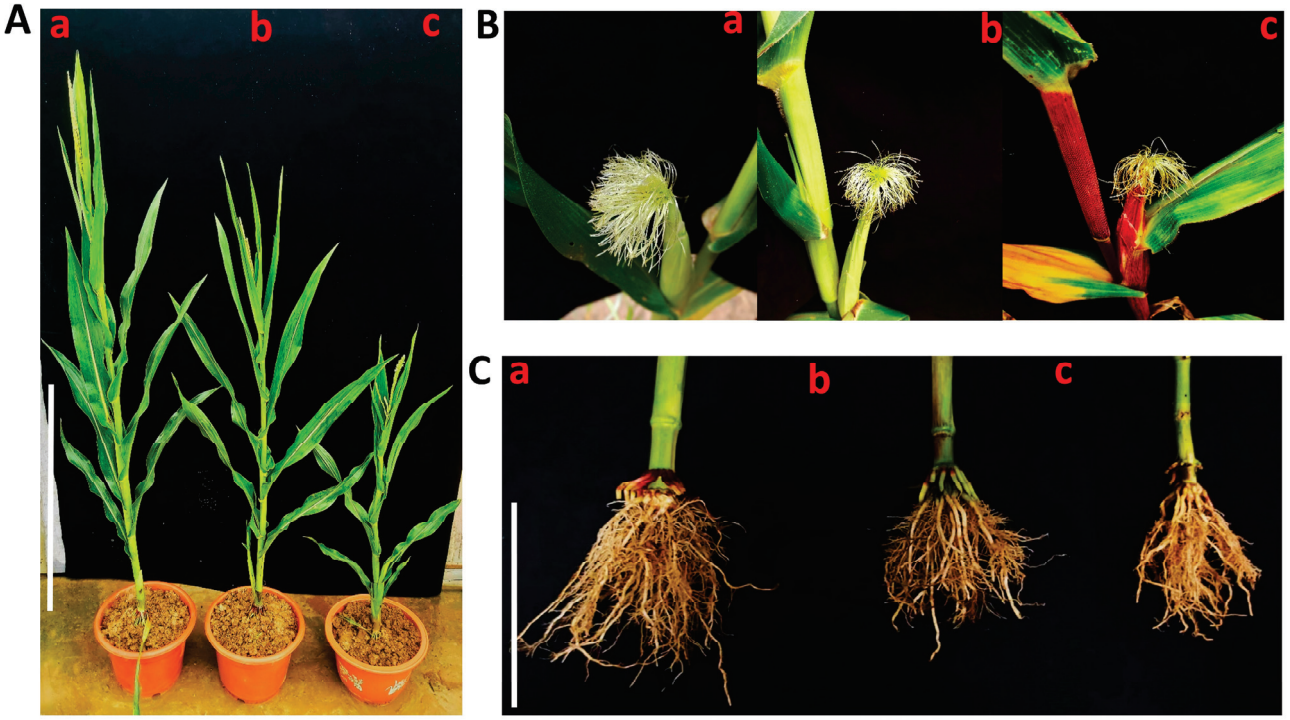


Figure 6. Comparative impact of paclitaxel (PTX) versus colchicine on the morphology of haploid maize plants. (**A**) Impact on plant and ear height; Scale = 85 cm. (**a**) CK (control), (**b**) PTX, (**c**) Colchicine. (**B**) Impact on silks number/ear (**a**) CK, (**b**) PTX, (**c**) Colchicine. (**C**) Root Growth (**a**) CK, (**b**) PTX, (**c**) colchicine; scale = 15 cm.

Comparative Plant Weight and Root Growth Treated by PTX versus Colchicine and CK

Our experiment measured plant weight significantly less for PTX and colchicine than CK in both genotypes. However, PTX treatment was significantly better than colchicine treatment for plant weight (Figure 7D), which helps to deduce that PTX produced fewer toxic effects on plant growth. Plant weight measurements per plant ranged between 95.52 and 110.11 g, 63.92 and 85.56 g, and 129.95 and 147.54 g for PTX, colchicine, and CK, respectively, for Genotype V1. Plant weight measurements per plant ranged between 116.40 and 130.74 gm, 97.520 and 107.62 gm, and 140.12 and 160.24 gm for PTX, colchicine,

and CK, respectively, for Genotype V2 (Supplementary Tables S5 and S6). Moreover, root growth was better for PTX than colchicine (Figure 6C).

2.5.2. Physiological Studies

Comparative Photosynthetic Pigments of Treated Seedlings by PTX versus Colchicine and CK

The chlorophyll *a* (Chl_a), chlorophyll *b* (Chl_b), total chlorophyll (Chl_T), and carotenoid (C_x) contents were significantly lower for both PTX and colchicine than CK in both the genotypes. However, PTX treatment was significantly better than colchicine treatment for Chl_a, Chl_b, Chl_T (Figure 7E,F), and C_x (Figure 7G).

Comparative Germination (%) Impact on Treated Seeds by PTX, CAF–T versus Colchicine

Moreover, to SR, we studied the effects of PTX and CAF–T versus colchicine on seed germination. However, the impact of colchicine on treated haploid maize seeds has not been reported. We selected the induced genotype (GO927 × 986) with Stock 6 inducer for this experiment because of its higher germination of about 97%.

1. Germination (%) Impact on Treated Seeds by PTX

Across all treatments (T1 to T12) of PTX, germination percentage decreased with increasing concentrations and processing time (Figure 8A). PTX-treated seed germination percentage was found to range from 95.7% to 86.8% across all respective seed-soaked treatments. PTX showed a non-significant effect on seed germination percentage from T1–T6 versus CK.

2. Germination (%) Impact on Treated Seeds by CAF–T

CAF–T-treated seed germination percentage was found between 96.2% and 77.8% across all respective seed-soaked treatments. CAF–T exhibited a non-significant impact on seed germination percentages on T1 and T2 versus CK (Figure 8B).

3. Germination (%) Impact on Treated Seeds by Colchicine

The colchicine-treated seed germination percentage ranged from 90.8% to 80.5% across all respective seed-soaked treatments (Figure 8C). In contrast, no colchicine treatment showed a non-significant effect versus CK; therefore, colchicine proved its more toxic effects on seed germination.

In conclusion, PTX revealed a lesser impact on seeds germination percentage versus colchicine.

2.6. Experiment 6

This experiment aimed to validate the chromosomes doubling induced by PTX and CAF–T in the field studies through the microscope.

2.6.1. PTX Induced Chromosome Doubling Signals Detected by Fluorescence in Situ Hybridization (FISH) Using Knob-2 as Probe

PTX Induced Chromosome Doubling by Seed-Soaking Method (M1)

The M1 revealed DR ranging from 60% to 80% across all treatments of PTX based on chromosome doubling count under the microscope using root tip cells.

Maximum DR in M1 achieved 80% by these treatments, i.e., PTX 800 µM 16 h, 24 h, and 48 h: PTX 200 µM 48 h (Figure 9A).

PTX-Induced Chromosome Doubling by Root-Immersion Method (M3)

The M3 showed DR ranging from 56.7% to 70.0% across all treatments of PTX. A maximum DR (70%) was achieved through M3 at concentrations of PTX 400 µM and 800 µM for 16 h (Figure 9B).

In conclusion, M1 exhibited better frequency and DR than M3. In addition, we noticed 48 h treatments were not exceptionally better than 24 h treatments both in M1 and M3. Therefore, it helps deduce that 48 h treatments cannot improve DR further but decrease SR.

The CK treatment showed a spontaneous doubling ratio of about 3%. Hence, PTX proved effective in doubling chromosomes, as shown in Figure 10A,B.

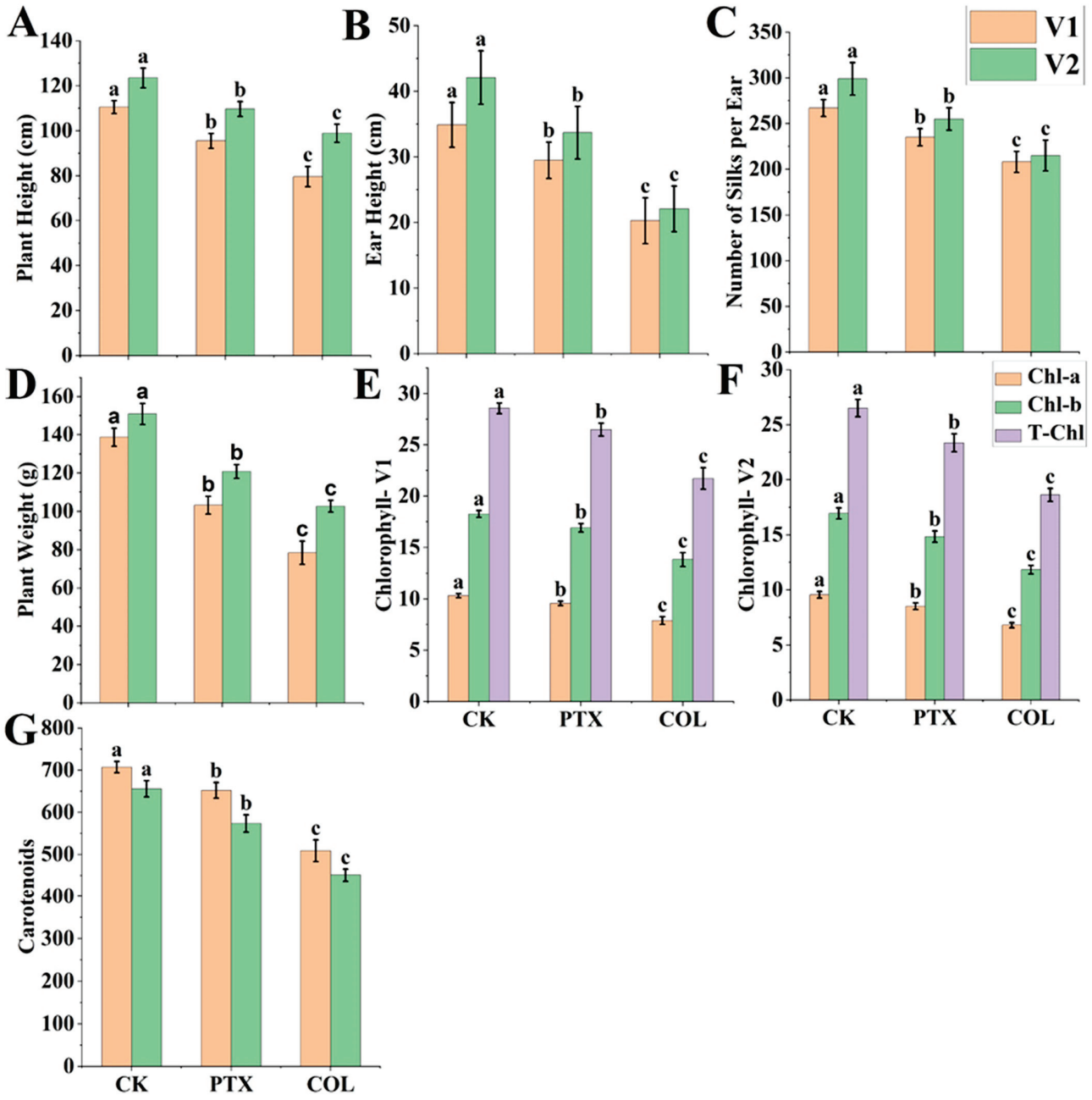


Figure 7. Comparative impact of paclitaxel (PTX) versus colchicine on morphology and physiology of haploid maize plants. (**A**) Plant height, (**B**) Ear height, (**C**) Number of silks/ears, (**D**) Plant weight, (**E**) Chlorophyll *a* (Chl_a), Chlorophyll *b* (Chl_b), and total chlorophyll (Chl_T)—variety 1 (V1), (**F**) Chl_a, Chl_b and Chl_T—variety 2 (V2), (**G**) Carotenoid contents. COL = colchicine; CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced PTX, colchicine, tween-80, and Di-methyl Sulfoxide (DMSO) solution. Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard error of the mean ($n = 3$ or $n = 4$).

2.7. Experiment 7

2.7.1. CAF–T-Induced Chromosome-Doubling Signals Detected by FISH Using Knob-2 as Probe

CAF–T-Induced Chromosome Doubling by Seed-Soaking Method (M1)

The M1 revealed DR ranging from 53.3% to 86.7% across all treatments of CAF–T based on chromosome doubling count under the microscope using root tip cells. The maximum DR in M1 was achieved at 86.7% by caffeine 2 g/L + taurine 12 g/L, 16 h (Figure 9C).

CAF–T-Induced Chromosome Doubling by Root-Immersion Method (M3)

The M3 disclosed DR ranging from 33.3% to 53.3% across all treatments of CAF–T. The maximum DR (53.3%) was achieved by M3 at a concentration of caffeine at 2 g/L and taurine at 12 g/L for 8 h; caffeine at 6 g/L and taurine at 36 g/L for 8 h; caffeine at 4 g/L and taurine at 24 g/L for 24 h; caffeine at 6 g/L and taurine at 36 g/L for 24 h; no further increases in ratios were observed (Figure 9D).

In conclusion, CAF–T-based M1 exhibited better frequency and DR than M3. In addition, we observed that 48 h treatments cannot improve DR further but decrease SR. Hence, CAF–T verified its efficacy in doubling chromosomes, as shown in Figure 10C.

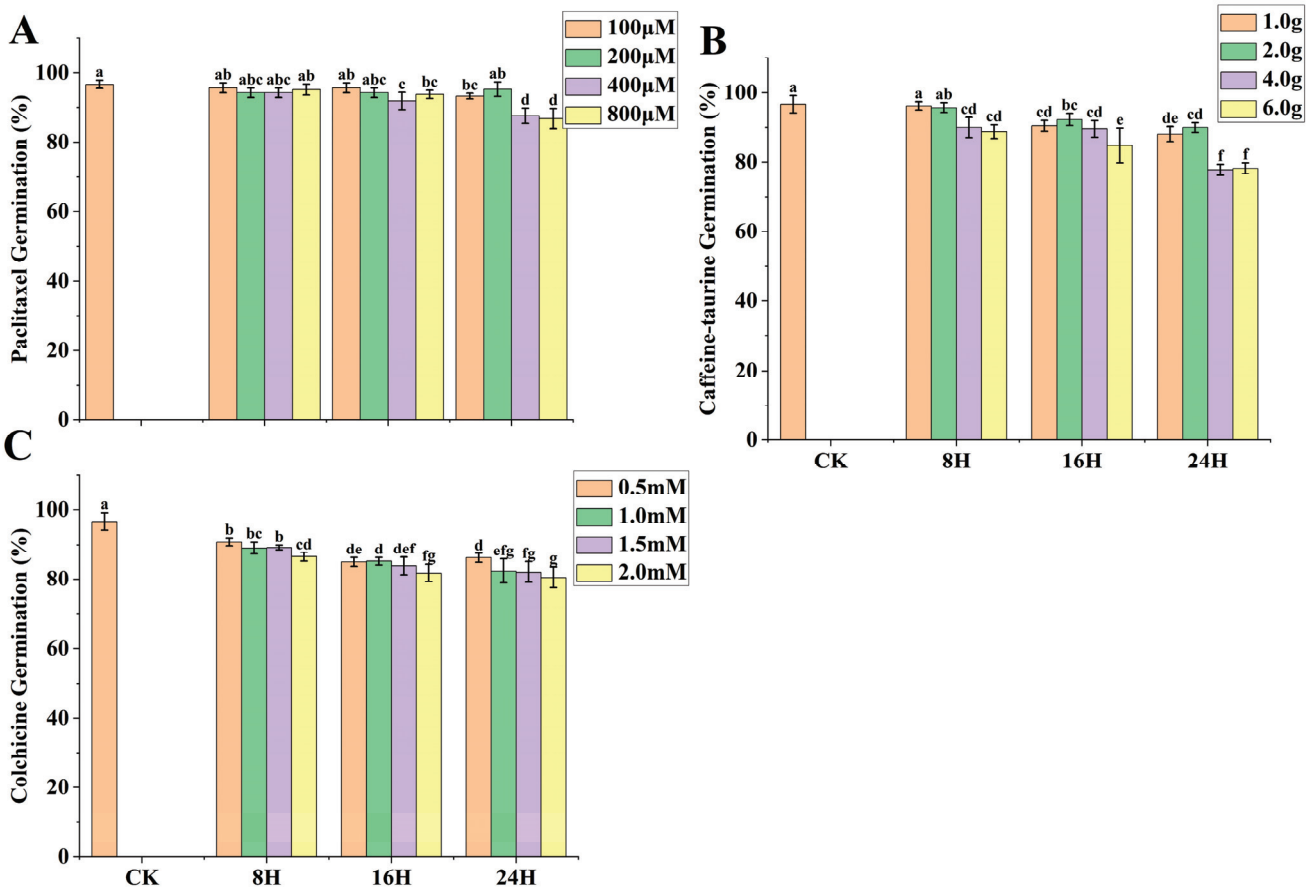


Figure 8. Comparative impact of paclitaxel (PTX), caffeine–taurine (CAF–T), and colchicine on seed germination with different concentrations and treatment time. (**A**) PTX impact on seed germination; (**B**) CAF–T impact on seed germination; (**C**) Colchicine impact on seed germination. CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced PTX, colchicine, tween-80, and di-methyl Sulfoxide (DMSO) solution. Different small letters on bars represent the significant differences within the treatments calculated using Tukey's HSD test at $p \leq 0.05$. Vertical bars on graphs indicate the standard error of the mean ($n = 3$ or $n = 4$).

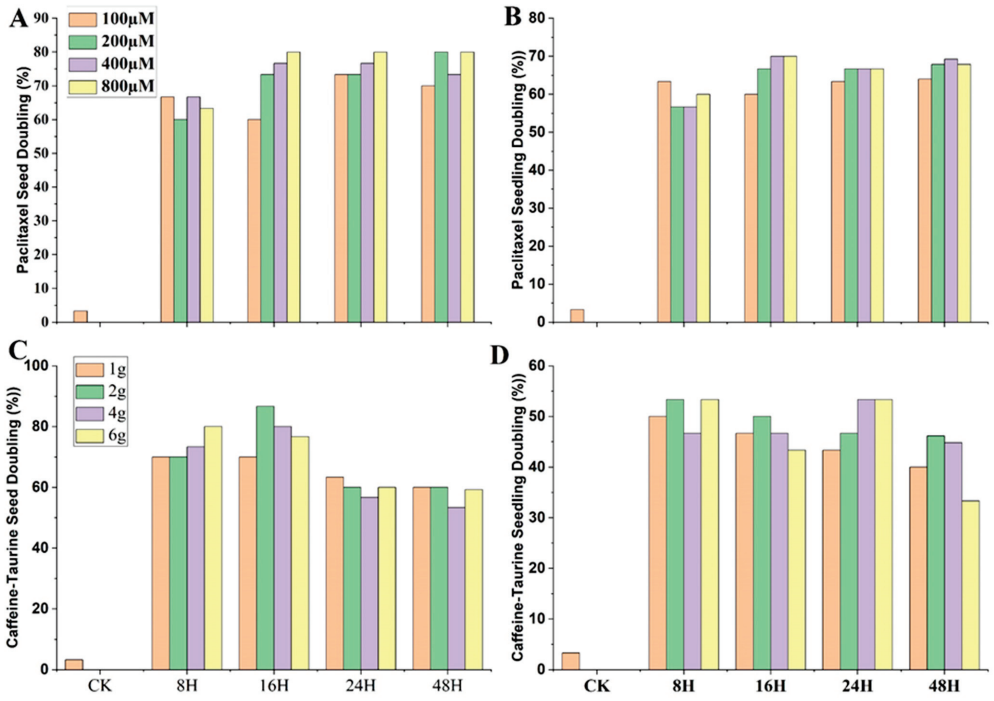


Figure 9. Confirmation of chromosome-doubling ratio (DR) under microscope induced by paclitaxel (PTX) and caffeine–taurine (CAF–T). (**A**) PTX-treated seeds; (**B**) PTX-treated seedlings (roots only); (**C**) CAF–T-treated seeds; (**D**) CAF–T-treated seedlings (roots only). CK = control; this treatment is identical to all other treatments, but distilled deionized water (ddH_2O) replaced PTX, colchicine, tween-80, and Di-methyl Sulfoxide (DMSO) solution.

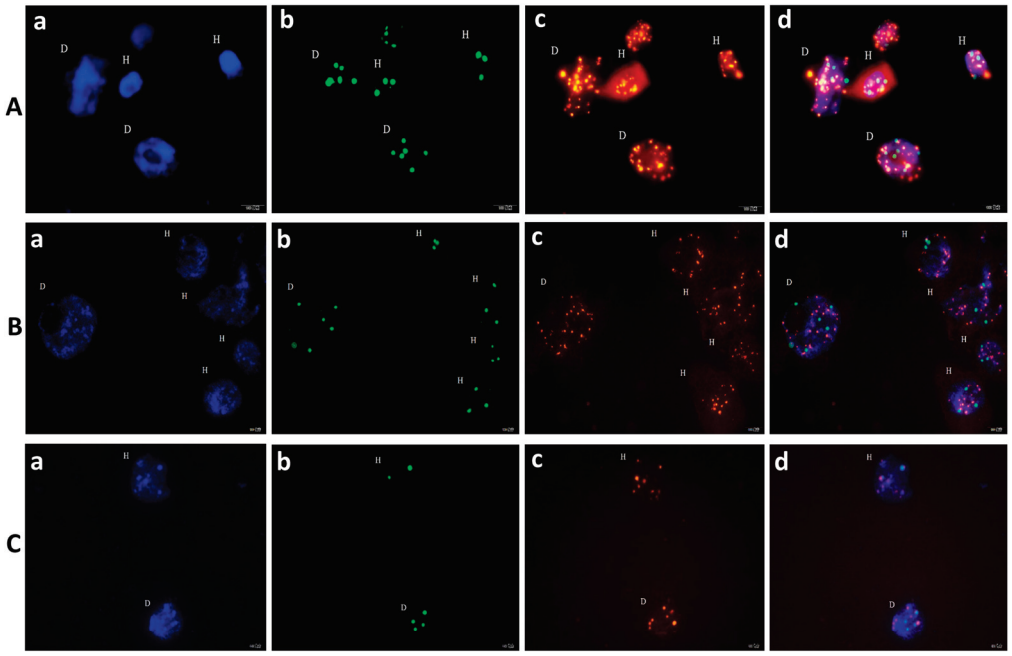


Figure 10. Fluorescence in situ hybridization (FISH) using knob-2 probe to detect chromosomal doubling signals induced by paclitaxel (PTX) and caffeine–taurine (CAF–T). (**A**) PTX-induced chromosomes doubled in the treated seed; (**B**) PTX-induced chromosomes doubled in treated seedling (roots only); (**C**) CAF–T-induced chromosomes doubled in the treated seed; (D) Diploid cells; (H) Haploid cells; (**a**) DAPI (4′,6-diamidino-2-phenylindole) (Vector) (blue); (**b**) Knob-2 (green); (**c**) Gypsy (red); (**d**) Merger of a, b, and c.

3. Discussion

No standard set criteria are used to assess chromosomal doubling agents' efficacy. Some studies were based on the microscopic chromosomal count [43], several results were established on the basis of seed production [5,44], many experiments were based on flow cytometry [24], and numerous researchers considered pollen production [45] as an indicator of successful chromosome doubling. We generated results based on field evaluation and assessment, as well as further confirmation through the microscope. Most researchers did not bother about the mortality of treated plants with toxic chemicals. However, [46] suggested three parameters named survival rate (SR), reproduction rate (RR), and overall success rate (OSR). Basically, to determine the efficacy of chromosome doubling agent and to achieve economically cheaper DH lines, two or three basic parameters are required, i.e., SR, fertility rate, and/or seed setting. Naturally, the haploid plants are weak and prone to several abiotic and biotic stresses [47]. The poor seed setting due to any biotic and abiotic factors, such as disease, temperature/heat, sunlight, irrigation, or rainfall, etc., can impact the efficacy of the chemical agent. Hence, we concentrated on two basic parameters, i.e., the SR and fertility rate, which determine the ADR = SR $\times$ DR. The OSR suggested by [46], and the ADR indicated by [24], are similar but differ in that OSR is based on survival and seed production, whereas ADR is based on survival and fertility. To assess the fertility and efficacy of chemical agents, we further divided the fertility rate into AER, PFR, and CFR. According to [48,49], AER is an effective parameter to determine fertility in haploids, but pollen-less anthers cannot pollinate. However, it can be considered as these plants have a potential probability of being fertile, but this might be due to genotypic or treatment effects that could not double all chromosomes. TFR predicts the success probability or potential of any treatment that might be underrated due to genotypic impact, treatment, and spontaneous chromosomal doubling. The genotypic influence [48] and spontaneous chromosomal doubling [50–52] impact on the efficacy of artificially induced chromosome doubling have been reported.

In our experiments, the CAF–T-based M1 disclosed SR ranges from 80.4% to 99% across all treatments. The PTX-based M1 and M2 showed SR ranges from 73.4% to 98.5% and 56.4% to 90.1%, respectively, whereas colchicine-based M1 and M2 revealed SR varies from 69% to 90.8% and 29% to 82.6%, respectively, across all treatments. After treatment with colchicine, the established seedlings can be only 40 to 80%, in which 10–30% diploids or false positives may be observed in the field [53]. The higher SR of PTX-treated maize seedlings revealed weaker inhibitory action of PTX than colchicine. Colchicine has a toxic effect that depends upon the germplasm's background and the treatment method/process, leading to the loss of many treated seedlings during chromosomal doubling. In conclusion, M1 proved better SR than M2 in our studies and endorsed [24]. The CAF–T disclosed the highest SR and declared it the safest chromosome-doubling chemical for maize haploid seeds compared to PTX and colchicine.

The spontaneous doubling rate was 1.4% in Experiments 1, 2, and 3, which was very low and almost equal to the reported 1.5% [24]. However, germplasm used in Experiments 4, 6, and 7 revealed a spontaneous doubling rate of 3.1% (V1, Exp. 4), 5% (V2, Exp. 4), and 3.3% (Exp. 6 and 7), which was significantly much lower in CK than treated treatments by PTX, colchicine, and CAF–T. Spontaneous fertile plants from diverse maize germplasm groups revealed DR ranging from 0 to 16.7% in Iodent heterotic groups, Lancaster, and intra-pool crosses from Stiff Stalk [48]. The spontaneous fertility depends upon the maize genotypes [48,50,53].

Overall, M2 was more effective than M1 in achieving higher DR in our field experiments because many stem cells were exposed to chemical treatments leading to tassel fertility. M2 also proved better in tetraploids than M1 [54]. However, M1 was observed to be more effective in Experiments 6 and 7 than M3 because a considerable number of seed cells are in the phase of mitosis [24]. The difference in results of field experiments 1 and 2 versus microscopic experiments 6 and 7 is based on two reasons: seeds of two different genotypes were used in these experiments, and, in the microscopic Experiments 6

and 7, root cells were studied under the microscope. The comparative study of M2 versus M3 using antimitotic herbicides [21,46] and colchicine revealed poor success for M3 in field experiments [19,55,56]. Both immersions of roots and crown region exposing the shoot apical meristem cells were recorded as being more successful in achieving a higher OSR [5]. In our Experiments 1 and 3, overall, M1 revealed poor results compared to M2, similar to the studies conducted by [25,46]. Chalyk [45] also reported no fertility using M1. Successful colchicine treatment depends on exposure time, concentration, and contact with meristematic cells [54].

In M1, PTX, colchicine, and CAF–T exhibited an AER range from 4.9% to 16.6%, 4.7% to 32.8%, and 4.5% to 12.4% across all treatments. In M2-based experiments, PTX and colchicine displayed AER range from 5.9% to 28.7% and 7.8% to 35.9%, respectively, across all treatments. In diverse China and US germplasm, the reported AER with significant genetic variance varied from 9.8 to 89.8%. [51].

In the LS field efficacy-validation experiment, PTX and colchicine-treated maize seedlings showed an ADR of 37.7% and 19.2%, respectively, in V1, 20% and 14.5%, respectively, in V2, whereas [24] reported 20.64% of ADR by colchicine. The DR and AER are inversely proportional to each other in both genotypes because if DR is higher, AER may be reduced. However, if AER is higher, DR may be lower due to sectoral diploidization. According to [53], colchicine treatment may or may not ensure that all cells' chromosomes are doubled, which is known as sectoral diploidization. The sectoral diploidization effect varies from genotype to genotype and colchicine application.

Haploid plant tassels cover a wide range of fertility, from one or two anthers shedding pollen to the complete tassel [22]. Therefore, we divided the fertility or DR into two types, i.e., PFR and CFR, to assess the doubling efficacy of the chemical agents. In LS, PTX showed PFR and CFR 29.4% and 24.5%, respectively; in V1, 16.6% and 11.2%, respectively, in V2, whereas colchicine showed PFR and CFR 17.1% and 20.7%, respectively, in V1, and 13.9% and 16.5%, respectively, in V2. Hence, PTX showed more CFR in V1 as compared to colchicine. There is a probability that 0–40% of colchicine-treated plants can have both silks and the ability to produce pollen (DR) for successful pollination among the remaining true haploids [53]. In Experiments 1 and 3, PTX and colchicine based M2 showed a maximum of DR 54.3% and 43.8%, respectively. In LS, PTX and colchicine showed DR 53.9% and 37.8% in V1, respectively, whereas PTX and colchicine-treated seedlings showed DR 27.8% and 30.4% in V2, respectively. Next, [24] calculated the highest doubling rate of 29.7% using colchicine, whereas it was reported in the same publication the colchicine chromosome doubling rates by studies of Wen's and Liu's 48.35% and 23.0%, respectively.

Colchicine, PTX, and CAF–T-treated M1 revealed a maximum DR of 21.2%, 29.6%, (0.08% for colchicine concentration for 8 h), and 20.9%, respectively, but colchicine induced 18% of DR at 0.06% colchicine for 12 h treatment [57–60]. CAF–T did not show doubling efficiency in treatments of 4 g and 6 g, as well as for 16 h and all 24 h treatments, which might reveal the impact of higher concentrations of CAF–T and longer time exposure. In addition, Ref. [61] also suggested caffeine as a fertility-inducing agent by root dipping of wheat with best-achieved results at the application of 3 g/L for 24 h.

PTX and CAF–T produced comparatively higher seed counts than colchicine. According to [53], only 30–50% of colchicine-treated haploid plants produce seeds. Another fact is that plants treated with chromosome-doubling chemicals like colchicine produced fewer seeds than spontaneously doubled haploid plants [62]. Toxic chemicals like colchicine further stress weak haploid plants [22]. Hence, we can conclude that bio-safe or less toxic chemicals are suitable for a higher seed setting as in our experiment PTX, and CAF–T produced more seeds per D1 ear. D1 ears' ability to produce seeds varies widely. However, on average, four seeds per D1 ear by colchicine have been reported [48]. At CIMMYT, according to unpublished data, 40–60% of the D0 plants treated by colchicine produced more than 25 seed grains depending on the season, whereas the rest exhibited poor seed setting [22].

The colchicine effects on the morphology and physiology of treated maize haploid seeds or seedlings have not been reported. However, only SR of both maize haploid seeds and seedlings have been widely reported. In addition to SR, PTX unveiled less impact on seeds germination percentages as compared to colchicine. Severe effects of different colchicine concentrations were also reported in *Phlox drummondi* [63]. In *Prunella vulgaris* and *Dendrocalumus brandisii*, a higher colchicine concentration with longer processing times decreased the germination of seeds [64,65]. Colchicine-treated okra seeds' germination was calculated 79.3%, which is significantly lower than untreated plants at 94.8% [66]. We found PTX treatment to be better than colchicine for plant weight, the number of silks/ear, and plant and ear height in both the genotypes. Then, Ref. [67] studied the impact of colchicine on plant height in Jimsonweed (*Datura stramonium* L.), which decreased linearly by increasing incubation time and the concentration of colchicine. A decrease in plant height in *Abelmoschus esculentusalso* (okra) and *Phlox drummondi* was measured due to colchicine effects [63,66]. Colchicine disclosed more adverse effects on the growth of seedlings than on root growth and seed germination in *Dendrocalumus brandisii*. Comparatively, colchicine revealed more negative effects regarding the degradation of sucrose than starch during the germination of seeds and seedling growth [65].

Chlorophyll is the most essential pigment for photosynthesis because it largely determines the process's capacity and plant growth. The Chl_T directly impacts the plants' photosynthesis capacity [68]. The carotenoids have three considerable roles in photosynthesis: (a) A pigment to harvest accessory light, (b) Wavelengths range extension to facilitate light for the photosynthetic process, (c) Protection of achlorophyllous pigments from photo-destructive reaction [69]. Damage in chlorophyll and carotenoid contents decreases the efficacy of plants. Therefore, PTX damage to chlorophyll and carotenoids were significantly lesser than toxic colchicine in our study. Single-Stranded Oligonucleotides (SSONs) serve as probes that are a popular method for chromosome detection, painting, and identification [70] because of their versatility, high resolution, sensitivity, and cost effectivity [71]. SSON probes are used after labelling with detectable stable signals, and in our experiments, these signals confirmed the efficacy of PTX and CAF–T regarding the doubling of chromosomes.

4. Materials and Methods

4.1. Germplasm, Chemicals, Instruments and Experimental Locations

Maize haploids seeds were produced at the College of Agriculture, Guizhou University (26°25′21″ N, 106°40′09″ E), China. Haploid seeds were separated visually from diploids by an *R1–nj* anthocyanin marker expression on kernels. The required sample sizes were obtained from 16,400 putative haploid seeds for all experiments in this embodiment. The details of genotypes and experimental locations are demonstrated in Table S7.

All treatments were replicated thrice in Experiments 1, 2, and 3. Each replication has an average of 70 putative haploids—some treatments have more than 70 seeds/seedlings to ensure equal stand with lower-dosed treatments because higher doses affect plant stands. All three replications of each treatment were blocked in a randomized complete block design (RCBD), including Experiment 4. PTX was purchased from Peptide Biotechnology Co. Ltd., Xi'an, China, CAF–T from BANNY DEER at Wuhan East–Lake High Tech Region, China, and colchicine (98%) from Shanghai Macklin Biochemical Co. Ltd., Shanghai Chemical Industry Park, Shanghai, China. MAPADA P1 UV–Visible Spectrophotometer was purchased from Shanghai Mapada Instruments Co., Ltd. Shanghai, China.

4.2. Treatments & Concentrations

4.2.1. Experiment 1

The putative maize haploid seeds and seedlings (including roots and stem) were treated with PTX at 28 °C for 8, 16, and 24 h. The stock solution concentrations of PTX were used for chromosomal doubling, including 100 μM (0.085 g/L), 200 μM (0.170 g/L), 400 μM (0.341 g/L), and 800 μM (0.682 g/L). PTX is highly lipophilic and has a poor water

solubility of less than 0.1 μg/mL, seriously affecting its bioavailability [72]. The method adopted to make PTX soluble was suggested by [73], with a surfactant tween-80 at the ratio of 1:3 diluted in required volume in addition to 2% DMSO (same for each treatment) as a penetrating agent; this was adjusted by distilled deionized water (ddH_2O).

4.2.2. Experiment 2

The putative maize haploid seeds were soaked in CAF–T at 28 °C for 8, 16, and 24 h. The concentrations of CAF–T were used for chromosomal doubling, including caffeine at 1 g/L and taurine at 6 g/L, caffeine at 2 g/L and taurine at 12 g/L, caffeine at 4 g/L and taurine at 24 g/L, and caffeine at 6 g/L and taurine at 36 g/L, with 0.1% DMSO, which was adjusted by ddH_2O.

4.2.3. Experiment 3

The putative maize haploid seeds and seedlings (including roots and stem) were treated with colchicine at 28 °C for 8, 16, and 24 h. The different concentrations of colchicine were used for chromosome doubling, including 0.5 mM (0.2 g/L), 1 mM (0.4 g/L), 1.5 mM (0.6 g/L), and 2 mM (0.8 g/L) [24] with 0.1% of DMSO, which was adjusted by ddH_2O.

4.2.4. Experiments 4 and 5

LS field efficacy-based comparative study was designed to further authenticate the best treatment (concluded in Experiment 1) of the new chemical, i.e., PTX (400μM and 2% of DMSO and 1.02 mL/L of tween-80 for 16 h). The morphological and physiological measurements were also taken. The putative maize haploid seedlings from two populations, i.e., V1 and V2, were used to compare PTX with colchicine treatment (2 mM and 0.1% of DMSO for 8 h treatment) as the best treatment in Experiment 3 regarding TFR and also suggested by [5] and control (ddH_2O).

4.2.5. Experiments 6 and 7

To verify the doubling of chromosomes under the microscope, the same concentrations of PTX and CAF–T (see Section 4.2.1) were used for seed soaking and root immersion for 8, 16, 24, and 48 h.

4.3. Methods

4.3.1. Seed-Soaking (Method 1/M1)

The putative maize haploid seeds were first soaked in ddH_2O for 6 h at 28 °C, and then soaked in three chemical agents, i.e., PTX, CAF–T, and colchicine for 8, 16, and 24 h at 28 °C. After chemical treatments, seeds were rinsed with ddH_2O for 30 min and shifted on a wet germination paper in an incubator with a maintained temperature of 28 °C [24]. Afterwards, the seedlings of 1–1.5 cm in length were transferred to trays.

4.3.2. Seedling-Immersion (Method 2/M2)

This protocol is a "standard seedling-immersion method" defined in the 1990s [19,43]. The putative maize haploid seeds were soaked in ddH_2O at 28 °C for 6 h and shifted on a moist germination paper for 72–96 h in an incubator temperature maintained at 28 °C for germination. When the roots achieved their length up to 1 cm, a 28 °C temperature decreased to 24 °C in demand to make shoots grow. When shoots' sizes increased up to 2 cm, the coleoptiles' top 2–3 mm [24] and roots above 2 cm [53] were cut and dipped in distilled water for 30 min for stress recovery. Then, these seedlings were immersed in PTX and colchicine for 8, 16, and 24 h at 28 °C. Later, they were rinsed well, soaked in ddH_2O for 30 min, and then shifted into trays for growing in the greenhouse.

4.3.3. Root-Immersion (Method 3/M3)

We followed the protocol of root immersion as suggested by [24]. The roots were immersed in the agents (PTX and CAF–T) for 8, 16, 24, and 48 h. Thirty seeds and seedlings

(roots only) were treated to confirm the efficacy of each treatment under the microscope using Knob-2 probe.

4.3.4. Acclimatization of Seedlings in Greenhouse and Transplanting in Field

After respective treatments, seedlings were transplanted into trays filled with suitable soil composition, sand, peat, and compost with pH 6 to 6.5. Treated seedlings were grown in a greenhouse for 15 h of photoperiod during the day and maintained at about 26–28 °C and 24–28 °C at night. Nutrients and water were provided on a regular need basis. Gradually transferring the seedlings from lower light intensity to higher light intensity and higher-to-lower humidity helped acclimate the treated seedlings [9].

When survived seedlings achieved V2 leaf stage [5] more or less after 15 days (depending upon growth), they were transplanted [42,53] in the well-ploughed and prepared field. Transferring treated seedlings from the greenhouse to the field at V2 was a cautious and delicate process. Survived seedlings' frequency mainly depends upon the treatment with chromosomal doubling agents and on growing condition, medium, genetic effect, technical equipment, and human skills [9]. The chimeric and weak maize seedlings required much-needed special care and attention, but these seedlings were not found to be useful. As a rule, these seedlings degenerated and did not survive in the greenhouse and field. Losses of seedlings lead due to hardening of seedlings [9]. The survived seedlings were grown to harvest and provided the best agronomy as recommended by [53]. Haploid plants are commonly weak [47], and chemical treatments enforce further stress. Therefore, they were handled with extreme care during and after chemical treatments and transplanting to harvesting. Pollen-producing plants were self-pollinated, and harvested seeds were visually observed at plant maturity for the marker (should have no marker color) to identify them as double haploids.

4.3.5. Experiments 6 and 7

The knob-2 probe used in our experiment was developed from tandem repeats of plasmid clones. The 180-base pair (bp) knob sequence was obtained by dividing the 180 bp knob sequence into three non-overlapping parts, each containing 59 nucleotides, and then modifying the 5′ ends with fluorescein amidites (FAM). The knob-2 probe sequence and maize chromosomes were prepared according to the protocol provided by [70]. This is the knob-2 probe sequence:

GAAGGCTAACACCTACGGATTTTTGACCAAGAAATGGTCTCCACCAGAAATCC AAAAAT

The fluorescence in situ hybridization (FISH) procedure was followed as described by [74].

4.4. Assessment Methodology and Data Collection Procedure

4.4.1. Experiments 1–4

In Experiments 1–4, the following parameters were documented: (a) Number of seeds or seedlings treated; (b) Germination percentage; (c) Number of survived Do plants at pollination; (d) Count of Do plants showed anthers only; (e) Sum of Do plants that exhibited partial fertility; (f) Number of plants presented complete fertility; (g) Number of false/wrong diploid plants; (h) Number of haploid plants. The following calculations were made in Experiments 1–4 and expressed in percentages (%).

1. Survival rate (SR) = count of survived plants at pollination/number of seedlings treated × 100 [46];
2. Doubling rate (DR) = sum of plants showed partial and/or complete fertility/count of haploid survived plants at pollination × 100;
3. Actual doubling rate (ADR) = SR × DR/100 [24];
4. Anthers emergence rate (AER) = number of plants showed anthers only/number of haploid plants survived at pollination × 100 [48,49];

5. Partial fertility rate (PFR) = number of plants that exhibited partial fertility/number of haploid plants that survived at pollination × 100;
6. Complete fertility rate (CFR) = number of plants presented complete fertility/number of haploid plants that survived at pollination ×100;
7. Total fertility rate (TFR) = AER + PFR + CFR.

In addition, seed setting data were also collected in Experiments 1–3 and expressed as follows; (a) Percentage of DH lines with 1–5 seeds; (b) Percentage of DH lines containing 6–25 seeds; Percentage of DH lines containing > 25 seeds [5]; Reproduction rate (RR); Overall success rate (OSR) [46].

1. Percentage of DH lines with 1–5 seeds = number of D_0 plants/D_1 ears produced 1–5 seeds/number of haploids survived plants at pollination × 100;
2. Percentage of DH lines containing 6–25 seeds = number of D_0 plants/D_1 ears produced 6–25 seeds/number of haploid survived plants at pollination × 100;
3. Percentage of DH lines containing > 25 seeds = number of D_0 plants/D_1 ears produced > 25 seeds/number of haploid survived plants at pollination × 100;
4. RR = number of D_0 plants produced seeds/number of D_0 survived haploid plants at pollination × 100;
5. OSR = number of D_0 plants produced seeds/number of putative seeds/seedlings treated with chemical agent × 100.

4.4.2. Experiment 5

Morphological Studies

1. Plant and Ear Height, Plant Weight, and Number of Silks/Ear

Five plants per repeat were randomly selected (evading three plants at each plot beginning and end to exclude any impact) from each treatment and genotype. Plant and ear height were measured randomly from haploid and diploid (false positive) plants. Plant height was measured from the plant base or soil to the last collar leaf. Ear height measurements were obtained from the plant's base (ground/soil) to the node-bearing upper ear. Plant and ear height data were measured from three replications and analyzed as means. Cut the plants from the ground surface and weigh them on a scale in grams (g). The mean plant weight and number of silks/ears were calculated from four replications.

Physiological Studies

1. Measurement of Photosynthetic Pigments (Chl_a, Chl_b, Chl_T, and C_x)

Maize leaf samples were obtained randomly at V2 leaf stage in each sample area and sealed in aluminum foil to measure in the laboratory via MAPADA P1 UV–Visible Spectrophotometer. Five random plants were selected from four different replicated plots from each treatment and genotype and analyzed as mean Chl_a, Chl_b, Chl_T, and C_x. With the aid of liquid nitrogen, 0.2 g of fresh leaf sample were ground into a powder and mixed into 1.5 mL of 95% ethanol [75,76]. The solution was mixed well and placed in darkness for 10 min. Afterwards, the solution was centrifuged at 8000–10,000 rpm for 8–10 min and the supernatant was separated at 1.2 mL in another glass tube for the spectrophotometer. Then, 1.2 mL 95% of ethanol were employed as a blank. The collected supernatant was measured for absorbance at 665, 649, and 470 nm in the spectrophotometer. The concentrations of Chl_a, Chl_b, Chl_T, and C_x were calculated in mg. g^{-1} using the following equations:

$$Chl_a = (13.95 \times A_{665} - 6.88 \times A_{649}) \times 1.5 \times 10^{-3} \div 0.18,$$

$$Chl_b = (24.96 \times A_{649} - 7.32 \times A_{665}) \times 1.5 \times 10^{-3} \div 0.18,$$

$$Chl_T = Chl_a + Chl_b,$$

$$C_x = (1000 \times A_{470} - 2.05 \times Chl_a - 114.8\, Chl_b) \div 245 \times 1.5 \times 10^{-3} \div 0.18,$$

The leaf samples were obtained in such a way as to avoid unbiased results; 20 leaf samples were obtained from four replicated plots.

2. Germination Percentages

The impact of PTX and CAF–T versus colchicine on seed germination was assessed and calculated as the count of germinated seedlings/total number of treated seeds × 100.

4.4.3. Experiments 6 and 7

For each maize seed or root treated with PTX or CAF–T, a slide of root cells was prepared, and 300 cells/slide were observed for maize chromosomes. Knob-2 probe signals counted under Olympus BX53 fluorescence microscope. The images were obtained with Olympus DP80 camera. The following data parameters were observed and recorded: (a) Total number of maize haploid seeds/roots were treated; (b) The number of treated seeds or roots doubled with 5–10% cells. The percentage of treated seeds/roots doubled was calculated as the sum of treated seeds or roots with cells doubled up to 5–10% and divided by the number of seeds or roots treated × 100.

4.5. Statistical Analysis and Graphics Improvement

In Experiments 1 to 5, the results of all three or four replications ($n = 3$ or $n = 4$) were expressed in graphs based on standard error (SE). The mean comparison was made by the "Tukey's Honestly Significant Difference (HSD)" test at $p \leq 0.05$, following the Analysis of Variance (ANOVA). All replicated data were analyzed via IBM SPSS Statistics 21.0 software. The graphics software known as "Origin" was applied to improve graphics (Version 2022, Origin Lab. Corporation, Northampton, MA, USA).

5. Conclusions

PTX produced better results than the widely used toxic traditional chemical agent (colchicine) to improve chromosome-doubling success rates and survivability. The ADR for PTX and colchicine in maize haploid seedlings were 42.1% (400 M, 16 h treatment) and 31.9% (0.5 mM, 24 h treatment), respectively. The ADR induced in maize haploid seeds by colchicine, CAF–T, and PTX were 26% (2.0 mM, 8 h treatment), 20% (caffeine 2 g/L + taurine 12 g/L, 16 h), and 19.9% (100 μM for 24 h treatment), respectively. PTX and CAF–T offer occupational health, operational ease, and disposal. PTX guarantees a cost-effective and better rate of doubling haploid plants at lower concentrations. Optimizing chromosomal doubling protocols based on less toxic or bio-safe chemicals like PTX and CAF–T will further enhance the overall efficiency of DH production. The morphological and physiological effects produced in haploid plants by PTX were significantly lower than toxic colchicine. PTX will improve the quantity of D0 lines seeds, which may exterminate the requirement to increase D1 lines seeds, thereby saving time and cost of the additional cycle to multiply seeds.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241914659/s1.

Author Contributions: The work presented here was a collaborative study of all the authors. S.A. prepared and executed the experiments and wrote and kept working on the revised version of the manuscript. Q.A. and M.W. helped with the data analysis and graphics improvement. G.M. helped with manuscript modification. Z.M. and Y.Y. designed the project, revised the manuscript, and obtained the funds to support this project. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (31571580), the Fundamental Research Funds for the Central Universities (KYTZ201402), and the Outstanding Scientific Innovation Team Program for Jiangsu Universities (2015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are already discussed in the main manuscript and Supplementary Files.

Acknowledgments: We are grateful to Guangxiao Bai (College of Agriculture, Guizhou University, China), who provided the haploid seeds for the experiments of this study. Thanks to Hong Ren (Guizhou Institute of Upland Food Crops, Guizhou Academy of Agricultural Sciences, China), helped with manuscript modification.

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

References

1. Hallauer, A.R.; Carena, M.J.; Miranda Filho, J.B.D. *Quantitative Genetics in Maize Breeding*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 6.
2. Jacquier, N.M.A.; Gilles, L.M.; Martinant, J.-P.; Rogowsky, P.M.; Widiez, T. Maize in planta haploid inducer lines: A cornerstone for doubled haploid technology. In *Doubled Haploid Technol.*; Humana: New York, NY, USA, 2021; pp. 25–48.
3. Chaikam, V.; Prasanna, B.M. Doubled Haploid Technology for Rapid and Efficient Maize Breeding. In *Accelerated Plant Breeding, Volume 1: Cereal Crops*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 257–292.
4. Yan, G.; Liu, H.; Wang, H.; Lu, Z.; Wang, Y.; Mullan, D.; Hamblin, J.; Liu, C. Accelerated generation of selfed pure line plants for gene identification and crop breeding. *Front. Plant Sci.* **2017**, *8*, 1786. [CrossRef]
5. Chaikam, V.; Gowda, M.; Martinez, L.; Ochieng, J.; Omar, H.A.; Prasanna, B.M. Improving the efficiency of colchicine-based chromosomal doubling of maize haploids. *Plants* **2020**, *9*, 459. [CrossRef]
6. Böhm, J.; Schipprack, W.; Utz, H.F.; Melchinger, A.E. Tapping the genetic diversity of landraces in allogamous crops with doubled haploid lines: A case study from European flint maize. *Theor. Appl. Genet.* **2017**, *130*, 861–873. [CrossRef]
7. Brauner, P.C.; Schipprack, W.; Utz, H.F.; Bauer, E.; Mayer, M.; Schön, C.-C.; Melchinger, A.E. Testcross performance of doubled haploid lines from European flint maize landraces is promising for broadening the genetic base of elite germplasm. *Theor. Appl. Genet.* **2019**, *132*, 1897–1908. [CrossRef]
8. Melchinger, A.E.; Böhm, J.; Utz, H.F.; Müller, J.; Munder, S.; Mauch, F.J. High-throughput precision phenotyping of the oil content of single seeds of various oilseed crops. *Crop Sci.* **2018**, *58*, 670–678. [CrossRef]
9. Kanbar, O.Z.; Lantos, C.; Pauk, J. In vitro anther culture as efficiently applied technique for doubled haploid production of wheat (*Triticum aestivum* L.). *Ratar. I Povrt.* **2021**, *58*, 31–45. [CrossRef]
10. Ravi, M.; Chan, S.W.L. Haploid plants produced by centromere-mediated genome elimination. *Nature* **2010**, *464*, 615–618. [CrossRef]
11. Sorrells, M.E.; Gustafson, J.P.; Somers, D.; Chao, S.; Benscher, D.; Guedira-Brown, G.; Huttner, E.; Kilian, A.; McGuire, P.E.; Ross, K. Reconstruction of the Synthetic W7984× Opata M85 wheat reference population. *Genome* **2011**, *54*, 875–882. [CrossRef]
12. Murovec, J.; Bohanec, B. Haploids and doubled haploids in plant breeding. In *Plant Breeding*; Abdurakhmonov, I., Ed.; InTech Europe: Rijeka, Croatia, 2012; pp. 87–106. ISBN 978-953-307-932-5.
13. Hao, M.; Chen, J.; Zhang, L.; Luo, J.; Yuan, Z.; Yan, Z.; Zhang, B.; Chen, W.; Wei, Y.; Zhang, H. The genetic study utility of a hexaploid wheat DH population with non-recombinant A-and B-genomes. *SpringerPlus* **2013**, *2*, 131. [CrossRef]
14. Shi, Y.G.; Lian, Y.; Shi, H.W.; Wang, S.G.; Fan, H.; Sun, D.Z.; Jing, R.L. Dynamic analysis of QTLs for green leaf area duration and green leaf number of main stem in wheat. *Cereal Res. Commun.* **2019**, *47*, 250–263. [CrossRef]
15. Dunwell, J.M. Haploids in flowering plants: Origins and exploitation. *Plant Biotechnol. J.* **2010**, *8*, 377–424. [CrossRef]
16. Chaikam, V.; Mahuku, G.; Prasanna, B.M. *Chromosome Doubling of Maternal Haploids*; CIMMYT: Texcoco, Mexico, 2012.
17. Dhooghe, E.; Van Laere, K.; Eeckhaut, T.; Leus, L.; Van Huylenbroeck, J. Mitotic chromosome doubling of plant tissues in vitro. *Plant Cell Tissue Organ Cult. (PCTOC)* **2011**, *104*, 359–373. [CrossRef]
18. Chase, S.S. Monoploids and monoploid-derivatives of maize (*Zea mays* L.). *Bot. Rev.* **1969**, *35*, 117–168. [CrossRef]
19. Deimling, S.; Röber, F.K.; Geiger, H.H. Methodology and genetics of in vivo haploid induction in maize. *Vortr. Pflanzenzüchtg* **1997**, *38*, 203–224.
20. Wan, Y.; Duncan, D.R.; Rayburn, A.L.; Petolino, J.F.; Widholm, J.M. The use of antimicrotubule herbicides for the production of doubled haploid plants from anther-derived maize callus. *Theor. Appl. Genet.* **1991**, *81*, 205–211. [CrossRef] [PubMed]
21. Melchinger, A.E.; Brauner, P.C.; Böhm, J.; Schipprack, W. In vivo haploid induction in maize: Comparison of different testing regimes for measuring haploid induction rates. *Crop Sci.* **2016**, *56*, 1127–1135. [CrossRef]
22. Chaikam, V.; Molenaar, W.; Melchinger, A.E.; Boddupalli, P.M. Doubled haploid technology for line development in maize: Technical advances and prospects. *Theor. Appl. Genet.* **2019**, *132*, 3227–3243. [CrossRef]
23. Duncan, D.R.; Widholm, J.M. Differential response to potassium permanganate of regenerable and of non-regenerable tissue cell walls from maize callus cultures. *Plant Sci.* **1989**, *61*, 91–103. [CrossRef]
24. Ren, X. Doubling effect of anti-microtubule herbicides on the maize haploid. *Emir. J. Food Agric.* **2018**, *30*, 903–908. [CrossRef]
25. Molenaar, W.S.; Schipprack, W.; Melchinger, A.E. Nitrous Oxide-Induced Chromosome Doubling of Maize Haploids. *Crop Sci.* **2018**, *58*, 650–659. [CrossRef]
26. Kato, A.; Geiger, H.H. Chromosome doubling of haploid maize seedlings using nitrous oxide gas at the flower primordial stage. *Plant Breed.* **2002**, *121*, 370–377. [CrossRef]

27. Morejohn, L.C.; Bureau, T.E.; Tocchi, L.P.; Fosket, D.E. Resistance of Rosa microtubule polymerization to colchicine results from a low-affinity interaction of colchicine and tubulin. *Planta* **1987**, *170*, 230–241. [CrossRef] [PubMed]
28. Morejohn, L.C.; Bureau, T.E.; Tocchi, L.P.; Fosket, D.E. Tubulins from different higher plant species are immunologically nonidentical and bind colchicine differentially. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 1440–1444. [CrossRef] [PubMed]
29. Khanna, C.; Rosenberg, M.; Vail, D.M. A review of paclitaxel and novel formulations including those suitable for use in dogs. *J. Vet. Intern. Med.* **2015**, *29*, 1006–1012. [CrossRef] [PubMed]
30. Fong, K.-W.; Leung, J.W.-C.; Li, Y.; Wang, W.; Feng, L.; Ma, W.; Liu, D.; Songyang, Z.; Chen, J. MTR120/KIAA1383, a novel microtubule-associated protein, promotes microtubule stability and ensures cytokinesis. *J. Cell Sci.* **2013**, *126*, 825–837. [CrossRef]
31. Snyder, J.P.; Nettles, J.H.; Cornett, B.; Downing, K.H.; Nogales, E. The binding conformation of Taxol in β-tubulin: A model based on electron crystallographic density. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5312–5316. [CrossRef]
32. Yan-Hua, Y.; Jia-Wang, M.A.O.; Xiao-Li, T.A.N. Research progress on the source, production, and anti-cancer mechanisms of paclitaxel. *Chin. J. Nat. Med.* **2020**, *18*, 890–897.
33. Lourenco, R.; Camilo, M.E. Taurine: A conditionally essential amino acid in humans? An overview in health and disease. *Nutr Hosp* **2002**, *17*, 262–270.
34. Mates, J.M.; Segura, J.A.; Alonso, F.J.; Marquez, J. Sulphur-containing non enzymatic antioxidants: Therapeutic tools against cancer. *Front. Biosci.-Sch.* **2012**, *4*, 722–748. [CrossRef]
35. Kilb, W.; Fukuda, A. Taurine as an essential neuromodulator during perinatal cortical development. *Front. Cell. Neurosci.* **2017**, *11*, 328. [CrossRef]
36. Morales-Borge, R.H.; González, M.J.; Gupta, R.C.; Ayeotan, O. Taurine as Anticancer and Antiviral: Case Report and Prospective Update. *Cancer Case Rep.* **2020**, *1*, 1–14. [CrossRef]
37. He, F.; Ma, N.; Midorikawa, K.; Hiraku, Y.; Oikawa, S.; Mo, Y.; Zhang, Z.; Takeuchi, K.; Murata, M. Anti-cancer mechanisms of taurine in human nasopharyngeal carcinoma cells. In *Taurine 11*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 533–541.
38. Yousuf, M.; Shamsi, A.; Mohammad, T.; Azum, N.; Alfaifi, S.Y.M.; Asiri, A.M.; Mohamed Elasbali, A.; Islam, A.; Hassan, M.I.; Haque, Q.M.R. Inhibiting Cyclin-Dependent Kinase 6 by Taurine: Implications in Anticancer Therapeutics. *ACS Omega* **2022**, *7*, 25844–25852. [CrossRef] [PubMed]
39. Zhou, L.; Lu, R.; Huang, C.; Lin, D. Taurine protects C2C12 myoblasts from impaired cell proliferation and myotube differentiation under cisplatin-induced ROS exposure. *Front. Mol. Biosci.* **2021**, *8*, 685362. [CrossRef]
40. Broughton, S.; Castello, M.; Liu, L.; Killen, J.; Hepworth, A.; O'Leary, R. The effect of caffeine and trifluralin on chromosome doubling in wheat anther culture. *Plants* **2020**, *9*, 105. [CrossRef]
41. Yasuhara, H. Caffeine inhibits callose deposition in the cell plate and the depolymerization of microtubules in the central region of the phragmoplast. *Plant Cell Physiol.* **2005**, *46*, 1083–1092. [CrossRef] [PubMed]
42. Prigge, V.; Melchinger, A.E. Production of haploids and doubled haploids in maize. In *Plant cell culture protocols*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 161–172.
43. Gayen, P.; Madan, J.K.; Kumar, R.; Sarkar, K.R. Chromosome doubling in haploids through colchicine. *Maize Genet. Coop. Newsl.* **1994**, *68*, 65.
44. Eder, J.; Chalyk, S. In vivo haploid induction in maize. *Theor. Appl. Genet.* **2002**, *104*, 703–708. [CrossRef] [PubMed]
45. Chalyk, S.T. Obtaining fertile pollen in maize maternal haploids. *Maize Genet. Coop. Newsl.* **2000**, *74*, 17–18.
46. Melchinger, A.E.; Molenaar, W.S.; Mirdita, V.; Schipprack, W. Colchicine alternatives for chromosome doubling in maize haploids for doubled-haploid production. *Crop Sci.* **2016**, *56*, 559–569. [CrossRef]
47. Mahuku, G. *Putative DH Seedlings: From the Lab to the Field*; CIMMYT: Texcoco, Mexico, 2012.
48. Kleiber, D.; Prigge, V.; Melchinger, A.E.; Burkard, F.; San Vicente, F.; Palomino, G.; Gordillo, G.A. Haploid fertility in temperate and tropical maize germplasm. *Crop Sci.* **2012**, *52*, 623–630. [CrossRef]
49. Ren, J.; Wu, P.; Tian, X.; Lübberstedt, T.; Chen, S. QTL mapping for haploid male fertility by a segregation distortion method and fine mapping of a key QTL qhmf4 in maize. *Theor. Appl. Genet.* **2017**, *130*, 1349–1359. [CrossRef]
50. Chaikam, V.; Gowda, M.; Nair, S.K.; Melchinger, A.E.; Boddupalli, P.M. Genome-wide association study to identify genomic regions influencing spontaneous fertility in maize haploids. *Euphytica* **2019**, *215*, 1–14. [CrossRef] [PubMed]
51. Wu, P.; Ren, J.; Tian, X.; Lübberstedt, T.; Li, W.; Li, G.; Li, X.; Chen, S. New insights into the genetics of haploid male fertility in maize. *Crop Sci.* **2017**, *57*, 637–647. [CrossRef]
52. Ma, H.; Li, G.; Würschum, T.; Zhang, Y.; Zheng, D.; Yang, X.; Li, J.; Liu, W.; Yan, J.; Chen, S. Genome-wide association study of haploid male fertility in maize (*Zea mays* L.). *Front. Plant Sci.* **2018**, *9*, 974. [CrossRef] [PubMed]
53. Prasanna, B.M.; Chaikam, V.; Mahuku, G. *Doubled Haploid Technology in Maize Breeding: Theory and Practice*; CIMMYT: Texcoco, Mexico, 2012.
54. Ghotbi Ravandi, E.; Rezanejad, F.; Zolala, J.; Dehghan, E. The effects of chromosome-doubling on selected morphological and phytochemical characteristics of *Cichorium intybus* L. *J. Hortic. Sci. Biotechnol.* **2013**, *88*, 701–709. [CrossRef]
55. Jensen, C. Chromosome Doubling Techniques in Haploids. In *Haploids in Higher Plants: Advances and Potential*; The University of Guelph: Guelph, ON, Canada, 1974; pp. 151–190.
56. Inagaki, M.N. Doubled haploid production in wheat through wide hybridization. In *Doubled Haploid Production in Crop Plants*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 53–58.

57. Davarnejad, R.; Moraveji, M.K.; Havaie, M. Integral technique for evaluation and optimization of Ni (II) ions adsorption onto regenerated cellulose using response surface methodology. *Arab. J. Chem.* **2018**, *11*, 370–379. [CrossRef]
58. Jiang, M.; Wu, H.; Li, Z.; Ji, D.; Li, W.; Liu, Y.; Yuan, D.; Wang, B.; Zhang, Z. Highly selective photoelectrochemical conversion of carbon dioxide to formic acid. *ACS Sustain. Chem. Eng.* **2018**, *6*, 82–87. [CrossRef]
59. Khaleel, C.; Tabanca, N.; Buchbauer, G. α-Terpineol, a natural monoterpene: A review of its biological properties. *Open Chem.* **2018**, *16*, 349–361. [CrossRef]
60. Sankarganesh, M.; Rajesh, J.; Kumar, G.G.V.; Vadivel, M.; Mitu, L.; Kumar, R.S.; Raja, J.D. Synthesis, spectral characterization, theoretical, antimicrobial, DNA interaction and in vitro anticancer studies of Cu(II) and Zn(II) complexes with pyrimidine-morpholine based Schiff base ligand. *J. Saudi Chem. Soc.* **2018**, *22*, 416–426. [CrossRef]
61. Thomas, J.; Chen, Q.; Howes, N. Chromosome doubling of haploids of common wheat with caffeine. *Genome* **1997**, *40*, 552–558. [CrossRef]
62. Chaudhary, H.K.; Badiyala, A.; Jamwal, N.S. New frontiers in doubled haploidy breeding in wheat. *Agric. Res. J.* **2015**, *52*, 1–12. [CrossRef]
63. Tiwari, A.K.; Mishra, S.K. Effect of colchicine on mitotic polyploidization and morphological characteristics of Phlox drummondi. *Afr. J. Biotechnol.* **2012**, *11*, 9336–9342. [CrossRef]
64. Kwon, S.-J.; Roy, S.K.; Cho, K.-Y.; Moon, Y.-J.; Woo, S.-H.; Kim, H.-H. Effect of Colchicine on the Induction of Prunella vulgaris for. albiflora Nakai. *Korean J. Crop Sci.* **2015**, *60*, 107–113. [CrossRef]
65. Lv, Z.; Zhu, F.; Jin, D.; Wu, Y.; Wang, S. Seed Germination and Seedling Growth of Dendrocalumus brandisii in vitro, and the Inhibitory Mechanism of Colchicine. *Front. Plant Sci.* **2021**, *12*, 784581. [CrossRef] [PubMed]
66. Megbo, B.C. The physiological effects of colchicine in Okra, Hibiscus esculentus L, plant growth and development. *Int. J. Sci. Eng. Res.* **2010**, *1*, 29–33.
67. Amiri, S.; Kazemitabaar, S.K.; Ranjbar, G.; Azadbakht, M. The effect of trifluralin and colchicine treatments on morphological characteristics of jimsonweed (*Datura stramonium* L.). *Trakia J. Sci.* **2010**, *8*, 47–61.
68. Li, Y.; He, N.; Hou, J.; Xu, L.; Liu, C.; Zhang, J.; Wang, Q.; Zhang, X.; Wu, X. Factors influencing leaf chlorophyll content in natural forests at the biome scale. *Front. Ecol. Evol.* **2018**, *6*, 64. [CrossRef]
69. Cogdell, R.J. Carotenoids in photosynthesis. *Philos. Trans. R. Soc. London. B Biol. Sci.* **1978**, *284*, 569–579. [CrossRef]
70. Zhu, M.; Du, P.; Zhuang, L.; Chu, C.; Zhao, H.; Qi, Z. A simple and efficient non-denaturing FISH method for maize chromosome differentiation using single-strand oligonucleotide probes. *Genome* **2017**, *60*, 657–664. [CrossRef]
71. Han, Y.; Zhang, T.; Thammapichai, P.; Weng, Y.; Jiang, J. Chromosome-specific painting in Cucumis species using bulked oligonucleotides. *Genetics* **2015**, *200*, 771–779. [CrossRef]
72. Surapaneni, M.S.; Das, S.K.; Das, N.G. Designing Paclitaxel drug delivery systems aimed at improved patient outcomes: Current status and challenges. *Int. Sch. Res. Not.* **2012**, *2012*, 623139. [CrossRef]
73. Shanmugam, S.; Im, H.T.; Sohn, Y.T.; Kim, Y.-I.; Park, J.-H.; Park, E.-S.; Woo, J.S. Enhanced oral bioavailability of paclitaxel by solid dispersion granulation. *Drug Dev. Ind. Pharm.* **2015**, *41*, 1864–1876. [CrossRef] [PubMed]
74. Wang, Y.Z. Development and Characterization of Small Segmental Translocations of Thinopyrum Bessarabicum and Cytological Mapping of Interesting Genes. Master's Thesis, Nanjing Agricultural University, Nanjing, China, 2013.
75. Su, S.; Zhou, Y.; Qin, J.G.; Yao, W.; Ma, Z. Optimization of the method for chlorophyll extraction in aquatic plants. *J. Freshw. Ecol.* **2010**, *25*, 531–538. [CrossRef]
76. Webb, D.J.; Burnison, B.K.; Trimbee, A.M.; Prepas, E.E. Comparison of chlorophyll a extractions with ethanol and dimethyl sulfoxide/acetone, and a concern about spectrophotometric phaeopigment correction. *Can. J. Fish. Aquat. Sci.* **1992**, *49*, 2331–2336. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

The Change in Whole-Genome Methylation and Transcriptome Profile under Autophagy Defect and Nitrogen Starvation

Yunfeng Shi [1,†], Baiyang Yu [1,2,†], Shan Cheng [1], Weiming Hu [1,*] and Fen Liu [1,*]

[1] Lushan Botanical Garden, Jiangxi Province and Chinese Academy of Sciences, Jiujiang 332000, China; shiyf@lsbg.cn (Y.S.); 12216011@zju.edu.cn (B.Y.); chengshan@lsbg.cn (S.C.)
[2] Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China
* Correspondence: huwm@lsbg.cn (W.H.); liuf@lsbg.cn (F.L.)
† These authors contributed equally to this work.

Abstract: Through whole-genome bisulfite sequencing and RNA-seq, we determined the potential impact of autophagy in regulating DNA methylation in *Arabidopsis*, providing a solid foundation for further understanding the molecular mechanism of autophagy and how plants cope with nitrogen deficiency. A total of 335 notable differentially expressed genes (DEGs) were discovered in wild-type *Arabidopsis* (Col-0-N) and an autophagic mutant cultivated under nitrogen starvation (*atg5-1*-N). Among these, 142 DEGs were associated with hypomethylated regions (hypo-DMRs) and were upregulated. This suggests a correlation between DNA demethylation and the ability of *Arabidopsis* to cope with nitrogen deficiency. Examination of the hypo-DMR-linked upregulated DEGs indicated that the expression of MYB101, an ABA pathway regulator, may be regulated by DNA demethylation and the recruitment of transcription factors (TFs; ERF57, ERF105, ERF48, and ERF111), which may contribute to the growth arrest induced by abscisic acid (ABA). Additionally, we found that DNA methylation might impact the biosynthesis of salicylic acid (SA). The promoter region of ATGH3.12 (PBS3), a key enzyme in SA synthesis, was hypomethylated, combined with overexpression of PBS3 and its potential TF AT3G46070, suggesting that autophagy defects may lead to SA-activated senescence, depending on DNA demethylation. These findings suggest that DNA hypomethylation may impact the mechanism by which *Arabidopsis* autophagy mutants (*atg5-1*) respond to nitrogen deficiency, specifically in relation to ABA and SA regulation. Our evaluation of hormone levels verified that these two hormones are significantly enriched under nitrogen deficiency in *atg5-1*-N compared to Col-0-N.

Keywords: autophagy; nitrogen starvation; whole-genome bisulfite sequencing (WGBS); RNA-seq; hormone

Citation: Shi, Y.; Yu, B.; Cheng, S.; Hu, W.; Liu, F. The Change in Whole-Genome Methylation and Transcriptome Profile under Autophagy Defect and Nitrogen Starvation. *Int. J. Mol. Sci.* **2023**, *24*, 14047. https://doi.org/10.3390/ijms241814047

Academic Editor: David A. Gewirtz

Received: 15 August 2023
Revised: 10 September 2023
Accepted: 12 September 2023
Published: 13 September 2023

1. Introduction

As a universal mechanism in eukaryotes that promotes cell longevity and nutrient recycling, autophagy helps plants cope with various biotic/abiotic stresses [1]. Autophagy in plants primarily encompasses three types: microautophagy, macroautophagy, and mega-autophagy. In the process of microautophagy, the separation and extraction of components in vacuoles are realized by direct wrapping within vacuole/lysosome membranes. This is related to the accumulation of anthocyanidin in vacuoles and the elimination of damaged chloroplasts in plants [2]. Mega-autophagy is implicated in both programmed cell death (PCD) that occurs during developmental processes and pathogenic invasion [3]. During macroautophagy (hereafter referred to as autophagy), cargos are trapped in newly formed cytoplasmic vesicles, which are created when a cup-shaped phagophore (also known as an isolation membrane) expands and envelops the cytoplasm, eventually sealing off to form the autophagosome, a double-membrane-bound structure. Subsequently, the outer membrane of the autophagosome fuses with the tonoplast, leading to the release of the

internal vesicle as an autophagic body [3]. The autophagy pathway is regulated by a sophisticated mechanism for which over 40 ATG genes have been identified in plants [4–8]. In recent years, various studies have indicated that autophagy plays a crucial role in diverse biological processes in plants, including growth, development, degradation of starch, senescence, and control of lipid metabolism [9]. In *Arabidopsis*, autophagy also affects the growth of pollen tubes. Through the specific mediation of ATG8-family interacting motif (AIM) recognition sites, ATG8e and mitochondria can bind and interact spatiotemporally, thereby inducing mitochondrial autophagy [10]. In addition, the relationship between autophagy and epigenetics in animals has also been validated. Under nutritional deficiency, fibroblast growth factor-21 (FGF21) signaling activates the global expression of autophagy-related genes through the histone H3K27-ME3 demethylase Jumonji-D3 (JMJD3/KDM6B), thus mediating lipid degradation [11]. The autophagy regulatory network may also be regulated by DNA methylation. It has been reported that MAP1LC3 is a key component of the core mechanism of autophagy. DNA methylation mediated by the de novo DNA methyltransferase DNMT3A at the MAP1LC3 locus leads to continuous downregulation of the transcription level of MAP1LC [12]. The relationship between autophagic function and epigenetic regulation has been explored in animals. However, it remains unknown whether plant autophagy and/or nitrogen starvation affect whole-genome DNA methylation.

Numerous studies have examined the function of DNA methylation in the abiotic stress response. DNA methylation resulting from abiotic stress is not limited to specific gene regions; it can occur throughout the genome. Moreover, the genes that are susceptible to modification in response to stress are ostensibly linked to methylation regulation [13]. In research conducted by Jiang, C. et al. on the methylation profile of *Arabidopsis* cultivated in soil with high salt concentration, in comparison to the control group, the stressed lineages amassed approximately 45% more differentially methylated cytosine positions (DMPs) at CG sites (CG DMPs), and the majority of these DMPs were observed to be inherited [14]. In *Arabidopsis*, exposure to low levels of Pi results in a significant alteration in the methylation level of the entire genome, which in turn is linked with the regulation of gene expression for responding to Pi starvation [15]. In a study of DNA methylation of the VRN-A1 gene of winter wheat, Abdul Rehman Khan et al. reported methylation at both CG and non-CG sites. Furthermore, they observed that site-specific hypermethylation induced by cold was transmitted through mitosis at non-CG sites [16]. To date, comprehensive research has been conducted on the response of various plants and crops to abiotic stress. This includes studies on the correlation between autophagy and abiotic stress and the regulatory mechanism of DNA methylation in plants or crops for resistance to abiotic stress [17–21].

Based on the above research background, the response of plants to abiotic stress seems to have a profound relationship with DNA methylation, so we hypothesized that DNA methylation may be involved in the response of autophagy-deficient mutants to nitrogen deficiency. However, up to now, no relevant research has been conducted on the relationship between autophagy and DNA methylation in plants and how to regulate gene expression by affecting the DNA methylation level. Thus, in this study, whole-genome bisulfite sequencing (WGBS) was performed on *Arabidopsis* autophagic mutants (*atg5-1*) developing under nitrogen starvation to address relevant concerns, in conjunction with transcriptome sequencing (RNA-seq).

2. Results

2.1. The Whole-Genome Methylation Level of Autophagic Mutants Was Strongly Affected under Nitrogen Deficiency Conditions

To investigate whether autophagy influences plant growth under nitrogen starvation stress through DNA methylation/demethylation, we performed WGBS and RNA-seq for *Arabidopsis* wild-type (Col-0) and autophagic mutant (*atg5-1*) lines under MS growth conditions (MS liquid culture medium) or nitrogen starvation conditions (MS-N liquid culture medium) (Figure 1A). Over 4G clean reads (paired-end reads) were produced for each sample in the WGBS experiment.

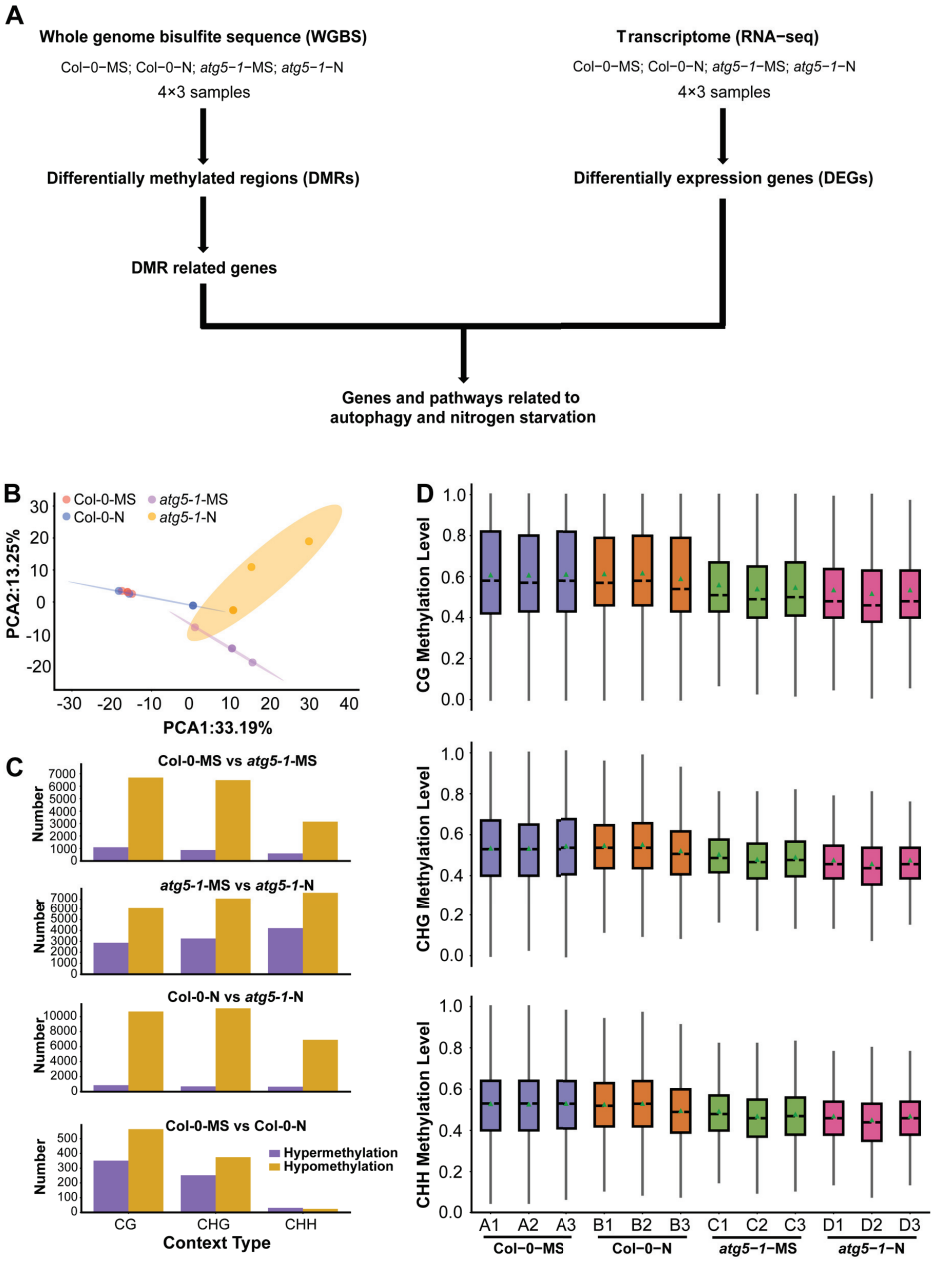


Figure 1. Autophagy function defects lead to changes in genome-wide methylation levels. (**A**) Data analysis workflow for whole-genome bisulfite sequencing (WGBS) and transcriptome analysis (RNA-seq). (**B**) PCA of the DNA methylation level in each sample. PCA of Col-0 and *atg5-1* genotypes, displaying obvious differences under MS and nitrogen starvation conditions. (**C**) The number of hypomethylated and hypermethylated DMRs in each comparison group. (**D**) The distribution of DNA methylation levels. The distribution plot of the methylation level of common sites showed that DNA methylation was affected by autophagy.

The average mapping ratios for alignment with the reference genome for the different treatment groups were 94.06% (Col-0-MS), 92.50% (Col-0-N), 87.64% (*atg5-1*-MS), and 85.06% (*atg5-1*-N), and the conversion ratios were >99.5% (Table S1). We then separated DNA methylation into three context types (CG, CHG, and CHH) and detected methylation sites based on the above context types [22,23]. According to the methylation level of common sites in all samples, a principal component analysis (PCA) was performed and showed obvious differentiation between wild-type *Arabidopsis* and the autophagic mutant (Figure 1B).

To characterize the methylation changes under nitrogen starvation conditions and in the autophagy mutant, differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) were identified in different comparison groups (Col-0-MS vs. Col-0-N, *atg5-1*-MS vs. *atg5-1*-N, Col-0-MS vs. *atg5-1*-MS, and Col-0-N vs. *atg5-1*-N).

In Col-0-MS vs. Col-0-N, there were 351, 252, and 31 upregulated DMRs and 565, 375, and 24 downregulated DMRs in the CG, CHG, and CHH contexts, respectively, and among the comparison groups, the fewest DMRs were identified in Col-0-MS vs. Col-0-N. Apparently, wild-type *Arabidopsis* was not significantly impacted by nitrogen starvation, and relatively obvious differences were observed in the comparison between the wild-type (Col-0-MS) and autophagic mutant (*atg5-1*-MS) under normal conditions (1088, 874, and 602 upregulated DMRs and 6690, 6483, and 3149 downregulated DMRs). Autophagy seems to have a greater influence on the whole-genome methylation level than nitrogen starvation, although autophagy is one of the most important pathways for responding to N stress in plants.

By comparing the DMR numbers of *atg5-1* samples in MS and nitrogen starvation conditions (*atg5-1*-MS vs. *atg5-1*-N), 2868, 3265, and 4213 hypermethylated DMRs and 6052, 7070, and 7603 hypomethylated DMRs were identified in the CG, CHG, and CHH contexts, respectively, showing a dramatic difference in Col-0-MS vs. Col-0-N. Significant differences in DMR number were observed in *atg5-1* genotype samples and wild-type samples growing in MS and nitrogen starvation conditions, which seemingly indicated that in wild-type samples, some compensation effects probably resisted the influence of nitrogen starvation on methylation levels; however, this kind of compensation effect might be inhibited due to autophagy defects. Under the same nitrogen starvation conditions, 863, 698, and 659 hypermethylated DMRs and 10691, 11110, and 6913 hypomethylated DMRs were identified in Col-0-N vs. *atg5-1*-N in the CG, CHG, and CHH contexts, respectively (Table S2) (Figure 1C). We examined the common methylation site for each treatment group, and the results showed that the overall methylation level of the *atg5-1* sample was significantly lower than that of the wild-type sample, which further suggested that the decrease in the DNA methylation level could be largely attributed to the autophagy defect (Figure 1D).

In summary, nitrogen starvation may not have a significant impact on *Arabidopsis* in terms of methylation. However, when autophagy in *Arabidopsis* produces defects, the impact of nitrogen starvation may be amplified.

2.2. Autophagy Stimulates Global Expression of the Arabidopsis Genome under Nitrogen Starvation Conditions

To examine the gene expression changes involved in resistance to autophagy and nitrogen starvation in *Arabidopsis*, transcriptome profiles under different conditions were generated by RNA-seq (Table S3), and more than 6G clean reads (paired-end reads) were obtained for each RNA-seq sample.

The number of upregulated genes was obviously greater than the number of downregulated genes in the Col-0-MS vs. Col-0-N comparison group, suggesting that global gene expression was activated by nitrogen starvation to maintain growth. In *atg5-1*-MS vs. *atg5-1*-N, there were similar changes in differential gene expression, indicating that in the case of autophagy-related functional defects, nitrogen starvation may still activate the expression of many genes in response. However, a minimum number of DEGs appeared in Col-0-MS vs. *atg5-1*-MS, indicating that autophagy-related functional defects probably mildly impact plant growth under normal conditions (Figure 2A,B).

In Col-0-MS vs. Col-0-N, the genes associated with the integral component of the membrane, the intrinsic component of the membrane, and the defense response were the most significantly enriched among the upregulated DEGs. KEGG analysis showed that genes associated with ABC transporters were significantly enriched in membrane transport-related pathways, suggesting that ABC transporters probably play an important role when *Arabidopsis* suffers nitrogen starvation. The genes associated with the chloroplast thylakoid membrane, the plastid thylakoid membrane, and the photosynthetic membrane

were the most significantly enriched among the downregulated DEGs. Interestingly, most of the top 10 GO enrichment annotations of the downregulated DEGs seemed to be related to thylakoids. We performed a KEGG analysis for DEGs annotated in thylakoid-related functional modules. The DEGs were mainly involved in pathways related to photosynthesis, ATPase, and NADH dehydrogenase (Table S4).

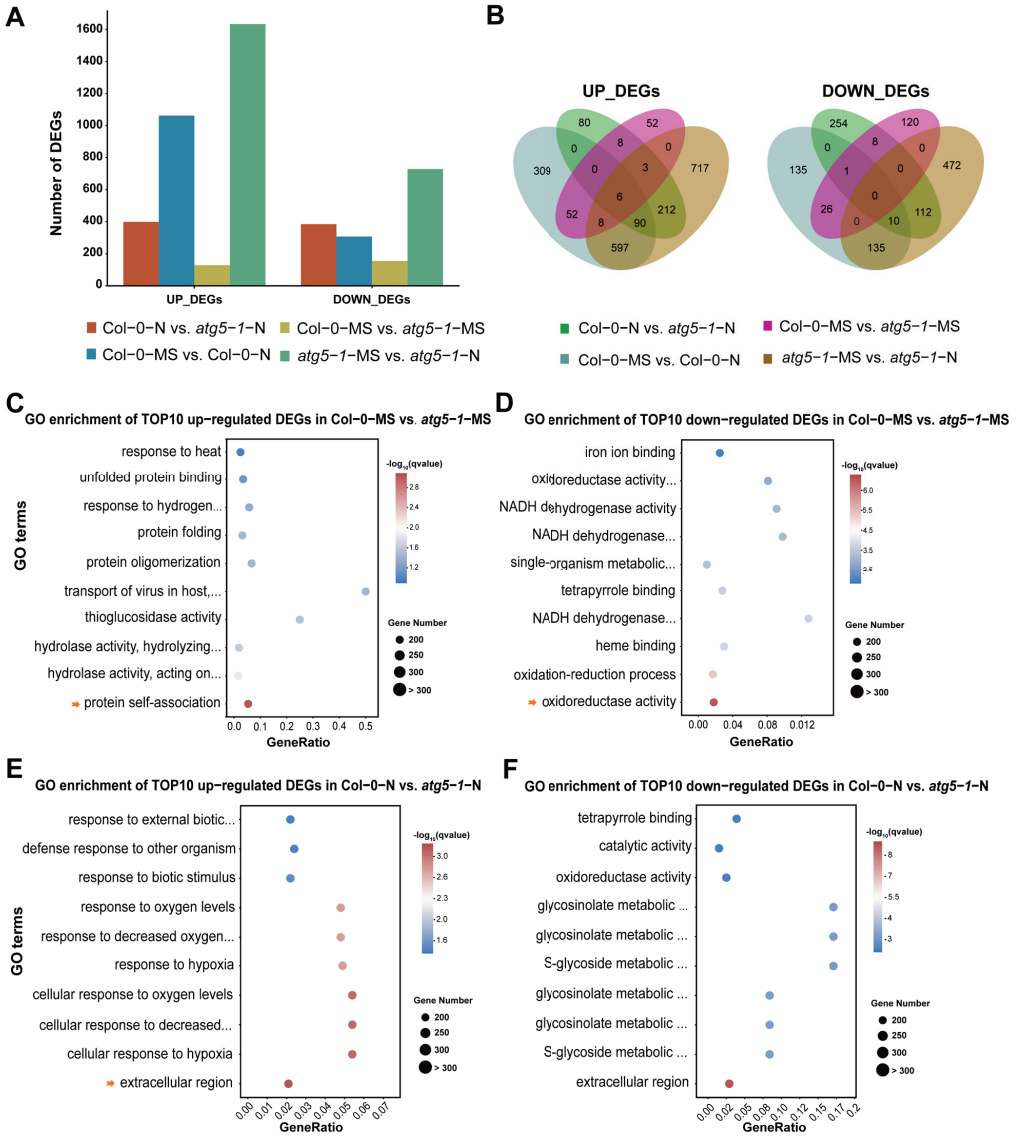


Figure 2. Nitrogen starvation activates differential gene expression across the whole genome of Col-0 and autophagic mutants. (**A**) DEG statistics. Gene expression was activated in the *atg5-1* and Columbia genotype groups under nitrogen starvation, showing the notable influence of nitrogen starvation. (**B**) Venn plot of DEGs in different comparison groups. (**C**–**F**) GO enrichment of the top 10 upregulated DEGs in Col-0-MS vs. *atg5-1*-MS, Col-0-MS vs. *atg5-1*-MS, Col-0-N vs. *atg5-1*-N, and Col-0-N vs. *atg5-1*-N. The arrow represents the GO terms with the most significant enrichment or that were meaningful for our research.

In Col-0-MS vs. *atg5-1*-MS, the upregulated DEGs were extremely enriched in protein self-association, and most of the functional modules in the top 10 enriched GO terms were associated with protein structure, unfolded protein processing, and hydrolase activity. Understandably, in autophagy, degradation, and recycling of unwanted content and misfolded proteins are the main functions (Figure 2C) [24,25]. A KEGG analysis was performed for these DEGs, which showed that the protein processing in the endoplasmic reticulum and

tryptophan metabolism pathways were significantly enriched, and the DEGs enriched in these two pathways were associated with heat-shock-protein- and myrosinase-related genes. (Table S5). GO enrichment analysis was performed for the downregulated DEGs and showed that oxidoreductase activity and the oxidation–reduction process were the most significantly enriched, while other significantly enriched GO terms were mainly related to the binding of substances and NADH dehydrogenase activity (Figure 2D). KEGG results produced based on GO enrichment showed that most DEGs were associated with oxidative reactions and encoded glutathione transferase (Table S6).

To explore the functions that were influenced differentially between Col-0-N and *atg5-1*-N, we performed GO enrichment analysis of the upregulated DEGs and removed the DEGs that were present in Col-0-MS vs. *atg5-1*-MS. The results clearly showed that most of the top 10 GO terms were related to oxygen levels and the response to hypoxia, suggesting that plants with autophagy defects were more sensitive to oxygen under nitrogen starvation conditions (Figure 2E). Additionally, we annotated the pathways associated with these upregulated genes using KEGG enrichment. The most upregulated DEGs, which were enriched in oxygen-level-related functions, were associated with the amino acid metabolism pathway, while carotenoid biosynthesis, the MAPK signaling pathway, and the biosynthesis of other secondary metabolites were also enriched (Table S7). Subsequently, exploration of gene function was carried out through GO and KEGG enrichment for the downregulated DEGs, from which, for functional analysis, DEGs that genes were also present in Col-0-MS vs. *atg5-1*-MS were removed. The DEGs associated with the extracellular region, the glucosinolate metabolic process, and the glucosinolate catabolic process were significantly enriched in metabolic pathways and the biosynthesis of secondary metabolites, mainly including genes encoding various enzymes (Figure 2F and Table S8).

To explore functional changes between *atg5-1*-MS and *atg5-1*-N, the DEGs that were also present in Col-0-MS vs. Col-0-N were removed. The GO results for the upregulated genes were similar to those for Col-0-N vs. *atg5-1*-N. The top 10 GO terms were mainly related to the response to changes in oxygen levels, and the KEGG results for DEGs enriched in the GO terms mentioned above showed that the significantly enriched terms were mainly related to metabolic and biosynthetic pathways of various substances, which mainly involved genes associated with a cytidine/deoxycytidylate deaminase family protein and genes encoding a cytidine deaminase, a gamma-glutamyltransferase and glycosyl hydrolase family protein (Tables S9 and S10). For downregulated DEGs, GO enrichment showed that extracellular function and oxygenase activity-related function were significantly enriched. KEGG annotation indicated that the genes with the functions mentioned above were mainly enriched in biosynthesis- and metabolism-related pathways (Tables S11 and S12).

2.3. Autophagy Deficiency May Induce Downregulation of DNA Methylation Levels, Thus Activating Gene Expression Related to Cell Death and Catabolic Pathways

To investigate the potential influence of DNA hypomethylation induced by autophagy defects, we counted the number of hypo-DMRs associated with upregulated DEGs. The results showed that most of the DEGs were identified in Col-0-MS vs. Col-0-N, but only a few of these DEGs were regulated by DNA methylation, which indicates that DNA methylation is not closely involved in the mechanism underlying the response to nitrogen deficiency in the growth process of wild-type *Arabidopsis* under nitrogen deficiency conditions.

In Col-0-MS vs. *atg5-1*-MS, *atg5-1*-MS vs. *atg5-1*-N, and Col-0-N vs. *atg5-1*-N, there were 38, 73, 142 hypo-DMRs associated with upregulated DEGs, respectively, which was comparable to the number of upregulated DEGs identified in the comparison groups mentioned above. The number of upregulated DEGs regulated by methylation was also considerable (Figure 3A). To explain the relationship between DNA methylation and autophagy and the possible mechanism by which DNA methylation regulates autophagy mutants' resistance to nitrogen deficiency, we mainly focused on Col-0-MS vs. *atg5-1*-MS and Col-0-N vs. *atg5-1*-N in the analysis process.

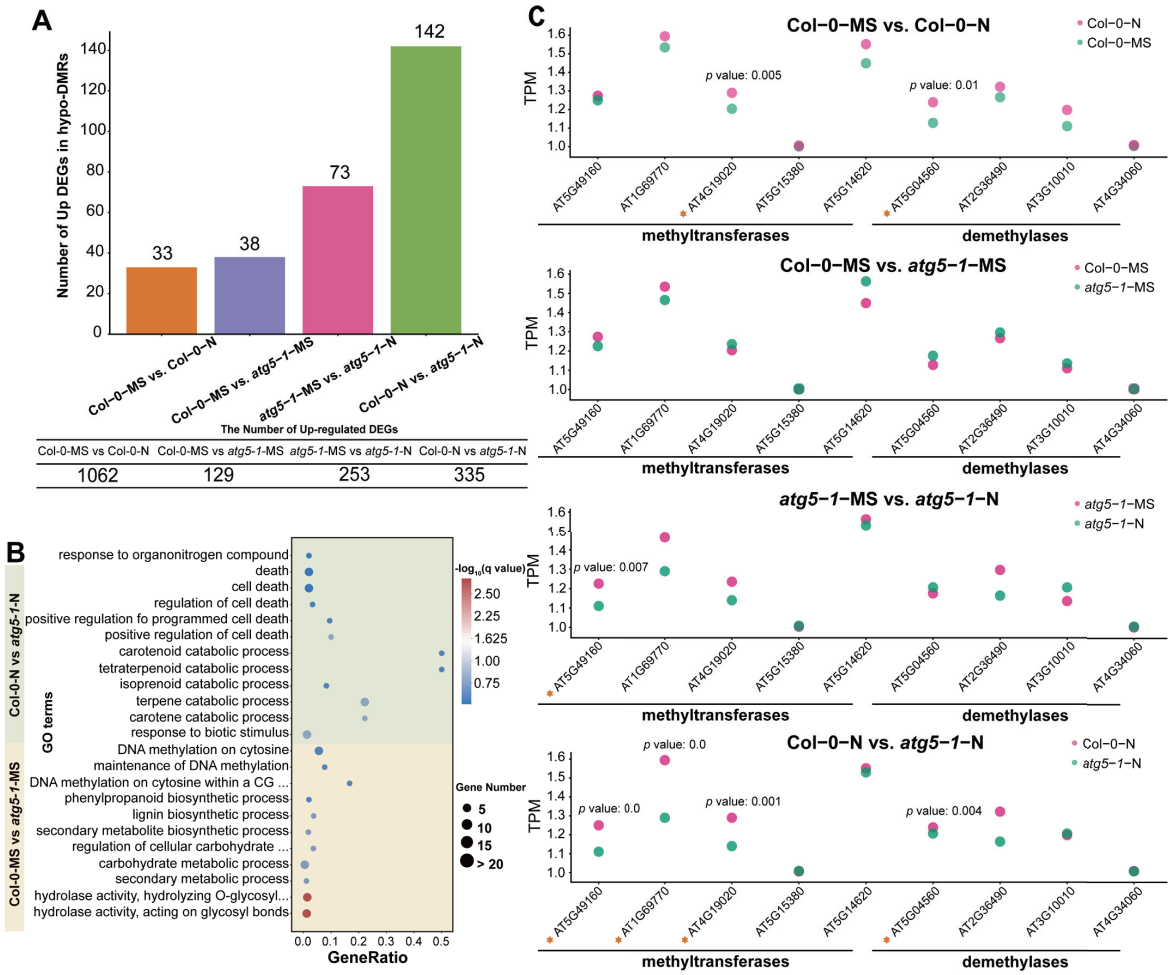


Figure 3. DNA methylation affects the response of autophagic mutants to nitrogen deficiency. (**A**) The number of hypo-DMRs associated with upregulated DEGs. For Col-0-N vs. *atg5-1*-N, half of the upregulated DEGs were associated with DMRs. (**B**) GO enrichment of hypo-DMRs associated with upregulated DEGs. (**C**) The expression of methyltransferases and demethylases calculated by RNA-seq indicated that DNA methylation was involved in the regulation of gene expression. The asterisks represent the transcription factors.

GO enrichment analysis of hypo-DMRs associated with the upregulation of DEGs in different comparison groups showed that DNA hypomethylation may have a regulatory effect on nitrogen starvation and autophagy defects. This effect may be controlled by specific genes in different biological pathways. The GO enrichment of hypo-DMRs associated with upregulated DEGs in Col-0-MS vs. *atg5-1*-MS showed that the upregulated DEGs were most significantly enriched in GO terms associated with hydrolase activity and meta-hydroxylase activity, suggesting that DNA hypomethylation significantly regulated hydrolase activity under autophagy defects (Figure 3B). The genes encoding cellulase (glycosyl hydrolase family 5) protein (AT1G13130), glycosyl hydrolase 28 (GH28) family polygalacturonase (PG) protein, cell wall invertase 5, disease resistance protein (TIR–NBS–LRR class) family and tricoumaroylspermidine meta-hydroxylase had the highest fold change. Surprisingly, several hypo-DMR-associated DEGs were enriched in GO terms related to various substance metabolic processes and biosynthetic processes, such as the secondary metabolite biosynthetic process, the lignin biosynthetic process, the lignin metabolic process, and the negative regulation of starch metabolic process, suggesting that metabolic and biosynthetic activity is induced by autophagy defects. Interestingly, the QQS (AT3G30720), ATMYB29 (AT5G07690), XTH6 (AT5G65730), AT4G08160, and BGLU34 (AT1G47600) genes were enriched in multiple metabolism-related GO terms, and

in biosynthesis-related GO terms, ATCAD8 (AT4G37990), LAC11 (AT5G03260), and AT-MYB29 (AT5G07690) were also enriched. In addition, the DNA methylation processes were probably regulated. For example, DNA methylation on cytosine within a CG sequence, maintenance of DNA methylation, DNA methylation on cytosine, and all GO terms related to methylation were enriched, including only ORTH3 (AT1G57800), which indicates that ORTH3 probably plays an important role in regulating the methylation process induced by autophagy defects (Table S13).

To explore the functions regulated by the autophagy mutant compared with the wild type under nitrogen starvation, GO enrichment was performed for Col-0-N vs. *atg5-1*-N. The results indicated that hypomethylation greatly affected the response to biotic stimuli (Figure 3B). The DEGs that exhibited the highest fold change were RLP3 (AT1G17250), AT1G32763, AIG1 (AT1G33960), LECRK-III.2 (AT2G29250), AT5G10530 (LECRK-IX.1), AT5G42223, and SAG12 (AT5G45890). In addition, the catabolic processes for many substances were significantly enriched, and five hypo-DMR-related upregulated DEGs, AT4G18350, AT4G32810, AT4G39650, AT5G24540, and AT4G29640, were mainly involved in these processes. AT4G29640 and AT4G39650 exhibited the highest fold change, suggesting that these two genes probably play an important role in resistance when autophagic mutants are subjected to nitrogen starvation stress. This response mechanism may be due to the deficiency of autophagy function, which leads to *Arabidopsis* being unable to maintain the original level of DNA methylation under nitrogen deficiency, thus activating the decomposition of some substances. Meanwhile, the regulatory pathways related to cell death were also affected, and AT2G32460, AT5G10530, AT5G13320, and AT5G45890 were significantly enriched in these GO terms, suggesting that they may be key for establishing the relationship between DNA methylation and premature aging caused by autophagy deficiency. Similar to the catabolic process for many substances, the downregulation of DNA methylation levels may activate gene expression regulating senescence or death in *Arabidopsis*. This regulatory process may be activated by nitrogen deficiency, as there was no enrichment observed for many cell-death-related functions in the Col-0-MS vs. Col-0-N comparison group. Therefore, we speculate that defects in autophagy may affect the level of DNA methylation, and this effect reduces the tolerance of *Arabidopsis* to nitrogen deficiency.

Additionally, compared with the result for Col-0-MS vs. *atg5-1*-MS, the response to organonitrogen compounds was exclusively enriched when *Arabidopsis* was under nitrogen starvation, and it was probably influenced by AT2G14610, AT2G46400, AT3G44350, and AT5G59820, which likely indicates that under nitrogen deficiency, changes in the level of DNA methylation may be involved in mediating the utilization of nitrogen by *Arabidopsis* when the autophagy function has broken down.

We performed Student's *t*-tests for the TPM (transcripts per million) value of several methyltransferases and demethylases of different comparison groups. A maximum number of methyltransferases and demethylases were differentially expressed in Col-0-N vs. *atg5-1*-N, which indicates that DNA methylation might be regulated by autophagy under nitrogen starvation conditions (Figure 3C).

2.4. Nitrogen Deficiency May Induce Transcription Factors of the ERF and C2H2 Type Zinc Finger Families to Regulate ABA-Induced Growth Arrest and SA Biosynthesis in atg5-1 Autophagic Mutants

A Venn diagram was plotted for different comparison groups, and most of the upregulated DEGs were methylated exclusively in Col-0-N vs. *atg5-1*-N (Figure 4A). Hypo-DMR is usually associated with upregulated expression of genes through the recruitment of transcription factors (TFs). It has been reported that low-methylation regions (LMRs) can be used to identify transcriptional enhancers, and TF binding sites can usually be detected in LMRs, which tend to be occupied by cell-type-specific TFs [26,27]. Thus, we predicted TFs that can bind to hypo-DMRs for Col-0-MS vs. *atg5-1*-MS using PlantTFDB, and a total of 53 TFs were identified, which included 17 types of TF families, the most abundant among which were the bZIP, Dof, and GATA families [28,29]. However, the TFs associated with hypo-DMRs did not exhibit significant differences in the expression profiles determined by

RNA-seq. In Col-0-N vs. *atg5-1*-N, a total of 26 types of TF families were predicted, the most abundant of which were the ERF, NAC, MYB, C2H2, LBD, GATA, and Trihelix families. By analyzing TF expression, *AT1G02230* (*ANAC004*), *AT2G40340* (*ATERF48*), *AT3G04060* (*ANAC046*), *AT3G12910*, *AT3G46070* (C2H2-type zinc finger family protein), *AT5G18270* (*ANAC087*), *AT5G51190* (*ERF105*), *AT5G64750* (*ERF111*), and *AT5G65130* (*ERF57*) were shown to be significantly differentially expressed (p-value < 0.05; FC $\geq$ 1.5), and most of them belong to the NAC and ERF families (Table S14 and Figure 4B). Among the predicted genes, TFs showed the possibility of binding to the DMRs in the *AT2G32460* (*MYB101*), *AT2G29250*, *AT1G32763*, *AT5G13320* (*PBS3*) and *AT5G59820* (*RHL41*) promoter regions; GO enrichment analysis showed that these genes are involved in biotic stimulus, cell death and response to organonitrogen compounds, indicating that these functions were probably regulated by hypo-DMR-recruiting TFs.

MYB101 has been confirmed to be associated with ABA-induced plant growth arrest, and transcription factors of the ERF family also play an important role in the mechanism of plant death. Our results predicted that the transcription factors binding to *MYB101* belong to the ERF family, and these transcription factors were also differentially upregulated in the RNA-seq-based differential expression analysis. This seems to indicate that under nitrogen deficiency, the methylation level of the promoter region of *MYB101* is downregulated, and the introduction of ERF family transcription factors binding to this region activates the ABA-induced plant growth arrest process. In addition, the transcription factor prediction results indicated that *AT3G46070* (C2H2 type zinc finger family protein) could bind to *ATGH3.12* (*PBS3*). *PBS3* is the rate-limiting enzyme in the last two steps of SA biosynthesis, which controls plant aging. Therefore, we speculate that autophagic deficiency can cause significant downregulation of the DNA methylation level in the *PBS3* promoter region, thus recruiting *AT3G46070* to combine with this low methylation region to promote SA biosynthesis and control the premature senescence of *Arabidopsis*.

2.5. The qRT–PCR Experiments on Key Genes Involved in the SA and ABA Synthesis Pathways Confirmed the WGBS and RNA-seq Analysis Results

Based on our hypothesis, DNA methylation is involved in the process of ABA-induced growth arrest of *Arabidopsis* in Col-0-N vs. *atg5-1*-N. We thus performed qRT–PCR for the DEGs mentioned above (Figure 5A). We speculate that the level of ABA would increase if it was involved in the premature senescence of the autophagy mutant under nitrogen deficiency, and this is also fundamental evidence that ABA-induced senescence is impacted by DNA methylation. Therefore, we also verified the expression levels of key genes related to ABA biosynthesis, and the results also corresponded to the RNA-seq profile and our hypothesis (Figure 5B).

SA biosynthesis was probably also associated with DNA methylation in our WGBS and RNA-seq analysis results, and DNA methylation might affect the expression of ATGH3.12 (PBS3), which participates in the last two steps of SA biosynthesis. We also performed qRT–PCR for the genes involved in the SA biosynthesis pathway, and the results showed that the expression of ATGH3.12 (PBS3) increased dramatically and that of key genes of the ABA biosynthesis pathway, such as ICS1, also changed significantly (Figure 5C).

The qPCR results of key genes suggested that the levels of the hormones SA and ABA were significantly upregulated, consistent with our hypothesis. This confirms the accuracy of the RNA-seq results and indicates that differential DNA methylation is highly likely related to the mechanism by which ABA and SA regulate plant growth arrest or premature senescence (Table S15).

2.6. The Levels of the Hormones SA and ABA in the atg5-1 Mutant Significantly Increased under Nitrogen Deficiency

We hypothesize that the ABA and SA pathways may be regulated by DNA methylation to affect premature plant senescence. Therefore, we measured the levels of SA and ABA. The results of ABA determination showed that the hormone levels were significantly higher in the *atg5-1*-N group than in the Col-0-N group (Figure 6A). The level of SA showed a

significant increase in *atg5-1*-N, exhibiting a more than eight-fold change compared with Col0-N (Figure 6B).

Hormonal measurements indicated that deficiencies in autophagy function may reduce the tolerance of *Arabidopsis* to nitrogen starvation, thus directly or indirectly changing the DNA methylation level, thereby triggering an increase in SA and ABA levels and leading to premature aging.

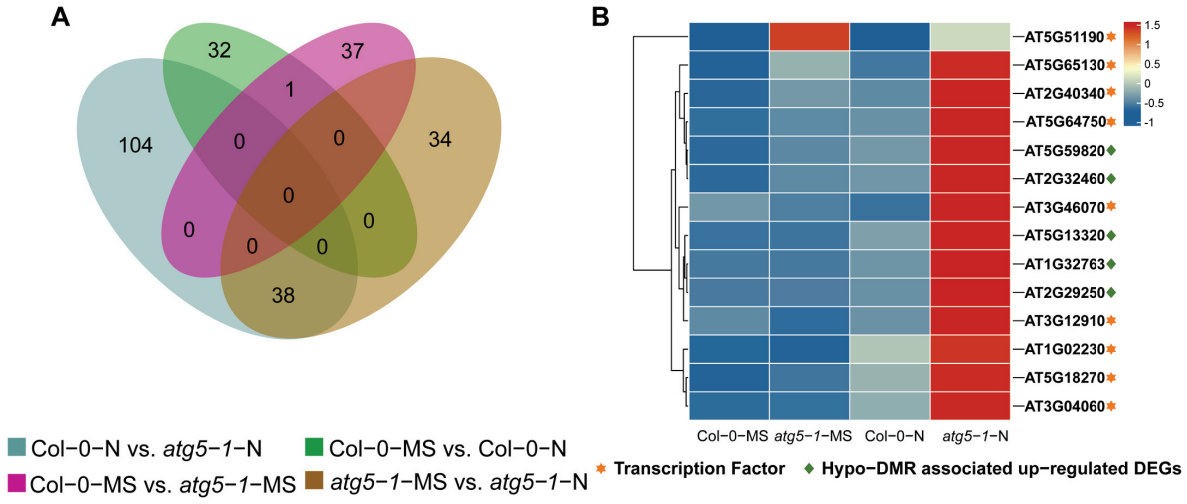


Figure 4. Senescence-related genes were strongly influenced by DNA methylation under nitrogen deficiency. (**A**) Venn plot for hypo-DMRs in different comparison groups. (**B**) Transcription factors (TFs) predicted to bind to the promoters of hypo-DMRs associated with upregulated DEGs. Most TFs belonged to the ERF family and NAC family, and most of them were highly expressed exclusively in *atg5-1*-N, which shows that these TFs and hypo-DMRs associated with upregulated DEGs probably play a key role in resisting nitrogen starvation in the *atg5-1* genotype. The asterisks and diamond represent the transcription factors and key genes related to plant death.

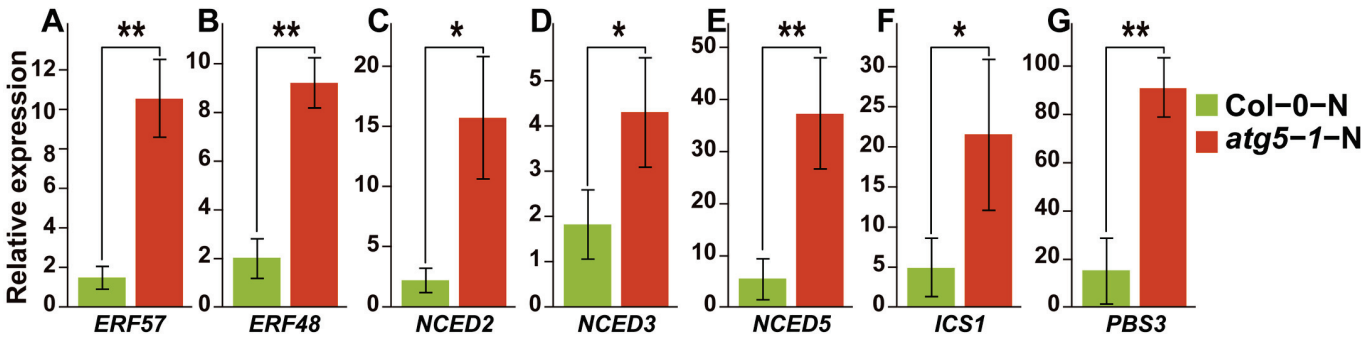
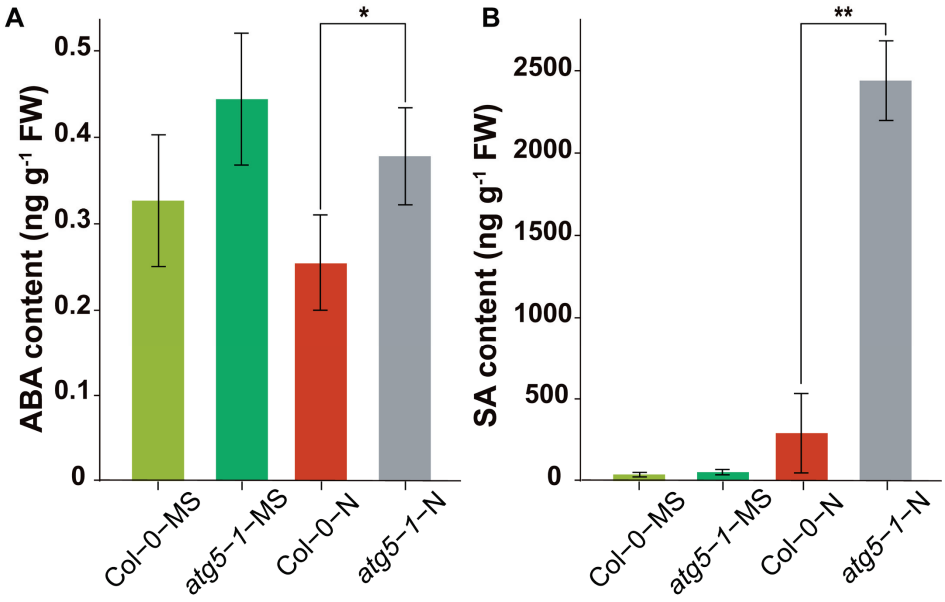


Figure 5. qRT–PCR results of key genes and predicted transcription factors that regulate the SA and ABA hormone pathways. ** Significant difference at $p < 0.01$ and * significant difference at $p < 0.05$. Vertical bars represent the standard deviation. (**A**–**E**) qRT–PCR of key genes in ABA biosynthesis pathway and predicted transcription factors that may be associated with ABA-induced growth arrest. (**F**,**G**) The expression levels of key genes regulated by DNA methylation in the SA biosynthesis pathway.



The mean level of hormones (ng g^{-1} FW)

Phytohormone	Col−0−MS	Col−0−N	*atg5−1*−MS	*atg5−1*−N
ABA	0.327	0.254	0.444	0.378
SA	54.460	291.380	25.589	2439.473

Figure 6. Quantification of the phytohormones ABA and SA. ** Significant difference at $p < 0.01$ and * significant difference at $p < 0.05$. Vertical bars represent the standard deviation. The levels of the hormones ABA (**A**) and SA (**B**) in the autophagic mutant were significantly higher than those in the wild type under nitrogen starvation.

3. Discussion

DNA methylation is an epigenetic modification that commonly occurs during plant growth and development and plays an important role in regulating various biological processes [30], including leaf senescence and plant cell death [31–33]. In research on autophagy in plants, it has been proven that the response to nutrition deficiency, senescence, intracellular degradation and recycling of amino acids, and other stresses from various environments can be regulated by autophagy [34,35]. Nitrogen is a crucial nutrient element required for plant growth, and autophagy, which is associated with plant aging, is activated as a response to nitrogen deprivation [36–38]. Nevertheless, it remains ambiguous whether DNA methylation plays a role in autophagy triggered by nitrogen deprivation. In previous research, *Arabidopsis* Col-0 and *atg5-1* mutant lines exhibited distinct phenotypic characteristics while growing under MS and nitrogen starvation conditions. Notably, the *atg5-1* genotype exhibited faster senescence and cell death under nitrogen starvation conditions, implying that autophagy potentially expedites senescence and cell death in comparison to Col-0-N. The dynamic changes in DNA methylation in Col-0-N and *atg5-1*-N corresponded to phenotype, wherein the number of hypo-DMRs was highest in the CG, CHG, and CHH contexts (Figure 1C), and among all the groups, 142 upregulated DEGs associated with hypo-DMRs were the most abundant in this comparison group (Figure 3A). The findings suggest that under conditions of nitrogen starvation, the *atg5-1* genotype may reduce the expression of methylated regions and potentially attract transcription factors to regulate gene expression and react to nitrogen starvation [39,40]. Based on the above results, the prediction of TFs that bind to hypo-DMRs was performed on Col-0-N vs. *atg5-1*-N, identifying nine TFs (*ERF105*, *ERF57*, *ATERF48*, *ERF111*, *AT3G46070*, *AT3G12910*, *ANAC004*, *ANAC087*, *ANAC046*) and five upregulated DEGs (*ATZAT12*, *ATMYB101*, *ATGH3.12*, *AT1G32763*, *LECRK-III.2*) binding with the above nine TFs from the GO terms response to biotic stimulus, cell death, catabolic process, and response to organonitrogen compound.

3.1. DNA Methylation Participates in Arabidopsis Senescence through the Regulation of ABA Biosynthesis

The process of PCD, known as senescence, has been found to be regulated by endogenous ethylene production under the control of specific transcription factors [41,42]. In our study, hypo-DMR-associated upregulated DEGs that might be regulated by TFs were identified, including *MYB101* (*AT2G32460*) and *ATGH3.12* (*AT5G13320*). Previous research indicates that *MYB101* functions as a constructive regulatory element in ABA-stimulated growth cessation, and an ineffective *MYB101* mutant reduces the sensitivity of plants to ABA [43,44]. It has been observed that ABA has the potential to induce senescence in leaves through a pathway that is independent of ethylene. Various scientific reports have indicated that the genes that encode rate-limiting enzymes involved in the biosynthesis of ABA include NCED2, NCED3, NCED5, NCED6, and NCED9 [45,46]. These five genes were differentially expressed (p-value < 0.1) in our analysis of gene expression differences between Col-0-N and *atg5-1*-N. Notably, NCED3 was identified as the most crucial rate-limiting enzyme (Figure 5B). It can be inferred that *atg5-1* induces high expression of *MYB101* under nitrogen starvation conditions, which consequentially exerts a positive regulatory effect on ABA synthesis and consequentially regulates *Arabidopsis* senescence. This regulatory relationship may be due to the downregulation of the methylation of the *MYB101* promoter region, which is achieved by recruiting TFs. Our investigation revealed that there are five TFs that have the capacity to bind to hypo-DMRs within the promoter region of the *MYB101* gene. Notably, these genes belong predominantly to the ERF family. Several research studies have demonstrated the participation of transcription factors from the ERF family in the deterioration and mortality mechanisms of plants. Based on research conducted by Chen, Y. et al., ERF.F5 plays a pivotal role in the senescence of tomato leaves. Specifically, the study indicated that ERF.F5 can directly regulate the promoter activity of *ACS6* and interact with *MYC2*, which ultimately leads to the regulation of tomato leaf senescence [47]. The transcription factor EPI1, also belonging to the ERF family, plays a negative regulatory role in dark-induced and JA-stimulated aging [48]. Furthermore, Juanxu Liu and colleagues discovered a robust correlation between numerous ERF genes and an increase in ethylene levels in petals and pistils during their examination of petunia blooms. This correlation is intimately linked with the process of petunia flower aging [49]. Therefore, we speculate that *ERF57*, *ERF105*, *ERF48*, and *ERF111* may be potential transcription factors that affect the aging of autophagic mutants under nitrogen starvation conditions. They are highly likely to bind to the hypo-DMR of the *MYB101* promoter, which affects the premature aging of the *atg5-1* mutant by upregulating the expression of *MYB101*.

3.2. SA Biosynthesis Was Probably Regulated by DNA Methylation

The ATGH3.12/PBS3 enzyme belongs to the GH3 acyl adenosylase enzyme family and plays a vital role in the accumulation of SA. Its main function is facilitating the conversion of isochorismate to unstable intermediates. Together with EPS1, it forms a two-step metabolic pathway in *Arabidopsis* to form SA, and this pathway is involved in the last two steps of SA formation [50]. SA plays an important role in regulating innate immunity and various developmental processes in plants. Research conducted by Niu, F. et al. showed that the transcription factor WRKY42 can regulate the aging process in *Arabidopsis* leaves by regulating the synthesis of SA [51,52]. According to reports, ICS1 is the key enzyme controlling SA synthesis, controlling 90% of this process [53,54]. Our study has demonstrated that there are significant changes in gene expression for two crucial enzymes, namely, *ATGH3.12*(*PBS3*) and *ICS1* (p-value < 0.05; FC > 1.5) (Figure 5), which means that SA biosynthesis is likely to be regulated, and DNA methylation may be involved in regulating the expression level of *ATGH3.12/PBS3*, thereby affecting SA biosynthesis. We also predicted a transcription factor (AT3G46070) that can probably bind to the hypo-DMR of the promoter of *ATGH3.12* (*PBS3*) and is differentially expressed; this transcription factor belongs to the C2H2-type zinc finger family of proteins.

Based on hormone level determination, our findings indicate that DNA methylation has a greater impact on SA biosynthesis, which probably contributes greatly to the premature senescence of *Arabidopsis* (Figure 7).

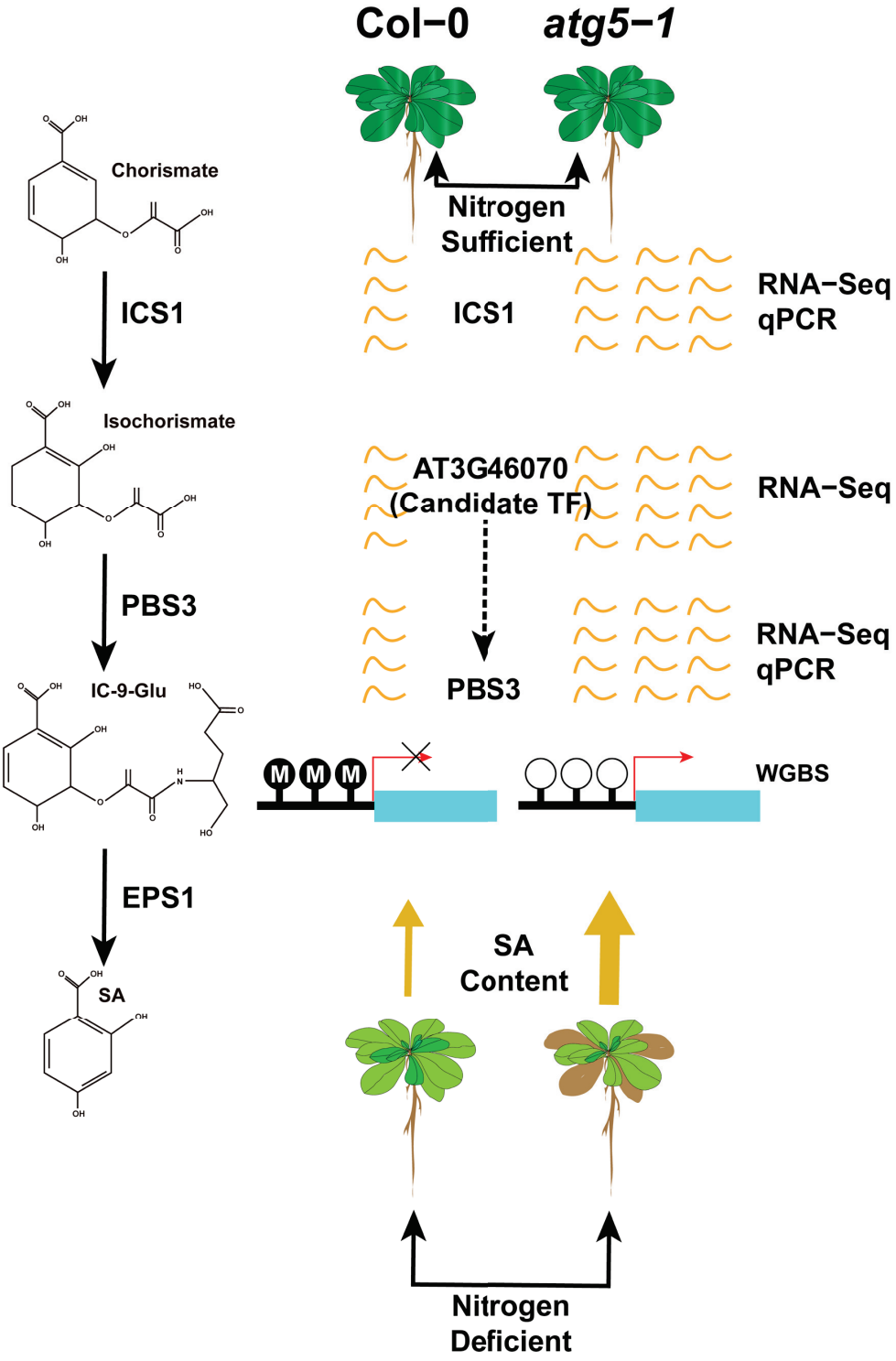


Figure 7. DNA methylation may regulate premature aging of *atg5-1* mutants under nitrogen starvation conditions by mediating the expression level of *PBS3*. By predicting transcription factors in the regions where hypomethylation occurs in the *PBS3* promoter, it is speculated that *AT3G46070* expressing transcription factors may be recruited to the hypomethylation regions, leading to an upregulation of PBS3 expression levels.

4. Materials and Methods

4.1. Plant Materials

The Arabidopsis thaliana Col-0 ecotype was used to elucidate methylation and transcription changes under nitrogen starvation and autophagy defects. The seeds were first cultivated in MS liquid culture medium for 7 days (16 h light/8 h darkness; constant temperature 22 °C; LED lighting; 120 μmol m^{-2} s^{-1}) and then transferred into MS nitrogen starvation culture medium (MS-N) (MSP21-50LT) and cultivated for another 60 h (16 h light/8 h darkness; constant temperature 22 °C; LED lighting; 120 μmol m^{-2} s^{-1}). The samples were dehydrated and snap-frozen in liquid nitrogen for subsequent WGBS, RNA-seq, and qRT–PCR analyses.

4.2. WGBS and Data Analysis

The DNA of the samples was extracted, and then an ultrasonic DNA fragmentation instrument was used to cleave the DNA sample (adding a certain proportion of Lambda DNA). A QSeq400 was used to verify cleavage (200–400 bp), and magnetic beads were used for purification. The purified product was subjected to bisulfite treatment and purification using the Method Gold Kit (ZYMO; Irvine, CA, USA). Finally, the product was amplified by PCR and sequenced using the Illumina platform at igenebook (www.ignenbook.com). The sequence of the sample was compared with that of lambda DNA to calculate the bisulfite conversion rate.

We used fastqc to evaluate raw data, and cutadapt [55] was used to remove low-quality bases from the 3′ ends, keeping the N base at no more than 5% of the total length and the length of reads at not less than 30 bp. In the mapping process, clean data were mapped to the *Arabidopsis thaliana* [56] reference genome by BSMAP [57]. The bisulfite conversion rate was calculated according to the result of mapping with lambda DNA. The coverage of cytosine bases across the genome and methylation level of three types of contexts (CG, CHG, CHH) were determined by CGmaptools. A PCA for each treatment group was performed according to the methylation level of each cytosine in the whole genome to evaluate the methylation level of each sample. The DMCs and DMRs were calculated using the CGmaptools dms and CGmaptools dmr tools (p-value < 0.001, $\Delta mC \geq 0.2$) [58]. For DMR-related genes, the differentially expressed genes (DEGs) identified from RNA-seq data ($|\log_2 FC| \geq 2$ and FDR < 0.01) and located within 2000 bp upstream or downstream of the DMR were defined as DMR-related genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were established by the Omicshare cloud platform with a Q-value cutoff of 0.05 (https://www.omicshare.com/tools/ accessed on 9 September 2023).

4.3. Transcriptome Sequencing (RNA-seq) and Data Analysis

Total RNA was extracted from each sample to prepare RNA libraries, and then RNA sequencing libraries were sequenced by IGENEBOOK Biotechnology Co., Ltd. (Wuhan, China) Genes with $|\log_2 FC| \geq 2$ and p-value < 0.05 were defined as differentially expressed genes (DEGs) [59]. The Q-values obtained by Fisher's exact test were adjusted for multiple comparisons with the false discovery rate [60].

4.4. qRT–PCR and Determination of Hormone Content

Col-0 and *atg5-1* plants were grown in complete Murashige and Skoog (MS) medium for 1 week, which was then changed to −N or +N liquid medium. Sixty hours later, plants were harvested from different groups (three biological replicates). Total RNA was extracted by a FastPure Plant Total RNA Isolation Kit (Vazyme, Nanjing, China) and reverse transcribed to cDNA by HiScript III RT SuperMix (Vazyme) according to the manufacturer's protocol. qPCR was performed using a CFX Connect PCR system (Bio-Rad, Hercules, CA, USA) according to the SYBR protocol (ChamQ Universal SYBR qPCR Master Mix, Vazyme), and the expression level was determined by the $2 - \Delta\Delta CT$ method.

An amount of 50 mg of plant sample was weighed into a 2 mL plastic microtube, frozen in liquid nitrogen, and dissolved in 1 mL of methanol/water/formic acid (15:4:1,

v/*v*/*v*). Ten microliters of internal standard mixed solution (100 ng/mL) were added to the extract as an internal standard (IS) for quantification. The mixture was vortexed for 10 min and then centrifuged for 5 min (12,000 r/min, 4 °C). The supernatant was transferred to a clean plastic microtube, evaporated to dryness, dissolved in 100 μL of 80% methanol (*v*/*v*), and passed through a 0.22 μM membrane filter for further LC–MS/MS analysis [61–64].

5. Conclusions

In this study, DNA methylation sequencing and RNA-seq sequencing of the whole genome of *Arabidopsis thaliana* showed that the expression of many genes was widely activated under nitrogen deficiency, especially when the autophagy function was defective, and the level of DNA methylation changed significantly. Under nitrogen deficiency, according to the analysis of DNA hypomethylation positions in autophagic mutants, many hypo-DMRs were found in the promoter regions of genes, which were largely related to the growth and death mechanisms of *Arabidopsis*. In our study, we found significant hypomethylation in the promoter regions of *MYB101* and *ATGH3.12* (*PBS3*). Based on the hypo-DMR sequence, we predicted transcription factors that may bind to the genes mentioned above (*ERF57*, *ERF48*, *ERF111*, *ERF105*, *and AT3G46070*) and verified differential upregulation of *MYB101*, *ATGH3.12* (*PBS3*), and related transcription factors through RNA-seq and qRT–PCR. *MYB101* and *ATGH3.12* (*PBS3*) have been confirmed to be associated with the biosynthesis of ABA and SA and are involved in regulating plant aging. Therefore, we applied LC–MS/MS technology to determine the hormone content of SA and ABA, and the results showed that under nitrogen-deficient growth conditions, the levels of SA and ABA in autophagic mutants were significantly upregulated compared to those in the wild type. By combining the results of WBGS, RNA-seq, and qRT–PCR, we speculated that under nitrogen deficiency, the premature senescence of the autophagy mutant of *Arabidopsis* was probably due to the change in DNA methylation level in the promoter region of *MYB101* and *ATGH3.12* (*PBS3*), leading to the recruitment of ERF family and C2H2-type zinc-finger family transcription factors (AT3G46070) to activate their own expression and then regulate ABA and SA to promote premature senescence.

In addition, in subsequent studies, we should focus on the molecular mechanism of autophagy affecting DNA methylation and regulating key genes in SA- and ABA-related pathways under nitrogen deficiency conditions and evaluate the impact of DNA methylation during plant growth.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ijms241814047/s1.

Author Contributions: The authors confirm their contributions to this work: W.H. and F.L. in conceptualization; Y.S. in original draft preparation, data analysis, and writing; B.Y. and S.C. in sample preparation and qRT–PCR; and W.H. and F.L. in revising and review. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Natural Science Foundation of China (32100297 to F.L.), the Jiangxi Provincial Natural Science Foundation of China (20224BAB215002 to F.L.), the Jiangxi Provincial Introduced Intelligence Program (20212BCJ25024 to F.L.), and the Jiangxi Provincial International Science and Technology Cooperation Program (S2023KJHZH0040 to F.L.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012212 for WGBS; GSA: CRA012214 for RNA-seq) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa/s/InLHU0Ef and https://ngdc.cncb.ac.cn/gsa/s/W6sx307x, accessed on 9 September 2023.

Acknowledgments: We apologize to the authors whose works are not cited because of space limitations.

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

References

1. Zhou, Y.; Manghwar, H.; Hu, W.; Liu, F. Degradation Mechanism of Autophagy-Related Proteins and Research Progress. *Int. J. Mol. Sci.* **2022**, *23*, 7301. [CrossRef] [PubMed]
2. Sienko, K.; Poormassalehgoo, A.; Yamada, K.; Goto-Yamada, S. Microautophagy in Plants: Consideration of Its Molecular Mechanism. *Cells* **2020**, *9*, 887. [CrossRef] [PubMed]
3. Marshall, R.S.; Vierstra, R.D. Autophagy: The Master of Bulk and Selective Recycling. *Annu. Rev. Plant Biol.* **2018**, *69*, 173–208. [CrossRef] [PubMed]
4. Zhang, X.; Ding, X.; Marshall, R.S.; Paez-Valencia, J.; Lacey, P.; Vierstra, R.D.; Otegui, M.S. Reticulon proteins modulate autophagy of the endoplasmic reticulum in maize endosperm. *Elife* **2020**, *9*, e51918. [CrossRef]
5. Suttangkakul, A.; Li, F.; Chung, T.; Vierstra, R.D. The ATG1/ATG13 protein kinase complex is both a regulator and a target of autophagic recycling in *Arabidopsis*. *Plant Cell* **2011**, *23*, 3761–3779. [CrossRef] [PubMed]
6. Thirumalaikumar, V.P.; Gorka, M.; Schulz, K.; Masclaux-Daubresse, C.; Sampathkumar, A.; Skirycz, A.; Vierstra, R.D.; Balazadeh, S. Selective autophagy regulates heat stress memory in *Arabidopsis* by NBR1-mediated targeting of HSP90.1 and ROF1. *Autophagy* **2021**, *17*, 2184–2199. [CrossRef]
7. Shangguan, L.; Fang, X.; Chen, L.; Cui, L.; Fang, J. Genome-wide analysis of autophagy-related genes (ARGs) in grapevine and plant tolerance to copper stress. *Planta* **2018**, *247*, 1449–1463. [CrossRef]
8. Eshkiki, E.M.; Hajiahmadi, Z.; Abedi, A.; Kordrostami, M.; Jacquard, C. In Silico Analyses of Autophagy-Related Genes in Rapeseed (*Brassica napus* L.) under Different Abiotic Stresses and in Various Tissues. *Plants* **2020**, *9*, 1393. [CrossRef] [PubMed]
9. Cheng, S.; Wang, Q.; Manghwar, H.; Liu, F. Autophagy-Mediated Regulation of Different Meristems in Plants. *Int. J. Mol. Sci.* **2022**, *23*, 6236. [CrossRef]
10. Yan, H.; Zhuang, M.; Xu, X.; Li, S.; Yang, M.; Li, N.; Du, X.; Hu, K.; Peng, X.; Huang, W.; et al. Autophagy and its mediated mitochondrial quality control maintain pollen tube growth and male fertility in *Arabidopsis*. *Autophagy* **2023**, *19*, 768–783. [CrossRef]
11. Byun, S.; Seok, S.; Kim, Y.C.; Zhang, Y.; Yau, P.; Iwamori, N.; Xu, H.E.; Ma, J.; Kemper, B.; Kemper, J.K. Fasting-induced FGF21 signaling activates hepatic autophagy and lipid degradation via JMJD3 histone demethylase. *Nat. Commun.* **2020**, *11*, 807. [CrossRef]
12. Gonzalez-Rodriguez, P.; Cheray, M.; Fullgrabe, J.; Salli, M.; Engskog-Vlachos, P.; Keane, L.; Cunha, V.; Lupa, A.; Li, W.; Ma, Q.; et al. The DNA methyltransferase DNMT3A contributes to autophagy long-term memory. *Autophagy* **2021**, *17*, 1259–1277. [CrossRef] [PubMed]
13. Zhang, H.; Lang, Z.; Zhu, J.K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 489–506. [CrossRef] [PubMed]
14. Jiang, C.; Mithani, A.; Belfield, E.J.; Mott, R.; Hurst, L.D.; Harberd, N.P. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis* thaliana mutations and epimutations. *Genome Res.* **2014**, *24*, 1821–1829. [CrossRef]
15. Yong-Villalobos, L.; Gonzalez-Morales, S.I.; Wrobel, K.; Gutierrez-Alanis, D.; Cervantes-Perez, S.A.; Hayano-Kanashiro, C.; Oropeza-Aburto, A.; Cruz-Ramirez, A.; Martinez, O.; Herrera-Estrella, L. Methylome analysis reveals an important role for epigenetic changes in the regulation of the *Arabidopsis* response to phosphate starvation. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E7293–E7302. [CrossRef]
16. Khan, A.R.; Enjalbert, J.; Marsollier, A.C.; Rousselet, A.; Goldringer, I.; Vitte, C. Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (VRN-A1) locus in hexaploid winter wheat. *BMC Plant Biol.* **2013**, *13*, 1–16. [CrossRef] [PubMed]
17. Li, J.; Huang, Q.; Sun, M.; Zhang, T.; Li, H.; Chen, B.; Xu, K.; Gao, G.; Li, F.; Yan, G.; et al. Global DNA methylation variations after short-term heat shock treatment in cultured microspores of *Brassica napus* cv. Topas. *Sci. Rep.* **2016**, *6*, 38401. [CrossRef]
18. Li, S.; Xia, Q.; Wang, F.; Yu, X.; Ma, J.; Kou, H.; Lin, X.; Gao, X.; Liu, B. Laser Irradiation-Induced DNA Methylation Changes Are Heritable and Accompanied with Transpositional Activation of mPing in Rice. *Front. Plant Sci.* **2017**, *8*, 363. [CrossRef]
19. Narsai, R.; Secco, D.; Schultz, M.D.; Ecker, J.R.; Lister, R.; Whelan, J. Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (*Oryza sativa*) coleoptiles under anoxia and re-oxygenation. *Plant J.* **2017**, *89*, 805–824. [CrossRef]
20. Sanchez, D.H.; Paszkowski, J. Heat-induced release of epigenetic silencing reveals the concealed role of an imprinted plant gene. *PLoS Genet.* **2014**, *10*, e1004806. [CrossRef]
21. Bocchini, M.; Bartucca, M.L.; Ciancaleoni, S.; Mimmo, T.; Cesco, S.; Pii, Y.; Albertini, E.; Del Buono, D. Iron deficiency in barley plants: Phytosiderophore release, iron translocation, and DNA methylation. *Front. Plant Sci.* **2015**, *6*, 514. [CrossRef] [PubMed]
22. Henderson, I.R.; Jacobsen, S.E. Epigenetic inheritance in plants. *Nature* **2007**, *447*, 418–424. [CrossRef]
23. Li, S.; Tollefsbol, T.O. DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods* **2021**, *187*, 28–43. [CrossRef] [PubMed]
24. Wang, P.; Nolan, T.M.; Yin, Y.; Bassham, D.C. Identification of transcription factors that regulate ATG8 expression and autophagy in *Arabidopsis*. *Autophagy* **2020**, *16*, 123–139. [CrossRef]

25. Yang, X.; Srivastava, R.; Howell, S.H.; Bassham, D.C. Activation of autophagy by unfolded proteins during endoplasmic reticulum stress. *Plant J.* **2016**, *85*, 83–95. [CrossRef] [PubMed]
26. Stadler, M.B.; Murr, R.; Burger, L.; Ivanek, R.; Lienert, F.; Scholer, A.; van Nimwegen, E.; Wirbelauer, C.; Oakeley, E.J.; Gaidatzis, D.; et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **2011**, *480*, 490–495. [CrossRef]
27. Fernandez-Jimenez, N.; Sklias, A.; Ecsedi, S.; Cahais, V.; Degli-Esposti, D.; Jay, A.; Ancey, P.B.; Woo, H.D.; Hernandez-Vargas, H.; Herceg, Z. Lowly methylated region analysis identifies EBF1 as a potential epigenetic modifier in breast cancer. *Epigenetics* **2017**, *12*, 964–972. [CrossRef]
28. Zou, J.-J.; Cai, X.; Yang, J.; Zeng, X.; Liu, D.-X.; Huang, S.; Chen, X.; Yang, Q.-Y.; Wang, C.; Chen, H. DNA hypomethylation mediates flower opening and senescence in sweet osmanthus through auxin and ethylene responsive pathways. *Postharvest Biol. Technol.* **2023**, *198*, 112250. [CrossRef]
29. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [CrossRef]
30. He, L.; Huang, H.; Bradai, M.; Zhao, C.; You, Y.; Ma, J.; Zhao, L.; Lozano-Duran, R.; Zhu, J.K. DNA methylation-free *Arabidopsis* reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat. Commun.* **2022**, *13*, 1335. [CrossRef]
31. Vatov, E.; Zentgraf, U.; Ludewig, U. Moderate DNA methylation changes associated with nitrogen remobilization and leaf senescence in *Arabidopsis*. *J. Exp. Bot.* **2022**, *73*, 4733–4752. [CrossRef]
32. Manna, I.; Sahoo, S.; Bandyopadhyay, M. Dynamic changes in global methylation and plant cell death mechanism in response to NiO nanoparticles. *Planta* **2023**, *257*, 93. [CrossRef] [PubMed]
33. Nie, W.F. DNA methylation: From model plants to vegetable crops. *Biochem. Soc. Trans.* **2021**, *49*, 1479–1487. [CrossRef] [PubMed]
34. Li, W.; Chen, M.; Wang, E.; Hu, L.; Hawkesford, M.J.; Zhong, L.; Chen, Z.; Xu, Z.; Li, L.; Zhou, Y.; et al. Genome-wide analysis of autophagy-associated genes in foxtail millet (*Setaria italica* L.) and characterization of the function of SiATG8a in conferring tolerance to nitrogen starvation in rice. *BMC Genom.* **2016**, *17*, 797. [CrossRef] [PubMed]
35. Wang, S.; Hu, W.; Liu, F. Autophagy in the Lifetime of Plants: From Seed to Seed. *Int. J. Mol. Sci.* **2022**, *23*, 11410. [CrossRef]
36. Guiboileau, A.; Yoshimoto, K.; Soulay, F.; Bataille, M.P.; Avice, J.C.; Masclaux-Daubresse, C. Autophagy machinery controls nitrogen remobilization at the whole-plant level under both limiting and ample nitrate conditions in *Arabidopsis*. *New Phytol.* **2012**, *194*, 732–740. [CrossRef]
37. Wang, Y.; Cao, J.J.; Wang, K.X.; Xia, X.J.; Shi, K.; Zhou, Y.H.; Yu, J.Q.; Zhou, J. BZR1 Mediates Brassinosteroid-Induced Autophagy and Nitrogen Starvation in Tomato. *Plant Physiol.* **2019**, *179*, 671–685. [CrossRef]
38. Liu, K.; Sutter, B.M.; Tu, B.P. Autophagy sustains glutamate and aspartate synthesis in Saccharomyces cerevisiae during nitrogen starvation. *Nat. Commun.* **2021**, *12*, 57. [CrossRef]
39. Spitz, F.; Furlong, E.E. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **2012**, *13*, 613–626. [CrossRef]
40. Deng, S.; Zhang, J.; Su, J.; Zuo, Z.; Zeng, L.; Liu, K.; Zheng, Y.; Huang, X.; Bai, R.; Zhuang, L.; et al. RNA m(6)A regulates transcription via DNA demethylation and chromatin accessibility. *Nat. Genet.* **2022**, *54*, 1427–1437. [CrossRef]
41. Bolt, S.; Zuther, E.; Zintl, S.; Hincha, D.K.; Schmulling, T. ERF105 is a transcription factor gene of *Arabidopsis* thaliana required for freezing tolerance and cold acclimation. *Plant Cell Env. Environ.* **2017**, *40*, 108–120. [CrossRef] [PubMed]
42. Zhang, Y.; Gao, Y.; Wang, H.L.; Kan, C.; Li, Z.; Yang, X.; Yin, W.; Xia, X.; Nam, H.G.; Li, Z.; et al. Verticillium dahliae secretory effector PevD1 induces leaf senescence by promoting ORE1-mediated ethylene biosynthesis. *Mol. Plant* **2021**, *14*, 1901–1917. [CrossRef] [PubMed]
43. Reyes, J.L.; Chua, N.H. ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J.* **2007**, *49*, 592–606. [CrossRef] [PubMed]
44. Wyrzykowska, A.; Bielewicz, D.; Plewka, P.; Soltys-Kalina, D.; Wasilewicz-Flis, I.; Marczewski, W.; Jarmolowski, A.; Szweykowska-Kulinska, Z. The MYB33, MYB65, and MYB101 transcription factors affect *Arabidopsis* and potato responses to drought by regulating the ABA signaling pathway. *Physiol. Plant* **2022**, *174*, e13775. [CrossRef] [PubMed]
45. Auler, P.A.; Nogueira do Amaral, M.; Rossatto, T.; Lopez Crizel, R.; Milech, C.; Clasen Chaves, F.; Maia Souza, G.; Bolacel Braga, E.J. Metabolism of abscisic acid in two contrasting rice genotypes submitted to recurrent water deficit. *Physiol. Plant* **2021**, *172*, 304–316. [CrossRef]
46. Chen, K.; Li, G.J.; Bressan, R.A.; Song, C.P.; Zhu, J.K.; Zhao, Y. Abscisic acid dynamics, signaling, and functions in plants. *J. Integr. Plant Biol.* **2020**, *62*, 25–54. [CrossRef]
47. Chen, Y.; Feng, P.; Tang, B.; Hu, Z.; Xie, Q.; Zhou, S.; Chen, G. The AP2/ERF transcription factor SlERF.F5 functions in leaf senescence in tomato. *Plant Cell Rep.* **2022**, *41*, 1181–1195. [CrossRef]
48. Chung, K.; Nakano, T.; Fujiwara, S.; Mitsuda, N.; Otsuki, N.; Tsujimoto-Inui, Y.; Naito, Y.; Ohme-Takagi, M.; Suzuki, K. The ERF transcription factor EPI1 is a negative regulator of dark-induced and jasmonate-stimulated senescence in *Arabidopsis*. *Plant Biotechnol.* **2016**, *33*, 235–243. [CrossRef]
49. Liu, J.; Li, J.; Wang, H.; Fu, Z.; Liu, J.; Yu, Y. Identification and expression analysis of ERF transcription factor genes in petunia during flower senescence and in response to hormone treatments. *J. Exp. Bot.* **2011**, *62*, 825–840. [CrossRef]
50. Torrens-Spence, M.P.; Bobokalonova, A.; Carballo, V.; Glinkerman, C.M.; Pluskal, T.; Shen, A.; Weng, J.K. PBS3 and EPS1 Complete Salicylic Acid Biosynthesis from Isochorismate in *Arabidopsis*. *Mol. Plant* **2019**, *12*, 1577–1586. [CrossRef]

51. Niu, F.; Cui, X.; Zhao, P.; Sun, M.; Yang, B.; Deyholos, M.K.; Li, Y.; Zhao, X.; Jiang, Y.Q. WRKY42 transcription factor positively regulates leaf senescence through modulating SA and ROS synthesis in *Arabidopsis thaliana*. *Plant J.* **2020**, *104*, 171–184. [CrossRef] [PubMed]
52. Huang, W.; Hickson, L.J.; Eirin, A.; Kirkland, J.L.; Lerman, L.O. Cellular senescence: The good, the bad and the unknown. *Nat. Rev. Nephrol.* **2022**, *18*, 611–627. [CrossRef] [PubMed]
53. Garcion, C.; Lohmann, A.; Lamodiere, E.; Catinot, J.; Buchala, A.; Doermann, P.; Metraux, J.P. Characterization and biological function of the ISOCHORISMATE SYNTHASE2 gene of *Arabidopsis*. *Plant Physiol.* **2008**, *147*, 1279–1287. [CrossRef]
54. Wildermuth, M.C.; Dewdney, J.; Wu, G.; Ausubel, F.M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **2001**, *414*, 562–565. [CrossRef] [PubMed]
55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *3*, 10–12. [CrossRef]
56. Cheng, C.Y.; Krishnakumar, V.; Chan, A.P.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the *Arabidopsis* thaliana reference genome. *Plant J.* **2017**, *89*, 789–804. [CrossRef]
57. Xi, Y.; Li, W. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinform.* **2009**, *10*, 232. [CrossRef]
58. Guo, W.; Zhu, P.; Pellegrini, M.; Zhang, M.Q.; Wang, X.; Ni, Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **2018**, *34*, 381–387. [CrossRef]
59. Vandenbrink, J.P.; Herranz, R.; Poehlman, W.L.; Alex Feltus, F.; Villacampa, A.; Ciska, M.; Javier Medina, F.; Kiss, J.Z. RNA-seq analyses of *Arabidopsis* thaliana seedlings after exposure to blue-light phototropic stimuli in microgravity. *Am. J. Bot.* **2019**, *106*, 1466–1476. [CrossRef]
60. Liu, N.; Cheng, F.; Zhong, Y.; Guo, X. Comparative transcriptome and coexpression network analysis of carpel quantitative variation in Paeonia rockii. *BMC Genom.* **2019**, *20*, 683. [CrossRef]
61. Cai, B.D.; Zhu, J.X.; Gao, Q.; Luo, D.; Yuan, B.F.; Feng, Y.Q. Rapid and high-throughput determination of endogenous cytokinins in Oryza sativa by bare Fe_3O_4 nanoparticles-based magnetic solid-phase extraction. *J. Chromatogr. A* **2014**, *1340*, 146–150. [CrossRef]
62. Flokova, K.; Tarkowska, D.; Miersch, O.; Strnad, M.; Wasternack, C.; Novak, O. UHPLC-MS/MS based target profiling of stress-induced phytohormones. *Phytochemistry* **2014**, *105*, 147–157. [CrossRef] [PubMed]
63. Li, Y.; Zhou, C.; Yan, X.; Zhang, J.; Xu, J. Simultaneous analysis of ten phytohormones in Sargassum horneri by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry. *J. Sep. Sci.* **2016**, *39*, 1804–1813. [CrossRef] [PubMed]
64. Simura, J.; Antoniadi, I.; Siroka, J.; Tarkowska, D.; Strnad, M.; Ljung, K.; Novak, O. Plant Hormonomics: Multiple Phytohormone Profiling by Targeted Metabolomics. *Plant Physiol.* **2018**, *177*, 476–489. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Brief Report

Marine Bacterioplankton Community Dynamics and Potentially Pathogenic Bacteria in Seawater around Jeju Island, South Korea, via Metabarcoding

Hyun-Jung Kim [1,2], Kang Eun Kim [1,3], Yu Jin Kim [1,3], Hangoo Kang [4], Ji Woo Shin [1], Soohyun Kim [1], Sang Heon Lee [2], Seung Won Jung [1,3,*] and Taek-Kyun Lee [3,5,*]

1 Library of Marine Samples, Korea Institute of Ocean Science & Technology, Geoje 53201, Republic of Korea; hjkim8845@kiost.ac.kr (H.-J.K.); rkddmssl@kiost.ac.kr (K.E.K.); rladbwls06069@kiost.ac.kr (Y.J.K.); sjw3003@kiost.ac.kr (J.W.S.); soohyunk1122@kiost.ac.kr (S.K.)
2 Department of Oceanography and Marine Research Institute, Pusan National University, Busan 46241, Republic of Korea; sanglee@pusan.ac.kr
3 Department of Ocean Science, University of Science & Technology, Daejeon 34113, Republic of Korea
4 Vessel Operation & Observation Team, Korea Institute of Ocean Science & Technology, Geoje 53201, Republic of Korea; hgkang@kiost.ac.kr
5 Ecological Risk Research Department, Korea Institute of Ocean Science & Technology, Geoje 53201, Republic of Korea
* Correspondence: diatoms@kiost.ac.kr (S.W.J.); tklee@kiost.ac.kr (T.-K.L.)

Abstract: Understanding marine bacterioplankton composition and distribution is necessary for improving predictions of ecosystem responses to environmental change. Here, we used 16S rRNA metabarcoding to investigate marine bacterioplankton diversity and identify potential pathogenic bacteria in seawater samples collected in March, May, September, and December 2013 from two sites near Jeju Island, South Korea. We identified 1343 operational taxonomic units (OTUs) and observed that community diversity varied between months. *Alpha-* and *Gamma-proteobacteria* were the most abundant classes, and in all months, the predominant genera were *Candidatus Pelagibacter*, *Leisingera*, and *Citromicrobium*. The highest number of OTUs was observed in September, and *Vibrio* (7.80%), *Pseudoalteromonas* (6.53%), and *Citromicrobium* (6.16%) showed higher relative abundances or were detected only in this month. Water temperature and salinity significantly affected bacterial distribution, and these conditions, characteristic of September, were adverse for *Aestuariibacter* but favored *Citromicrobium*. Potentially pathogenic bacteria, among which *Vibrio* (28 OTUs) and *Pseudoalteromonas* (six OTUs) were the most abundant in September, were detected in 49 OTUs, and their abundances were significantly correlated with water temperature, increasing rapidly in September, the warmest month. These findings suggest that monthly temperature and salinity variations affect marine bacterioplankton diversity and potential pathogen abundance.

Keywords: metabarcoding; bacterioplankton community; 16S rRNA; *Candidatus Pelagibacter*; *Pseudoalteromonas*; *Vibrio*; pathogenic bacteria

Citation: Kim, H.-J.; Kim, K.E.; Kim, Y.J.; Kang, H.; Shin, J.W.; Kim, S.; Lee, S.H.; Jung, S.W.; Lee, T.-K. Marine Bacterioplankton Community Dynamics and Potentially Pathogenic Bacteria in Seawater around Jeju Island, South Korea, via Metabarcoding. *Int. J. Mol. Sci.* **2023**, *24*, 13561. https://doi.org/10.3390/ijms241713561

Academic Editor: Wajid Zaman

Received: 10 August 2023
Revised: 23 August 2023
Accepted: 24 August 2023
Published: 1 September 2023

1. Introduction

Marine bacterioplankton have essential ecosystem functions owing to their ecological connections with other organisms, and their distribution and abundance in marine environments are affected by several environmental and geographic factors [1,2]. Understanding how these environmental factors affect their composition and distribution may help to elucidate their relationships with other organisms and improve predictions of ecosystem responses to environmental changes [3]. In particular, examining the effects of seasonal variations in environmental and geographic factors on community dynamics is essential to elucidate their ecological roles [4–6].

Temporal and spatial changes in bacterioplankton communities in response to environmental change have been examined in recent studies [7,8]. Marine bacterioplankton communities, which are finely structured, exhibit discrete phylogenetic clustering and abundant diversity [9]. Metabarcoding, which characterizes taxonomic diversity and community structure based on the analysis of 16S rRNA gene sequences, has been used to reveal the high diversity of marine bacterioplankton [10–12]. We previously applied metabarcoding to investigate bacterioplankton communities and species composition, monthly changes, and potential pathogenic bacteria [13–18]. For instance, in Goseong Bay (South Korea), *Bacteroidetes* and *Actinobacteria* were identified as the most abundant phyla [19]. Studies on year-round bacterioplankton diversity may help in forecasting seasonal changes [6], and metabarcoding may provide insights into the intricate relationships within bacterioplankton communities owing to changes in various environmental factors [5].

To elucidate these factors, we used metabarcoding to investigate marine bacterioplankton community dynamics and identify potentially pathogenic bacteria in seawater around Jeju Island, South Korea, in four months in 2013. Our findings elucidated the distribution of marine bacterioplankton in the waters around this site and established a baseline for further studies. These findings will improve the prediction of marine bacterioplankton responses to changes in environmental factors.

2. Results and Discussion

2.1. Seawater Environmental Characteristics

We sampled seawater around Jeju Island, South Korea, in March, May, September, and December 2013 to characterize the bacterioplankton community in this area, considering seasonal changes in seawater temperature and salinity. Water temperature (range 17.0–25.3 °C) varied seasonally, consistent with the patterns commonly observed for temperate environments: highest in summer and lowest in winter [20,21]. Salinity ranged from 18.5 to 34.4 and was lowest in September. In South Korea, the summer–autumn rainy season corresponds to the period from mid-August to mid-September, when typhoons affect the Korean Peninsula directly or indirectly [22]. The large input of rainwater during this period rapidly reduces seawater salinity, and this affects the microbial community in this area [23,24].

2.2. Metabarcoding Results

2.2.1. OTU Diversity

Following quality control filtering, we obtained 8003 reads (average number of reads per sample) (Table 1). These reads corresponded to 1343 bacterioplankton operational taxonomic units (OTUs). Ranking by month showed the highest number of OTUs in September, followed by May, December, and March (Figure 1). The high taxonomic diversity and richness in September may be related to the low salinity and high temperature of the seawater within this month. These findings are consistent with those of other studies, which revealed that bacterial species diversity is particularly high in September owing to these climate conditions [25,26]. In autumn, interspecific interactions between bacterioplankton weaken, resulting in lower diversity [24]. Further, the active mixing of the water column during this period alters the water column's physical and chemical parameters; this, in turn, leads to significant seasonal differences in diversity and composition [27].

Table 1. Mean read counts, observed operational taxonomic units (OTUs), and alpha-diversity (Chao1, Shannon, and Simpson indices) for marine bacterioplankton in seawater around Jeju Island, South Korea, in four months in 2013.

Month	Trimmed Reads	OTUs	Chao1	Shannon	Simpson
March	6673	272	406.1	2.92	0.88
May	9782	393	603.4	3.51	0.91
September	9813	567	784.2	4.09	0.93
December	5745	369	625.1	3.16	0.86

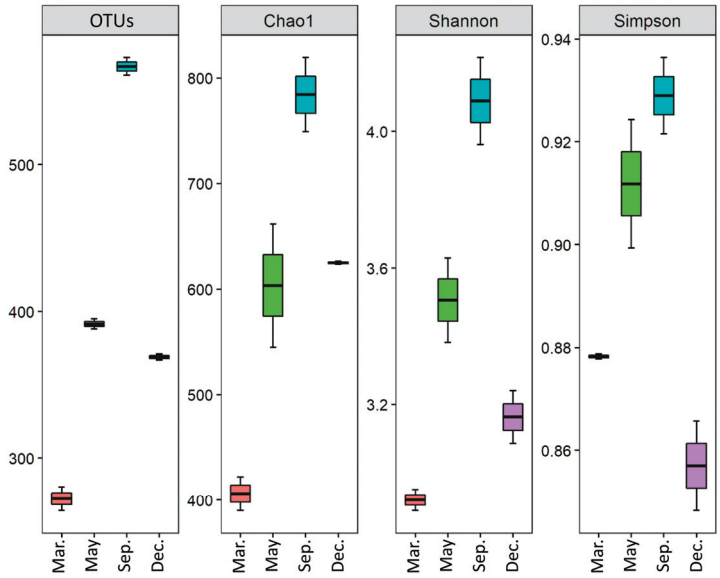


Figure 1. Variation of the number of sequences, operational taxonomic units, and diversity indices (Chao1, Shannon, and Simpson indices) for the marine bacterioplankton community in seawater around Jeju Island, South Korea, in four months in 2013.

2.2.2. Diversity of Bacterioplankton

The identified bacterioplankton OTUs were assigned to 21 phyla. The most abundant phyla, accounting for 98.25% of the total relative abundance, were *Proteobacteria*, *Bacteroidetes*, and *Cyanobacteria* (Figure 2a). The relative abundance of *Proteobacteria* was slightly lower in May (at 78.86%) but was >85% in the other months. The second most abundant phylum was *Bacteroidetes*. Its abundance increased from 12.54% in March to 19.23% in May and declined to 6.39% in September. Further, *Cyanobacteria* were two to five times more abundant in September (3.62%) than in the other months.

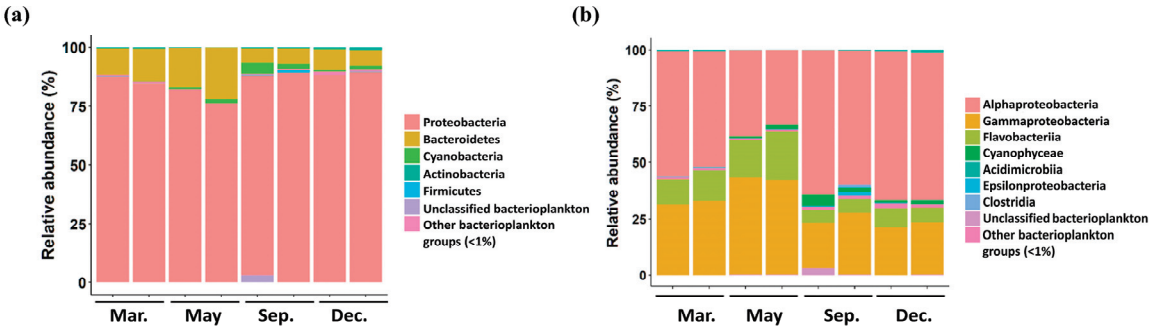


Figure 2. Monthly variation of bacterioplankton community relative abundance in seawater around Jeju Island, South Korea, in four months in 2013. (**a**) Phylum and (**b**) class level. Taxa with relative abundance < 1.0% were categorized as other bacterioplankton groups.

Class-level diversity varied considerably between months (Figure 2b). Specifically, 47 classes were identified, with *Alphaproteobacteria*, *Gammaproteobacteria*, *Flavobacteriia*, and *Cyanophyceae* showing dominance in all of the months. *Alphaproteobacteria* was more abundant in September (61.84%) and December (62.83%) than in May (35.58%) and March (53.61%). *Gammaproteobacteria* and *Flavobacteriia* were more abundant in May (42.66% and 19.03%, respectively) than in the other months. *Cyanophyceae* exhibited 1.48% relative abundance in May and was more abundant in September (3.59%). However, in December, its abundance decreased (1.25%).

2.2.3. Changes in Common Bacterioplankton

The genus-level heat map of the bacterioplankton community revealed distinct changes in diversity over the months (Figure 3a). Of the 719 genera identified, 12 accounted for >5% of the total bacterioplankton relative abundance. Via nonmetric multidimensional scaling (NMDS) analysis, the bacterioplankton community was classified into two groups, considering a 35% similarity level (Figure 3b). Thus, *Candidatus Pelagibacter* (*Alphaproteobacteria*) was identified as the most abundant genus in all the months and was most abundant in December (29.24%). *Aestuariibacter* (*Gammaproteobacteria*), the second most abundant genus, was detected in all months except September and was most abundant in March (23.18%). Further, the highest abundances of *Lentibacter* (8.33%), *Nereida* (7.20%), *Loktanella* (5.54%), and *Aliiroseovarius* (5.44%), belonging to *Alphaproteobacteria*, were recorded in March, and the highest abundances of *Marinomonas* (11.17%; *Gammaproteobacteria*) and *Aurantivirga* (10.16%; *Bacteroidetes*) were recorded in May. September showed the highest abundances of *Vibrio* (7.80%) and *Pseudoalteromonas* (6.53%), which are *Gammaproteobacteria*. Furthermore, *Citromicrobium* (6.16%; *Alphaproteobacteria*) was detected only in September, and December showed the highest abundance of *Leisingera* (22.0%; *Alphaproteobacteria*).

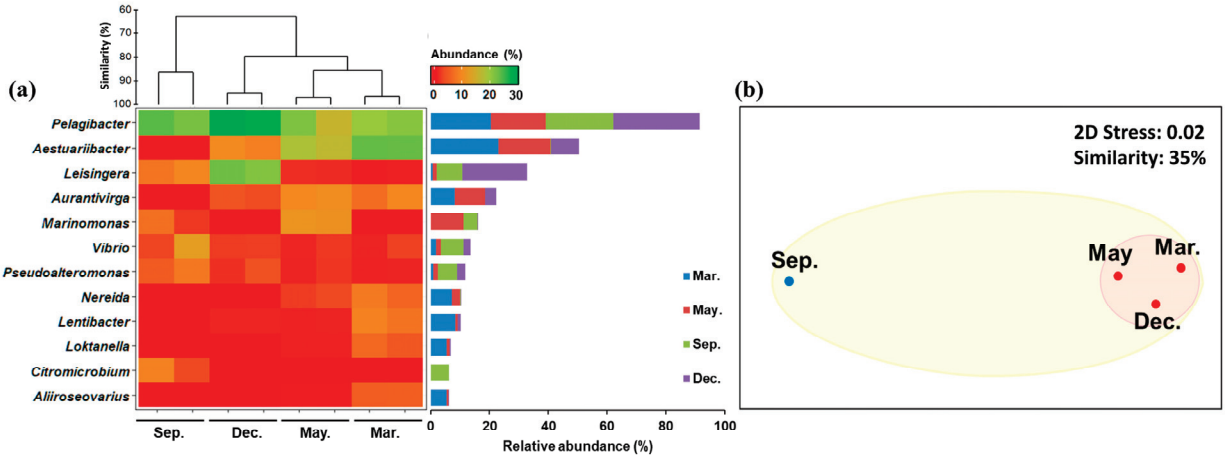


Figure 3. Heatmap of the relative abundances of the operational taxonomic units (OTUs) of common bacterioplankton (for OTUs with relative abundance > 1% in at least one sample) in seawater samples collected around Jeju Island, South Korea, in four different months in 2013. (**a**) Hierarchical agglomerative clustering results based on the group average of the relative abundances of the bacterioplankton OTUs. (**b**) Nonmetric multidimensional scaling (NMDS) plot for the bacterioplankton community based on the Bray–Curtis dissimilarity method. The relative abundances were square-root normalized.

Candidatus Pelagibacter ubique, which accounts for most species of the genus *Pelagibacter*, is distributed in marine environments at temperatures in the range of 12–15 °C and can be cultured at 18 °C [28,29]. In this study area, the water is slightly warmer in September, but still within this range; however, salinity varies substantially, in the range of 18–34. Reportedly, *Candidatus Pelagibacter ubique* grows under oligohaline–mesohaline conditions at salinity values $\geq$ 30 [29,30]. Thus, the temperature and salinity conditions of the seawater at our study site favor the growth of species of the genus *Pelagibacter*.

For *Aestuariibacter halophilus*, the most abundant species of this genus, the optimum temperature and pH for growth are 40 °C and pH 7–8, respectively, but its growth is still possible at 15–40 °C. This species exhibits strict halophilicity; hence, it requires the saline conditions of seawater for growth [31]. Low salinity, therefore, limited its growth in this region in the month of September. *Citromicrobium*, detected only in September, comprised one species, *C. bathyomarinum*, which grows under temperature, salinity, and pH conditions in the ranges 20–42 °C, 0–10, and 6.0–8.0, respectively [32]. Thus, the water around Jeju Island is most suitable for *C. bathyomarinum* in September. Given that bacterial growth

is affected by various environmental factors, it is difficult to clarify the effects of water temperature and salinity in this regard. In this study, however, these two factors were identified as the most important factors affecting bacterial growth and distribution.

We performed LEfSe (Linear discriminant analysis Effect Size) analyses to identify the abundant bacterioplankton taxon based on a comparison between September and the other months (March, May, and December) (Figure 4; Table S1). Based on the results obtained, *Cyanophyceae*, *Alphaproteobacteria* (*Kiloniellales*, *Rhodospirillales*, *Sphingomonadales*), *Saprospirales*, and *Pseudomonadales* were identified as the major bacterioplankton in September. Conversely, *Bacteroidetes* and *Gammaproteobacteria* (*Pseudomonadales*, *Alteromonadales*, and *unclassified Gammaproteobacteria*) constituted the common taxa in Jeju seawaters in the other months.

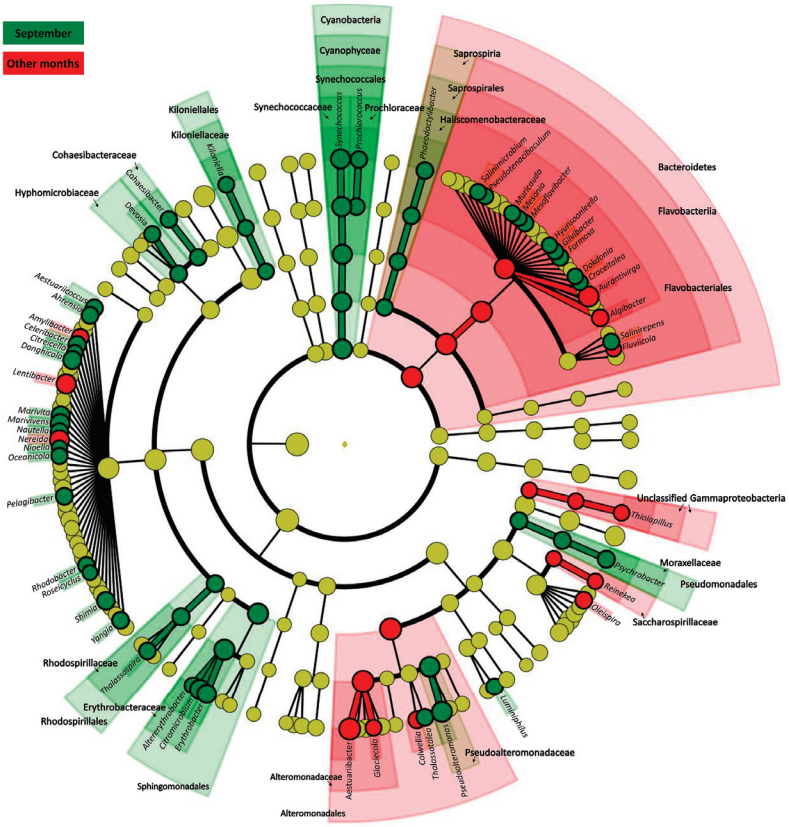


Figure 4. Cladograms of attached bacterioplankton lineages differ significantly between September and other months. The associated bacterioplankton groups at phylum to genus levels are listed from the center to the outside. The circle diameters are proportional to bacterioplankton taxon abundance. Significant discriminatory nodes are colored, and branch areas are shaded according to the highest-ranked group for the given taxon. Green and red areas indicate September and other months, respectively.

2.3. Identification of Potential Pathogens

At the class level, we identified the following potentially pathogenic bacteria: *Gammaproteobacteria* (44 OTUs), *Flavobacteriia* (three OTUs), *Betaproteobacteria* (one OTU), and *Epsilonproteobacteria* (one OTU) (Table S2). In particular, *Vibrio* (28 OTUs) and *Pseudoalteromonas* (six OTUs) were the most abundant potentially pathogenic genera (Figure 5, Table S2), and were most abundant in September. Further, potential pathogenic bacteria showed significantly increased relative abundances in September ($p < 0.05$, one-way ANOVA).

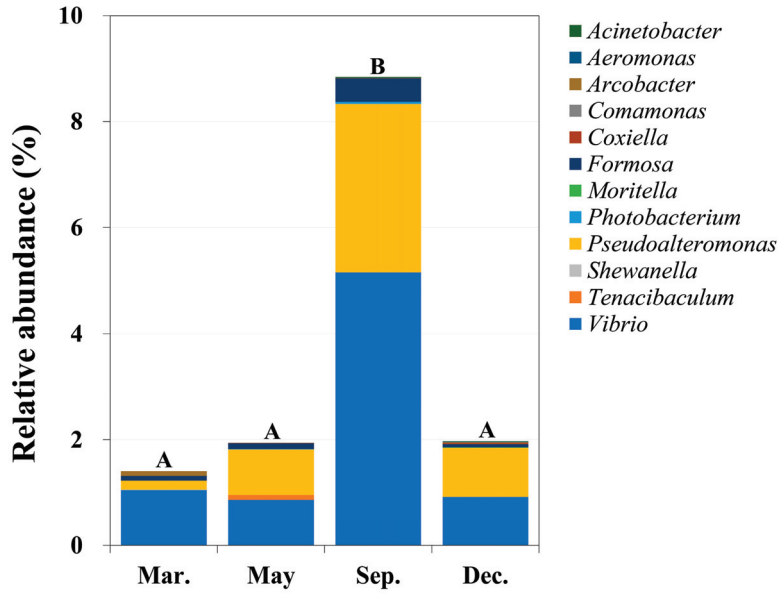


Figure 5. Genus-level relative abundances of potentially pathogenic bacteria in seawater around Jeju Island, South Korea, in four months in 2013. For each genus listed, >1% of the species in the genus are potentially pathogenic. These results were obtained by performing one-way ANOVA and Scheffe's post hoc test. The letters A and B indicate significant differences among months ($p < 0.05$).

The relative abundance of potentially pathogenic bacteria was significantly and positively correlated with water temperature (Spearman correlation analysis; $p < 0.05$). In particular, the relative abundance of *Pseudoalteromonas* showed a highly significant correlation with water temperature ($p < 0.01$, $\rho = 0.835$). For *Pseudoalteromonas*, the optimal temperature for growth is >20 °C [33]. Further, *Vibrio* spp. were 3.5–14 times more abundant in September than in the other months. Reportedly, *Vibrio* grows at temperatures > 20 °C (similar to the temperature of the seawater around Jeju Island in September) and can tolerate salinity in the range 3–37 [34]. It has also been reported that *Vibrio* abundance is strongly associated with water temperature [35,36]. *Vibrio* and *Pseudoalteromonas* were the most common potential pathogens detected in the waters of Kaneohe Bay, Hawaii [37]. These potentially pathogenic bacteria can infect various marine plants and animals, including humans (Table S2). In particular, *V. kanaloae*, *V. tasmaniensis*, *V. chagasii*, *P. tetraodonis*, and *P. nigrifaciens* infect marine animals, such as lobsters, fish, oysters, and sea cucumbers.

3. Materials and Methods

3.1. Sample Collection

Seawater samples (1 L) were collected from a 1 m depth at two sites around Jeju Island (Site 1, 33°22′89″ N, 126°56′53″ E; Site 2, 33°23′53″ N, 126°56′53″ E) on 10 March, 21 May, 8 September, and 10 December 2013 (Figure 6). The seawater temperature and salinity at these sampling points were obtained from the Marine Environment Information System of Korea (http://www.meis.go.kr, accessed on 15 May 2014). Thereafter, 1 L samples of the collected seawater were each passed through a polycarbonate filter (3 μm pore size) (TSTP04700, Millipore, Ireland), harvested, and further passed through a 0.2 μm filter (A020A047A, Advantec MFS Inc., Tokyo, Japan). Next, they were stored at 4 °C until genomic DNA extraction.

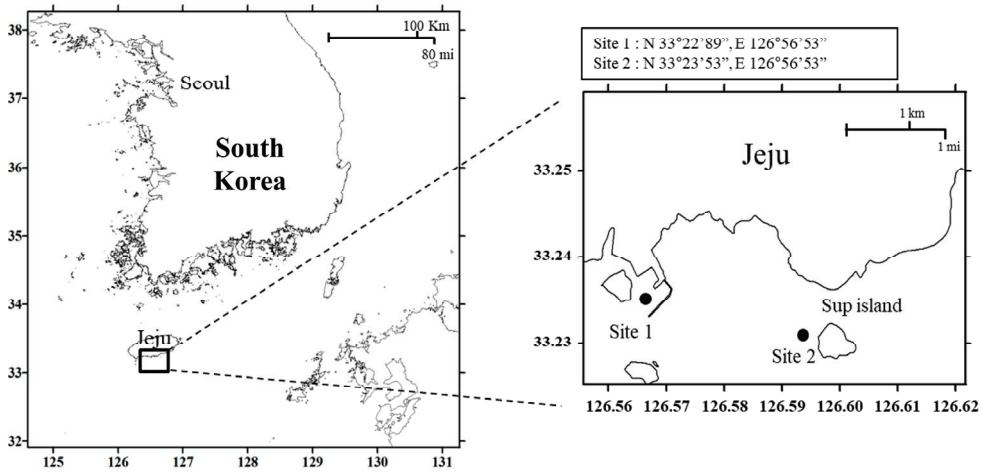


Figure 6. Seawater sampling sites around Jeju Island, South Korea, in four different months in 2013.

3.2. Metabarcoding Analysis

The filters containing the microbes were cut into eight pieces before DNA extraction. Thereafter, total microbial DNA was isolated using the PowerSoil DNA Isolation Kit (MoBio, Solana Beach, CA, USA) according to the manufacturer's instructions. The extracted genomic DNA was then subjected to PCR amplification using primers targeting the V1–V3 region of the 16S rRNA gene [18], i.e., 27F forward primer, 5′-GAG TTT GAT CMT GGC TCA G-3′ and 518R reverse primer, 5′-ATT ACC GCG GCT GCT GG-3′. Partial 16S rDNA gene sequences were used to analyze bacterial diversity. Given that the divergence level varies between the regions of the 16S rDNA gene, the choice of partial sequence regions can significantly affect analysis results [38]. Thus, it was important to determine whether a partial 16S rDNA sequence region could support bacterial characterization as reliably as nearly full-length 16S rDNA genes.

Each primer was tagged using multiplex identifier (MID) adaptors (Roche, Mannheim, Germany) following the manufacturer's instructions. The use of MID adaptors enabled the automatic sorting of metabarcoding-derived sequencing reads. Further, amplification was performed under the following conditions: pre-denaturation at 95 °C for 5 min; 30 cycles of denaturation at 95 °C for 30 s, primer annealing at 55 °C for 30 s, elongation at 72 °C for 30 s; and final elongation at 72 °C for 5 min. The PCR products were purified using a QIAquick PCR Purification Kit (cat. 28106; Qiagen, Hilden, Germany). Similar amounts of extracted PCR products were pooled, and short fragments (non-target products) were eliminated using the AMPure Bead Kit (Agencourt Bioscience, Beverly, MA, USA). PCR product size and quality were further evaluated using a Bioanalyzer 2100 device (Agilent, Palo Alto, Foster City, CA, USA). Thereafter, sequencing was conducted using the 454 GS Junior Sequencing System (Roche Applied Science, Penzberg, Germany) following the manufacturer's instructions.

3.3. Bioinformatics Analysis

Following metabarcoding, bioinformatics analysis was performed as previously described [18]. After sequencing, the quality check was performed to remove short sequence reads (<150 bp), low-quality sequences (quality score < 25), singletons, chloroplast sequences, non-bacterial ribosomal sequences, and chimeras [39,40]. Then, using the Basic Local Alignment Search Tool (BLAST v. 2.14.0), the sequence reads obtained were compared to the sequences in the Silva rRNA database. Similar sequence reads (E-value < 0.0001) were considered partial 16S rDNA sequences, and the taxonomic level (class or genus) of the most similar sequence in the rRNA database was assigned to each of the identified sequence reads. To analyze OTUs, CD-HIT-OTU software was used for clustering [41], while Mothur platform (v 1.35.1) was used to estimate Shannon–Weaver diversity and

Chao1 richness [42]. The taxonomy of the sequence with the highest similarity was assigned to the sequence read (species or genus levels with >98 or >94%, respectively).

3.4. Statistical Analysis and Selection of Potentially Pathogenic Bacteria

Data were presented as the mean of samples from two sampling sites. To compare bacterioplankton community abundances among the four months, hierarchical clustering analysis was performed via "group average" clustering using the Bray–Curtis dissimilarity method. This generated a ranked similarity matrix in which the rows represented the rankings of the column cases based on their similarity to the corresponding row case. OTU class and family relative abundances were square-root normalized for comparison. Normal distributions were assessed using the Kolmogorov–Smirnov test. We also considered OTUs with a relative abundance > 1% in at least one sample as 'abundant' and pooled the remainder.

To examine the relationships between measured parameters, Spearman correlation analysis was employed. An ordination plot was produced via nonmetric multidimensional scaling (NMDS) using a ranked similarity matrix. Further, the clustering, NMDS, and correlation analyses were performed using PRIMER 6 (v 6.1.13). Alpha diversity metrics and OTUs were analyzed using the 'vegan' package [43] in R Studio (v. 1.3.959). Heat maps of the relative abundances of the most abundant OTUs were plotted using ggplot2 [44] in R Studio. The investigated months were compared via one-way analysis of variance (ANOVA), followed by Scheffe's post hoc test. p values < 0.05 were considered statistically significant.

To perform linear discriminant analysis (LDA), the Kruskal–Wallis test was employed to assess differences among classes ($\alpha < 0.05$), while the pairwise Wilcoxon test, acting as a non-parametric analog, was conducted to compare subclasses ($\alpha < 0.05$) [45,46]. The threshold on the logarithmic LDA score was set to 2.0 to determine discriminative features [47,48]. After the analysis was completed, the number of differentially abundant OTUs identified using each tool was assessed at an α level of 0.05. Finally, the potential presence of taxonomic groups that may explain the difference between bacterioplankton communities in different samples was explored using Linear discriminant analysis Effect Size (LEfSe, v 1.0) in the Galaxy framework.

The selection of potentially pathogenic bacteria was based on the literature. The characteristics and references on the basis of which these potential pathogenic bacteria were chosen are shown in Table S2.

4. Conclusions

In this study, we examined the variation in bacterioplankton diversity, community composition, and the presence of potentially pathogenic bacteria in the waters around Jeju Island for four months in 2013. Thus, we observed that community diversity was highest in September, within which the highest abundance of potentially pathogenic bacteria was also observed. This is the first study describing species-level diversity and several pathogenic bacteria in seawater in the study region. These findings regarding potentially pathogenic bacteria can be used to develop an early warning system with respect to marine pathogens.

Supplementary Materials: The supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ijms241713561/s1. References [49–104] are cited in the supplementary materials.

Author Contributions: Conceptualization: S.W.J., and T.-K.L.; data curation: H.-J.K., K.E.K., Y.J.K., H.K., J.W.S., S.K. and S.H.L.; writing—original draft preparation: H.-J.K., S.H.L., S.W.J. and T.-K.L.; formal analysis: H.-J.K., K.E.K., Y.J.K., H.K., J.W.S. and S.K.; data visualization: H.-J.K., K.E.K., Y.J.K., H.K. and J.W.S.; funding acquisition: S.W.J. and T.-K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation of South Korea, grant number NRF-2020R1A2C2005970, and a project titled "Diagnosis, treatment and control technology based on big data of infectious virus in the marine environment" by the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries, South Korea (Ref. No. 21210466).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study can be found in online repositories. The repository and accession number PRJNA999943 (https://www.ncbi.nlm.nih.gov/genbank, accessed on 1 August 2023).

Acknowledgments: We deeply appreciate the Library of Marine Samples of the Korea Institute of Ocean Science and Technology (KIOST), Republic of Korea, for storing the genomic DNA samples.

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

References

1. Nemergut, D.R.; Costello, E.K.; Hamady, M.; Lozupone, C.; Jiang, L.; Schmidt, S.K.; Fierer, N.; Townsend, A.R.; Cleveland, C.C.; Stanish, L. Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* **2011**, *13*, 135–144. [CrossRef]
2. Baek, S.H.; Son, M.; Kim, D.; Choi, H.-W.; Kim, Y.-O. Assessing the ecosystem health status of Korea Gwangyang and Jinhae bays based on a planktonic index of biotic integrity (P-IBI). *Ocean Sci.* **2014**, *49*, 291–311. [CrossRef]
3. Lee, O.O.; Wang, Y.; Yang, J.; Lafi, F.F.; Al-Suwailem, A.; Qian, P.-Y. Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *ISME J.* **2011**, *5*, 650–664. [CrossRef] [PubMed]
4. Chow, C.-E.T.; Sachdeva, R.; Cram, J.A.; Steele, J.A.; Needham, D.M.; Patel, A.; Parada, A.E.; Fuhrman, J.A. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. *ISME J.* **2013**, *7*, 2259–2273. [CrossRef]
5. Cram, J.A.; Chow, C.-E.T.; Sachdeva, R.; Needham, D.M.; Parada, A.E.; Steele, J.A.; Fuhrman, J.A. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* **2015**, *9*, 563–580. [CrossRef] [PubMed]
6. Gilbert, J.A.; Steele, J.A.; Caporaso, J.G.; Steinbrück, L.; Reeder, J.; Temperton, B.; Huse, S.; McHardy, A.C.; Knight, R.; Joint, I. Defining seasonal marine microbial community dynamics. *ISME J.* **2012**, *6*, 298–308. [CrossRef]
7. Schauer, M.; Balagué, V.; Pedrós-Alió, C.; Massana, R. Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquat. Microb. Ecol.* **2003**, *31*, 163–174. [CrossRef]
8. Pinhassi, J.; Hagström, Å. Seasonal succession in marine bacterioplankton. *Aquat. Microb. Ecol.* **2000**, *21*, 245–256. [CrossRef]
9. Acinas, S.G.; Klepac-Ceraj, V.; Hunt, D.E.; Pharino, C.; Ceraj, I.; Distel, D.L.; Polz, M.F. Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **2004**, *430*, 551–554. [CrossRef]
10. Kirchman, D.L.; Cottrell, M.T.; Lovejoy, C. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ. Microbiol.* **2010**, *12*, 1132–1143. [CrossRef] [PubMed]
11. Ladau, J.; Sharpton, T.J.; Finucane, M.M.; Jospin, G.; Kembel, S.W.; O'dwyer, J.; Koeppel, A.F.; Green, J.L.; Pollard, K.S. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **2013**, *7*, 1669–1677. [CrossRef] [PubMed]
12. Aravindraja, C.; Viszwapriya, D.; Karutha Pandian, S. Ultradeep 16S rRNA sequencing analysis of geographically similar but diverse unexplored marine samples reveal varied bacterial community composition. *PLoS ONE* **2013**, *8*, e76724. [CrossRef]
13. Jeong, G.; Kim, H.-J.; Kim, K.E.; Kim, Y.J.; Lee, T.-K.; Shim, W.J.; Jung, S.W. Selective attachment of prokaryotes and emergence of potentially pathogenic prokaryotes on four plastic surfaces: Adhesion study in a natural marine environment. *Mar. Pollut. Bull.* **2023**, *193*, 115149. [CrossRef]
14. Kim, H.-J.; Jeoung, G.; Kim, K.E.; Park, J.S.; Kang, D.; Baek, S.H.; Lee, C.Y.; Kim, H.; Cho, S.; Lee, T.-K. Co-variance between free-living bacteria and *Cochlodinium polykrikoides* (Dinophyta) harmful algal blooms, South Korea. *Harmful Algae* **2023**, *122*, 102371. [CrossRef]
15. Kim, H.-J.; Park, J.S.; Lee, T.-K.; Kang, D.; Kang, J.-H.; Shin, K.; Jung, S.W. Dynamics of marine bacterial biofouling communities after initial *Alteromonas genovensis* biofilm attachment to anti-fouling paint substrates. *Mar. Pollut. Bull.* **2021**, *172*, 112895. [CrossRef] [PubMed]
16. Jung, S.W.; Kang, J.; Park, J.S.; Joo, H.M.; Suh, S.-S.; Kang, D.; Lee, T.-K.; Kim, H.-J. Dynamic bacterial community response to *Akashiwo sanguinea* (Dinophyceae) bloom in indoor marine microcosms. *Sci. Rep.* **2021**, *11*, 6983. [CrossRef] [PubMed]
17. Kang, J.; Park, J.S.; Jung, S.W.; Kim, H.J.; Joo, H.M.; Kang, D.; Seo, H.; Kim, S.; Jang, M.C.; Lee, K.W. Zooming on dynamics of marine microbial communities in the phycosphere of *Akashiwo sanguinea* (Dinophyta) blooms. *Mol. Ecol.* **2021**, *30*, 207–221. [CrossRef]

18. Kim, H.J.; Jung, S.W.; Lim, D.-I.; Jang, M.-C.; Lee, T.-K.; Shin, K.; Ki, J.-S. Effects of temperature and nutrients on changes in genetic diversity of bacterioplankton communities in a semi-closed bay, South Korea. *Mar. Pollut. Bull.* **2016**, *106*, 139–148. [CrossRef] [PubMed]
19. Suh, S.-S.; Park, M.; Hwang, J.; Kil, E.-J.; Jung, S.W.; Lee, S.; Lee, T.-K. Seasonal dynamics of marine microbial community in the South Sea of Korea. *PLoS ONE* **2015**, *10*, e0131633. [CrossRef] [PubMed]
20. Jang, S.H.; Jeong, H.J.; Du Yoo, Y. Gambierdiscus jejuensis sp. nov., an epiphytic dinoflagellate from the waters of Jeju Island, Korea, effect of temperature on the growth, and its global distribution. *Harmful Algae* **2018**, *80*, 149–157. [CrossRef]
21. Kim, S.H.; Choi, B.K.; Kim, E. Study on the behavior of the water temperature inversion layer in the northern East China Sea. *J. Mar. Sci. Eng.* **2020**, *8*, 157. [CrossRef]
22. Qian, W.; Kang, H.-S.; Lee, D.-K. Distribution of seasonal rainfall in the East Asian monsoon region. *Theor. Appl. Climatol.* **2002**, *73*, 151–168. [CrossRef]
23. Yu, B.; Xie, G.; Shen, Z.; Shao, K.; Tang, X. Spatiotemporal variations, assembly processes, and co-occurrence patterns of particle-attached and free-living bacteria in a large drinking water reservoir in China. *Front. Microbiol.* **2023**, *13*, 1056147. [CrossRef] [PubMed]
24. Jiao, C.; Zhao, D.; Zeng, J.; Guo, L.; Yu, Z. Disentangling the seasonal co-occurrence patterns and ecological stochasticity of planktonic and benthic bacterial communities within multiple lakes. *Sci. Total Environ.* **2020**, *740*, 140010. [CrossRef] [PubMed]
25. Rooney-Varga, J.N.; Giewat, M.W.; Savin, M.C.; Sood, S.; LeGresley, M.; Martin, J. Links between phytoplankton and bacterial community dynamics in a coastal marine environment. *Microb. Ecol.* **2005**, *49*, 163–175. [CrossRef]
26. Riemann, L.; Winding, A. Community dynamics of free-living and particle-associated bacterial assemblages during a freshwater phytoplankton bloom. *Microb. Ecol.* **2001**, *42*, 274–285. [CrossRef] [PubMed]
27. Zhang, Y.; Chen, W.; Yang, D.; Jiang, J. The current water environment of Tianmuhu Lake and countermeasures for the sustainable development of the ecological tourism. *Ecol. Environ.* **2003**, *12*, 405–408.
28. Rappé, M.S.; Connon, S.A.; Vergin, K.L.; Giovannoni, S.J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **2002**, *418*, 630–633. [CrossRef]
29. Lankiewicz, T.S.; Cottrell, M.T.; Kirchman, D.L. Growth rates and rRNA content of four marine bacteria in pure cultures and in the Delaware estuary. *ISME J.* **2016**, *10*, 823–832. [CrossRef]
30. Herlemann, D.P.; Woelk, J.; Labrenz, M.; Jürgens, K. Diversity and abundance of "Pelagibacterales"(SAR11) in the Baltic Sea salinity gradient. *Syst. Appl. Microbiol.* **2014**, *37*, 601–604. [CrossRef]
31. Yi, H.; Bae, K.S.; Chun, J. *Aestuariibacter salexigens* gen. nov., sp. nov. and *Aestuariibacter halophilus* sp. nov., isolated from tidal flat sediment, and emended description of *Alteromonas macleodii*. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 571–576. [CrossRef]
32. Yurkov, V.V.; Krieger, S.; Stackebrandt, E.; Beatty, J.T. *Citromicrobium bathyomarinum*, a novel aerobic bacterium isolated from deep-sea hydrothermal vent plume waters that contains photosynthetic pigment-protein complexes. *J. Bacteriol.* **1999**, *181*, 4517–4525. [CrossRef]
33. Sanchez-Porro, C.; Mellado, E.; Bertoldo, C.; Antranikian, G.; Ventosa, A. Screening and characterization of the protease CP1 produced by the moderately halophilic bacterium *Pseudoalteromonas* sp. strain CP76. *Extremophiles* **2003**, *7*, 221–228. [CrossRef]
34. Brinkmeyer, R.; Knittel, K.; Jurgens, J.; Weyland, H.; Amann, R.; Helmke, E. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl. Environ. Microbiol.* **2003**, *69*, 6610–6619. [CrossRef] [PubMed]
35. Park, K.-I.; Paillard, C.; Le Chevalier, P.; Choi, K.-S. Report on the occurrence of brown ring disease (BRD) in Manila clam, *Ruditapes philippinarum*, on the west coast of Korea. *Aquaculture* **2006**, *255*, 610–613. [CrossRef]
36. Osaka, K.; Komatsuzaki, M.; Takahashi, H.; Sakano, S.; Okabe, N. *Vibrio vulnificus* septicaemia in Japan: An estimated number of infections and physicians' knowledge of the syndrome. *Epidemiol. Infect.* **2004**, *132*, 993–996. [CrossRef]
37. Aeby, G.; Ross, M.; Williams, G.; Lewis, T.; Work, T. Disease dynamics of Montipora white syndrome within Kaneohe Bay, Oahu, Hawaii: Distribution, seasonality, virulence, and transmissibility. *Dis. Aquat. Org.* **2010**, *91*, 1–8. [CrossRef] [PubMed]
38. Kunin, V.; Engelbrektson, A.; Ochman, H.; Hugenholtz, P. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **2010**, *12*, 118–123. [CrossRef]
39. Gontcharova, V.; Youn, E.; Wolcott, R.D.; Hollister, E.B.; Gentry, T.J.; Dowd, S.E. Black box chimera check (B2C2): A windows-based software for batch depletion of chimeras from bacterial 16S rRNA gene datasets. *Open Microbiol. J.* **2010**, *4*, 47–52. [CrossRef]
40. Huse, S.M.; Huber, J.A.; Morrison, H.G.; Sogin, M.L.; Welch, D.M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **2007**, *8*, R143. [CrossRef]
41. Li, W.; Fu, L.; Niu, B.; Wu, S.; Wooley, J. Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief Bioinform.* **2012**, *13*, 656–668. [CrossRef]
42. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef] [PubMed]
43. Oksanen, J.; Simpson, G.L.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; et al. "Package 'Vegan'," in Community Ecology Package. 2022. Available online: https://github.com/vegandevs/vegan (accessed on 17 June 2023).
44. Schwarzer, G. Meta-Analysis in R. In *Systematic Reviews in Health Research: Meta-Analysis in Context*, 3rd ed.; Matthias, E., Julian, P.T.H., George, D.S., Eds.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2022; pp. 510–534.

45. Wilcoxon, F. Individual comparisons by ranking methods. In *Breakthroughs in Statistics: Methodology and Distribution*; Kotz, S., Johnson, N.L., Eds.; Springer: New York, NY, USA, 1992; pp. 196–202.
46. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]
47. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188. [CrossRef]
48. Audic, S.; Claverie, J.-M. The significance of digital gene expression profiles. *Genome Res.* **1997**, *7*, 986–995. [CrossRef]
49. Yang, M.; Wang, Q.; Chen, J.; Wu, H. The occurrence of potential pathogenic bacteria on international ships' ballast water at Yangshan Port, Shanghai, China. *Mar. Pollut. Bull.* **2022**, *184*, 114190. [CrossRef]
50. Houf, K.; Stephan, R. Isolation and characterization of the emerging foodborn pathogen Arcobacter from human stool. *J. Microbiol. Methods* **2007**, *68*, 408–413. [CrossRef]
51. Johnson, L.G.; Murano, E.A. Arcobacter Isolates from Various Sources. *J. Food Prot.* **2002**, *65*, 1789–1795. [CrossRef] [PubMed]
52. Levican, A.; Alkeskas, A.; Günter, C.; Forsythe, S.J.; Figueras, M.J. Adherence to and Invasion of Human Intestinal Cells by Arcobacter Species and Their Virulence Genotypes. *Appl. Environ. Microbiol.* **2013**, *79*, 4951–4957. [CrossRef] [PubMed]
53. Goecke, F.; Labes, A.; Wiese, J.; Imhoff, J.F. Phylogenetic analysis and antibiotic activity of bacteria isolated from the surface of two co-occurring macroalgae from the Baltic Sea. *Eur. J. Phycol.* **2013**, *48*, 47–60. [CrossRef]
54. Piñeiro-Vidal, M.; Riaza, A.; Santos, Y. Tenacibaculum discolor sp. nov. and Tenacibaculum gallaicum sp. nov., isolated from sole (*Solea senegalensis*) and turbot (*Psetta maxima*) culture systems. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 21–25. [CrossRef]
55. Fernández-Álvarez, C.; Santos, Y. Phenotypic and Molecular Characterization of Lacinutrix venerupis Isolated from Atlantic Horse Mackerel Trachurus trachurus. *J. Aquat. Anim. Health* **2019**, *31*, 320–327. [CrossRef]
56. Kozińska, A.; Paździor, E.; Pękala, A.; Niemczuk, W. Acinetobacter johnsonii and Acinetobacter lwoffii—The emerging fish pathogens. *J. Vet. Res.* **2014**, *58*, 193–199. [CrossRef]
57. Seifert, H.; Strate, A.; Schulze, A.; Pulverer, G. Vascular Catheter—Related Bloodstream Infection Due to *Acinetobacter johnsonii* (Formerly *Acinetobacter calcoaceticus* var. lwoffii): Report of 13 Cases. *Clin. Infect. Dis.* **1993**, *17*, 632–636. [CrossRef]
58. Manan, H.; Rosland, N.A.; Deris, Z.M.; Hashim, N.F.C.; Kasan, N.A.; Ikhwanuddin, M.; Suloma, A.; Fauzan, F. 16S rRNA sequences of *Exiguobacterium* spp. bacteria dominant in a biofloc pond cultured with whiteleg shrimp, *Penaeus vannamei*. *Aquac Res.* **2022**, *53*, 2029–2041. [CrossRef]
59. Jalali, S.; Kohli, S.; Latka, C.; Bhatia, S.; Vellarikal, S.K.; Sivasubbu, S.; Scaria, V.; Ramachandran, S. Screening Currency Notes for Microbial Pathogens and Antibiotic Resistance Genes Using a Shotgun Metagenomic Approach. *PLoS ONE* **2015**, *10*, e0128711. [CrossRef]
60. Kazar, J. Coxiella burnetii infection. *Ann. N. Y. Acad. Sci.* **2005**, *1063*, 105–114. [CrossRef] [PubMed]
61. Løvoll, M.; Wiik-Nielsen, C.; Tunsjø, H.; Colquhoun, D.; Lunder, T.; Sørum, H.; Grove, S. Atlantic salmon bath challenged with Moritella viscosa—Pathogen invasion and host response. *Fish Shellfish. Immunol.* **2009**, *26*, 877–884. [CrossRef] [PubMed]
62. Martins, P.; Cleary, D.F.R.; Pires, A.C.C.; Rodrigues, A.M.; Quintino, V.; Calado, R.; Gomes, N.C.M. Molecular Analysis of Bacterial Communities and Detection of Potential Pathogens in a Recirculating Aquaculture System for Scophthalmus maximus and Solea senegalensis. *PLoS ONE* **2013**, *8*, e80847. [CrossRef] [PubMed]
63. Urbanczyk, H.; Ogura, Y.; Hendry, T.A.; Gould, A.L.; Kiwaki, N.; Atkinson, J.T.; Hayashi, T.; Dunlap, P.V. Genome Sequence of Photobacterium mandapamensis Strain *svers.* 1.1, the Bioluminescent Symbiont of the Cardinal Fish Siphamia versicolor. *J. Bacteriol.* **2011**, *193*, 3144–3145. [CrossRef]
64. Fichi, G.; Cardeti, G.; Perrucci, S.; Vanni, A.; Cersini, A.; Lenzi, C.; De Wolf, T.; Fronte, B.; Guarducci, M.; Susini, F. Skin lesion-associated pathogens from Octopus vulgaris: First detection of Photobacterium swingsii, Lactococcus garvieae and betanodavirus. *Dis. Aquat. Org.* **2015**, *115*, 147–156. [CrossRef] [PubMed]
65. Li, J.; Weinberger, F.; Saha, M.; Majzoub, M.E.; Egan, S. Cross-Host Protection of Marine Bacteria Against Macroalgal Disease. *Microb. Ecol.* **2021**, *84*, 1288–1293. [CrossRef]
66. Weiland-Bräuer, N.; Pinnow, N.; Langfeldt, D.; Roik, A.; Güllert, S.; Chibani, C.M.; Reusch, T.B.H.; Schmitz, R.A. The Native Microbiome is Crucial for Offspring Generation and Fitness of *Aurelia aurita*. *mBio* **2020**, *11*. [CrossRef] [PubMed]
67. Li, J.; Mou, Z.; Yang, H.; Mao, Y.; Yan, Y.; Mo, Z. Isolation and identification the pathogen of Pyropia yezoensis green spot disease. *Prog. Fish. Sci.* **2019**, *40*, 140–146.
68. Wang, Y. Etiology of skin ulcer syndrome in cultured juveniles of Apostichopus japonicus and analysis of reservoir of the pathogens. *J. Fish. Sci. China* **2006**, *13*. [CrossRef]
69. Simidu, U.; Kita-Tsukamoto, K.; Yasumoto, T.; Yotsu, M. Taxonomy of four marine bacterial strains that produce tetrodotoxin. *Int. J. Syst. Evol. Microbiol.* **1990**, *40*, 331–336. [CrossRef]
70. Ivanova, E.P.; Romanenko, L.A.; Matte, M.H.; Matte, G.R.; Lysenko, A.M.; Simidu, U.; Kita-Tsukamoto, K.; Sawabe, T.; Vysotskii, M.V.; Frolova, G.M. Retrieval of the species Alteromonas tetraodonis Simidu et al. 1990 as *Pseudoalteromonas tetraodonis* comb. nov. and emendation of description. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 1071–1078. [CrossRef] [PubMed]
71. Liu, H.; Zheng, F.; Sun, X.; Hong, X.; Dong, S.; Wang, B.; Tang, X.; Wang, Y. Identification of the pathogens associated with skin ulceration and peristome tumescence in cultured sea cucumbers Apostichopus japonicus (Selenka). *J. Invertebr. Pathol.* **2010**, *105*, 236–242. [CrossRef]

72. Popović, N.T.; Kazazić, S.; Bilić, B.; Babić, S.; Bojanić, K.; Bujak, M.; Bujak, I.T.; Jadan, M.; Strunjak-Perović, I.; Kepec, S.; et al. Shewanella spp. from wastewater treatment plant-affected environment: Isolation and characterization. *Environ. Sci. Pollut. Res.* **2022**, *29*, 82986–83003. [CrossRef]
73. Labreuche, Y.; Lambert, C.; Soudant, P.; Boulo, V.; Huvet, A.; Nicolas, J.-L. Cellular and molecular hemocyte responses of the Pacific oyster, Crassostrea gigas, following bacterial infection with *Vibrio aestuarianus* strain 01/32. *Microbes Infect.* **2006**, *8*, 2715–2724. [CrossRef]
74. Costa, J.C.C.P.; Floriano, B.; Villegas, I.M.B.; Rodríguez-Ruiz, J.P.; Posada-Izquierdo, G.D.; Zurera, G.; Pérez-Rodríguez, F. Study of the microbiological quality, prevalence of foodborne pathogens and product shelf-life of Gilthead sea bream (*Sparus aurata*) and Sea bass (*Dicentrarchus labrax*) from aquaculture in estuarine ecosystems of Andalusia (Spain). *Food Microbiol.* **2020**, *90*, 103498. [CrossRef]
75. Li, S.-Y.; Huang, Y.-E.; Chen, J.-Y.; Lai, C.-H.; Mao, Y.-C.; Huang, Y.-T.; Liu, P.-Y. Genomics of Ochrobactrum pseudogrignonense (newly named Brucella pseudogrignonensis) reveals a new bla OXA subgroup. *Microb. Genom.* **2021**, *7*, 000626. [CrossRef]
76. Li, S.; Wang, S.; Xie, L.; Liu, Y.; Chen, H.; Feng, J.; Ouyang, L. Identification and Optimization of the Algicidal Activity of a Novel Marine Bacterium Against Akashiwo sanguinea. *Front. Mar. Sci.* **2022**, *9*. [CrossRef]
77. Wang, L.; Chen, Y.; Huang, H.; Huang, Z.; Chen, H.; Shao, Z. Isolation and identification of Vibrio campbellii as a bacterial pathogen for luminous vibriosis of Litopenaeus vannamei. *Aquac. Res.* **2015**, *46*, 395–404. [CrossRef]
78. Austin, B.; Austin, D.; Sutherland, R.; Thompson, F.; Swings, J. Pathogenicity of vibrios to rainbow trout (*Oncorhynchus mykiss*, Walbaum) and Artemia nauplii. *Environ. Microbiol.* **2005**, *7*, 1488–1495. [CrossRef]
79. Pérez-Cataluña, A.; Lucena, T.; Tarazona, E.; Arahal, D.R.; Macián, M.C.; Pujalte, M.J. An MLSA approach for the taxonomic update of the Splendidus clade, a lineage containing several fish and shellfish pathogenic *Vibrio* spp. *Syst. Appl. Microbiol.* **2016**, *39*, 361–369. [CrossRef]
80. Bruto, M.; James, A.; Petton, B.; Labreuche, Y.; Chenivesse, S.; Alunno-Bruscia, M.; Polz, M.F.; Le Roux, F. Vibrio crassostreae, a benign oyster colonizer turned into a pathogen after plasmid acquisition. *ISME J.* **2016**, *11*, 1043–1052. [CrossRef] [PubMed]
81. Song, J.; Liu, X.; Wu, C.; Zhang, Y.; Fan, K.; Zhang, X.; Wei, Y. Isolation, identification and pathogenesis study of Vibrio diabolicus. *Aquaculture* **2020**, *533*, 736043. [CrossRef]
82. Ramamurthy, T.; Chowdhury, G.; Pazhani, G.P.; Shinoda, S. Vibrio fluvialis: An emerging human pathogen. *Front. Microbiol.* **2014**, *5*, 91. [CrossRef]
83. Wang, X.; Zhang, Y.; Qin, G.; Luo, W.; Lin, Q. A novel pathogenic bacteria (*Vibrio fortis*) causing enteritis in cultured seahorses, Hippocampus erectus Perry, 1810. *J. Fish Dis.* **2016**, *39*, 765–769. [CrossRef] [PubMed]
84. Roux, F.L.; Goubet, A.; Thompson, F.; Faury, N.; Gay, M.; Swings, J.; Saulnier, D. *Vibrio gigantis* sp. nov., isolated from the haemolymph of cultured oysters (*Crassostrea gigas*). *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 2251–2255. [CrossRef] [PubMed]
85. Sicuro, B.; Luzzana, U. The state of Seriola spp. other than yellowtail (*S. quinqueradiata*) farming in the world. *Rev. Fish. Sci. Aquac.* **2016**, *24*, 314–325. [CrossRef]
86. Harrison, J.; Nelson, K.; Morcrette, H.; Morcrette, C.; Preston, J.; Helmer, L.; Titball, R.W.; Butler, C.S.; Wagley, S. The increased prevalence of Vibrio species and the first reporting of Vibrio jasicida and Vibrio rotiferianus at UK shellfish sites. *Water Res.* **2022**, *211*, 117942. [CrossRef]
87. Huang, B.; Zhang, X.; Wang, C.; Bai, C.; Li, C.; Li, C.; Xin, L. Isolation and characterization of Vibrio kanaloae as a major pathogen associated with mass mortalities of Ark Clam, Scapharca broughtonii, in cold season. *Microorganisms* **2021**, *9*, 2161. [CrossRef] [PubMed]
88. Schaeck, M.; Duchateau, L.; Broeck, W.V.D.; Van Trappen, S.; De Vos, P.; Coulombet, C.; Boon, N.; Haesebrouck, F.; Decostere, A. Vibrio lentus protects gnotobiotic sea bass (*Dicentrarchus labrax* L.) larvae against challenge with Vibrio harveyi. *Vet. Microbiol.* **2016**, *185*, 41–48. [CrossRef] [PubMed]
89. Farto, R.; Armada, S.; Montes, M.; Guisande, J.; Pérez, M.; Nieto, T. Vibrio lentus associated with diseased wild octopus (*Octopus vulgaris*). *J. Invertebr. Pathol.* **2003**, *83*, 149–156. [CrossRef] [PubMed]
90. Andree, K.B.; Carrasco, N.; Carella, F.; Furones, D.; Prado, P. Vibrio mediterranei, a potential emerging pathogen of marine fauna: Investigation of pathogenicity using a bacterial challenge in Pinna nobilis and development of a species-specific PCR. *J. Appl. Microbiol.* **2021**, *130*, 617–631. [CrossRef] [PubMed]
91. Torres, M.; Reina, J.C.; Fuentes-Monteverde, J.C.; Fernández, G.; Rodríguez, J.; Jiménez, C.; Llamas, I. AHL-lactonase expression in three marine emerging pathogenic Vibrio spp. reduces virulence and mortality in brine shrimp (*Artemia salina*) and Manila clam (*Venerupis philippinarum*). *PLoS ONE* **2018**, *13*, e0195176. [CrossRef]
92. Mondal, S.K.; Lijon, B.; Reza, R.; Ishika, T. Isolation and identification of Vibrio nereis and Vibrio harveyi in farm raised Penaeus monodon marine shrimp. *Int. J. Biosci.* **2016**, *8*, 55–61.
93. Ruiz, P.; Balado, M.; Fuentes-Monteverde, J.C.; Toranzo, A.E.; Rodríguez, J.; Jiménez, C.; Avendaño-Herrera, R.; Lemos, M.L. The Fish Pathogen *Vibrio ordalii* Under Iron Deprivation Produces the Siderophore Piscibactin. *Microorganisms* **2019**, *7*, 313. [CrossRef]
94. Amin, A.M.R.; Feng, G.; Al-Saari, N.; Meirelles, P.M.; Yamazaki, Y.; Mino, S.; Thompson, F.L.; Sawabe, T.; Sawabe, T. The First Temporal and Spatial Assessment of Vibrio Diversity of the Surrounding Seawater of Coral Reefs in Ishigaki, Japan. *Front. Microbiol.* **2016**, *7*, 1185. [CrossRef] [PubMed]
95. Villamil, L.; Figueras, A.; Aranguren, R.; Novoa, B. Non-specific immune response of turbot, *Scophthalmus maximus* (L.), experimentally infected with a pathogenic Vibrio pelagius. *J. Fish Dis.* **2003**, *26*, 321–329. [CrossRef] [PubMed]

96. Guisande, J.A.; Lago, E.P.; Prado, S.; Nieto, T.P.; Seguín, R.F. Genotypic Diversity of Culturable Vibrio Species Associated with the Culture of Oysters and Clams in Galicia and Screening of Their Pathogenic Potential. *J. Shellfish. Res.* **2008**, *27*, 801–809. [CrossRef]
97. Verschuere, L.; Heang, H.; Criel, G.; Sorgeloos, P.; Verstraete, W. Selected bacterial strains protect *Artemia* spp. from the pathogenic effects of Vibrio proteolyticus CW8T2. *Appl. Environ. Microbiol.* **2000**, *66*, 1139–1146. [CrossRef]
98. Ray, A.; Kinch, L.N.; Santos, M.d.S.; Grishin, N.V.; Orth, K.; Salomon, D. Proteomics Analysis Reveals Previously Uncharacterized Virulence Factors in Vibrio proteolyticus. *mBio* **2016**, *7*. [CrossRef]
99. Gökbulak, F. Effect of American bison (*Bison bison* L.) on the recovery and germinability of seeds of range forage species. *Grass Forage Sci.* **2002**, *57*, 395–400. [CrossRef]
100. Qiao, G.; Jang, I.-K.; Won, K.M.; Woo, S.H.; Xu, D.-H.; Park, S.I. Pathogenicity comparison of high- and low-virulence strains of Vibrio scophthalmi in olive flounder Paralichthys olivaceus. *Fish. Sci.* **2012**, *79*, 99–109. [CrossRef]
101. Duperthuy, M.; Schmitt, P.; Garzón, E.; Caro, A.; Rosa, R.D.; Le Roux, F.; Lautrédou-Audouy, N.; Got, P.; Romestand, B.; de Lorgeril, J.; et al. Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2993–2998. [CrossRef]
102. Kesarcodi-Watson, A.; Kaspar, H.; Lategan, M.; Gibson, L. Two pathogens of Greenshell™ mussel larvae, Perna canaliculus: Vibrio splendidus and a *V. coralliilyticus*/neptunius-like isolate. *J. Fish Dis.* **2009**, *32*, 499–507. [CrossRef] [PubMed]
103. Robino, E.; Poirier, A.C.; Amraoui, H.; Le Bissonnais, S.; Perret, A.; Lopez-Joven, C.; Auguet, J.C.; Rubio, T.P.; Cazevieille, C.; Rolland, J.L. Resistance of the oyster pathogen Vibrio tasmaniensis LGP32 against grazing by *Vannella* sp. marine amoeba involves Vsm and CopA virulence factors. *Environ. Microbiol.* **2020**, *22*, 4183–4197. [CrossRef]
104. Ho, H.T.; Lipman, L.J.; Gaastra, W. Arcobacter, what is known and unknown about a potential foodborne zoonotic agent! *Vet. Microbiol.* **2006**, *115*, 1–13. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Article

Analyzing the Antinociceptive Effect of Interleukin-31 in Mice

Iwao Arai [1,2,*], Minoru Tsuji [1], Kohei Takahashi [1], Saburo Saito [2] and Hiroshi Takeda [1]

[1] Department of Pharmacology, International University of Health and Welfare, 2600-1 Kitakanemaru, Ohtawara 324-8510, Japan; takahashi-k@iuhw.ac.jp (K.T.)
[2] Division of Environmental Allergy, The Jikei University School of Medicine, 3-25-8 Nishi-Shinbashi, Tokyo 105-8461, Japan; misaburo@jikei.ac.jp
* Correspondence: i.arai@iuhw.ac.jp; Tel.: +81-287-24-3481

Abstract: The theory that an itch inhibits pain has been refuted; however, previous research did not investigate this theory for an interleukin-31 (IL-31)-induced itch. Previously, we have found that morphine-induced antinociception was partially reduced in IL-31 receptor A (IL-31RA)-deficient (IL-31RAKI) mice, indicating that IL-31RA may play an important role in morphine-induced peripheral antinociception. In the present study, we evaluated the effect of IL-31-induced analgesia on a 2,4,6-trinitrochlorobenzene (TNCB)-sensitized mice using a hot-plate test. This test evaluated the antinociceptive activity of morphine and non-steroidal anti-inflammatory drugs (NSAIDs). Repeated pretreatment with IL-31 showed significant antinociceptive action. Furthermore, its combination with morphine, but not with NSAIDs, increased the analgesic action. In contrast, treatment with TNCB and capsaicin decreased antinociception. Moreover, TNCB increased IL-31RA expression in the dorsal root ganglia at 24 h, whereas capsaicin inhibited it. The comparative action of several analgesics on TNCB or capsaicin was evaluated using a hot-plate test, which revealed that the antinociceptive activity was decreased or disappeared in response to capsaicin-induced pain in IL-31RAKI mice. These results indicate that the analgesic action of IL-31 involves the peripheral nervous system, which affects sensory nerves. These results provide a basis for developing novel analgesics using this mechanism.

Keywords: analgesia; alloknesis; antinociception; interleukin-31 (IL-31); interleukin receptor A (IL-31RA); IL-31 receptor A-deficient (IL-31RAKI) mice; itch; pain

Citation: Arai, I.; Tsuji, M.; Takahashi, K.; Saito, S.; Takeda, H. Analyzing the Antinociceptive Effect of Interleukin-31 in Mice. *Int. J. Mol. Sci.* **2023**, *24*, 11563. https://doi.org/10.3390/ijms241411563

Academic Editor: Wajid Zaman

Received: 9 May 2023
Revised: 4 July 2023
Accepted: 11 July 2023
Published: 17 July 2023

1. Introduction

Interleukin-31 (IL-31), a possible mediator of itching, induces severe pruritus and dermatitis in mice [1]. Elevated cutaneous IL-31 expression levels have been observed in lesional skin atopic dermatitis [2,3]. Moreover, the repeated administration of IL-31 causes itch-associated scratching behavior, which is significantly increased with the increased expression of IL-31 receptor A (IL-31RA) in the dorsal root ganglia (DRG) [4]. The interaction between cutaneous IL-31 and neuronal DRG IL-31RA causes severe itch-associated scratching behavior (long-lasting scratching, LLS) [5].

The central and peripheral nervous systems (CNS and PNS, respectively) are well-known sites of action for analgesia. Morphine, an opioid, induces analgesia by acting on both the CNS and PNS [6]. It inhibits the release of neurotransmitters from the primary afferent terminals in the spinal cord and activates the descending inhibitory controls in the midbrain [7,8]. However, the site of action of morphine in the PNS and its role in regulating pain transmission remain unclear. The administration of morphine increases the scratch counts within 10 min, followed by a return to basal levels approximately 90 min after administration. A close correlation has been observed between time-course changes in morphine-induced LLS counts and antinociceptive activity [9]. Moreover, IL-31 causes characteristic LLS upon administration [5]. In IL-31RA-deficient (IL-31 receptor $A^{tLacZ/+}$ knock-in) mice, the administration of morphine results in the disappearance of LLS and

the partial disappearance of antinociceptive activity [9]. Additionally, recent findings have indicated that IL-31 is partially involved in the peripheral analgesic mechanism. As IL-31 and IL-31RA are not expressed in the CNS, the endogenous opioid system is activated under pathological conditions. Moreover, morphine-induced antinociceptive activity in the peripheral site is better than that in the central site of morphine action because central untoward adverse effects, such as respiratory depression, somnolence, and addiction, are avoided [10,11]. Although non-steroidal anti-inflammatory drugs (NSAIDs) also mediate a peripheral analgesic effect, they inhibit peripheral inflammation. In contrast, morphine has no anti-inflammatory action on the peripheral site.

The hot-plate test using mice is simple and easy to perform [12] and is generally used for strong-acting analgesics, such as opioids, but not for peripherally acting drugs. Currently, a few experimental models can evaluate peripherally acting analgesic drugs in mice. A well-known model for analgesic action for NSAIDs is paw inflammation induced by an injection of Freund's complete adjuvant [13] or carrageenin [14] and a subsequent hyperalgesia assessment in rats [15,16]. However, it is substantially more difficult to perform long-term tests in mice. Moreover, although these methods cause inflammation, they can be used to evaluate even weak analgesics, such as NSAIDs and acetaminophen, if the pain threshold is lowered. Acute contact dermatitis is induced after a single application of hapten, dinitrochrolbenzen (DNCB) [17]. A modified hot-plate test on the DNCB derivative 2,4,6-trinitrochlorobenzene (TNCB) has been previously performed [18]. TNCB was applied to the limbs of mice to reduce the pain threshold, which allowed the evaluation of antinociceptive activity at a temperature lower than that used in the hot-plate test [19]. Additionally, it was possible to evaluate the antinociceptive activity of NSAIDs and IL-31 using TNCB.

Itching elicits a strong desire to scratch; therefore, scratching behavior count is a useful index to evaluate itching [20]. Mice exhibit two types of scratching behavior, long-lasting scratching (LLS, scratching behavior lasting more than 1.0 s) and short-lasting scratching (SLS, scratching behavior lasting from 0.3 to 1.0 s). Current studies on itching are based on the human perceptive sense, and these nociceptive stimuli have no discernible differences. However, the sensory perception of a foreign substance and true itching could be differentiated by dividing the scratching behavior of mice into LLS and SLS [21]. The number of spontaneous scratches can be automatically detected and objectively evaluated via a computer. Based on our evaluation standard, we have previously found that histamine is not a pruritogen [22].

Capsaicin acts on the transient receptor potential cation channel vanilloid subfamily V member 1 (TRPV1), which modulates nociceptive inputs to the spinal cord and brain stem centers integrating diverse painful stimuli [23,24]. The sensation of itching can be reduced by the painful sensations caused by scratching [25]. The inhibition of itching through painful stimuli has been experimentally demonstrated using various painful stimuli. NC/Nga mice, an animal model of atopic dermatitis with chronic itching, spontaneously develop skin lesions [26]. We have previously demonstrated that cutaneous prostaglandins (PGs) levels are significantly elevated upon scratching the mouse skin with a stainless-steel wire brush (mechanical scratching), and these PGs suppressed LLS in skin-lesioned NC/Nga mice [27,28]. Because PGs are associated with inflammation, their administration enhances pain [29]. Notably, although the pain-induced suppression of itching is temporary, the effect of capsaicin lasts more than 72 h, suggesting that this effect may not only be due to pain but also another action of capsaicin. The partial activation of TRPV1 by capsaicin results in pain, which, in turn, may partially suppress itching sensations [30]. Paradoxically, the application of capsaicin produces transient burning pain and induces analgesia for neuropathic pain. The mechanisms underlying these opposing actions of capsaicin on the onset of pain and the induction of analgesia have not yet been explained. The application of capsaicin causes the desensitization of TRPV1 [31], the inhibition of nociceptor firing [32], and a decrease in mechanotransduction [33]. These early effects of capsaicin on the function of primary afferents might contribute to the analgesic

effects immediately after capsaicin application. However, these effects may be unrelated to suppressing itching sensations because there are no reports of an analgesic action inhibiting itching for more than 72 h following capsaicin application. Recently, we reported that capsaicin suppresses LLS and SLS for more than 72 h; at this point, the expression of DRG IL-31RA mRNA is significantly decreased, whereas that of cutaneous IL-31RA shows no significant change. Thus, capsaicin may suppress LLS by inhibiting IL-31RA mRNA expression in the DRG [34].

Our previous report suggested the partial involvement of IL-31 in the antinociceptive action of morphine [9]. Moreover, we investigated the effects of pretreatment with IL-31 on morphine-induced antinociceptive activity using a modified hot-plate test and found a close correlation between morphine-induced LLS and antinociceptive effect. We also investigated LLS as an indicator of itching by morphine-induced itching and found that the repeated administration of IL-31 gradually promoted LLS in a dose-dependent manner and increased DRG neuronal IL-31RA expression. These data show that cutaneously-injected IL-31 induces LLS and promotes DRG IL-31RA expression [4].

The present study aimed to reveal the mechanism of action of IL-31 on TNCB- or capsaicin-induced pain models and assess the long-term physiological effects of its application. Our results provide important insights into the development of a new type of algesic based on the antinociceptive mechanism of IL-31.

2. Results

2.1. Effect of IL-31 on TNCB Applied Hot-Plate Test in BALB/c Mice

In the conventional hot-plate test (51 °C), IL-31 did not show significant antinociceptive activity. In contrast, in the TNCB-applied (3%, 45 °C) hot-plate test [19], the latency rapidly decreased 30 min after the TNCB application (Figure 1a, red line). Moreover, a single pretreatment injection with IL-31 (50 μg/kg, intraperitoneal) showed slight antinociception, which was not significant compared with that of the vehicle (phosphate-buffered saline, PBS, 10 mL/kg)-treated group. However, repeated pretreatment with IL-31 (50 μg/kg, intraperitoneal, every 12 h for 3 days) significantly increased the antinociceptive action on the TNCB-applied hot-plate test at 45 °C (Figure 1a, green line) and the total antinociceptive index of IL-31 after 0.5–6 h ($AUC_{0.5–6\,h}$) of TNCB application compared with the vehicle (acetone–ethanol mixed liquor, AEM)-treated group (Figure 1b, red column). These results suggest that the repeated administration of IL-31 was a suitable condition for the evaluation of the antinociceptive effect. Therefore, we performed subsequent experiments with repeated administration of IL-31 (50 μg/kg; Figure 1a, intraperitoneal, every 12 h for 3 days).

In the vehicle (PBS, 10 mL/kg, intraperitoneal) + AEM-treated group, a few LLS counts were observed (Figure 1c, blue line). And, these LLS counts decreased after TNCB, which was not significant compared with those of the AEM-treated group (Figure 1c, yellow line). On the other hand, a single large dose injection of IL-31 (1 mg/kg, intraperitoneal) increased LLS counts, which gradually increased 4 h after IL-31 administration. This increase in LLS counts showed a circadian rhythm; in particular, the LLS counts significantly increased at nighttime. Moreover, the application of 0.2 mL of TNCB to the dorsal skin surface significantly decreased LLS counts immediately after or within 6 h after application (Figure 1c, red line). The total LLS counts also significantly decreased in the IL-31 + TNCB group compared with those in the IL-31 + AEM-treated group (Figure 1d, yellow column). However, the total LLS counts of IL-31 + TNCB were not significantly changed compared with those in the PBS + TNCB-treated group.

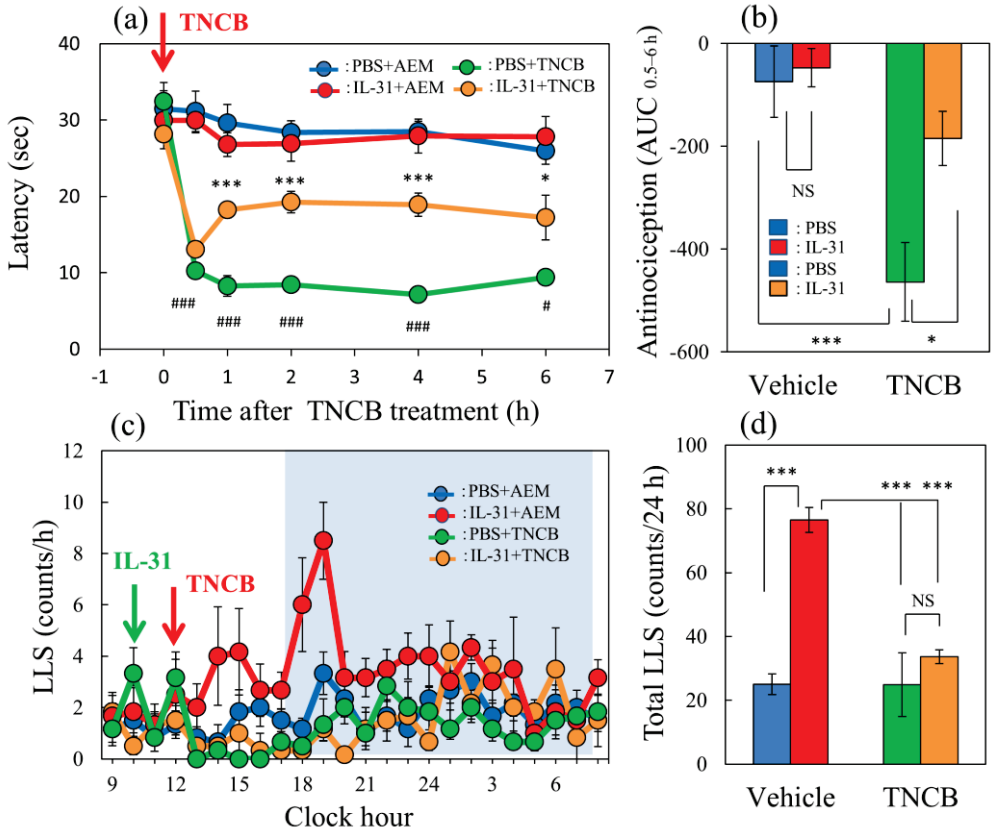


Figure 1. Effect of IL-31 on TNCB-applied hot-plate test in BALB/c mice. (**a**) Effect of repeated pre-treatment of IL-31 (50 μg/kg, intraperitoneally, every 12 h for three days) on TNCB (3%, 0.04 mL/each limb, total 0.16 mL/site)-applied hot-plate (45 °C) test. $^{\#}$ $p < 0.05$, $^{\#\#\#}$ $p < 0.001$ compared with the corresponding values in the saline + vehicle-treated group. (**b**) Total antinociceptive index of IL-31 after 0.5 to 6 h ($AUC_{0.5–6\,h}$). (**c**) Time-course change of itch-associated scratching behavior (LLS, counts/h) after the application of TNCB (3%, 0.2 mL/dorsal site). The blue line indicates the PBS + AEM-treated group; the red line indicates IL-31 + AEM-treated group; the green line indicates PBS + TNCB-treated group; the yellow line indicates IL-31 + TNCB-treated group. The blue column indicates the PBS + AEM-treated group; the red column indicates the IL-31 + AEM-treated group; the green column indicates the PBS + TNCB-treated group; the yellow column indicates the IL-31 + TNCB-treated group. (**a**,**b**) * $p < 0.05$, *** $p < 0.001$ compared with the corresponding values in the PBS + TNCB-treated group. The lateral axis indicates the clock hour, and the shaded area represents nighttime (dark phase, 7:00 pm to 7:00 am). (**d**) Total LLS counts for 24 h. The blue column indicates the IL-31 + AEM-treated group; the red column indicates the IL-31 + TNCB-treated group; the green column indicates the PBS + TNCB-treated group; the yellow column indicates the IL-31 + TNCB-treated group. The red arrow indicates the AEM or TNCB application point. The blue arrow indicates the IL-31 (1 mg/kg, intraperitoneally) administration point. Each value represents the mean ± standard error (S.E.) from 6 mice (total 36 mice). *** $p < 0.001$ compared with the vehicle (PBS) + TNCB-treated group. Three-way ANOVA: group × treatment × time: $F(5, 120) = 2.760$, $p = 0.0214$, (**a**); group × treatment × time: $F(23, 480) = 1.230$, $p = 0.2124$, (**c**). Two-way ANOVA: group × treatment: $F(1, 20) = 4.303$, $p = 0.0512$, group: $F(1, 20) = 18.68$, $p = 0.0003$, treatment: $F(1, 20) = 6.296$, $p = 0.0208$, (**b**); group × treatment: $F(1, 20) = 13.95$, $p = 0.0013$, group: $F(1, 20) = 14.16$, $p = 0.0012$, treatment: $F(1, 20) = 27.89$, $p < 0.0001$, (**d**).

2.2. Effects of the Combination of Morphine and IL-31 on Hot-Plate Test in BALB/c Mice

In a conventional hot-plate test (51 °C), morphine (MP, 3 mg/kg, subcutaneously) significantly increased the latency (Figure 2a, green line) and total antinociceptive index 15–120 min after administration ($AUC_{15–120\,min}$) (Figure 2b, green column) compared with

the saline-treated group (Figure 2b, blue column). Morphine (3 mg/kg, subcutaneously) significantly increased the latency (Figure 2a, green line) and total antinociceptive index 15–120 min after administration ($AUC_{15–120\ min}$) (Figure 2b, green column) compared with the saline-treated group (Figure 2b, blue column) in a conventional hot-plate test (51 °C). In contrast, repeated pretreatment with IL-31 decreased latency (Figure 2a, red line) and the total antinociceptive index after 15 to 120 min ($AUC_{15–120\ min}$) (Figure 2b, red column); however, no significant difference was observed. Moreover, the effect of the combination of morphine and repeated pretreatment with IL-31 significantly enhanced morphine-induced increasing latency (Figure 2a, yellow line) and the total antinociceptive index ($AUC_{15–120\ min}$) (Figure 2b, yellow column).

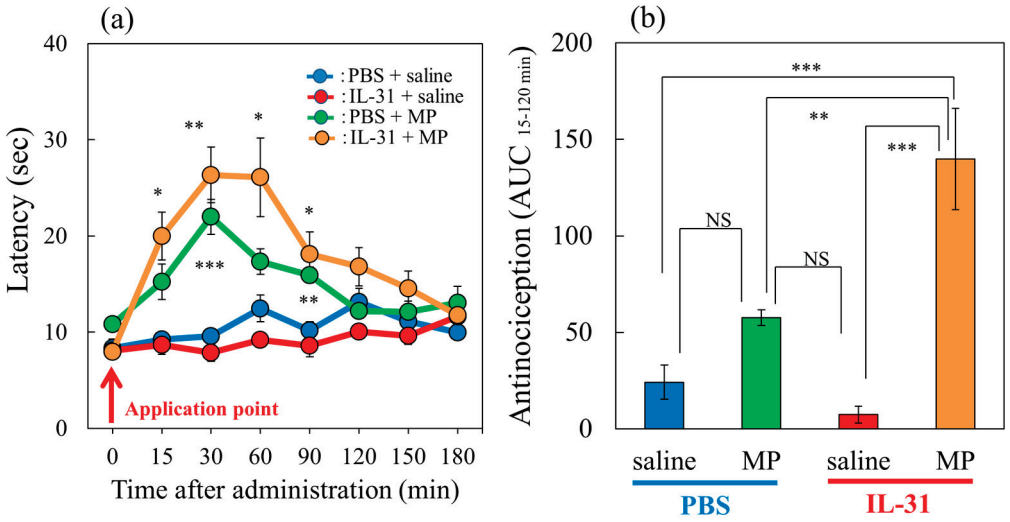


Figure 2. Effects of IL-31 and morphine on conventional hot-plate (51 °C) test in BALB/c mice. (**a**) Effect of repeated pretreatment of IL-31 (50 μg/kg, intraperitoneally, every 12 h for three days) and a single dose of morphine (MP, 3 mg/kg, subcutaneously) on conventional hot-plate (51 °C) test. The blue line indicates the phosphate-buffered saline (PBS) + saline-treated group; the red line indicates the IL-31 + saline-treated group; the green line indicates PBS + morphine-treated group; the yellow line indicates IL-31 + morphine-treated group. Each value represents the mean ± standard error (S.E.) from 6 mice (24 mice). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ compared with the corresponding values in the PBS + saline-treated group. (**b**) Total antinociceptive index of morphine after 15 to 120 min ($AUC_{15–120\ min}$). The blue column indicates PBS + saline-treated group; the green column indicates PBS + morphine-treated group. The red column indicates the IL-31 + saline-treated group; the yellow column indicates IL-31 + morphine-treated group. Each value represents the mean ± standard error (S.E.) from 6 mice (total 24 mice). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ compared with each group. Three-way ANOVA: group × treatment × time: $F(7, 168) = 2.360$, $p = 0.0252$, (**a**). Two-way ANOVA: group × treatment: $F(1, 20) = 12.11$, $p = 0.0024$, group: $F(1, 20) = 34.12$, $p < 0.0001$, treatment: $F(1, 20) = 5.28$, $p = 0.0325$, (**b**).

2.3. Effects of the Combination of IL-31 and Loxoprofen on TNCB-Applied Hot-Plate Test in BALB/c Mice

In the group where the vehicle (acetone–ethanol mixed liquor, AEM, 0.04 mL/each limb) was applied to the limbs of the mice, latency was not changed during the experimental period on the modified hot-plate test (45 °C) (Figure 3a, black line). In the TNCB (1%)-applied hot-plate (45 °C) test, the latency significantly decreased from 30 to 120 min after TNCB application (Figure 3a, blue line). Loxoprofen (LXP, 15 mg/kg, oral) significantly increased antinociception (Figure 3a, green line) and the total antinociceptive index after 30–120 min ($AUC_{30–120\ min}$) (Figure 3b, green column) compared with the vehicle (carboxymethyl cellulose sodium, CMC, oral)-treated group (Figure 3a, blue line: Figure 3b, blue column). Repeated pretreatment with IL-31 also increased latency (Figure 3a, red line) and total antinociceptive index ($AUC_{30–120\ min}$) (Figure 3b, red column). However, the

combination of loxoprofen and IL-31 did not significantly enhance latency ($AUC_{30–120\ min}$) (Figure 3a, yellow line) and total antinociceptive index ($AUC_{30–120\ min}$) (Figure 3b, yellow column) compared with treatment with loxoprofen alone.

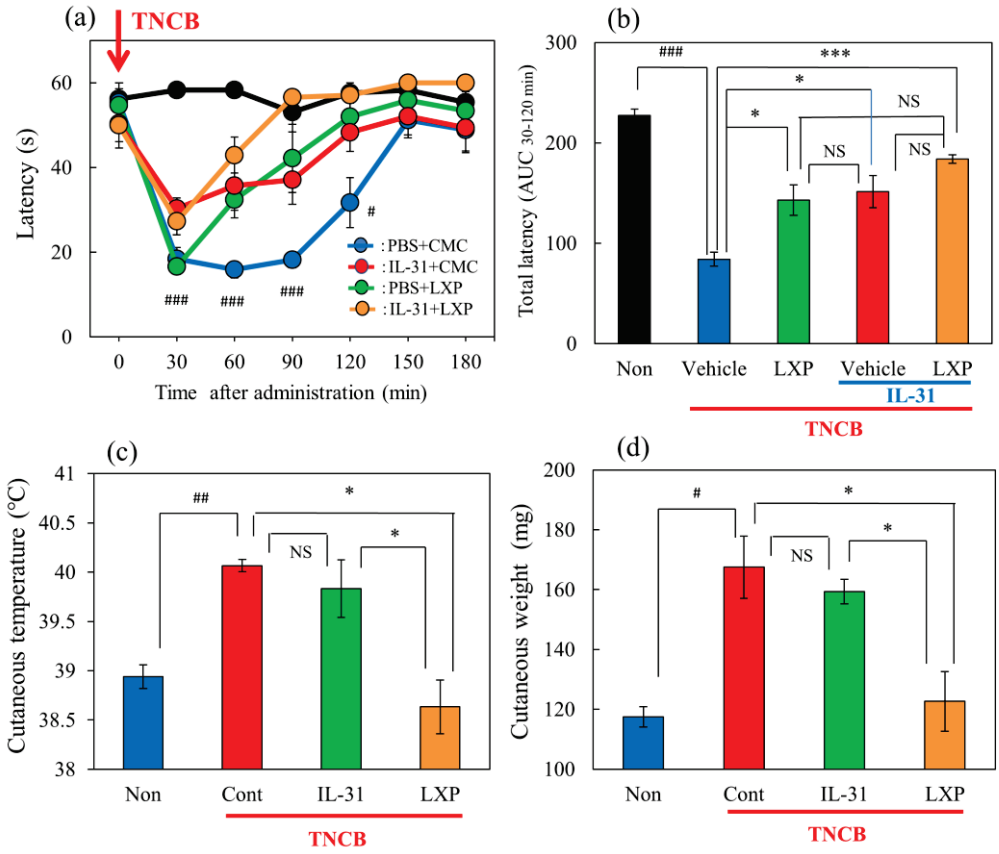


Figure 3. Effects of IL-31 and loxoprofen (LXP) on TNCB–applied the hot-plate test in BALB/c mice. (**a**) Effects of IL-31 and loxoprofen (LXP) on TNCB-applied hot-plate (45 °C) test in mice. The black line and column indicate the non-TNCB-treated group; the blue line indicates phosphate-buffered saline (PBS) + vehicle (carboxymethyl cellulose sodium, CMC, 10 mL/kg, oral)-treated group; the red line indicates IL-31 + CMC-treated group; the green line indicates PBS + LXP-treated group; the yellow line indicate IL-31 + LXP treated group. The red arrow indicates the AEM or TNCB application point. $^{\#}$ $p < 0.05$, $^{\#\#\#}$ $p < 0.001$ compared with the corresponding values in the saline + vehicle-treated group. (**b**) Total antinociceptive index of LXP after 30–120 min ($AUC_{30–120\ min}$). The black line and column indicate the non-treated group; the blue column indicates the PBS + CMC-treated group; the green column indicates the PBS + LXP-treated group; the red column indicates the IL-31 + CMC treated-group; the yellow column indicates the IL-31 + LXP-treated group. Each value represents the mean ± standard error (S.E.) from 6 mice (total 54 mice). NS, not significant, $^{\#\#\#}$ $p < 0.001$ compared with the corresponding values in the saline + vehicle-treated group. * $p < 0.05$, *** $p < 0.001$ compared with the corresponding values in the vehicle + TNCB treated group; (**c**) Effects of IL-31 and LXP on TNCB-induced increasing cutaneous temperature (fever) after 24 h application of TNCB. (**d**) Effects of IL-31 and LXP on TNCB-induced increasing cutaneous weight (swelling) after 24 h application of TNCB. The blue column indicates saline + vehicle-treated group; the red column indicates vehicle + TNCB-treated group; the green column indicates IL-31 + TNCB-treated group; the yellow column indicates LXP + TNCB-treated group. (**c**,**d**) Each value represents the mean ± S.E. from 6 mice (total 24 mice), NS, not significant, $^{\#}$ $p < 0.05$, $^{\#\#}$ $p < 0.01$ compared with the non-TNCB-treated group. * $p < 0.05$ when compared with each group. Two-way ANOVA: time × group: F (24, 175) = 5.382, $p < 0.0001$, time: F (6, 175) = 34.22, $p < 0.0001$, group: F (4, 175) = 37.29, $p < 0.0001$, (**a**). One-way ANOVA: F (4, 25) = 111, $p < 0.0001$, (**b**); F (3, 20) = 10.65, $p = 0.0002$, (**c**); F (3, 20) = 10.87, $p = 0.0002$, (**d**).

In the preliminary study, we assessed the time-course change of cutaneous temperature as an indicator of fever, and skin weight as an indicator of swelling, after the 1% TNCB application to determine the optimal experimental conditions. The cutaneous temperature decreased within 1 to 6 h after TNCB application, then significantly increased after 24 h of application and returned to basal level after 72 h. Skin tissue weight significantly increased between 6 and 72 after TNCB application.

Therefore, we measured the cutaneous weight and temperature for 24 h after TNCB application in subsequent experiments. Then, IL-31 did not change the cutaneous temperature or swelling. In contrast, loxoprofen (LXP) significantly decreased TNCB-induced fever and swelling (Figure 3c,d).

2.4. Comparison of the Effects of TNCB or Capsaicin on Modified Hot-Plate Test in BALB/c Mice

The application of TNCB (3%, 0.04 mL/site of limb) or capsaicin (1%, 0.04/site of limb) rapidly decreased latency on the hot-plate (45 °C) test within 0.5 to 6 h after their topical application (Figure 4a, red and green lines). Cutaneous prostaglandin D_2 (PGD_2) contents significantly increased 1 h after TNCB application, whereas this effect was not observed in the capsaicin-applied group (Figure 4b). Capsaicin significantly suppressed IL-31RA expression in the DRG but not the TNCB applications groups (Figure 4c).

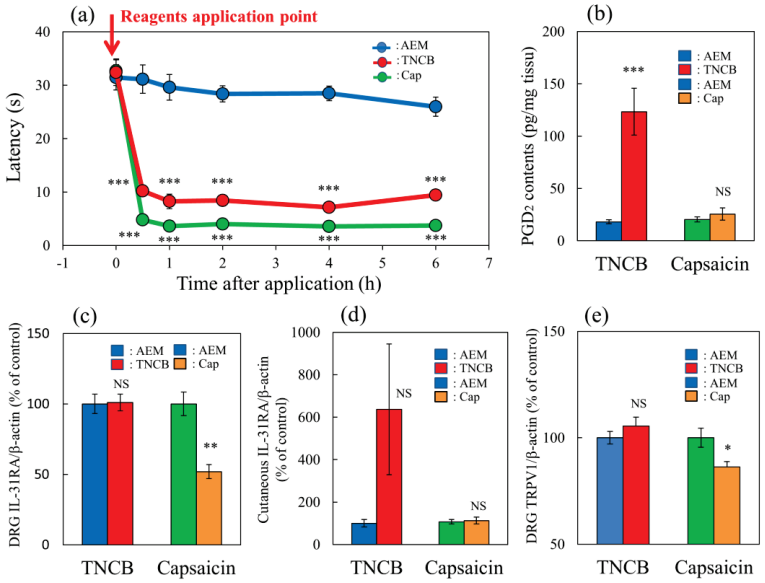


Figure 4. Comparative effects of TNCB or capsaicin on latency in modified hot-plate test and several pain-related parameters in BALB/c mice. (**a**) Time-course changes of latency caused by topical application of TNCB (3%) or capsaicin (Cap, 1%) in modified hot-plate (45 °C) test. The red arrow indicates the vehicle (acetone-ethanol mixed liquor, AEM, 0.2 mL/site) or TNCB or capsaicin application point. The values represent the means ± standard error (S.E.) from 6 mice. *** $p < 0.001$ compared with the respective value of the AEM-treated group. (**b**) Effects of TNCB and capsaicin on cutaneous PGD_2 contents 1 h after application. (**c**) Effects of TNCB or capsaicin on DRG neuronal IL-31RA mRNA expression 6 h after application. (**d**) Effects of TNCB or capsaicin on cutaneous IL-31 mRNA expression 1 h after application. € Effects of TNCB or capsaicin on DRG neuronal TRPV1 mRNA expression 6 h after application. The values represent the means ± standard error (S.E.) from 6 mice (total 42 mice). (**a**–**d**) NS, not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ compared with the respective value of the AEM-treated group. Two-way ANOVA: time × group: F (10, 90) = 14.22, $p < 0.0001$, time: F (5, 90) = 74.75, $p < 0.0001$, group: F (2, 90) = 307.4, $p < 0.0001$, (**a**); group × treatment: F (1, 20) = 18.51, $p = 0.0003$, group: F (1, 20) = 22.51, $p = 0.0001$, treatment: F (1, 20) = 16.9, $p = 0.0005$, (**b**); group × treatment: F (1, 20) = 10.65, $p = 0.0039$, group: F (1, 20) = 9.817, $p = 0.0052$, treatment: F (1, 20) = 10.65, $p = 0.0039$, (**c**); group × treatment: F (1, 20) = 3.016, $p = 0.0978$, group: F (1, 20) = 3.014, $p = 0.0979$, treatment: F (1, 20) = 3.016, $p = 0.0978$, (**d**); group × treatment: F (1, 20) = 6.93, $p = 0.0160$, group: F (1, 20) = 1.273, $p = 0.2725$, treatment: F (1, 20) = 6.932, $p = 0.0159$, (**e**).

TNCB markedly increased cutaneous IL-31 mRNA expression 1 h after TNCB application; however, the difference was not significant (Figure 4d). Moreover, capsaicin significantly decreased TRPV1 expression in the DRG but not the TNCB application group at 6 h after these applications (Figure 4e). The TNCB and capsaicin groups exhibited pain and had shortened latency from 0.5 to 6 h after application, similarly to the effects observed at 6 h; however, many differences were observed in the expression of several pain-related factors. Therefore, we investigated the site of action of IL-31 on PGD_2- or capsaicin-induced pain in mice.

2.5. Effect of IL-31 on ProstaglandinD$_2$- or Capsaicin-Applied Modified Hot-Plate Test in BALB/c Mice

In the group where the vehicle (ethanol, 0.04 mL/each limb) was applied to the limbs of the mice, latency was not changed during the experimental period on the modified hot-plate (45 °C) test (Figure 5a,c, blue line). Similar to the group receiving repeated pretreatment with IL-31, latency was also not changed during the experimental period (Figure 5a,c, red line). In contrast, prostaglandin D_2 (PGD_2) significantly decreased latency from 0.5 to 4 h after PGD_2 application (Figure 5a, green line). After repeated pretreatment with IL-31 in the PGD_2-applied group, the PGD_2-induced decreased latency was significantly increased from 0.5 to 1 h after PGD_2 application (Figure 5a, yellow line). The total antinociceptive index of IL-31 after 30 to 120 min ($AUC_{30–120\ min}$) after vehicle (PBS) application was not significantly different between the PBS and IL-31 pretreatment groups (Figure 5b, left side). In contrast, the total antinociceptive index of IL-31 after 30 to 120 min ($AUC_{30–120\ min}$) in the PGD_2-applied group showed a significant difference between the PBS and IL-31 pretreatment groups (Figure 5b, right side).

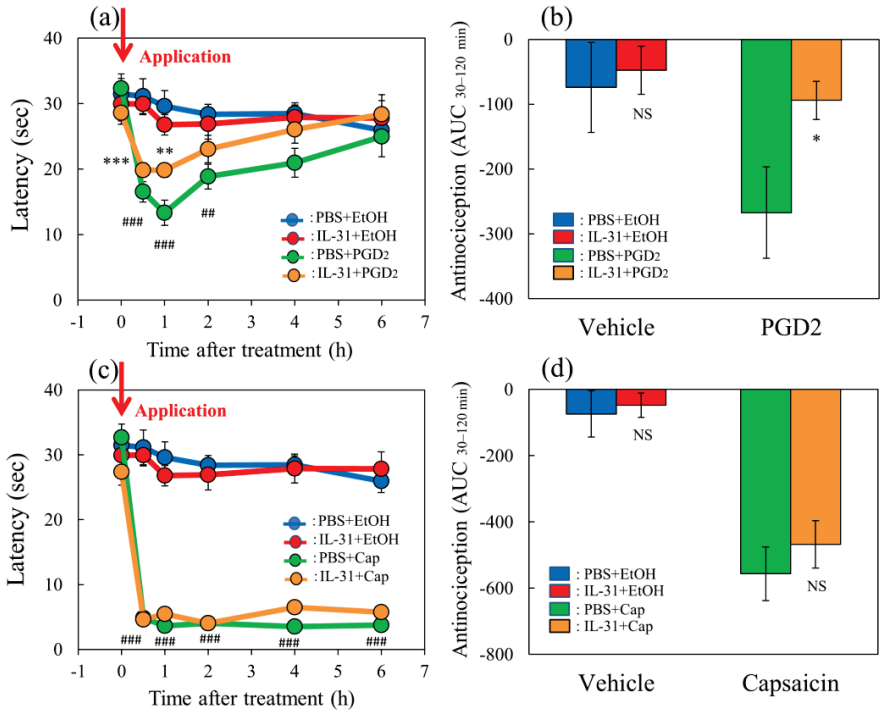


Figure 5. Effects of IL-31 on PGD_2- or capsaicin-applied hot-plate test in BALB/c mice. (**a**) Effect of IL-31 on PGD_2-induced decreasing latency on hot–plate (45 °C) test. ** $p < 0.01$, *** $p < 0.001$ compared with the corresponding values in the vehicle (ethanol, EtOH, 0.2 mL/site)-treated group; ## $p < 0.01$, ### $p < 0.001$ compared with the corresponding values in the vehicle-treated group. (**b**) Total antinociceptive index from 30 to 120 min ($AUC_{30–120\ min}$) after PGD_2 application. NS, not significant, * $p < 0.05$ compared with the corresponding values in the vehicle-treated group. (**c**) Effect of IL-31 on capsaicin (Cap)-induced decreasing latency on modified hot-plate (45 °C) test. ### $p < 0.001$ compared with the corresponding values in the vehicle-treated group. (**d**) After

capsaicin application, the total antinociceptive index changes from 30 to 120 min ($AUC_{30–120\ min}$). The blue line indicates the vehicle (phosphate-buffered saline, PBS, intraperitoneal, every 12 h for 3 days) + EtOH-group; the red line indicates the IL-31-pretreated (50 μg/kg, intraperitoneal, every 12 h for 3 days) + EtOH-treated group; the green line indicates PBS + PGD_2 (0.1%, 0.04 mL/each limb)- or capsaicin (3%, 0.04 mL/each limb)-treated group; the yellow line indicates IL-31 + PGD_2- or capsaicin-treated group. The red arrow indicates the vehicle (ethanol, EtOH)- or PGD_2 or capsaicin application point. Values represent the mean ± S.E. from 6 mice (total of 48 mice). NS, not significant, compared with the respective values of the vehicle-treated group. Three-way ANOVA: group × treatment × time: F (5, 120) = 0.885, $p = 0.4934$, group × treatment: F (1, 120) = 5.631, $p = 0.0192$, group: F (1, 120) = 47.782, $p < 0.0001$, treatment: F (1, 120) = 1.591, $p = 0.2096$, time: F (5, 120) = 7.123, $p < 0.0001$, (**a**); group × treatment × time: F (5, 126) = 0.824, $p = 0.5347$, group: F (1, 126) = 903.432, $p < 0.0001$, treatment: F (1, 126) = 0.315, $p = 0.5759$, time: F (5, 126) = 50.026, $p < 0.0001$, (**c**). Two-way ANOVA: group × treatment: F (1, 20) = 1.79, $p = 0.1959$, group: F (1, 20) = 3.317, $p = 0.0835$, treatment: F (1, 20) = 4.791, $p = 0.0406$, (**b**); group × treatment: F (1, 21) = 0.2122, $p = 0.6498$, group: F (1, 21) = 44.57, $p < 0.0001$, treatment: F (1, 21) = 0.727, $p = 0.4035$, (**d**).

Capsaicin significantly decreased latency from 0.5 to 6 h after application (Figure 5c, green line). After repeated pretreatment with IL-31 in the capsaicin-applied group, the capsaicin-induced decreased latency did not change (Figure 5c, yellow line). The total antinociceptive index of the vehicle (PBS)-applied to the group and the IL-31 pretreatment group were not significantly different (Figure 5d, left side). Therefore, we investigated the effects of several analgesics on TNCB- or capsaicin-induced pain in wild-type and IL-31RAKI mice.

2.6. Effects of Morphine on TNCB- or Capsaicin-Applied Hot-Plate Test in Wild-Type and IL-31RAKI Mice

In the TNCB-applied hot-plate test (45 °C), the latency rapidly decreased 30 min after TNCB (3%, 0.04 mL/each limb) and capsaicin (1%, 0.04 mL/each limb) application (Figure 6a, blue and green lines) in BALB/c mice. Morphine (3 mg/kg, subcutaneously) delayed the latency and the increased antinociceptive index 30 to 120 min after administration ($AUC_{30–120\ min}$) compared with those of the vehicle (saline)-treated group (Figure 6a, red and yellow lines). The total antinociceptive index of morphine after 30 to 120 min ($AUC_{30–120\ min}$) was approximately the same as in those of the TNCB- and capsaicin-applied hot-plate test groups. However, there was a significant effect for the TNCB-applied hot-plate test (Figure 6b, left side), but not the capsaicin-applied hot-plate test (Figure 6b, right side).

In the conventional hot-plate test (51 °C), the latency did not change during the experimental period (Figure 6c, blue and green lines) in wild-type (C57BL/6) and IL-31RAKI mice (C57BL/6 genetic background). Morphine (3 mg/kg, subcutaneously) delayed the latency and increased antinociception compared with the vehicle-treated group (Figure 6c, red and yellow lines). The total antinociceptive index of morphine after 30 to 120 min ($AUC_{30–120\ min}$) was decreased in IL-31RAKI mice compared with wild-type mice in the conventional hot-plate test. In contrast, a significant effect was not observed in IL-31RAKI mice (Figure 6d, right side).

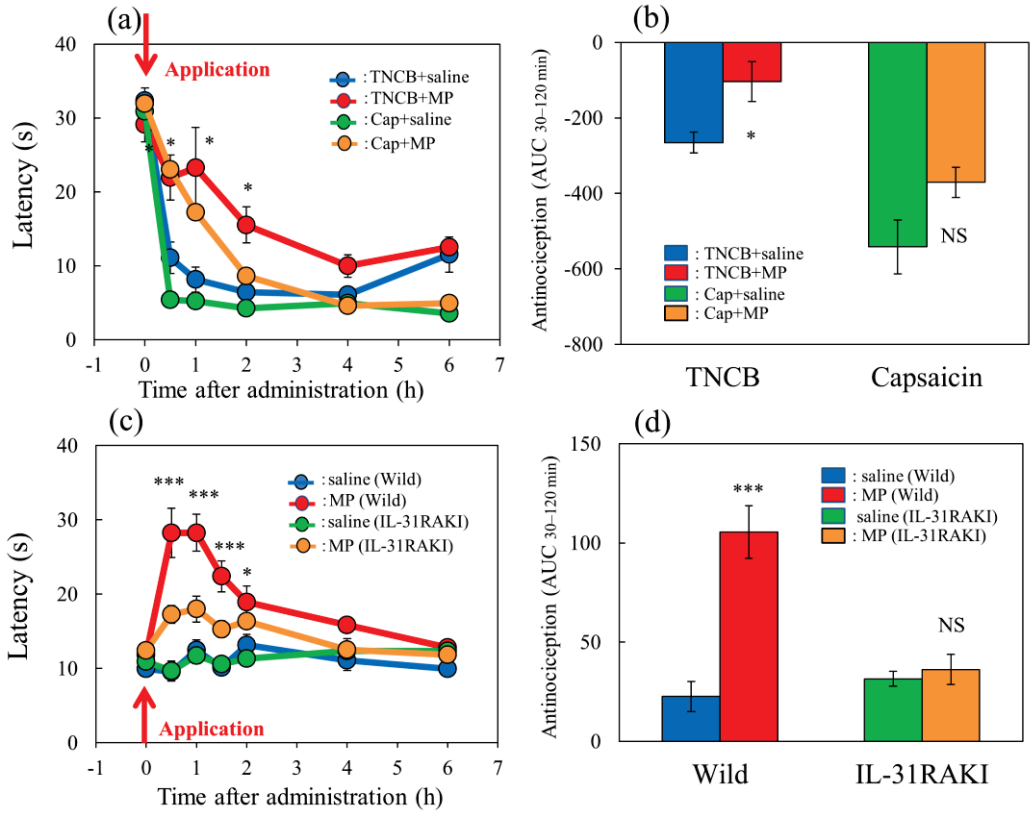


Figure 6. Effect of morphine on TNCB- or capsaicin-applied hot-plate test in wild-type or IL-31RAKI mice. (**a**) Effect of morphine (3 mg/kg, subcutaneous) on TNCB- or capsaicin-applied hot-plate (45 °C) test in BALB/c mice. Each value represents the mean ± standard error (S.E.) from 6 mice * $p < 0.05$ compared with the respective values in vehicle (saline)-treated mice. (**b**) Total antinociceptive index of morphine after 30 to 120 min ($AUC_{30–120\ min}$) as per the TNCB or capsaicin-treated hot-plate test. NS, not significant, * $p < 0.05$ compared with the respective values in vehicle-treated mice. The blue line and column indicate the TNCB + vehicle (saline)-treated group; the red line and column indicates the TNCB + morphine-treated group; the green line and column indicates the capsaicin + saline-treated group; the yellow line and column indicates the capsaicin + morphine-treated group. The red arrow indicates the TNCB or capsaicin application point. (**c**) Effect of morphine on conventional hot-plate (51 °C) test in wild-type (C57BL/6) and IL-31RAKI (C57BL/6 genetic background) mice. * $p < 0.05$, and *** $p < 0.001$ compared with the respective values in vehicle-treated mice. (**d**) Total antinociceptive index of morphine after 30 to 120 min ($AUC_{30–120\ min}$) of wild-type and IL-31RAKI mice. The blue line and column indicate the vehicle (saline)-treated group in wild-type mice; the red line and column indicates the morphine-treated group in wild-type mice; the green line and column indicates the saline-treated group in IL-31RAKI mice; the yellow line and column indicates the morphine-treated group in IL-31RAKI mice. Wild, C57BL/6 mice; IL-31RAKI, IL-31 receptor A-deficient mice (C57BL/6 genetic background). Each value represents the mean ± standard error (S.E.) from 6 mice (total 48 mice). *** $p < 0.001$ compared with the respective values in vehicle-treated mice. Three-way ANOVA: group × treatment × time: $F(5, 120) = 1.275$, $p = 0.2792$, group: $F(1, 120) = 17.296$, $p < 0.0001$, treatment: $F(1, 120) = 49.027$, $p < 0.0001$, time: $F(5, 120) = 73.243$, $p < 0.0001$, (**a**); group × treatment × time: $F(7, 160) = 2.171$, $p = 0.0395$, (**c**). Two-way ANOVA: group × treatment: $F(1, 20) = 0.007463$, $p = 0.9320$, group: $F(1, 20) = 28.95$, $p < 0.0001$, treatment: $F(1, 20) = 10.85$, $p = 0.0036$, (**b**); group × treatment: $F(1, 20) = 19.79$, $p = 0.0002$, group: $F(1, 20) = 11.82$, $p = 0.0026$, treatment: $F(1, 20) = 24.87$, $p < 0.0001$, (**d**).

2.7. Effects of Loxoprofen on TNCB- or Capsaicin-Applied Hot-Plate Test in Wild-Type and IL-31RAKI Mice

The latency rapidly decreased 30 min after TNCB or capsaicin was applied in the hot-plate (45 °C) test in BALB/c mice (Figure 7a, blue and green line). Loxoprofen (LXP), which is a reversible cyclooxygenase inhibitor, (15 mg/kg, oral), significantly delayed latency 0.5 to 6 h after loxoprofen treatment on the TNCB (3%, 0.04 mL/each limb)-applied hot-plate test, compared with the respective value of the vehicle (carboxymethyl cellulose sodium, CMC, 10 mL/kg, oral)-treated group (Figure 7a, red lines).

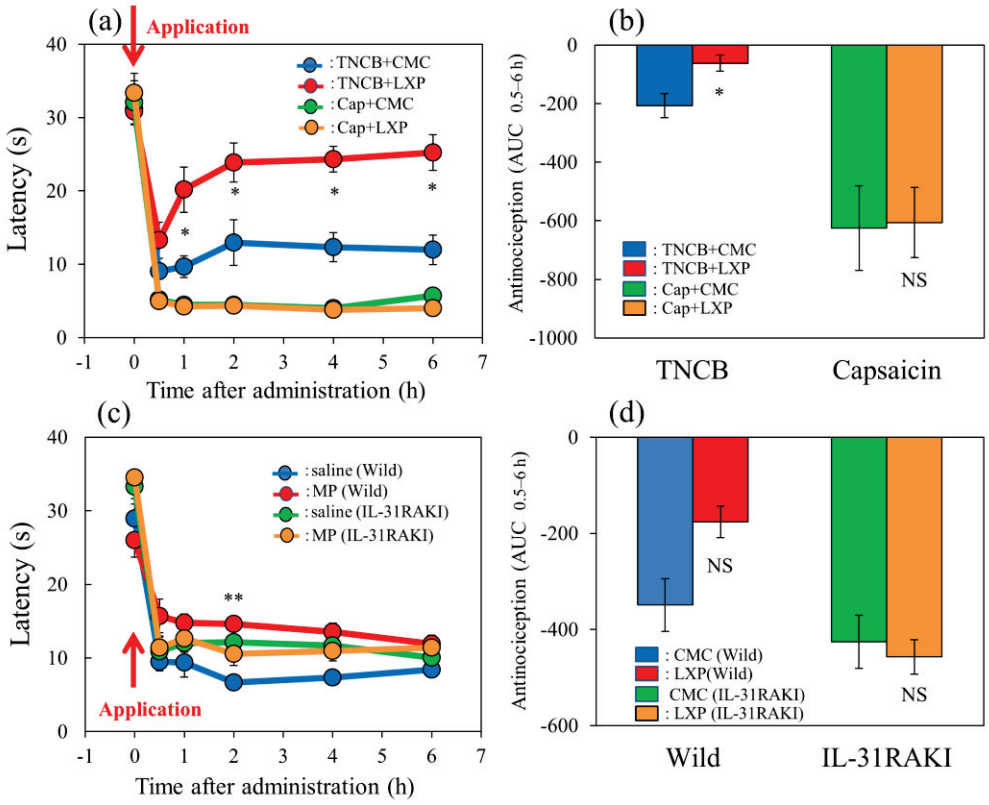


Figure 7. Effect of loxoprofen on TNCB- or capsaicin-applied hot-plate test in wild-type and IL-31RAKI mice. (**a**) Effect of loxoprofen on TNCB- or capsaicin-applied hot-plate (45 °C) test in BALB/c mice. Each value represents the mean ± standard error (S.E.). * $p < 0.05$ compared with the respective values in the vehicle (carboxymethyl cellulose sodium saline, CMC, 10 mL/kg, oral)-treated mice. (**b**) Total antinociceptive index of loxoprofen after 0.5 to 6 h ($AUC_{0.5–6\,h}$) as per the TNCB- or capsaicin-applied hot-plate test in BALB/c mice. Each value represents the mean ± standard error (S.E.) from 6 mice. NS, not significant, * $p < 0.05$ compared with the respective values in the CMC-treated mice. The blue line and column indicate the TNCB + vehicle (saline)-treated group; the red line and column indicates the TNCB + loxoprofen-treated group; the green line and column indicates the capsaicin + saline-treated group; the yellow line and column indicates the capsaicin + loxoprofen-treated group. (**c**) Effect of loxoprofen on TNCB-applied hot-plate (45 °C) test in wild-type (C57BL/6) and IL-31RAKI (C57BL/6 genetic background) mice. Each value represents the mean ± standard error (S.E.). ** $p < 0.01$ compared with the respective values in the CMC-treated mice. (**d**) Total antinociceptive index of loxoprofen ($AUC_{0.5–6\,h}$) of wild-type and IL-31RAKI mice. NS, not significant, compared with the respective values in vehicle-treated mice. The blue line indicates TNCB + CMC-and column indicate the vehicle (saline)-treated group in wild-type mice; the red line and column indicates TNCB +the loxoprofen-treated (15 mg/kg, oral) group in wild-type mice; the green line and column indicates capsaicin + CMCthe saline-treated group in IL-31RAKI mice; the yellow line and column indicates capsaicin +the loxoprofen -treated group in IL-31RAKI mice. The red arrow indicates the TNCB or

capsaicin application point. Wild, C57BL/6 mice; IL-31RAKI, IL-31 receptor A-deficient mice (C57BL/6 genetic background). Each value represents the mean ± standard error (S.E.) from 6 mice (total 48 mice). Three-way ANOVA: group × treatment × time: F (5, 120) = 2.430, $p = 0.0389$, (**a**); group × treatment × time: F (5, 96) = 1.095, $p = 0.3685$, group: F (1, 96) = 6.646, $p = 0.0115$, treatment: F (1, 96) = 14.570, $p = 0.0002$, time: F (5, 96) = 102.550, $p < 0.0001$, (**c**). Two-way ANOVA: group × treatment: F (1, 20) = 0.4223, $p = 0.5232$, group: F (1, 20) = 24.67, $p < 0.0001$, treatment: F (1, 20) = 0.7201, $p = 0.4062$, (**b**); group × treatment: F (1, 16) = 4.854, $p = 0.0426$, group: F (1, 16) = 10.21, $p = 0.0056$, treatment: F (1, 16) = 3.69, $p = 0.0727$, (**d**).

However, the total antinociceptive index of loxoprofen after 0.5 to 6 h ($AUC_{0.5–6\ h}$) was not significant increased compared with that of the CMC-treated group (Figure 7b, left side). However, loxoprofen did not affect the capsaicin-applied hot-plate test compared with the CMC-treated group (Figure 7a,b, red line and column). On the other hand, loxoprofen did not affect the capsaicin-applied hot-plate test (Figure 7a,b, yellow line and column, right side).

In wild-type (C57BL/6) and IL-31RAKI (C57BL/6 genetic background) mice, the latency time rapidly decreased 30 min after TNCB application (Figure 7c, blue and green lines). Loxoprofen (15 mg/kg, oral) significantly delayed the latency at 2 h after treatment and increased antinociception compared with the respective value of the CMC-treated group (Figure 7c, red line). The total antinociceptive index of loxoprofen after 0.5 to 6 h ($AUC_{0.5–6\ h}$) was significantly increased compared with that of the CMC-treated group (Figure 7d, left side). However, loxoprofen did not affect the TNCB-treated hot-plate test in IL-31RAKI mice (Figure 7c,d, yellow line and column, right side). These results were similar to those of capsaicin-induced pain on the modified hot-plate test (Figure 7b).

2.8. Effects of Acetaminophen on TNCB- or Capsaicin-Applied Hot-Plate Test in Wild-Type and IL-31RAKI Mice

The latency rapidly decreased 30 min after TNCB- or capsaicin-applied hot-plate (35 °C) tests in BALB/c mice (Figure 8a, blue and green lines). Acetaminophen (APAP, 300 mg/kg, oral) delayed the latency and increased antinociceptive activity on the TNCB-applied hot-plate test, compared with the vehicle (carboxymethyl cellulose sodium, CMC, 10 mL/kg, oral)-treated group (Figure 8a, red lines). The total antinociceptive index of acetaminophen after 30 to 120 min ($AUC_{30–120\ min}$) was significantly increased compared with that of the CMC-treated group (Figure 8b, red line and column, left side). However, acetaminophen did not affect the capsaicin-treated hot-plate test, compared with the CMC-treated group (Figure 8a,b, yellow line and column, right side).

In wild-type (C57BL/6) and IL-31RAKI (C57BL/6 genetic background) mice, the latency rapidly decreased 30 min after TNCB application on the modified hot-plate (35 °C) test (Figure 8c, blue and green lines). Acetaminophen (APAP, 300 mg/kg, oral) delayed the latency time and increased antinociception compared with the CMC-treated group (Figure 8c, red line). The total antinociceptive index of acetaminophen after 30–120 min ($AUC_{30–120\ min}$) was significantly increased compared with that of the CMC-treated group (Figure 8d, left side). However, acetaminophen did not affect the TNCB-applied hot-plate test in IL-31RAKI mice (Figure 8d, yellow line and column, right side). These results were similar to those of capsaicin-induced pain on the modified hot-plate test (Figure 8b).

2.9. Long-Term Changes after TNCB or Capsaicin Application on Sense of Itch and Pain

In the group where the vehicle (acetone-ethanol mixed liquor, AEM, 0.04 mL/each limb) was applied to the limbs of the mice, latency was not changed during the experimental period on the modified hot-plate test (45 °C) (Figure 9a, blue line). Both TNCB (3%, 0.04 mL/each limb) and capsaicin (1%, 0.04/each limb) induced a decreased latency after their application until 6 h. However, their effect differed 24 h after their topical application. The decreased latency was reversed in the TNCB group after 24 h and showed pain torpor until 72 h, whereas that in the capsaicin group continued to decrease until 72 h after application (Figure 9a, red and green lines). In IL-31RAKI mice, TNCB showed decreased latency

after application until 6 h (Figure 5a, yellow line) and then returned to basal levels after 24 h (Figure 9a, yellow line). The pain torpor was not observed in IL-31RAKI mice. In the TNCB group, IL-31RA expression in the DRG increased within 1 to 6 days after application, whereas that in the capsaicin group continued decreasing until 6 days after application (Figure 9b, red and green lines). TNCB showed a significant decrease in LLS counts from immediately after application until 6 h (Figure 9c), which then increased 3 days after administration, and gradually decreased to basal levels 12 days after administration (Figure 9c, red line). The capsaicin group also showed significantly decreased LLS counts immediately after its application until 6 days after (Figure 9c, green line), which did not recover during the experimental period. We measured IL-31-induced (1 mg/kg, intravenously) LLS counts for IL-31 reactivity. Before capsaicin application, LLS counts of the vehicle (PBS) and IL-31 administration groups were 70.3 ± 11.3 and 299.1 ± 39.7 counts/24 h, respectively (Figure 9, green and blue columns). The IL-31-induced LLS counts at 1, 24, 72, and 144 h (6 days) after capsaicin application were 30.1 ± 4.3, 56.8 ± 23.1, 184.6 ± 2.2, and 124.2 ± 26.2 counts/24 h, respectively (Figure 9d, red columns), which were significantly decreased compared with those before capsaicin application (Figure 9d, red columns).

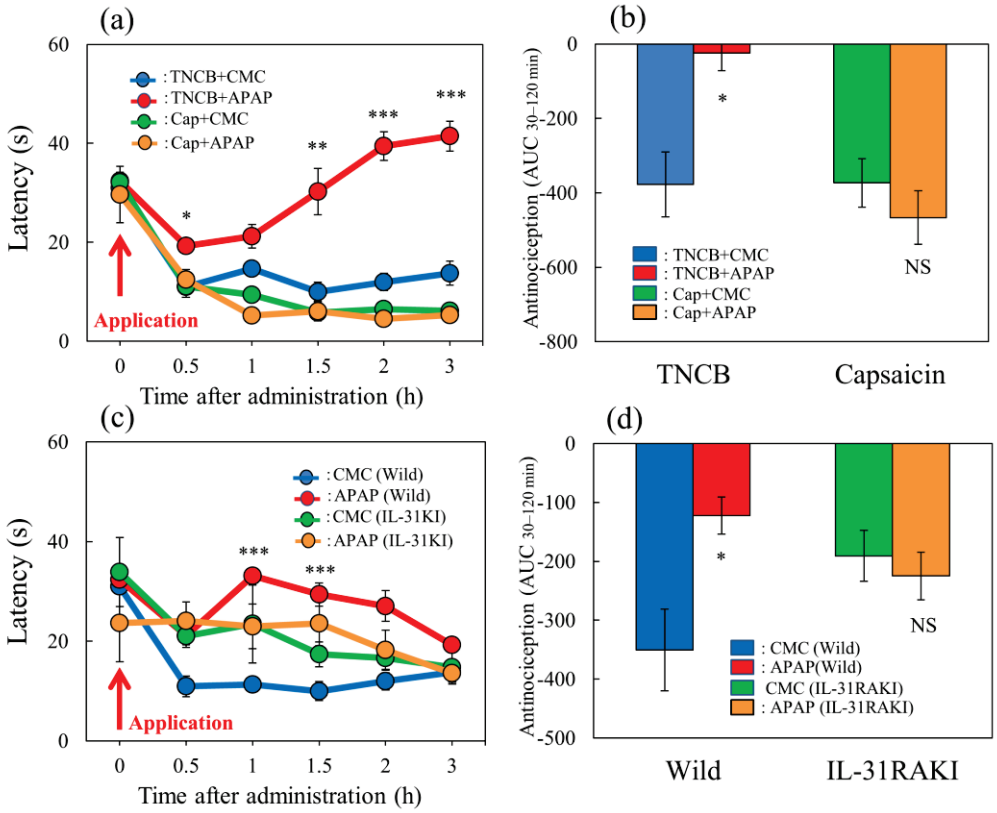


Figure 8. Effect of acetaminophen on TNCB– or capsaicin–applied hot-plate (35 °C) test in wild-type and IL-31RAKI mice. (**a**) Effect of acetaminophen (APAP, 300 mg/kg, oral) on TNCB- or capsaicin-applied hot-plate (35 °C) test in BALB/c mice. Each value represents the mean ± standard error (S.E.).* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ compared with the respective values in vehicle (carboxymethyl cellulose sodium, CMC, 10 mL/kg, oral)-treated mice. (**b**) Total antinociceptive index of acetaminophen after 30–120 min ($AUC_{30–120\ min}$) as per the TNCB- or capsaicin-applied hot-plate test in BALB/c mice. Each value represents the mean ± standard error (S.E.). NS, not significant, * $p < 0.05$ compared with the respective values in vehicle-treated mice. The blue line and column indicate the TNCB + vehicle (saline)-treated group; the red line and column indicates the TNCB + acetaminophen-treated group; the green line and column indicates the capsaicin + saline-treated group; the yellow line and column indicates the capsaicin + acetaminophen-treated group. (**c**) Effect

of acetaminophen on TNCB-applied hot-plate (35 °C) test in wild-type (C57BL/6) mice and IL-31RAKI (C57BL/6 genetic background) mice. Each value represents the mean ± standard error (S.E.). *** $p < 0.001$ compared with the respective values in the CMC-treated mice. (**d**) Total antinociceptive index of acetaminophen after 30–120 min ($AUC_{30–120\ min}$) of wild-type and IL-31RAKI mice. The blue line indicatesand column indicate the TNCB + CMC-vehicle (saline)-treated group in wild-type mice; the red line and column indicates the TNCB + acetaminophenAPAP-treated group in wild-type mice; the green line and column indicates the capsaicin + CMCsaline-treated group in IL-31RAKI mice; the yellow line and column indicates the capsaicin + acetaminophenAPAP-treated group in IL-31RAKI mice. The red arrow indicates the TNCB or capsaicin application point. Wild, C57BL/6 mice; IL-31RAKI, IL-31 receptor A-deficient mice (C57BL/6 genetic background). Each value represents the mean ± standard error (S.E.) from 6 mice (total 48 mice). NS, not significant, * $p < 0.05$ compared with the respective values in the CMC-treated mice. Three-way ANOVA: group × treatment × time: F (5, 126) = 1.200, $p = 0.3128$, group: F (1, 126) = 381.730, $p < 0.0001$, treatment: F (1, 126) = 6.316, $p = 0.0132$, time: F (5, 126) = 44.926, $p < 0.0001$, (**a**); group × treatment × time: F (5, 120) = 0.248, $p = 0.9400$, group: F (1, 120) = 2.609, $p = 0.1089$, treatment: F (1, 120) = 21.475, $p < 0.0001$, time: F (5, 120) = 4.594, $p = 0.0007$, (**c**). Two-way ANOVA: group × treatment: F (1, 20) = 10.33, $p = 0.0043$, group: F (1, 20) = 9.935, $p = 0.0050$, treatment: F (1, 20) = 3.516, $p = 0.0754$, (**b**); group × treatment: F (1, 20) = 10.33, $p = 0.0043$, group: F (1, 20) = 9.935, $p = 0.0050$, treatment: F (1, 20) = 3.516, $p = 0.0754$, (**d**).

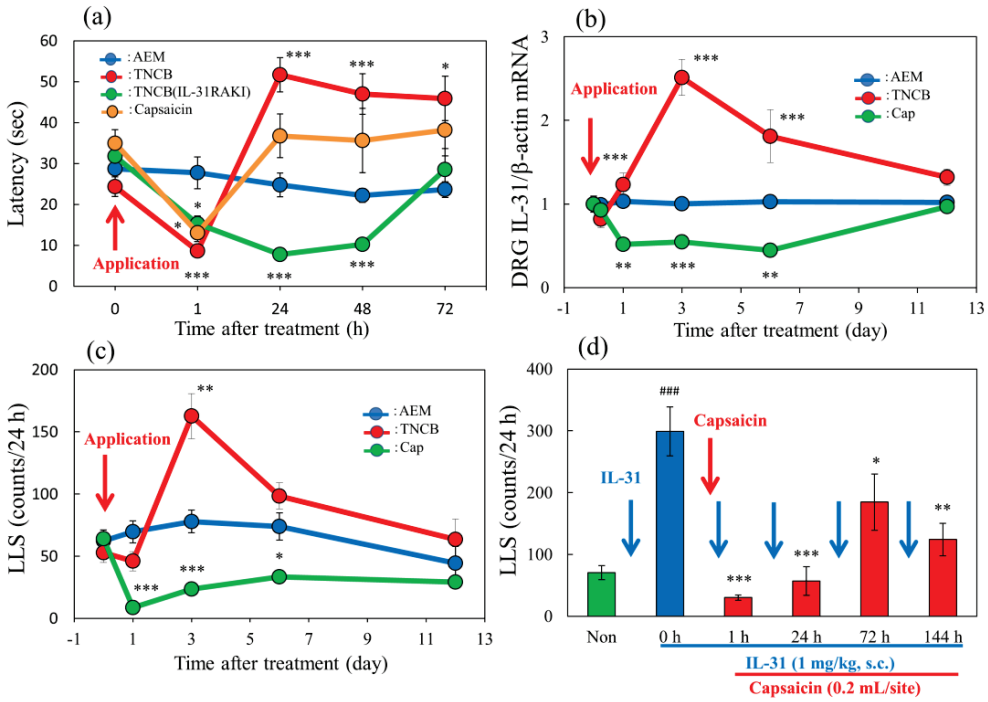


Figure 9. Long–term changes after TNCB or capsaicin application on the sense of itch and pain. (**a**) Effects of topical application of TNCB or capsaicin on modified hot-plate (45 °C) test in BALB/c mice and IL-31RAKI mice. The blue line indicates the vehicle (acetone-ethanol mixed liquor, AEM)-treated mice; the red line indicates the TNCB-treated mice; the green line indicates the capsaicin-treated mice; the yellow line indicates the TNCB-treated IL-31RAKI mice. (**b**) Effects of topical application of TNCB or capsaicin on IL-31RA expression in the DRG. The blue line indicates AEM-treated mice; the red line indicates TNCB-treated mice; the green line indicates capsaicin-treated mice. (**c**) Effects of topical application of TNCB or capsaicin on itch-associated behavior counts (LLS counts/24 h). The blue line indicates AEM-treated mice; the red line indicates TNCB-treated mice; the green line indicates capsaicin-treated mice. The red arrow indicates the AEM or TNCB or capsaicin application point. Each value represents the mean ± standard error (S.E.) from 6 mice (total 54 mice). * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ compared with the respective values of vehicle (AEM)-treated

group. (**d**) Time-course change of topical application of capsaicin on IL-31-induced LLS counts. The blue arrows indicate the IL-31 (1 mg/kg, intravenous) administration point. The red arrow indicates the capsaicin (1%, 0.2 mL/site) application point. Each value represents the mean ± standard error (S.E.) from 6 mice. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ compared with IL-31-induced LLS counts before capsaicin application. Two-way ANOVA: group × time: F (12, 100) = 7.899, $p < 0.0001$, group: F (3, 100) = 18.81, $p < 0.0001$, time: F (4, 100) = 12.45, $p < 0.0001$, (**a**); group × time: F (10, 90) = 15.2, $p < 0.0001$, group: F (2, 90) = 63.57, $p < 0.0001$, time: F (5, 90) = 6.516, $p < 0.0001$, (**b**); group × time: F (8, 75) = 8.204, $p < 0.0001$, group: F (2, 75) = 31.92, $p < 0.0001$, time: F (4, 75) = 9.405, $p < 0.0001$, (**c**). One-way ANOVA: F (5, 30) = 12.11, $p < 0.0001$, (**d**).

3. Discussion

The IL-31-caused scratching behavior is distinct from other pruritogens (e.g., histamine, serotonin)-induced scratching [21,35–37]. A single intradermal, or other administration route (e.g., intravenously), injection of IL-31 elicits LLS but not SLS, with LLS manifesting gradually approximately 1 h after injection and persisting over 24 h [35]. However, other pruritogens, injected only via an intradermal route, elicits SLS but not LLS, with SLS manifesting immediately after the intradermal injection and persisting for at least 30 min. IL-31 is a neuronal transmitter that causes aloknesis but is not a pruritogens, indicating that it causes alloknesis-induced itching [38,39]. In a previous study, we showed that repeated administration of IL-31 gradually increased the LLS counts. Our results indicated that IL-31-induced LLS counts depend on the interaction of blood IL-31 concentration, DRG neuronal IL-31RA expression induces LLS, and IL-31 promotes the onset of DRG neuronal IL-31RA [5]. In this study, a single pretreatment with IL-31 (50 μg/kg, intrapretoneal) showed no significant antinociceptive activity whereas repeated pretreatment with IL-31 showed significant antinociceptive activity on the TNCB-applied hot-plate test at 45 °C. The repeated administration of IL-31 enhanced the action of IL-31-induced increases in DRG neuronal IL-31RA expression. Moreover, a single large dose of IL-31 (1 mg/kg, intraperitoneal) elicited clear LLS, and TNCB-induced pain inhibited the IL-31-induced LLS. These results suggest that the IL-31-induced itch inhibits TNCB-induced pain. Although it is well-known that pain inhibits itching [40], the reverse phenomenon that an itch inhibits pain is controversial. Our results show that IL-31-induced alloknesis inhibits pain. These data suggest that the sensation of pain and itch may be regulated through their functional antagonism.

In the present study, the repeated pretreatment of IL-31 decreased latency; however, this result was insignificant in the conventional hot-plate test (51 °C). This result suggests that IL-31-induced itching is caused by heat stimulation; therefore, pain and itching actions are indistinguishable. However, morphine-induced antinociception was enhanced with pretreatment with IL-31, suggesting that the increased analgesic action was caused by the IL-31-induced peripheral action of morphine and its interaction with the central analgesic action of morphine. Morphine produces analgesia by acting on both the CNS and PNS, and it is a one of the few drugs to induce LLS counts and cause alloknesis simultaneously [41,42]. In addition, morphine-induced LLS and antinociceptive effects are closely correlated in mice; morphine-induced LLS and antinociceptive effects were completely and partially inhibited in IL-31RAKI mice, respectively [9]. Previously, we reported that IL-31 might play a significant role in the modulation of peripheral morphine-induced antinociception by sensory neurons in IL-31RAKI mice compared with wild-type mice [9]. In the present study, we demonstrated that repeated pretreatment with IL-31 showed antinociceptive activity. These results suggest that the antinociceptive activity of IL-31 partially contributes to the antinociceptive action of morphine.

In the present study, loxoprofen, a PGs inhibitor, significantly increased latency and total antinociceptive activity in the TNCB (1%)-applied hot plate (45 °C) test. The application of 3% TNCB will show almost the same result. Repeated pretreatment with IL-31 also increased latency and total nociception; however, the combined effect of loxoprofen and IL-31 was not significantly enhanced. These results suggest that the antinociceptive activity

of IL-31 was approximately as effective as that of loxoprofen (NSAIDs). An enhanced action was not observed because the compound characteristics have the same peripheral action site. Inflammatory reactions increased tissue PGs' (PGD_2, PGE_2, and PGI_2) biosynthesis and are characterized by redness (arterial extension), fever (increased body temperature), and edema (swelling). Therefore, we examined the comparative effects of IL-31 and loxoprofen on TNCB-induced fever (surface body temperature) and swelling (cutaneous weight) in mice. TNCB significantly increased the cutaneous temperature and skin tissue weight 24 h after application, whereas IL-31 did not change the cutaneous temperature and swelling. In contrast, loxoprofen significantly decreased TNCB-induced fever and swelling. These results show that the main action of IL-31 is not to inhibit PGs biosynthesis. Moreover, even if they share the same peripheral analgesic sites of action, NSAIDs and IL-31 have different mechanisms of action; NSAIDs suppress inflammation by inhibiting the biosynthesis of these PGs. However, these inflammatory reactions are spontaneous defense mechanisms of the body; in particular, fever and swelling are involved in the reproduction of wound tissue by vasodilatation. Pain is also a part of the defense mechanism; however, the reaction to stressful psychological damage is more serious than the reaction that induces fever and swelling. Therefore, NSAIDs, such as loxoprofen (PGs' inhibitor), are used to treat different inflammatory diseases. As the substance that specifically inhibits pain, IL-31 may be a more suitable analgesic drug for patients with various inflammatory diseases.

In the present study, the application of TNCB and capsaicin caused pain and decreased latency in the hot-plate test (45 °C) immediately or within 6 h after treatment. However, the onset of pain caused by the application of these materials was different. TNCB caused tissue injury and inflammation. Alternatively, the responses to noxious stimuli may be enhanced (hyperalgesia), or normally innocuous stimuli may produce pain (allodynia). PGs influence inflammation, and their administration induces the major signs of inflammation, including augmented pain [29,43]. However, the topical application of PGD_2, PGE_2, and PGI_2 significantly suppressed LLS in skin-lesioned NC/Nga mice, and their inhibitory activities are in the order of $PGD_2 > PGI_2 > PGE_2$ [28]. Moreover, PGs significantly enhanced nociceptive activity on hot-plate tests in mice [43]; PGD_2, PGE_2, and PGI_2 significantly increased the nociceptive effect in the order of $PGD_2 > PGI_2 > PGE_2$, which is the same as the order of their anti-pruritic activities. Although the scratching behavior strips off the epidermis and removes the alien substance invading the epidermis, it is a physiological reaction to directly control itching. In other words, scratching increases epidermal PGD_2 biosynthesis, and PGD_2-induced pain inhibits itching [27]. These previous and present findings suggest that cutaneous PGD_2 could be mainly produced in response to pain and play a critical role in regulating the sensation of pain [44]. The significant individual differences in cutaneous IL-31 expression may be a reaction to TNCB-induced pain. Cutaneous IL-31 expression is unstable and intermittent; however, DRG neuronal IL-31RA expression always shows a stable increase through stimulation [5]. This phenomenon may be a physiological reaction to inhibit pain.

In the present study, capsaicin did not affect cutaneous PGD_2 contents but significantly decreased DRG neuronal IL-31RA expression. These results indicate that the activation of TRPV1 and inhibition of IL-31RA mRNA expression in the DRG may mediate the antipruritic effects of capsaicin. The decrease in TRPV1 mRNA expression in the DRG triggered by capsaicin is hypothesized to respond to the onset of TRPV1-stimulated pain [30]. Therefore, we examined the effect of repeated administration of IL-31 on PGD_2- or capsaicin-induced hyperalgesia on the modified hot-plate (45 °C) test. Repeated pretreatment with IL-31 inhibited PGD_2-induced pain but not capsaicin-induced pain in the modified hot-plate test, suggesting that the site of action of IL-31 was after the PGs of the peripheral sensory nervous system. Furthermore, we assessed the effect of several analgesics, such as morphine, loxoprofen, and acetaminophen, on TNCB- or capsaicin-applied hot-plate (45 or 35 °C) tests to determine the effect of the IL-31-induced antinociceptive activity in wild-type and IL-31RAKI mice.

Morphine (3 mg/kg, subcutaneous) showed significant antinociception in the TNCB- or wild-type mice on the TNCB-applied hot-plate (45 °C) test but not in capsaicin- and IL-31RAKI mice. However, this was a partial inhibitory action as morphine-induced antinociceptive activity recovered in high doses (10 mg/kg, subcutaneously). Loxoprofen (15 mg/kg, oral) and acetaminophen (300 mg/kg, oral) showed significant antinociception on TNCB- or wild-type mice in the TNCB-applied hot-plate (35 °C) test; however, these antinociceptive activities were also completely inhibited by capsaicin- and IL-31RAKI mice. This result shows that the antinociceptive action of loxoprofen (NSAIDs) and acetaminophen does not develop without the activation of IL-31. Moreover, acetaminophen does not inhibit PG biosynthesis [45]. Although our present study suggested that the site of action of acetaminophen is the CNS, these collective results indicate that acetaminophen functions on the PNS, suggesting that IL-31 participates in the antinociceptive activity of these drugs. In addition, some antinociceptive activities remained in the morphine (opioid) group but not in the loxoprofen and acetaminophen groups. Chronic neuropathic pain shows strong resistance to treatment with opioids and NSAIDs [46]. This reaction is similar to drug action in chronic neuropathic pain. Therefore, the decrease in DRG neuronal IL-31RA may play a role in the onset of chronic neuropathic pain (Figure 10).

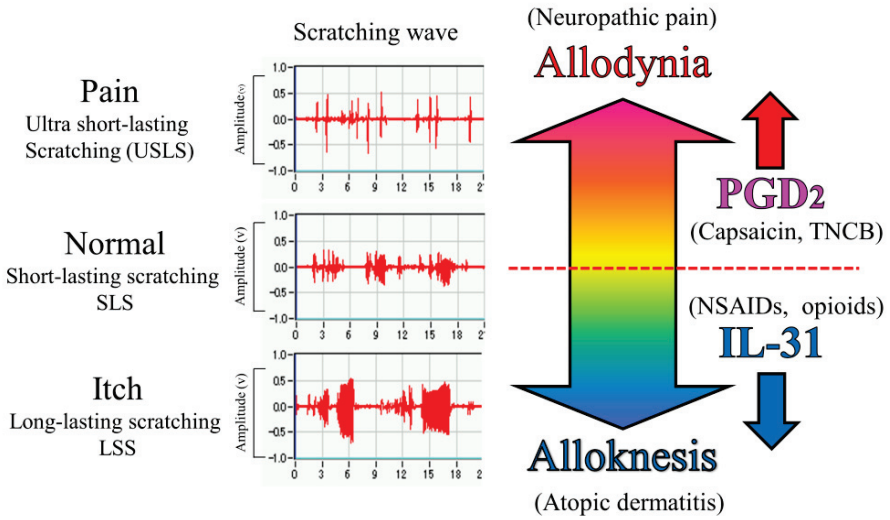


Figure 10. Schematic diagram of the putative roles of IL-31 and PGD_2 in the regulation of the sensation of cutaneous itch and pain in mice. We evaluated a wave pattern of scratching behavior from the movement of the hind leg of the mouse, objectively. In normal conditions, the mouse elicits 0.3–1.0 s lasting scratching behavior (SLS) caused by several stimulants. Under itchy conditions, the mouse elicits over 1.0 s lasting scratching behavior (LLS) caused by IL-31. If this stimulation continues for a long time, it may become atopic dermatitis. Under painful conditions, the mouse elicits below 0.3 s lasting scratching behavior (USLS) caused by chronic inflammatory stimulants. If this stimulation continues for a long time, it may become chronic neuropathic pain. IL-31 can change non-selective stimulation into itch stimulation. In contrast, PGD_2 can change non-selective stimulation into pain stimulation transmitted by the primary nerves of C-fibers and by second-order nerves and spinothalamic tract neurons in the spinal cord. This suggests that IL-31 and PGD_2 regulate the perception of sense (pain or itch) through their mutual functional antagonism.

In the present study, we objectively evaluated the distinction between pain and itch using a wave pattern of scratching behavior by the movement of the hind leg of the mouse [21]. When the mouse skin is in a normal condition, e.g., histamine-injected intradermally, the mouse elicits 0.3–1.0 s of lasting scratching behavior (LLS), suggesting that the mouse felt something on the epidermis. When mice experience an itch, e.g., when IL-31 is injected intradermally, they elicit over 1.0 s lasting scratching behavior (LLS). This itching strips off the epidermis and removes an alien substance. For example, when a mite irritates the skin, IL-31 is highly expressed in that region, and an itch sensation is produced in response to

various kinds of cutaneous stimulation [38]. If this stimulation continues for a long time, it may become atopic dermatitis due to the overexpression of DRG neuronal IL-31RA [39] (Figure 10). In contrast, when the mouse skin is experiencing a painful condition, e.g., TNCB (PGs inducer) application in the dorsal skin, the mouse elicits ultra-short-lasting scratching behavior (USLS) (less than 0.3 s) or no scratching behavior. Therefore, in the present study, we used a modified hot-plate test for a real pain evaluation. When the mouse was experiencing pain, it did not touch the skin, if possible. PGs is produced in response to inflammation, such as that in response to a burn, and may be involved in the onset of pain in response to various kinds of cutaneous stimulation. If this stimulation continues for a long time, it may induce chronic neuropathic pain due to the decreased expression of DRG neuronal IL-31RA (Figure 10).

In this study, we evaluated the effects of TNCB- or capsaicin-induced pain and itch after 24 h of this application on a modified hot-plate (45 °C) test. TNCB and capsaicin induced pain and inhibited itching after the application in a similar manner until 6 h, but they showed different reactions 24 h later. The TNCB application led to the development of pain torpor, which was not observed in IL-31RAKI mice. Therefore, this reaction was considered a result of endogenous IL-31. The itch (LLS counts) increase corresponded with the increase in the DRG neuronal IL-31RA expression. This reaction might be the cause of itching during the wound recovery period. The itch observed 24 h after TNCB application may be due to an increase in DRG neuronal IL-31RA expression after the restoration of the wound. In contrast, the application of capsaicin induced decreased latency 24 h after application and a significant decrease 6 days after. Strong pain is sometimes observed after the application of capsaicin cream when an individual takes a warm water shower. This pain can be explained by a decrease in DRG neuronal IL-31RA expression after capsaicin application. Although decreased latency caused by capsaicin was restored after 72 h of application, IL-31-induced (1 mg/kg, intravenously) LLS counts and DRG neuronal IL-31RA expression significantly decreased 144 h (6 days) after capsaicin application. IL-31RAKI mice lack a peripheral analgesic mechanism; therefore, the recovery from this latency involves the participation of the central analgesic mechanism.

These findings suggest that IL-31 may cause alloknesis; it alters the non-selective irritant stimulation into itch stimulation in mouse skin [47,48]. Therefore, it is possible that pain and itch are transmitted on the same nerve fibers, and a sensation is perceived as pain or itch depending on the operation of IL-31. However, PGD_2 decreased LLS counts induced by the cutaneous injection of pruritogens or algogens in IL-31-induced itchy skin. These results indicate that PGD_2 improves IL-31-induced alloknesis and it alters the non-selective irritant stimulation into pain stimulation in mouse skin [43]. Collectively, these data suggest that the sensation of pain and itch may be regulated by PGD_2 (allodynia-inducer) [49] and/or IL-31 (alloknesis-inducer) [38], through their functional antagonism (Figure 10).

The itch can be produced by a gentle touch, pressure, vibration, thermal, and electrical stimuli, such as transcutaneous or direct nerve stimulation. While itching may have different causes, it is conceived similarly by our senses. Our study indicates that the sensation of itch and pain depends on the type of stimulant of the peripheral sensory nerve ending. A strong pain will be felt, even if it is an itch if the burnt part is gently touched, suggesting that the sense of itch and pain are not stimulated from the outside, but from the inside. In this study, we were able to show that pain and itching sensations may be regulated by their functional antagonism as a phenomenon side, but the adjustment mechanism involved is still unknown and should be elucidated in future studies. And this study indicates that the analgesic action of IL-31 involves the PNS, which directly affects sensory nerves, providing a basis for developing novel analgesics using this mechanism.

4. Materials and Methods

4.1. Animals

Male BALB/c and C57BL/6 mice aged 6–8 weeks were purchased from SLC Japan (Shizuoka, Japan). Mice lacking IL-31RA (IL-31RA$^{-/-}$) were generated as described previ-

ously [50]. The IL-31RAKI (IL-31RA-deficient) mice used in this study were on a C57BL/6 genetic background and were obtained from hybrid mutant mice originally created based on a 129 SVJ-C57BL/6 background by backcrossing breeding over 15 generations. In this study, we used male homozygous (IL-31RAKI, IL-31RA$^{-/-}$) and wild-type (IL-31RA$^{+/+}$) mice matched for age. Therefore, we used C57BL/6 mice for the target group in the experiment using the IL-31RAKI mice. The animals were housed under a controlled temperature (23 ± 3 °C), humidity (50 ± 5%), and lighting (lights on from 7:00 am to 7:00 pm). All animals were provided free access to food and tap water. All procedures for animal experiments were approved by the Committee for Animal Experimentation at the International University of Health and Welfare following the Guidelines for Proper Conduct of Animal Experiments (Science Council of Japan, 2006).

4.2. Drugs

The mouse IL-31 cDNA-spanning amino acids 24–163 of IL-31 were cloned in a frame with pET30A (Novagen), and the construct was transformed into BL-21 cells (Novagen, Darmstadt, Germany). After induction with isopropyl-β-D-thiogalactopyranoside, IL-31 protein was purified under denaturation conditions using nickel-chelating sepharose (Qiagen, Benelux B.V., Hulsterweg, The Netherlands) and dialyzed in PBS [51]. IL-31, at a dose of 50 μg/kg, was used for subcutaneous or intravenous injection, as previously described [39]. Morphine hydrochloride (Takeda Pharmaceutical Co., Ltd., Osaka, Japan) was dissolved in saline and administered at 0.5 mL/kg, subcutaneously. Loxoprofen (Loxonin, Daiichi-Sankyou, Tokyo, Japan) and acetaminophen (Tylenol, Towa Pharmaceutical Co., Ltd., Toyama, Japan) were purchased as commercially available drugs. They were suspended in 0.3% carboxymethyl cellulose sodium saline (CMC), 0.5 mL/kg, for oral administration. TNCB (2,4,6-trinitrochlorobenzene, Tokyo Kasei, Tokyo, Japan) or capsaicin (Wako Jyunyaku, Tokyo, Japan) were dissolved in an acetone ethanol mixture (acetone: ethanol = 1:4, AEM) or ethanol and applied to the limbs of mice (0.04 mL/site) using a pipette. The application of 0.16 mL TNCB and capsaicin on the hot-plate test was sufficient for this method. Aliquots of 0.04 mL/site were applied to the site of application. Scratching counts can be measured by evaluating the expression of several mRNAs, the measurement of cutaneous PGD_2 contents tests, and the application of 0.2 mL TNCB or capsaicin/site using a pipette.

4.3. Hot-Plate Test

The hot plate test [12] was used to measure withdrawal latency as described previously [19]. The hot-plate test was improved to evaluate even weak analgesic drugs such as NSAIDs: loxoprofen. Mice were placed on a hot-plate maintained at 50 ± 0.5 °C, and the latency to either paw-lick or an attempt to escape by jumping was recorded. To prevent tissue damage, mice that showed no response within 60 s were removed from the hot-plate and assigned a score of 60 s. The percentage of nociception (nociceptive index) was calculated according to the formula: $[(T_1 - T_0)/(T_2 - T_0)] \times 100$, where T_0 and T_1 were the latencies observed before and after the drug administration, respectively, and T_2 was the cut-off time (60 s). The modified hot-plate test applied 2,4,6-trinitrochlorobenzene (TNCB, 0.3, 1, and 3 w/v %) to the limbs of mice at a lower temperature (30 or 45 °C) than that used in the hot-plate test. Animals were tested before and 30, 60, 90, 120, 150, and 180 min after drug administration. As the number of experiments in which pretreatment of TNCB, capsaicin, or PGD_2 application was over the cut-off level of 60 s was high, it was impossible to calculate many nociceptive indices at 35 or 45 °C. Therefore, the data were transcribed in latency.

4.4. Measurement of Scratching Counts

Scratching counts were measured as previously described [43]. A small magnet (1.0 mm diameter, 3.0 mm length) was implanted subcutaneously into both hind paws of isoflurane-anesthetized mice 24 h before the measurement. Each mouse was placed in an

observation chamber (11 cm diameter, 18 cm height) surrounded by a circular coil, through which electric current, induced by movement of the magnets attached to the hind paws, was amplified and recorded. The number of spontaneous scratches was automatically detected and objectively evaluated via a computer using MicroAct (Neuroscience, Tokyo, Japan) [52]. The analysis parameters for detecting waves were as follows: threshold, 0.1 V; event gap, 0.2 s; minimum duration, 0.3 or 1.0 s; maximum frequency, 20 Hz; and minimum frequency, 2 Hz.

4.5. Measurement of Dorsal Cutaneous Temperature and Skin Weight

Time-course changes of dorsal cutaneous surface temperature were measured after 0 (pre), 1-, 3-, 6-, 24-, and 72 h after 3 w/v% TNCB application in the dorsal skin of mice. The dorsal hair of the mouse was shaved off with an electric razor, and 0.2 mL of 3% TNCB solution was applied using a pipette. The cutaneous surface temperature was measured using a non-contact type thermometer (HPC-01, Harasawa Co. Ltd., Tokyo, Japan). Mice were killed by dislocation of the cervical vertebrae after 6, 24, and 36 h of TNCB application. The dorsal skin of each mouse was removed with a fixed circular area (Φ 1.6 mm) at the site of the TNCB application. The skin was weighed, and the oncotic index of skin swelling was measured.

4.6. Measurement of Cutaneous Prostaglandin D_2 Content

The mice were injected intravenously with indomethacin (10 mg/kg) 1 h after TNCB application, as well as 1, 3, 6, and 12 days after TNCB application to prevent further production of prostaglandins D_2 (PGD_2). Five minutes after the indomethacin injection, the mice were killed by dislocation of the cervical vertebrae, and approximately 100 mg of the back skin of each mouse was removed. The skin was minced and homogenized in ice-cold phosphate-buffered saline containing 0.1 mM indomethacin with a Polytron tissue homogenizer for 30 s on ice. Four milliliters of acetone were added to the sample and mixed. The precipitate was then removed by centrifugation at 2000× g for 10 min at 4 °C. The supernatant was carefully poured into a test tube, evaporated under a stream of nitrogen, and re-suspended in an enzyme immunoassay (EIA) buffer. The amount of PGD_2 in the suspension was measured using specific EIA kits (Cayman Chemical, Ann Arbor, MI, USA) following the manufacturer's instructions.

4.7. Real-Time Quantitative PCR

The gene expression levels of IL-31, IL-31RA, TRPV1, and β-actin were measured using real-time polymerase chain reaction (RT-PCR) in the DRG ($C_{4–7}$, $T_{1–4}$) neuron cell body from the shoulder and back of the BALB/c and C57BL/6 mice at each point. Total RNA was extracted from the dorsal skin of each mouse by Trizol (Invitrogen, Carlsbad, CA, USA) and digested using amplification-grade DNase I (Invitrogen), according to the manufacturer's instruction. cDNA was synthesized by the SuperScript III First-Strand Synthesis System (Invitrogen). Quantitative RT-PCR was performed with SYBR Green Master Mix, using an Applied Biosystems 7700 Sequence Detection System (Applied Biosystems, Foster City, CA, USA). The PCR primers for IL-31 were designed using PRIMER 3 software (v. 0.4.0), and primers for TRPV1 and β-actin were purchased from TAKARA BIO (Otsu, Shiga, Japan). Primer sequences were as follows: IL-31 (5′-ATA CAG CTG CCG TGT TTC AG-3′ and 5′ -AGC CAT CTT ATC ACC CAA GAA-3′), IL-31RA (5′-CCA GAA GCT GCC ATG TCG AA-3′ and 5′-TCT CCA ACT CGG TGT CCC AAC-3′), TRPV1 (5′-CAA CAA GAA GGG GCT TAC ACC-3′ and 5′ -TCT GGA GAA TGT AGG CCA AGA C-3′), and β—actin (5′ -TGA CAG GAT GCA GAA GGA GA-3′ and 5′-GCT GGA AGG TGG ACA GTG AG-3′). Relative expression levels were calculated using the relative standard curve method as outlined in the manufacturer's technical bulletin. A standard curve was generated using the fluorescence data obtained from four-fold serial dilutions of the total RNA of the sample with the highest expression. The curve was then used to calculate the relative amounts of

target mRNA in test samples. Quantities of all targets in the test samples were normalized to the corresponding β-actin RNA transcript in skin samples.

4.8. Statistical Analysis

All data were statistically analyzed using GraphPad InStat and GraphPad Prism (GraphPad Software, Version 7, Inc., La Jolla, CA, USA). Experimental values are expressed as means and standard errors (S.E.). Data on time-course changes in scratching counts or the percentage of mRNA expression were analyzed using one-way ANOVA followed by the Student–Newman–Keuls test. Then, other data were analyzed using one-way, two-way, or three-way ANOVA followed by Bonferroni or Tukey tests. In some cases, when a main effect was significant without interaction effect, we did an exploratory and limited pairwise post hoc comparison consistent with our a priori hypothesis. *p*-values less than 0.05 were considered statistically significant.

5. Conclusions

Itch and pain are common senses caused by several drugs and physical stimulations. However, the onset mechanisms of itch and pain remain unclear. We objectively evaluated itch and pain via a wave pattern of scratching behavior in mice. Scratching behavior in mice is divided into two types, LLS (itch-associated scratching) and SLS (hygiene behavior). Moreover, morphine is one of the few drugs that induce LLS. In IL-31RAKI mice, LLS disappeared upon the administration of morphine, accompanied by the partial disappearance of antinociceptive action. Moreover, we found that IL-31 was partially involved in the peripheral analgesic mechanism and that IL-31-induced alloknesis inhibited pain. The antinociceptive activity of IL-31 was approximately as effective as that of loxoprofen (NSAIDs). However, the site of action of IL-31 was in the peripheral sensory nerve after the PGs biosynthesis inhibition, unlike NSAIDs. In addition, cutaneous IL-31 expression increases at the onset of pain and DRG neuronal IL-31RA expression increased during recovery. Our results suggest that IL-31 functions via a physiological pain-inhibitory mechanism. These results indicate that the analgesic action of IL-31 involves the peripheral nervous system, which affects sensory nerves and provides a basis for the development of novel analgesics that utilize this mechanism.

Author Contributions: I.A. and M.T. performed the research, designed the study, conducted the study, collected the data, and wrote the paper, K.T. performed data analysis, S.S. contributed essential reagents, generated and provided genetically modified animals, and helped in the preparation of the manuscript, H.T. conducted the study and helped in the preparation of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank N. Akiyama, M. Sugimoto, and K. Miyagawa for technical assistance.

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

References

1. Dillon, S.R.; Sprecher, C.; Hammond, A.; Bilsborough, J.; Rosenfeld-Franklin, M.; Presnell, S.R.; Haugen, H.S.; Maurer, M.; Harder, B.; Johnston, J.; et al. Interleukin 31, a cytokine produced by activated T cells, induces dermatitis in mice. *Nat. Immunol.* **2004**, *5*, 752–760. [CrossRef] [PubMed]
2. Sonkoly, E.; Muller, A.; Lauerma, A.I.; Pivarcsi, A.; Soto, H.; Kemeny, L.; Alenius, H.; Dieu-Nosjean, M.-C.; Meller, S.; Rieker, J.; et al. IL-31: A new link between T cells and pruritus in atopic skin inflammation. *J. Allergy Clin. Immunol.* **2006**, *117*, 411–417. [CrossRef] [PubMed]

3. Bilsborough, J.; Leung, D.Y.; Maurer, M.; Howell, M.; Boguniewcz, M.; Yao, L.; Storey, H.; LeCiel, C.; Harder, B.; Gross, J.A. IL-31 is associated with cutaneous lymphocyte antigen–positive skin homing T cells in patients with atopic dermatitis. *J. Allergy Clin. Immunol.* **2006**, *117*, 418–425. [CrossRef] [PubMed]
4. Arai, I.; Tsuji, M.; Miyagawa, K.; Takeda, H.; Akiyama, N.; Saito, S. Repeated administration of IL-31 upregulates IL-31 receptor A (IL-31RA) in dorsal root ganglia and causes severe itch-associated scratching behaviour in mice. *Exp. Dermatol.* **2014**, *24*, 75–78. [CrossRef] [PubMed]
5. Arai, I.; Tsuji, M.; Miyagawa, K.; Takeda, H.; Akiyama, N.; Saito, S. Increased itching sensation depends on an increase in the dorsal root ganglia IL-31 receptor A (IL-31RA) expression in mice with atopic-like dermatitis. *Itch Pain* **2015**, *1*, 10. [CrossRef]
6. Shannon, H.E.; Lutz, E.A. Comparison of the peripheral and central effects of the opioid agonists loperamide and morphine in the formalin test in rats. *Neuropharmacology* **2001**, *42*, 253–261. [CrossRef]
7. Kuraishi, Y.; Harada, Y.; Aratani, S.; Satoh, M.; Takagi, H. Separate involvement of the spinal noradrenergic and serotonergic systems in morphine analgesia: The differences in mechanical and thermal algesic tests. *Brain Res.* **1983**, *273*, 245–252. [CrossRef]
8. Kuraishi, Y.; Hirota, N.; Satoh, M.; Takagi, H. Antinociceptive effects of intrathecal opioids, noradrenaline and serotonin in rats: Mechanical and thermal algesic tests. *Brain Res.* **1985**, *326*, 168–171. [CrossRef]
9. Tsuji, M.; Arai, I.; Miyagawa, K.; Miyagishi, H.; Saito, A.; Takeda, K.; Takeda, H.; Akiyama, N.; Saito, S. Involvement of interleukin-31 receptor A in morphine-induced itching and antinociception in mice. *Eur. J. Pain* **2018**, *23*, 378–388. [CrossRef]
10. Scholzen, T.E.; Steinhoff, M.; Sindrilaru, A.; Schwarz, A.; Bunnett, N.W.; Luger, T.A.; Armstrong, C.A.; Ansel, J.C. Cutaneous allergic contact dermatitis responses are diminished in mice deficient in neurokinin 1 receptors and augmented by neurokinin 2 receptor blockage. *FASEB J.* **2004**, *18*, 1007–1009. [CrossRef]
11. Smith, H.S. Peripherally-Acting Opioids. *Pain Physician* **2008**, *11*, S121–S132. [CrossRef] [PubMed]
12. Eddy, N.B.; Leimbach, D. Synthetic Analgesics II. Dithienylbutenyl- and Dithienylbutylamines. *J. Pharmacol. Exp. Ther.* **1953**, *107*, 385–393. [PubMed]
13. Xue, Y.; Dai, S.; Liang, J.; Ji, W. Effect of reactive oxygen species of the psoas major muscle in complete Freund's adjuvant-induced inflammatory pain in rats. *Mol. Pain* **2020**, *16*, 1744806920929246. [CrossRef]
14. Li, J.; Chen, Y.; Liu, J.; Zhang, D.; Liang, P.; Lu, P.; Shen, J.; Miao, C.; Zuo, Y.; Zhou, C. Elevated Expression and Activity of Sodium Leak Channel Contributes to Neuronal Sensitization of Inflammatory Pain in Rats. *Front. Mol. Neurosci.* **2021**, *14*, 723395. [CrossRef]
15. Randall, L.O.; Sellito, J.J. A Method for Measurement of Analgesic Activity on Inflamed Tissue. *Arch. Int. Pharmacodyn. Thér.* **1957**, *111*, 409–411.
16. Futaki, N.; Harada, M.; Sugimoto, M.; Hashimoto, Y.; Honma, Y.; Arai, I.; Nakaike, S.; Hoshi, K. The importance of brain PGE2 inhibition versus paw PGE2 inhibition as a mechanism for the separation of analgesic and antipyretic effects of lornoxicam in rats with paw inflammation. *J. Pharm. Pharmacol.* **2009**, *61*, 607–614. [CrossRef] [PubMed]
17. Garrigue, J.-L.; Nicolas, J.-F.; Fraginals, R.; Benezra, C.; Bour, H.; Schmitt, D. Optimization of the mouse ear swelling test for in vivo and in vitro studies of weak contact sensitizers. *Contact Dermat.* **1994**, *30*, 231–237. [CrossRef] [PubMed]
18. Takahashi, N.; Arai, I.; Honma, Y.; Hashimoto, Y.; Harada, M.; Futaki, N.; Sugimoto, M.; Nakaike, S. Scratching behavior in spontaneous- or allergic contact-induced dermatitis in NC/Nga mice. *Exp. Dermatol.* **2005**, *14*, 830–837. [CrossRef]
19. Arai, I.; Tsuji, M.; Saito, S.; Takeda, H. Analyzing the Antinociceptive Effect of Analgesics Using a Modified Hot-Plate Test in Mice. *J. Clin. Trials* 2023. [CrossRef]
20. Kuraishi, Y.; Nagasawa, T.; Hayashi, K.; Satoh, M. Scratching behavior induced by pruritogenic but not algesiogenic agents in mice. *Eur. J. Pharmacol.* **1995**, *275*, 229–233. [CrossRef]
21. Takano, N.; Arai, I.; Kurachi, M. Analysis of the spontaneous scratching behavior by NC/Nga mice: A possible approach to evaluate antipruritics for subjects with atopic dermatitis. *Eur. J. Pharmacol.* **2003**, *471*, 223–228. [CrossRef] [PubMed]
22. Hashimoto, Y.; Takano, N.; Nakamura, A.; Nakaike, S.; Yu, Z.; Endo, Y.; Arai, I. Scratching behavior in NC/Nga mice with dermatitis: Involvement of histamine-induced itching. *Allergol. Int.* **2004**, *53*, 349–358. [CrossRef]
23. Frias, B.; Merighi, A. Capsaicin, Nociception and Pain. *Molecules* **2016**, *21*, 797. [CrossRef] [PubMed]
24. Chung, M.-K.; Jung, S.J.; Oh, S.B. Role of TRP Channels in Pain Sensation. *Adv. Exp. Med. Biol.* **2010**, *704*, 615–636. [CrossRef]
25. Takaoka, A.; Arai, I.; Sugimoto, M.; Futaki, N.; Sakurai, T.; Honma, Y.; Nakaike, S. Role of scratch-induced cutaneous prostaglandin D_2 production on atopic-like scratching behaviour in mice. *Exp. Dermatol.* **2007**, *16*, 331–339. [CrossRef] [PubMed]
26. Matsuda, H.; Watanabe, N.; Geba, G.P.; Sperl, J.; Tsudzuki, M.; Hiroi, J.; Matsumoto, M.; Ushio, H.; Saito, S.; Askenase, P.W.; et al. Development of atopic dermatitis-like skin lesion with IgE hyperproduction in NC/Nga mice. *Int. Immunol.* **1997**, *9*, 461–466. [CrossRef]
27. Futaki, N.; Arai, I.; Sugimoto, M.; Honma, Y.; Hashimoto, Y.; Sakurai, T.; Nakanishi, Y.; Tanaka, M.; Nakaike, S. Role of prostaglandins on mechanical scratching-induced cutaneous barrier disruption in mice. *Exp. Dermatol.* **2007**, *16*, 507–512. [CrossRef]
28. Arai, I.; Takano, N.; Hashimoto, Y.; Futaki, N.; Sugimoto, M.; Takahashi, N.; Inoue, T.; Nakaike, S. Prostanoid DP1 receptor agonist inhibits the pruritic activity in NC/Nga mice with atopic dermatitis. *Eur. J. Pharmacol.* **2004**, *505*, 229–235. [CrossRef]
29. Wang, C.; Li, G.W.; Huang, L.Y.M. Prostaglandin E2 Potentiation of P2X3 Receptor Mediated Currents in Dorsal Root Ganglion Neurons. *Mol. Pain* **2007**, *10*, 22. [CrossRef]

30. Caterina, M.J.; Schumacher, M.A.; Tominaga, M.; Rosen, T.A.; Levine, J.D.; Julius, D. The capsaicin receptor: A heat-activated ion channel in the pain pathway. *Nature* **1997**, *389*, 816–824. [CrossRef]
31. Joseph, J.; Wang, S.; Lee, J.; Ro, J.Y.; Chung, M.-K. Carboxyl-terminal Domain of Transient Receptor Potential Vanilloid 1 Contains Distinct Segments Differentially Involved in Capsaicin- and Heat-induced Desensitization. *J. Biol. Chem.* **2013**, *288*, 35690–35702. [CrossRef] [PubMed]
32. Ma, X.-L.; Zhang, F.-X.; Dong, F.; Bao, L.; Zhang, X. Experimental Evidence for Alleviating Nociceptive Hypersensitivity by Single Application of Capsaicin. *Mol. Pain* **2015**, *11*, 22. [CrossRef]
33. Borbiro, I.; Badheka, D.; Rohacs, T. Activation of TRPV1 channels inhibits mechanosensitive Piezo channel activity by depleting membrane phosphoinositides. *Sci. Signal.* **2015**, *8*, ra15. [CrossRef] [PubMed]
34. Arai, I.; Tsuji, M.; Takeda, H.; Akiyama, N.; Saito, S. Capsaicin suppresses interleukin-31-induced itching partially involved in inhibiting the expression of dorsal root ganglion interleukin-31 receptor A in male mice. *Neurobiol. Pain* **2022**, *11*, 100088. [CrossRef]
35. Arai, I.; Tsuji, M.; Takeda, H.; Akiyama, N.; Saito, S. A single dose of interleukin-31 (IL-31) causes continuous itch-associated scratching behaviour in mice. *Exp. Dermatol.* **2013**, *22*, 669–671. [CrossRef]
36. Akiyama, T.; Merrill, A.W.; Carstens, M.I.; Carstens, E. Activation of Superficial Dorsal Horn Neurons in the Mouse by a PAR-2 Agonist and 5-HT: Potential Role in Itch. *J. Neurosci.* **2009**, *29*, 6691–6699. [CrossRef]
37. Inagaki, N.; Igeta, K.; Kim, J.F.; Nagao, M.; Shiraishi, N.; Nakamura, N.; Nagai, H. Involvement of unique mechanisms in the induction of scratching behavior in BALB/c mice by compound 48/80. *Eur. J. Pharmacol.* **2002**, *448*, 175–183. [CrossRef]
38. Arai, I.; Tsuji, M.; Miyagawa, K.; Takeda, H.; Akiyama, N.; Saito, S. Interleukin-31 (L-31) causes alloknesis: Pain-stimulation becomes itch-stimulation in mouse skin. *Itch Pain* **2015**, *2*, e924. [CrossRef]
39. Arai, I.; Saito, S. Interleukin-31 Receptor A Expression in the Dorsal Root Ganglion of Mice with Atopic Dermatitis. *Int. J. Mol. Sci.* **2023**, *24*, 1047. [CrossRef] [PubMed]
40. Davidson, S.; Giesler, G.J. The multiple pathways for itch and their interactions with pain. *Trends Neurosci.* **2010**, *33*, 550–558. [CrossRef]
41. Heyer, G.; Dotzer, M.; Diepgen, L.T.; Handwerker, O.H. Opiate and H1 antagonist effects on histamine induced pruritus and alloknesis. *Pain* **1997**, *73*, 239–243. [CrossRef] [PubMed]
42. Fjellner, B.; Hägermark, O. Potentiation of histamine-induced itch and flare responses in human skin by the enkephalin analogue FK 33-824, ?-endorphin and morphine. *Arch. Dermatol. Res.* **1982**, *274*, 29–37. [CrossRef] [PubMed]
43. Arai, I.; Tsuji, M.; Miyagawa, K.; Takeda, H.; Akiyama, N.; Saito, S. Prostaglandin D2 improves IL-31-induced alloknesis: Itch-stimulation becomes pain-stimulation in mouse skin. *Itch Pain* **2016**, *1*, e1138. [CrossRef]
44. Ohkubo, T.; Shibata, M.; Takahashi, H.; Inoki, R. Effect of prostaglandin D2 on pain and inflammation. *Jpn. J. Pharmacol.* **1983**, *33*, 264–266. [CrossRef]
45. Esh, C.J.; Chrismas, B.C.R.; Mauger, A.R.; Taylor, L. Pharmacological hypotheses: Is acetaminophen selective in its cyclooxygenase inhibition? *Pharmacol. Res. Perspect.* **2021**, *9*, e00835. [CrossRef] [PubMed]
46. Backonja, M.; Glanzman, R.L. Gabapentin dosing for neuropathic pain: Evidence from randomized, placebo-controlled clinical trials. *Clin. Ther.* **2003**, *25*, 81–104. [CrossRef] [PubMed]
47. Ikoma, A.; Rukwied, R.; Ständer, S.; Steinhoff, M.; Miyachi, Y.; Schmelz, M. Neuronal Sensitization for Histamine-Induced Itch in Lesional Skin of Patients with Atopic Dermatitis. *Arch. Dermatol.* **2003**, *139*, 1455–1458. [CrossRef]
48. Ikoma, A.; Fartasch, M.; Heyer, G.; Miyachi, Y.; Handwerker, H.; Schmelz, M. Painful stimuli evoke itch in patients with chronic pruritus: Central sensitization for itch. *Neurology* **2004**, *62*, 212–217. [CrossRef]
49. Eguchi, N.; Minami, T.; Shirafuji, N.; Kanaoka, Y.; Tanaka, T.; Nagata, A.; Yoshida, N.; Urade, Y.; Ito, S.; Hayaishi, O. Lack of tactile pain (allodynia) in lipocalin-type prostaglandin D synthase-deficient mice. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 726–730. [CrossRef]
50. Saito, S.; Aoki, A.; Arai, I.; Takaishi, S.; Ito, H.; Akiyama, N.; Kiyonari, H. Regulation of Th2 responses by different cell types expressing the interleukin-31 receptor. *Allergy Asthma Clin. Immunol.* **2017**, *13*, 23. [CrossRef]
51. Akiyama, N.; Ohno, Y.; Fukuda, T.; Manome, Y.; Saito, S. Enhancing activity of N-glycosylation for constitutive proteins secretions in non-polarized cells. *Biochem. Biophys. Res. Commun.* **2009**, *381*, 612–618. [CrossRef] [PubMed]
52. Inagaki, N.; Igeta, K.; Shiraishi, N.; Kim, J.; Nagao, M.; Nakamura, N.; Nagai, H. Evaluation and Characterization of Mouse Scratching Behavior by a New Apparatus, MicroAct. *Skin Pharmacol. Appl. Skin Physiol.* **2003**, *16*, 165–175. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Article

A New Strategy for the Old Challenge of Thalidomide: Systems Biology Prioritization of Potential Immunomodulatory Drug (IMiD)-Targeted Transcription Factors

Thayne Woycinck Kowalski [1,2,3,4,5,*], Mariléa Furtado Feira [1,3], Vinícius Oliveira Lord [3,5], Julia do Amaral Gomes [3], Giovanna Câmara Giudicelli [1,4], Lucas Rosa Fraga [2,3,6,7], Maria Teresa Vieira Sanseverino [1,2,8], Mariana Recamonde-Mendoza [4,9], Lavinia Schuler-Faccini [1,2] and Fernanda Sales Luiz Vianna [1,2,3,6,*]

1 Graduate Program in Genetics and Molecular Biology, Genetics Department, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, Brazil; mfeira@hcpa.edu.br (M.F.F.); ggiudicelli@hcpa.edu.br (G.C.G.); msanseverino@hcpa.edu.br (M.T.V.S.); lschuler@hcpa.edu.br (L.S.-F.)
2 Teratogen Information System (SIAT), Medical Genetics Service, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre 90035-903, Brazil; lrfraga@ufrgs.br
3 Laboratory of Genomic Medicine, Center of Experimental Research, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre 90035-903, Brazil; vlord@hcpa.edu.br (V.O.L.); juliagomes@hcpa.edu.br (J.d.A.G.)
4 Bioinformatics Core, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre 90035-903, Brazil; mrmendoza@inf.ufrgs.br
5 Biomedical Sciences Course, Centro Universitário CESUCA, Cachoeirinha 94935-630, Brazil
6 Post-Graduation Program in Medicine, Medical Sciences, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 90035-003, Brazil
7 Department of Morphological Sciences, Institute of Health Sciences, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 90010-150, Brazil
8 School of Medicine, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre 90619-900, Brazil
9 Post-Graduation Program in Computer Science, Institute of Informatics, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, Brazil
* Correspondence: tkowalski@hcpa.edu.br (T.W.K.); fvianna@hcpa.edu.br (F.S.L.V.)

Citation: Kowalski, T.W.; Feira, M.F.; Lord, V.O.; Gomes, J.d.A.; Giudicelli, G.C.; Fraga, L.R.; Sanseverino, M.T.V.; Recamonde-Mendoza, M.; Schuler-Faccini, L.; Vianna, F.S.L. A New Strategy for the Old Challenge of Thalidomide: Systems Biology Prioritization of Potential Immunomodulatory Drug (IMiD)-Targeted Transcription Factors. *Int. J. Mol. Sci.* **2023**, *24*, 11515. https://doi.org/10.3390/ijms241411515

Academic Editor: Wajid Zaman

Received: 1 June 2023
Revised: 6 July 2023
Accepted: 8 July 2023
Published: 15 July 2023

Abstract: Several molecular mechanisms of thalidomide embryopathy (TE) have been investigated, from anti-angiogenesis to oxidative stress to cereblon binding. Recently, it was discovered that thalidomide and its analogs, named immunomodulatory drugs (IMiDs), induced the degradation of C2H2 transcription factors (TFs). This mechanism might impact the strict transcriptional regulation of the developing embryo. Hence, this study aims to evaluate the TFs altered by IMiDs, prioritizing the ones associated with embryogenesis through transcriptome and systems biology-allied analyses. This study comprises only the experimental data accessed through bioinformatics databases. First, proteins and genes reported in the literature as altered/affected by the IMiDs were annotated. A protein systems biology network was evaluated. TFs beta-catenin (CTNNB1) and SP1 play more central roles: beta-catenin is an essential protein in the network, while SP1 is a putative C2H2 candidate for IMiD-induced degradation. Separately, the differential expressions of the annotated genes were analyzed through 23 publicly available transcriptomes, presenting 8624 differentially expressed genes (2947 in two or more datasets). Seventeen C2H2 TFs were identified as related to embryonic development but not studied for IMiD exposure; these TFs are potential IMiDs degradation neosubstrates. This is the first study to suggest an integration of IMiD molecular mechanisms through C2H2 TF degradation.

Keywords: lenalidomide; pomalidomide; cereblon; bioinformatics; zinc finger; microarray; RNA-seq; teratogen; transcriptome; graph

1. Introduction

The lack of proper research on and the understanding of reproductive toxicology had severe consequences in the 1960s. Marketed as a safe drug, thalidomide sales around the world were used as a panacea [1]; however, its use in early pregnancy occurred soon after its release associated with the outburst of babies born with a range of congenital anomalies [2], later named thalidomide embryopathy (TE). It is believed that TE affected ten-thousand children worldwide before it was withdrawn from the market in 1962 [1]. TE is especially characterized by limb anomalies; however, the drug can compromise the correct development of almost every organ and system [3,4]. Multidisciplinary research has focused on the attempt to encounter a safe alternative for thalidomide [5]. The need is urgent once thalidomide immunomodulatory and anti-angiogenic properties led to its approval for multiple-myeloma treatment worldwide [6], and the other immunomodulatory drugs (IMiDs) synthesized later, lenalidomide and pomalidomide, are also used for treating this condition and are teratogenic in animal models [7–9]. In Brazil, thalidomide has been mainly used for erythema nodosum for leprosy (ENL) treatment since 1965 [10–12]. Although the studies demonstrate most individuals in treatment with thalidomide for ENL are male [13,14], a percentage of the subjects refer to women of a reproductive age [13].

Past researchers have discovered relevant molecular mechanisms that might be involved in TE, including anti-angiogenesis [15,16], increased apoptosis through the induction of oxidative stress [17], and binding to the cereblon protein, which is part of an E3-ubiquitin–ligase complex, named CRL4-CRBN [18]. The latter has been the main focus of the research lately because of the identification that CRL4-CRBN induces the degradation of Spalt-like transcription factor 4 (SALL4) when in the presence of thalidomide [19,20]. SALL4 is a zinc-finger transcription factor (TF) containing a C2H2 domain, and it is known for its role in limb development [19,20]. Since the discovery of thalidomide-induced SALL4 degradation, several researchers have evaluated other C2H2 TFs, which might also be neosubstrates of IMiD-induced degradation, through the CRL4-CRBN complex [21,22]. Neosubstrates is a term that has recently been applied to the IMiD research field to denominate CRL4-CRBN targets that are only degraded in the presence of one of the IMiDs [22], meaning these proteins are not physiological substrates of the complex. However, degrome screening performed by Sievers et al. (2018) only identified 11 zinc-finger neosubstrates [23], despite having more than seven hundred C2H2 zinc-finger TFs registered [24]. Moreover, other studies have focused on non-zinc-finger neosubstrates, such as tumor protein 63 (p63) [21] and heart and neural crest derivatives expressed 2 (HAND2) [25], or even non-TF proteins, such as casein kinase 1 alpha 1 (CK1α) [26]. Hence, this study aims to prioritize TFs that might be neosubstrates of the IMiDs thalidomide, lenalidomide, and pomalidomide, through bioinformatic combined strategies. Systems biology and differential gene expression analyses were performed after rigorous, systematized literature and database research, only comprising the experimental data. The results demonstrate beta-catenin (CTNNB1) as a hub in the network of thalidomide-affected proteins and specificity protein 1 (SP1) as a feasible C2H2 thalidomide neosubstrate. New C2H2 potential neosubstrates were also identified, resulting from an IMiD widespread effect on gene expression.

2. Results

A scheme of the strategy developed in the present study is described in Figure 1.

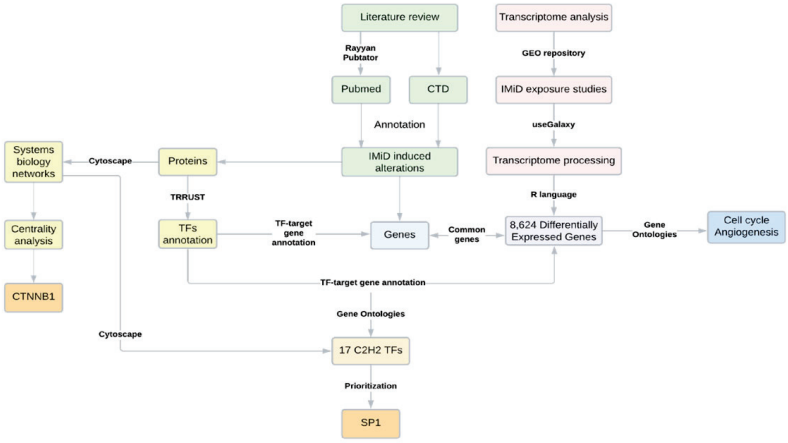


Figure 1. Scheme of the research strategy applied, comprising a literature review, bioinformatics, and systems biology analyses. CTD = comparative toxicogenomics database.

2.1. Literature Reports of 221 Genes and 80 Proteins Impacted by IMiD Exposure in Embryonic Cells/Tissues Were Identified

The literature review provided 407 entries from 48 manuscripts regarding the genes and proteins affected by the IMiDs during embryonic development (Table S1). Most of these entries (n = 339) were exposed to thalidomide, followed by pomalidomide (n = 42) and lenalidomide (n = 26). Impacts on genes were reported in 253 of the 407 entries (62.1%), and 133 of these entries were based on human embryonic stem cell (hESC) assays. The remaining entries (n = 154) referred to the impact on proteins, reported especially in human umbilical vein endothelial cells (HUVECs; n = 51). Across the studies, 23 genes and 34 proteins were replicated, meaning these genes/proteins were detected as altered by the IMiDs in two or more studies. Hence, excluding those replicates, there were 221 distinct genes and 80 distinct proteins impacted by IMiD exposure, according to the literature review; only 20 targets were reported to be affected both at the gene and protein levels. The most annotated gene in the literature review was vascular endothelial growth factor A (*VEGFA*) (n = 5), whilst the most reported protein was cereblon (n = 11). The VEGFA protein was also cited five times as being altered by IMiD exposure; however, no reports regarding the effects on the *CRBN* gene were encountered. Impacts on genes were mostly related to the expression profile, with 174 reports of downregulation and 78 reports of upregulation. Protein impacts were also related to downregulation (n = 52) and protein degradation (n = 38). The literature review demonstrated evidence that IMiDs' impacts on embryonic cells or tissues were more related to alterations in gene expressions, with *VEGFA* being the most studied gene. However, the reports showed that the studied IMiD effects on proteins were also considerably based on studying the neosubstrates' degrome mechanism, especially driven by the amount of research on the CRBN protein.

2.2. Beta-Catenin Is an Essential Protein in the Network of IMiD-Affected Proteins

The 80 proteins reported in the literature review (Table S1) to be affected by the IMiDs were inserted in the STRING tool to obtain a protein–protein interaction network, then transferred to Cytoscape v.3.7.2 software for topological analysis. Forty-six proteins (nodes) presented at least one interaction (edge), and 36 of these proteins were arranged in a main network with 51 edges, a clustering coefficient of 0.334, and an average number of neighbors of 2.833 (Figure S1). The full statistics of the main network are presented in Table S2. The ten proteins that did not interact with the nodes of the principal network were excluded from the subsequent analyses.

After the network topological analysis, beta-catenin (CTNNB1), a non-C2H2 TF, was considered the essential protein of the network, being first ranked in degree, closeness, betweenness, and maximal clique centrality (MCC) (Table 1); the complete results, for all the nodes, are presented in Table S3. This result indicates that CTNNB1 is (I) a hub based on the

high degree; (II) a node located where most of the information flows in the network, being a possible controller of this information, based on the high betweenness centrality; (III) a node through which the information flows faster, based on the high closeness centrality; and (IV) an essential protein to the network, based on the centralities' measures, including the maximal clique centrality (MCC). Other high-ranked proteins were CRBN and CUL4A, both members of the CRL4-CRBN complex.

Table 1. Topological analysis of the protein networks and centralities' definitions.

Topological Analysis	Definition	Highest-Ranked Nodes (Literature Network) [1]
Degree	Number of edges that connect to a node. Nodes with a high degree are defined as hubs.	CTNNB1 (13), CRBN (9), and AKT1 (5)
Closeness	Measures how fast the flow of information travels through the nodes. High closeness centrality scores indicate rapid information flow.	CTNNB1 (20.83), CUL4A (15.67), and CRBN (15.57)
Betweenness	Demonstrates how crowded a network is. High betweenness centrality score indicate nodes that can control the information flow.	CTNNB1 (713.67), CRBN (283), and CUL4A (220)
Maximal Clique Centrality (MCC)	Maximal clique indicates subsets of nodes that cannot be extended by adding additional nodes, because all the nodes in the mentioned subset are already interacting with each other. This centrality is proposed by the developers of cytoHubba as the best method to obtain the essential proteins in a network.	CTNNB1 (17), CRBN (12), and CUL4A (7)

[1] Network available in Figure 2.

Figure 2 presents the main network obtained. The sizes of the nodes are based on the MCC score and the colors of the nodes show the molecular mechanisms these proteins are related to; these mechanisms were obtained through the literature review. When analyzing the network configuration, it can be concluded that the CTNNB1 central position might contribute to the propagation of a systemic effect induced by the IMiDs. Hence, thalidomide binding to the CRL4-CRBN complex (represented by the yellow color) might induce alterations in molecular signaling pathways. From this perspective, through CTNNB1, these signaling alterations can be reflected in the angiogenesis mechanisms (orange) or cell cycle (blue). There is no evidence implicating that CTNNB1 is a neosubstrate of the CRL4-CRBN complex. However, its central position in the network suggests that molecular alterations induced by the IMiDs might be easily reflected in beta-catenin and, through this protein, be transmitted to other non-neosubstrate proteins. This mechanism helps to explain the IMiDs' systemic effects on embryonic development.

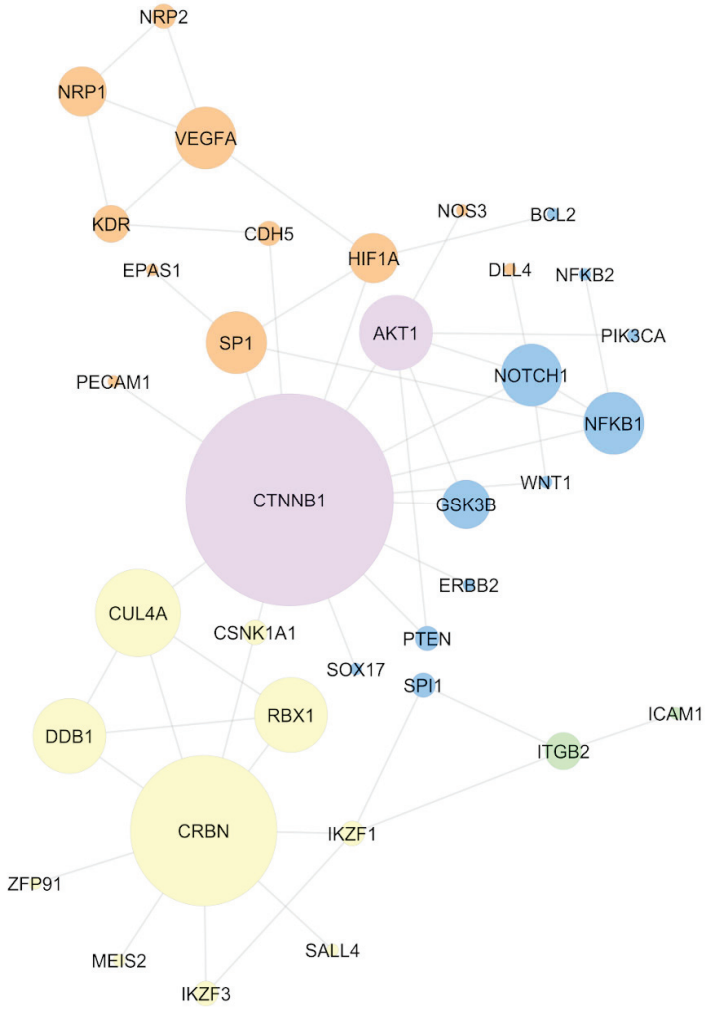


Figure 2. Network comprising experimental evidence of the protein–protein interaction, considering only IMiD-altered proteins, according to the literature review. Node size represents the maximal clique centrality (MCC) score. The colors of the nodes represent the mechanisms mainly described in the literature for the selected proteins: CRL4-CRBN binding (yellow), angiogenesis (orange), cell cycle (blue), and cell adhesion (green). AKT1 and CTNNB1 are presented in purple because of the association with multiple mechanisms.

2.3. IMiDs' Exposure Results in a Widespread Impact on Gene Expression

IMiD-induced effects on gene expression were studied using a literature review and by the analysis of transcriptomes available from the GEO repository. As presented before, the literature review provided 221 genes affected by the IMiDs, with *VEGFA* being the most annotated gene (Table S1). The search performed in GEO provided 166 datasets. Forty-five datasets met the eligibility for analysis: 11 with thalidomide exposure, 29 with lenalidomide exposure, and five with pomalidomide exposure. Twenty-two studies were further excluded because of the low quality of the samples, or because no significant differential gene expression (DGE) was detected (Table S4). The differentially expressed genes of the 23 remaining datasets were annotated, providing 8624 differentially expressed genes across the studies (Table S5). This number represents 30.4% of the genes contained in the reference genome (GRCh38/hg38) used for alignment (28,395 genes) (Table 2). The coagulation factor XIII A chain (*F13A1*) was the gene most frequently differentially expressed, significantly altered in nine datasets. Galectin-binding protein 3 (*LGALS3BP*), implicated in immune response, and Serpin family H member 1 (*SERPINH1*), a collagen-specific chaperone, were differentially expressed eight times. Of the 221 genes reported in the literature, 102 (46.1%) were differentially expressed in at least one dataset that was

analyzed (Figure S2). The DGE of the most reported literature gene, *VEGFA*, was found in three datasets that we analyzed. Following the literature, the *CRBN* gene did not present significant expression alterations in any of the datasets analyzed. A description of the functions of the genes differentially expressed in at least 25% of the studies (six times or more), representing 41 genes, can be observed in Table S6. The high number of differentially expressed genes concerning IMiD exposure demonstrated the widespread impact these drugs can exert in the human cell, altering several biological processes that may result in thalidomide embryopathy, when this exposure occurs during embryonic development. It is reasonable to hypothesize that this widespread effect on gene expression might result from the IMiD-induced TFs' degradations, unbalancing the transcription regulatory processes.

Table 2. Proportion of genes differentially expressed considering the reference genome and the number of datasets in which the genes presented differential expressions.

Number of Datasets with Differential Expressions	Number of Differentially Expressed Genes (%)
Total number of genes in the GRCh38 (human reference genome)	28,395 (100%)
1 or more	8624 (30.4%)
2 or more	2947 (10.2%)
3 or more	1041 (3.6%)
4 or more	334 (1.2%)
5 or more	118 (0.4%)
6 or more	41 (0.15%)
7 or more	16 (0.05%)
8 or more	3 (0.01%)
9	1 (0.003%)

Despite the focus of this study being the evaluation of TF proteins, non-coding RNA (ncRNA) data were not excluded from the transcriptome analysis. There were 154 long non-coding RNAs (lncRNAs), 12 small nucleolar RNAs (snoRNAs), five small nuclear RNAs (snRNAs), and two microRNAs (miRNAs) differentially expressed in at least two datasets. The lncRNAs, named *MALAT1* and *ST7-AS1*, were differentially expressed in five studies. *MALAT1* can act as a transcriptional regulator of genes involved in cell cycle and cell migration, whilst *ST7-AS1* does not have ontologies described, but has been associated with glioma.

2.4. Cell Cycle and Angiogenesis Are among the Top Biological Processes Altered in IMiD Exposure

To better understand the molecular roles of these differentially expressed genes, a gene-set enrichment analysis (GSEA) was conducted, ranking the genes by the number of times they were differentially expressed and evaluating the gene ontologies (GOs); the highest ranked genes were included. Regarding the biological processes, the mitogen-activated protein kinase 1 (MAPK) cascade was the more enriched ontology in the GSEA analysis (Table S7); this ontology refers to an important pathway of intracellular signaling transduction. Other processes already known to be affected by IMiDs were also highly enriched, such as the cell cycle and angiogenesis ontologies, blood vessel development, and vasculature development (Figure 3). The cellular component mostly enriched was the nuclear chromosome, while the molecular function was chromatin binding (Table S7). The GSEA analysis was consonant with the literature review, demonstrating that the main biological processes affected by IMiD exposure were the cell cycle- and angiogenesis-related mechanisms. In addition, the molecular function and cellular components enriched in the analysis helped to support the assumption of a systemic effect initiated by the degradation of TFs. This degradation might end up deregulating other biological processes that depend on strictly regulated transcriptions.

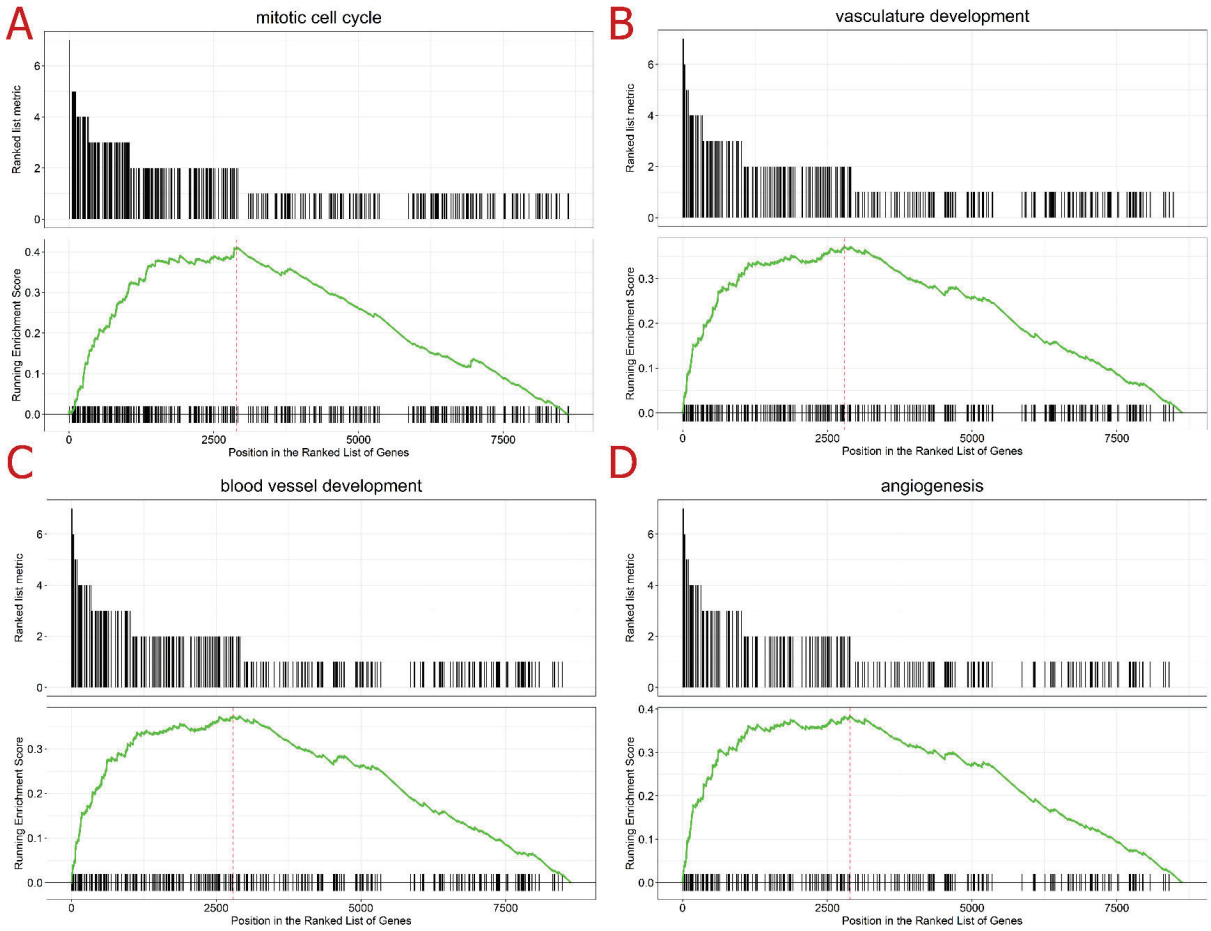


Figure 3. Gene set enrichment analysis (GSEA) top-ranked ontologies in the following order: (**A**) mitotic cell cycle (upper left), (**B**) vasculature development (upper right), (**C**) blood vessel development (lower left), and (**D**) angiogenesis (lower right). Legend: The first plot (ranked list metric) represents the number of datasets where the gene was differentially expressed, according to the 23 transcriptome datasets analyzed here; hence, the values are presented from 9 to 1. The second plot (running enrichment score) refers to the GSEA performed: the black lines along the x-axis represent the position of the gene in the ontology analyzed; the green line represents the running enrichment score for each of the genes; and the red line represents the maximum enrichment score for the analysis.

2.5. SP1, a C2H2 Transcription Factor, Is a Strong Candidate for an IMiD Neosubstrate

To evaluate the TFs that regulate differentially expressed genes, a TF-gene analysis was performed in the TRRUST database, which only comprised experimental data from humans. Since gene expression is very dynamic, genes differentially expressed in only one dataset, without previous literature reports of an IMiD-induced effect, were filtered out from the subsequent analyses. Hence, after these filters were applied, 3005 genes were queried in the TRRUST database, 2947 differentially expressed in two or more datasets, plus 58 genes differentially expressed once, but with a previous literature report regarding an IMiD-induced effect. As described previously, 119 of the genes provided in the literature were never differentially expressed in the 23 datasets analyzed, being also ruled out from the analysis (Figure S2). As a result, the TRRUST tool suggested 203 TFs experimentally known to regulate the expression of the 3005 queried genes. These 203 TFs were crossed with the list of proteins obtained from the literature review and with a list of C2H2 TFs, aiming to comprehend their function better; a Venn diagram presenting the intersection between the three lists is presented in Figure 4A. Two TFs suggested by TRRUST were C2H2 TFs already reported in the literature: SALL4 and SP1. The other four TFs suggested by TRRUST were non-C2H2 TFs; however, there are previous reports in the literature suggesting that

they can be affected by IMiD exposure: CTNNB1, hypoxia-induced factor subunit alpha (HIF1A), nuclear factor kappa B subunit 1 (NFKB1), and Spi-1 proto-oncogene (SPI1). As previously described, CTNNB1 was considered an essential protein in the network analysis. The other five TFs were also presented in the network generated (Figure 2): SALL4 was related to the CRBN-CRL4 degradation property (yellow), while SPI1 and NFKB1 were related to cell-cycle mechanisms (blue), and HIF1A and SP1 were related to angiogenesis mechanisms (orange). Of the 203 TFs suggested by TRRUST, SP1 had the greatest number of targets, reported to regulate 127 genes from the list provided to TRRUST. SP1 targets included *VEGFA*, the gene of chemokine ligand 2 (*CCL2*), which was differentially expressed in seven of the 23 datasets analyzed, and cytochrome P450 family 1 subfamily B member 1 (*CYP1B1*), deregulated six times in the DGE analysis. A gene ontology over-representation analysis was performed with SP1 target genes, pointing to an involvement of these targets in proliferation/apoptosis processes and angiogenesis. SP1 itself was involved in angiogenesis, as well as *VEGFA*, *CYP1B1*, and 25 more of its target genes.

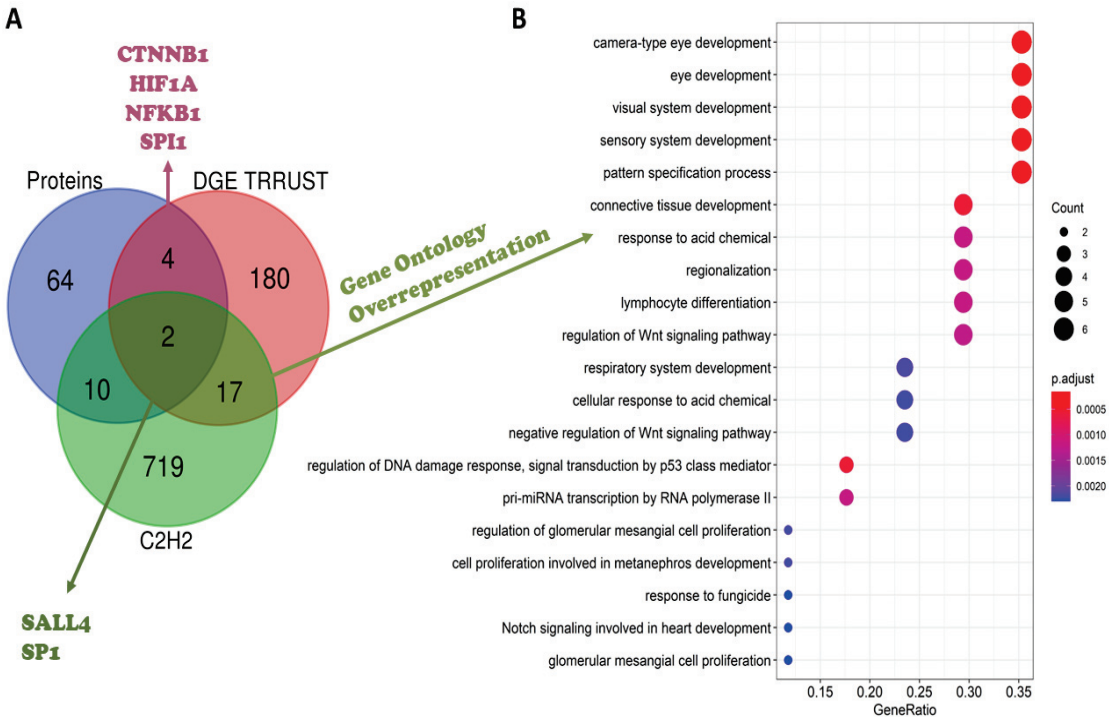


Figure 4. (**A**) Venn diagram considering the IMiD-affected proteins, according to the literature review (blue), transcription factors (TFs) provided by the TRRUST database (red), and the list of C2H2 transcription factors (green). (**B**) Over-representation analysis of the 17 C2H2 TFs suggested by the TRRUST analysis.

The TRRUST analysis pointed out SP1 as the TF that regulated the expression of the greatest number of genes encountered as being differentially expressed. SP1 is a C2H2 TF previously reported in the literature as being affected by IMiD exposure. Several of the SP1 target genes were related to angiogenesis and proliferation/apoptosis processes. Hence, it was reasonable to suggest SP1 as a probable neosubstrate of IMiD-induced degradation.

2.6. C2H2 Transcription Factors' Roles in Embryonic Development Must Be Prioritized in the Search for IMiDs Neosubstrates

As presented in Figure 4A, 17 TFs suggested by TRRUST are C2H2 TFs never-before reported in the literature as being affected by the IMiDs. To verify whether they played a role in embryonic development, a gene ontology over-representation analysis was performed (Table S8). It was observed that these TFs were overrepresented in the development

of several embryonic structures and organs, especially the eye and heart, which are known to be part of the spectrum of congenital anomalies reported in thalidomide embryopathy (Figure 4B). The analysis demonstrated these TFs were also included in relevant pathways in embryonic development, such as Wnt- and Notch-signaling pathways. A summary of the main processes related to these 17 TFs analyzed is available in Table 3. Interestingly, many of these TFs belong to the same families, such as the Kruppel-like family (KLF), here represented by KLF2, KLF4, KLF6, and KLF8; the snail family transcriptional repressor (SNAI), with the members SNAI1 and SNAI2; zinc-finger E-box binding homeobox (ZEB), with the members ZEB1 and ZEB2; and the specificity protein (SP), with the members SP2 and SP3, the same family from SP1, reported previously. This analysis provided several TFs that might be neosubstrates of the IMiDs, including members of the same family as SP1. The IMiD-induced degradation of these TFs would lead to several alterations in gene expression, a strictly regulated process in embryonic development; this mechanism could be the key to explaining the occurrence of IMiD-induced congenital anomalies in organs, such as the heart and eyes.

Table 3. Groups and functions of the 17 C2H2 prioritized transcription factors (TFs).

TFs	Chromosomal Location	Gene Groups [1]	Related Pathways or Ontologies [2]
BCL6	3q27.3	BTB domain	Cytokine signaling in immune system
CTCF	16q22.1	Cilia/flagella	Nervous system development
EGR1	5q31.2		Regulation of cell survival, proliferation, and cell death
GLI1	12q13.3		Signaling by Hedgehog
HIC1	17p13.3	BTB domain	Metabolism of proteins
KLF2	19p13.11	KLF transcription factors	Embryonic and induced pluripotent stem cells and lineage-specific markers
KLF4	9q31.2	KLF transcription factors	FOXO-mediated transcription
KLF6	10p15.2	KLF transcription factors	Adipogenesis
KLF8	Xp11.21	KLF transcription factors	Epithelial to mesenchymal transition
SNAI1	20q13.13	SNAG transcriptional repressors	Gastrulation
SNAI2	8q11.21	SNAG transcriptional repressors	Epithelial to mesenchymal transition
SP2	17q21.32	Sp transcription factors	Histone deacetylase binding
SP3	2q31.1	Sp transcription factors	Metabolism of proteins
WT1	11p13		Nervous system development
YY1	14q32.2	INO80 complex	ESR-mediated signaling
ZEB1	10p11.22	ZF class homeoboxes and pseudogenes	Cytokine signaling in immune system
ZEB2	2q22.3	ZF class homeoboxes and pseudogenes	TGFB-receptor signaling

[1] All the TFs described are also part of the zinc-finger C2H2 type group (HGNC, 2023). [2] Other than transcription regulation/gene expressions.

2.7. Beta-Catenin and SP1 Might Be Essential to Explaining the Systemic Effects of IMiDs

As described above, the TRRUST analysis provided six TFs previously reported in the literature, the ones with a C2H2 domain, SALL4 and SP1, and the non-C2H2 CTNNB1, HIF1A, NFKB1, and SPI1, all of them presented in the analyzed protein–protein interaction network (Figure 2). Here, a second network was created, including the proteins provided in the literature and the 17 non-reported C2H2 TFs that were related to embryonic development gene ontologies (Figure 5). This new network was assembled similar to the previous one, queried in STRING, and transferred to Cytoscape for the topology analysis. Of the 17 TFs presently included in the study, eight presented interactions with the other nodes of the network were labeled in blue in the image. EGR1, KLF4, KLF6, SP3, and YY1

interacted directly with SP1 and CTCF indirectly, through YY1. SNAI1 and SNAI2 were included through their interactions with GSK3B. The colors of the nodes in this network emphasized the TFs. The C2H2 TFs are represented in green, while the non-C2H2 TFs are in pink; the gray nodes are the non-TF proteins. A topology analysis was also conducted via cytoHubba and used to represent the sizes of the nodes. According to this analysis, CTNNB1 continued to be the hub of the network and the most essential protein, presenting the highest-ranking values in degree, betweenness, closeness, and MCC criteria (Table 4). In addition to CRBN and CUL4A, in this analysis, SP1 presented high scores in the centralities criteria and could be considered an essential protein as well. This change occurred because six of the eight TFs added interacted with the whole network through SP1, directly or indirectly. The network configuration demonstrated that an IMiD-induced degradation of SP1 could negatively impact several C2H2 TFs that were active in embryonic development. Likewise, SP1 degradation could easily reflect on non-C2H2 TFs, such as HIF1A, NFKB1, and, especially CTNNB1, hence being spread to other proteins involved in diverse molecular mechanisms. Hence, these results suggest SP1 as a feasible IMiD neosubstrate and, once again, beta-catenin is an essential protein in the systemic role of IMiDs; however, it is not necessarily a neosubstrate for drug degradation.

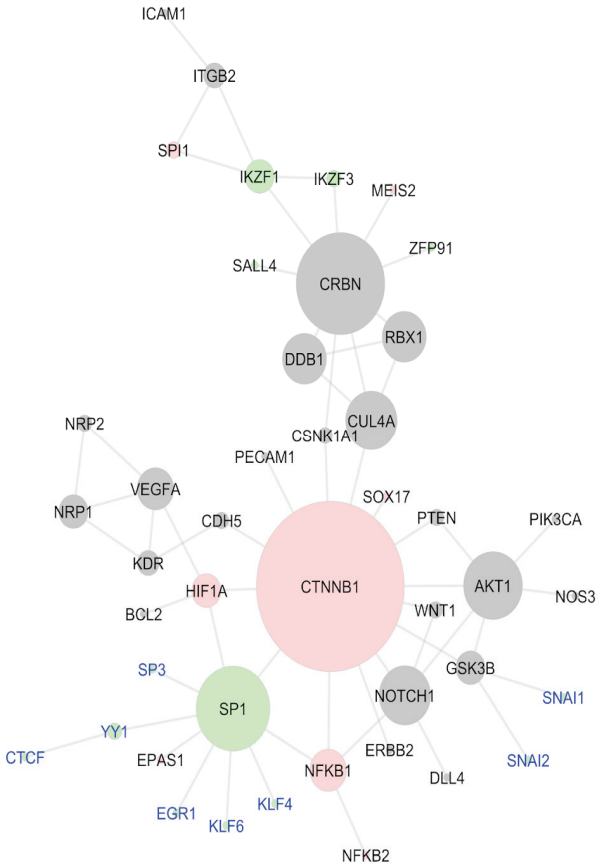


Figure 5. Systems biology network comprising the literature review proteins and the C2H2 transcription factors non-described as IMiD-affected that have a role in embryonic development; non-TF proteins (gray), non-C2H2 TFs (red), and C2H2 TFs (green). New TFs added are labeled in blue.

Table 4. Topological analysis of the protein networks and centralities' definitions.

Topological Analysis	Definition	Highest-Ranked Nodes (Literature + New TFs) [1]
Degree	Number of edges that connect to a node. Nodes with a high degree are defined as hubs.	CTNNB1 (14), CRBN (9), and SP1 (9)
Closeness	Measures how fast the flow of information travels through the nodes. High closeness centrality scores indicate rapid information flow.	CTNNB1 (26.37), SP1 (21.98), and CUL4A (19.67)
Betweenness	Demonstrates how crowded a network is. High betweenness centrality score indicate nodes that can control the information flow.	CTNNB1 (1325.67), CRBN (601), and SP1 (553.34)
Maximal Clique Centrality (MCC)	Maximal clique indicates subsets of nodes that cannot be extended by adding additional nodes, because all the nodes in the mentioned subset are already interacting with each other. This centrality is proposed by the developers of cytoHubba as the best method to obtain the essential proteins in a network.	CTNNB1 (20), CRBN (12), and SP1 (10)

[1] Network available in Figure 5.

3. Discussion

The present study, based on transcriptome and systems biology-allied strategies, aimed to prioritize TFs that could help to explain the effects of IMiDs on embryonic development. A systematized literature review was performed, annotating the proteins and genes altered/affected by the IMiDs. The annotated proteins were studied regarding their role as TFs through systems biology strategies, while the genes were evaluated together with 23 publicly available transcriptome datasets (Table S4). The main conclusions were: (I) IMiDs have a widespread effect on gene expression that might be explained by their induced TF degradation; (II) beta-catenin (CTNNB1) is an essential node and hub of the network of IMiD-altered proteins; (III) SP1 is a putative neosubstrate of IMiD-induced TF degradation; and (IV) there are 17 C2H2 TFs known to regulate the expression of genes altered by IMiDs that are active in embryonic development and have not been studied previously.

A transcription factor is a definition applied to describe proteins involved in the regulation of the transcription process; hence, they are able to affect the expression levels of a gene [27]. Several processes, such as embryonic development and cell differentiation, are regulated by TFs [28,29]. IMiD-induced alterations in some TF proteins have been known for a while [17,30,31]; however, only recently, studies have suggested an IMiD-induced degradation of these TFs, driven by the binding of drugs to the CRL4-CRBN complex [19,20]; this mechanism is believed to be mostly centered, but not exclusive, to TFs with a C2H2 protein domain [21,25]. However, few TFs have been identified when studying the C2H2 degrome induced by IMiDs [23]. This is an alarming result; however, it is a potential overestimation of the IMiDs' impact on gene expression. The evaluation of the several datasets included here can lead to an overestimation because of the high heterogeneity between the studies, inherent of the methodologies and study designs. Nevertheless, IMiDs' potential effects on gene expression must be evaluated with great attention because gene expression regulation during embryonic development is known to be synchronized, complex, and deeply regulated [32]. TF degradation, not necessarily strict to C2H2 TFs, is a plausible explanation for the IMiDs' impact on gene expression. Regarding

non-protein-coding genes, IMiDs' effects on ncRNAs have been little studied. According to the literature review, the genes altered by the IMiDs in this study were protein-coding genes (Table S1). Although transcriptome studies should cover the whole amount of RNA in the cell, usually, they are designed to target mRNA [33]. In this regard, microarray probes were mainly designed to cover the protein-coding genes [34]; RNA-Seq library preparation was preceded by RNA prioritization steps, which included poly-A selection and rRNA depletion [35]. Even though not all researchers performed poly-A selection, the abundance of ncRNAs was small in relation to the mRNA; hence, the ncRNA was poorly captured in transcriptome studies [35]. Because of this limitation, targeted ncRNA transcriptome analyses were conducted when these molecules were the main objective of the research [36]. Although we did not exclude ncRNA studies, we did not encounter transcriptome datasets that specifically targeted these molecules or that were registered in the GEO database as "non-coding RNA profiling" experiments in IMiD exposure. Nevertheless, 173 ncRNA genes were differentially expressed at least twice, according to the transcriptome analysis performed here. Since the evaluation of ncRNAs was impaired due to their abundance compared to the mRNA, the results presented here might be a small representation of the overall impact of the IMiDs in non-protein-coding genes.

From a systems biology perspective, TFs exert a protein–gene interaction by binding to sequence-specific DNA motifs. TFs are also known to be part of distinct regulatory networks, mediating protein–protein interactions [27,37] and defining both target gene selectivity and chromatin dynamics [37]. In the present study, systems biology analyses suggested beta-catenin as an essential protein in the network of IMiD-altered proteins. Thalidomide is known to diminish beta-catenin expression [31] and a pharmacogenetics study associated variants in the beta-catenin gene, *CTNNB1*, with lenalidomide adverse effects in multiple myelomas [38]. Beta-catenin is part of the Wnt canonical pathway, involved in a series of embryonic development events, including the establishment of the body axis and the orchestration of tissue and organ development [39]. This regulation is based upon the integration of two distinct beta-catenin functions: structural, as a cell-adhesion junction molecule, and signaling, as a TF [40]. In addition, the abundance of beta-catenin-binding partners provides an interaction with other TFs and signaling pathways [39]. Hence, beta-catenin deregulation can alter several regulatory networks, helping to explain the IMiDs' systemic effects (Figure 6). One protein that has a direct interaction with beta-catenin is SP1, a C2H2 TF. Thalidomide was shown to inhibit SP1 activity in endothelial cells [41], which led to two hypotheses of SP1 involvement in thalidomide embryopathy: SP1 can be a second messenger of growth factors involved in limb development [42], or SP1 transcriptional activity can be blocked by thalidomide, impeding this TF from binding to its motif [41]. Studies evaluating SP1 in light of the degrome evidence have not been encountered. Nevertheless, the findings in this study point to SP1 as a feasible neosubstrate of IMiD-induced TF degradation. First, there are 127 genes experimentally known to be regulated by SP1 that were differentially expressed in this study, including well-established genes, such as *VEGFA*; no other suggested TFs presented so many target genes with their expressions altered. Second, one of the main mechanisms involving altered genes is angiogenesis, a process that SP1 is known to be part of [41]. Third, SP1 interacts directly with other TFs known to be affected by IMiDs, including beta-catenin, NFKB1, and HIF1A, in addition to other suggested C2H2 TFs that can also be a part of the degrome.

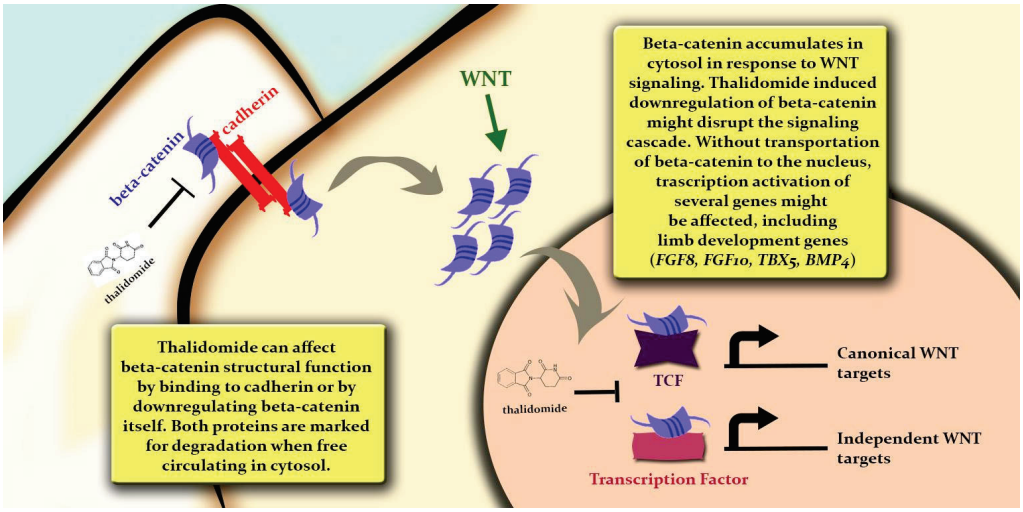


Figure 6. Representation of beta-catenin (CTNNB1) molecular mechanisms, with possible IMiD-altered effects (represented by thalidomide) and putative consequences on embryonic development.

According to Mackeh et al. (2018), 3% of human genes refer to C2H2 zinc-finger proteins, totalizing more than seven hundred proteins described [43]. In chicken embryos, a species well-established as an animal model of thalidomide embryopathy, C2H2 was suggested as one of the most dominant types of TFs in embryogenesis [44]. Despite the fact that the role of human C2H2 proteins is yet to be fully explained, most of the 17 C2H2 TFs suggested here were involved in embryonic development [45–47]. They are part of families known to have a role in this period, such as SNAI, ZEB, and KLF [45–47], in addition to other factors, such as GLI1 [48]. YY1 is also considered essential to embryogenesis [43]. For the completion of this study, SALL4 was the only C2H2 zinc-finger TF with a well-known role in limb development that was known to be degraded by the IMiDs [49]. SALL4 pathogenic variants lead to Duane-radial ray syndrome [50], an autosomal dominant syndrome that presents a pattern of limb anomalies very similar to the ones identified in thalidomide embryopathy [51,52]. SALL4 degradation might explain the typical limb anomalies caused by thalidomide exposure in utero; however, it is rational to affirm new C2H2 targets, named neosubstrates, and should be studied. From the perspective of an IMiD degrome, the TFs suggested here might help us to comprehend the effects of drugs in other organs and systems known to be affected by the embryopathies. Another interesting result was *CYP1B1* as a gene deregulated in six datasets analyzed. Despite being mainly involved in drug metabolism, CYP1B1 is known to promote angiogenesis by suppressing NF-kB activity [53]. In cancer cell studies, SP1 was considered a key mediator of CYP1B1 action, whilst CYP1B1 was shown to activate a epithelial to mesenchymal transition and Wnt/beta-catenin-signaling pathways, upregulating beta-catenin and other TFs, such as ZEB2 and SNAI1 [54]. These mechanisms must be studied in light of embryonic development to understand whether this regulation can also help to explain IMiDs' effects on embryos. Nevertheless, it was established that embryonic development and oncogenesis presented several similarities, especially regarding the biological processes and molecular pathways involved [55].

One limitation of the present study was the lack of experimental validation of the prioritized TFs suggested. Nevertheless, all the analyses conducted in our study, from the literature review to the networks and TF–gene interactions, were only based on the experimental evidence; thus, no in silico prediction was performed. All the conclusions presented here were also based on the literature evidence regarding the biological processes that the genes encountered were known to be a part of. It is relevant to state that bioinformatics approaches were conducted well to highlight molecular mechanisms under different medical conditions and, therefore, should not be observed as pure theoretical, computational modeling [56,57]. Bioinformatics tools improve experimental validations by

providing standardized pipelines to access cell and molecular mechanisms [56,57]. Molecular techniques using animal models and cellular methodologies can be very expensive and time-consuming [58]. Hence, the studies performed using public databases and providing a secondary analysis of omic data are helpful to better comprehend the biological processes altered after drug exposure, following the reduce, replace, and refine strategy in animal model research [59].

4. Materials and Methods

In summary, combined literature and database research was the basis for all the systems biology analyses, which included the evaluation of protein–protein interaction networks and protein–gene-targeted mechanisms. In protein–gene networks, the TFs were inserted as proteins and the gene its targets. Concomitantly, a second search was performed in the Gene Expression Omnibus (GEO) repository [60], annotating transcriptome studies that were evaluated regarding differential gene expressions (DGEs) in IMiD exposure. For a better comprehension, enrichment analyses and Venn diagrams were performed and created, respectively, throughout the study, using the data obtained from both DGE and systems biology strategies. Finally, the results were gathered through a network-assembling strategy, providing a protein–protein interaction network that presented the main TFs prioritized in the analyses. A detailed analysis description was presented in sequence. A scheme is presented in Figure 1.

4.1. Literature Review

A literature review was performed to annotate all the genes and/or proteins previously reported as disturbed by thalidomide, lenalidomide, or pomalidomide during embryonic development. The Rayyan tool [61] was used to annotate PubMed (Medline) and Embase manuscripts with the search terms available from Supplemental Material S1. The second search was performed in the PubTator repository [62], which was accessed through R language, using the *pubtatordb* package (https://github.com/cran/pubtatordb; URL accessed on 31 May 2023); Medical Subheading (MeSH) terms were applied to this research as follows: D013792 for thalidomide, D000077269 for lenalidomide, and C467566 for pomalidomide. The third literature search was performed using the Comparative Toxicogenomics Database (CTD) [63] by typing the name of the drug in the query search and downloading all the "References" provided.

Duplicates were excluded and the studies were primarily filtered by title and abstract screening. The selected manuscripts were fully read and the genes and/or proteins disrupted by the drugs we studied were annotated. For "disruption", we considered any type of effect driven by the drugs (i.e., expression alteration, binding, activity inhibition, and interaction). Hence, the inclusion criteria for genes and/or proteins disrupted were: evidence of an experimental assay using human embryonic cells, tissues, or organoids exposed to one of the IMiDs, thalidomides, lenalidomides, or pomalidomides, and only statistically significant disrupted genes/proteins were included, according to the cutoffs used in the study. The exclusion criteria were: transformed cells, knockout studies, exposure to two or more concomitant drugs, and studies that had abstracts, but not the full manuscript, available.

4.2. Systems Biology Analysis

The proteins annotated in the literature review or the subsequent TF analyses were inserted in the STRING tool [64], comprising only experimental evidence of interactions and co-expressions with a score > 0.400. A network of protein–protein interactions was obtained and transferred to Cytoscape v.3.7.2 [65]. Topological network analyses, such as centrality measures, were calculated through the cytoHubba plugin [66]. Essential proteins were obtained through the maximal clique centrality (MCC) criteria. The MCC score was used to indicate the size of the node, i.e., MCC = 1 was represented by a node of size 10 and MCC = 5 was represented by a node of size 50. The colors of the nodes were selected

manually, based on the network-relevant characteristics for each step of the analysis, and were detailed in the legends of figures.

4.3. Differential Gene Expression Analysis

Publicly available transcriptomes were downloaded from the GEO repository [60] and a DGE analysis was performed. Since gene expression was very dynamic and transcriptome analyses might result in several false-positive differentially expressed genes, a wide range of datasets was included. The terms "thalidomide", "lenalidomide", or "pomalidomide" were queried in GEO, filtered for *Homo sapiens*, and all the assays were annotated. The inclusion criteria were assays comprising human cells, tissues, or samples from patients who were exposed to one of the IMiDs (thalidomide, lenalidomide, or pomalidomide) at any age or stage of development. Datasets without raw data available, studies without a control group, knockout studies, and exposure assays or therapeutic schemes that used combined drugs in the same sample were excluded.

Microarray datasets were evaluated in R language by performing robust multi-averaging (RMA) normalization for Affymetrix studies with the *affy* package [67] and variance stabilizing normalization (VSN), followed by quantile normalization in Illumina or Agilent studies, using the NetworkAnalyst web interface [68] for sample processing. DGE was calculated through the *limma* package [69]. RNA-Seq samples were evaluated regarding quality control through FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; URL accessed on 31 May 2023) [70]. Alignment and transcript count were obtained through the useGalaxy server [71], using Bowtie2 [72] and featurecounts [73], respectively. The alignment of all the samples was performed using the GRCh38 (hg38) reference genome. In R, the *edgeR* package [74] was used to perform a trimmed mean normalization (TMM) and calculate the DGE. For all the datasets, microarrays, and RNA-Seqs, a surrogate variable analysis (SVA) normalization was performed to remove batch effects. Genes with a logFC > |1| and an adjusted p-value < 0.05 were annotated for their significant differential expressions; p-value adjustment was performed by the false discovery rate (FDR). If none of the genes met the statistical criteria, the whole dataset was excluded from the present study.

4.4. Transcription Factor Analysis

The Transcriptional Regulatory Relationships Unraveled by Sentence-based Textmining (TRRUST) database [75] was used for the TFs analyses. TRRUST is a curated database for human TFs and their target genes. Aiming to reduce false-positive results, it was included in the TF search, only for genes that (I) were differentially expressed in at least two datasets, or (II) that were differentially expressed in only one dataset but had evidence of IMiD-induced alterations. Two distinct approaches for TF analyses were conducted. Proteins obtained in the literature review were inserted in the database and their target genes were annotated or, contrariwise, the set of the genes obtained by literature review or DGE strategies were inserted in the tool, then having their regulatory TFs annotated. In the second approach, only regulations with a p-value < 0.05 were considered significant. The complete list of human TFs was obtained from the study of Lambert et al. (2018) [27], available on the website http://humantfs.ccbr.utoronto.ca/ (accessed on 31 May 2023). The list of C2H2 TFs was downloaded from the HUGO Gene Nomenclature Committee (HGNC) [76]. Venn diagrams used in the analyses were performed in the tool provided by the Bioinformatics Evolutionary Genomics website, from Ghent University (http://bioinformatics.psb.ugent.be/webtools/Venn/, URL accessed on 31 May 2023).

4.5. Enrichment and Overrepresentation Analyses

Proteins and genes were obtained through a literature review, and TFs and their target genes were evaluated regarding gene ontologies (GOs), signaling pathways, and, specifically for TFs, protein domains. GOs and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were accessed through the *clusterprofileR* package [77], using the over-representation analysis strategy.

A gene set enrichment analysis (GSEA) was performed with the differentially expressed genes to rank the ontologies and pathways mainly affected by the IMiDs. Genes were ranked according to the number of studies where they were significantly differentially expressed. Hence, a gene differentially expressed in four studies received a score = 4, while a gene differentially expressed only once received a score = 1; consequently, genes that never presented differential expressions received a score = 0. Three types of ontologies were accessed: biological processes, molecular functions, and cellular components. According to the GO resource, molecular functions referred to the molecular activities performed by the gene products that, when evaluated together, resulted in a biological process. Cellular component referred to the anatomical location of the protein in the cell.

5. Conclusions

In addition to the previous fundamental studies that identified crucial molecular mechanisms behind thalidomide embryopathy [15,17,18], the systems biology approach presented here allowed us to evaluate these hypotheses in a more integrative manner. Moreover, this was also the first study to provide a systematized, strictly performed literature review on the genes and proteins altered by IMID exposure in embryonic development. Furthermore, 35 transcriptome datasets were processed and evaluated regarding differential gene expressions, making this research the most thorough one known to address IMiDs' transcriptional effects. The transcriptomics strategy conducted here can also be applied to other drugs, aiming to evaluate the therapeutic or adverse effects, such as teratogenesis. In conclusion, not only did this study prioritize SP1 and beta-catenin as strong candidates for IMiD effects on embryonic development, it also analyzed the large amount of publicly available data, indicating there is much new knowledge to be integrated to understand an old challenge.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241411515/s1.

Author Contributions: Conceptualization, T.W.K. and F.S.L.V.; methodology, T.W.K., M.F.F., V.O.L., J.d.A.G. and G.C.G.; formal analysis, T.W.K., J.d.A.G. and G.C.G.; investigation, T.W.K., L.R.F., M.R.-M., M.T.V.S., L.S.-F. and F.S.L.V.; resources, M.R.-M., L.S.-F. and F.S.L.V.; data curation, T.W.K., M.F.F., V.O.L., J.d.A.G. and G.C.G.; writing—original draft preparation, T.W.K., M.F.F., V.O.L., J.d.A.G. and G.C.G.; writing—review and editing, L.R.F., M.R.-M., M.T.V.S., L.S.-F. and F.S.L.V.; supervision, M.R.-M., L.S.-F. and F.S.L.V.; project administration, F.S.L.V.; funding acquisition, M.R.-M., L.S.-F. and F.S.L.V. All authors have read and agreed to the published version of the manuscript.

Funding: The scholarships of the authors were funded by the Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant no. 23/2551-0000115-2; Hospital de Clínicas de Porto Alegre (HCPA)-Fundo de Incentivo à Pesquisa e Eventos (FIPE), grant no. 2017-0248, 2019-0792, and 2022-0567. T.W.K. is the recipient of a CNPq scholarship (grant no. 150181/2023-0). G.C.G. is the recipient of an HCPA scholarship (grant no. 23092.012897/2021-71). M.F.F. is the recipient of a CNPq scholarship (grant no. 165197/2022-6). M.R-M. is the recipient of a CNPq scholarship grant (grant no. 308075/2021-8).

Institutional Review Board Statement: This study only used data publicly available from databases and repositories. Nevertheless, it was submitted to the Ethics Committee in the Research of Hospital de Clínicas de Porto Alegre and it was approved under numbers 2017-0248 (CAAE 67015317.0.0000.5327) and 2019-0792 (39155220.9.0000.5327).

Informed Consent Statement: This study only used data publicly available in databases and repositories.

Data Availability Statement: The study only used data publicly available in databases and repositories. All the results generated are fully available in the Supplementary Material.

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

References

1. Lenz, W. A short history of thalidomide embryopathy. *Teratology* **1988**, *38*, 203–215. [CrossRef]
2. Lenz, W.; Knapp, K. Thalidomide embryopathy. *Arch. Environ. Health* **1962**, *5*, 100–105. [CrossRef]
3. Kowalski, T.W.; Sanseverino, M.T.; Schuler-Faccini, L.; Vianna, F.S. Thalidomide embryopathy: Follow-up of cases born between 1959 and 2010. *Birth Defects Res. A Clin. Mol. Teratol.* **2015**, *103*, 794–803. [CrossRef]
4. Smithells, R.W.; Newman, C.G. Recognition of thalidomide defects. *J. Med. Genet.* **1992**, *29*, 716–723. [CrossRef]
5. Vargesson, N. Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Res. C Embryo Today* **2015**, *105*, 140–156. [CrossRef] [PubMed]
6. Röllig, C.; Knop, S.; Bornhäuser, M. Multiple myeloma. *Lancet* **2015**, *385*, 2197–2208. [CrossRef] [PubMed]
7. Celgene. Revlimid (Lenalidomide) Prescribing Information. Food and Drug Administration Access Data. 2012. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/021880s028lbl.pdf (accessed on 31 May 2023).
8. Celgene. POMALYST (Pomalidomide) Prescribing Information. Food and Drug Administration Access Data. 2017. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/204026s019lbl.pdf (accessed on 31 May 2023).
9. Christian, M.S.; Laskin, O.L.; Sharper, V.; Hoberman, A.; Stirling, D.I.; Latriano, L. Evaluation of the developmental toxicity of lenalidomide in rabbits. *Birth Defects Res. B Dev. Reprod. Toxicol.* **2007**, *80*, 188–207. [CrossRef] [PubMed]
10. Paumgartten, F.J.; de Souza, N.R. Clinical use and control of the dispensing of thalidomide in Brasilia-Federal District, Brazil, from 2001 to 2012. *Cien. Saude Colet.* **2013**, *18*, 3401–3408. [CrossRef]
11. Sales Luiz Vianna, F.; de Oliveira, M.Z.; Sanseverino, M.T.; Morelo, E.F.; de Lyra Rabello Neto, D.; Lopez-Camelo, J.; Camey, S.A.; Schuler-Faccini, L. Pharmacoepidemiology and thalidomide embryopathy surveillance in Brazil. *Reprod. Toxicol.* **2015**, *53*, 63–67. [CrossRef] [PubMed]
12. Vianna, F.S.; Schüler-Faccini, L.; Leite, J.C.; de Sousa, S.H.; da Costa, L.M.; Dias, M.F.; Morelo, E.F.; Doriqui, M.J.; Maximino, C.M.; Sanseverino, M.T. Recognition of the phenotype of thalidomide embryopathy in countries endemic for leprosy: New cases and review of the main dysmorphological findings. *Clin. Dysmorphol.* **2013**, *22*, 59–63. [CrossRef]
13. Costa, P.; Maciel-Fiuza, M.F.; Kowalski, T.W.; Fraga, L.R.; Feira, M.F.; Camargo, L.M.A.; Caldoncelli, D.I.O.; Silveira, M.; Schuler-Faccini, L.; Vianna, F.S.L. Evaluation of the influence of genetic variants in Cereblon gene on the response to the treatment of erythema nodosum leprosum with thalidomide. *Mem. Inst. Oswaldo Cruz* **2022**, *117*, e220039. [CrossRef]
14. Maciel-Fiuza, M.F.; Costa, P.; Kowalski, T.W.; Schuler-Faccini, L.; Bonamigo, R.R.; Vetoratto, R.; Eidt, L.M.; de Moraes, P.C.; Silveira, M.; Camargo, L.M.A.; et al. Evaluation of Polymorphisms in Toll-Like Receptor Genes as Biomarkers of the Response to Treatment of Erythema Nodosum Leprosum. *Front. Med.* **2021**, *8*, 713143. [CrossRef]
15. D'Amato, R.J.; Loughnan, M.S.; Flynn, E.; Folkman, J. Thalidomide is an inhibitor of angiogenesis. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4082–4085. [CrossRef] [PubMed]
16. Therapontos, C.; Erskine, L.; Gardner, E.R.; Figg, W.D.; Vargesson, N. Thalidomide induces limb defects by preventing angiogenic outgrowth during early limb formation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8573–8578. [CrossRef] [PubMed]
17. Hansen, J.M.; Harris, C. A novel hypothesis for thalidomide-induced limb teratogenesis: Redox misregulation of the NF-kappaB pathway. *Antioxid. Redox Signal.* **2004**, *6*, 1–14. [CrossRef] [PubMed]
18. Ito, T.; Ando, H.; Suzuki, T.; Ogura, T.; Hotta, K.; Imamura, Y.; Yamaguchi, Y.; Handa, H. Identification of a primary target of thalidomide teratogenicity. *Science* **2010**, *327*, 1345–1350. [CrossRef]
19. Donovan, K.A.; An, J.; Nowak, R.P.; Yuan, J.C.; Fink, E.C.; Berry, B.C.; Ebert, B.L.; Fischer, E.S. Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome. *eLife* **2018**, *7*, e38430. [CrossRef]
20. Matyskiela, M.E.; Couto, S.; Zheng, X.; Lu, G.; Hui, J.; Stamp, K.; Drew, C.; Ren, Y.; Wang, M.; Carpenter, A.; et al. SALL4 mediates teratogenicity as a thalidomide-dependent cereblon substrate. *Nat. Chem. Biol.* **2018**, *14*, 981–987. [CrossRef]
21. Asatsuma-Okumura, T.; Ando, H.; De Simone, M.; Yamamoto, J.; Sato, T.; Shimizu, N.; Asakawa, K.; Yamaguchi, Y.; Ito, T.; Guerrini, L.; et al. p63 is a cereblon substrate involved in thalidomide teratogenicity. *Nat. Chem. Biol.* **2019**, *15*, 1077–1084. [CrossRef]
22. Gao, S.; Wang, S.; Fan, R.; Hu, J. Recent advances in the molecular mechanism of thalidomide teratogenicity. *Biomed. Pharmacother.* **2020**, *127*, 110114. [CrossRef]
23. Sievers, Q.L.; Petzold, G.; Bunker, R.D.; Renneville, A.; Słabicki, M.; Liddicoat, B.J.; Abdulrahman, W.; Mikkelsen, T.; Ebert, B.L.; Thomä, N.H. Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* **2018**, *362*, eaat0572. [CrossRef] [PubMed]
24. The HUGO Gene Nomenclature Committee Database. Zinc Fingers C2H2-Type. Available online: https://www.genenames.org/data/genegroup/#!/group/28 (accessed on 31 May 2023).
25. Khalil, A.; Tanos, R.; El-Hachem, N.; Kurban, M.; Bouvagnet, P.; Bitar, F.; Nemer, G. A HAND to TBX5 Explains the Link Between Thalidomide and Cardiac Diseases. *Sci. Rep.* **2017**, *7*, 1416. [CrossRef]
26. Krönke, J.; Udeshi, N.D.; Narla, A.; Grauman, P.; Hurst, S.N.; McConkey, M.; Svinkina, T.; Heckl, D.; Comer, E.; Li, X.; et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **2014**, *343*, 301–305. [CrossRef] [PubMed]
27. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [CrossRef] [PubMed]
28. Papavassiliou, K.A.; Papavassiliou, A.G. Transcription Factor Drug Targets. *J. Cell. Biochem.* **2016**, *117*, 2693–2696. [CrossRef]

29. Huilgol, D.; Venkataramani, P.; Nandi, S.; Bhattacharjee, S. Transcription Factors That Govern Development and Disease: An Achilles Heel in Cancer. *Genes* **2019**, *10*, 794. [CrossRef]
30. Knobloch, J.; Schmitz, I.; Götz, K.; Schulze-Osthoff, K.; Rüther, U. Thalidomide induces limb anomalies by PTEN stabilization, Akt suppression, and stimulation of caspase-dependent cell death. *Mol. Cell. Biol.* **2008**, *28*, 529–538. [CrossRef]
31. Knobloch, J.; Shaughnessy, J.D.; Rüther, U. Thalidomide induces limb deformities by perturbing the Bmp/Dkk1/Wnt signaling pathway. *FASEB J.* **2007**, *21*, 1410–1421. [CrossRef]
32. Shahbazi, M.N.; Zernicka-Goetz, M. Deconstructing and reconstructing the mouse and human early embryo. *Nat. Cell Biol.* **2018**, *20*, 878–887. [CrossRef]
33. Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: The teenage years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [CrossRef]
34. Raghavachari, N.; Garcia-Reyero, N. Overview of Gene Expression Analysis: Transcriptomics. *Methods Mol. Biol.* **2018**, *1783*, 1–6. [CrossRef]
35. Minnier, J.; Pennock, N.D.; Guo, Q.; Schedin, P.; Harrington, C.A. RNA-Seq and Expression Arrays: Selection Guidelines for Genome-Wide Expression Profiling. *Methods Mol. Biol.* **2018**, *1783*, 7–33. [CrossRef] [PubMed]
36. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [CrossRef]
37. Francois, M.; Donovan, P.; Fontaine, F. Modulating transcription factor activity: Interfering with protein-protein interaction networks. *Semin. Cell Dev. Biol.* **2020**, *99*, 12–19. [CrossRef] [PubMed]
38. Butrym, A.; Rybka, J.; Łacina, P.; Gębura, K.; Frontkiewicz, D.; Bogunia-Kubik, K.; Mazur, G. Polymorphisms within beta-catenin encoding gene affect multiple myeloma development and treatment. *Leuk. Res.* **2015**, *39*, 1462–1466. [CrossRef] [PubMed]
39. Valenta, T.; Hausmann, G.; Basler, K. The many faces and functions of β-catenin. *EMBO J.* **2012**, *31*, 2714–2736. [CrossRef] [PubMed]
40. Lyashenko, N.; Winter, M.; Migliorini, D.; Biechele, T.; Moon, R.T.; Hartmann, C. Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation. *Nat. Cell Biol.* **2011**, *13*, 753–761. [CrossRef]
41. Moreira, A.L.; Friedlander, D.R.; Shif, B.; Kaplan, G.; Zagzag, D. Thalidomide and a thalidomide analogue inhibit endothelial cell proliferation in vitro. *J. Neurooncol.* **1999**, *43*, 109–114. [CrossRef]
42. Stephens, T.D.; Fillmore, B.J. Hypothesis: Thalidomide embryopathy-proposed mechanism of action. *Teratology* **2000**, *61*, 189–195. [CrossRef]
43. Mackeh, R.; Marr, A.K.; Fadda, A.; Kino, T. C2H2-Type Zinc Finger Proteins: Evolutionarily Old and New Partners of the Nuclear Hormone Receptors. *Nucl. Recept. Signal.* **2018**, *15*, 1550762918801071. [CrossRef]
44. Liao, L.; Yao, Z.; Kong, J.; Zhang, X.; Li, H.; Chen, W.; Xie, Q. Transcriptomic analysis reveals the dynamic changes of transcription factors during early development of chicken embryo. *BMC Genom.* **2022**, *23*, 825. [CrossRef] [PubMed]
45. Nemer, M.; Horb, M.E. The KLF family of transcriptional regulators in cardiomyocyte proliferation and differentiation. *Cell Cycle* **2007**, *6*, 117–121. [CrossRef]
46. Bell, C.E.; Watson, A.J. SNAI1 and SNAI2 are asymmetrically expressed at the 2-cell stage and become segregated to the TE in the mouse blastocyst. *PLoS ONE* **2009**, *4*, e8530. [CrossRef] [PubMed]
47. Gheldof, A.; Hulpiau, P.; van Roy, F.; De Craene, B.; Berx, G. Evolutionary functional analysis and molecular regulation of the ZEB transcription factors. *Cell. Mol. Life Sci.* **2012**, *69*, 2527–2541. [CrossRef]
48. Hojo, H.; Ohba, S.; Yano, F.; Saito, T.; Ikeda, T.; Nakajima, K.; Komiyama, Y.; Nakagata, N.; Suzuki, K.; Takato, T.; et al. Gli1 protein participates in Hedgehog-mediated specification of osteoblast lineage during endochondral ossification. *J. Biol. Chem.* **2012**, *287*, 17860–17869. [CrossRef] [PubMed]
49. Asatsuma-Okumura, T.; Ito, T.; Handa, H. Molecular Mechanisms of the Teratogenic Effects of Thalidomide. *Pharmaceuticals* **2020**, *13*, 95. [CrossRef]
50. Kohlhase, J.; Heinrich, M.; Schubert, L.; Liebers, M.; Kispert, A.; Laccone, F.; Turnpenny, P.; Winter, R.M.; Reardon, W. Okihiro syndrome is caused by SALL4 mutations. *Hum. Mol. Genet.* **2002**, *11*, 2979–2987. [CrossRef]
51. Vargesson, N. The teratogenic effects of thalidomide on limbs. *J. Hand Surg. Eur. Vol.* **2019**, *44*, 88–95. [CrossRef]
52. Gomes, J.D.A.; Kowalski, T.W.; Fraga, L.R.; Macedo, G.S.; Sanseverino, M.T.V.; Schuler-Faccini, L.; Vianna, F.S.L. The role of ESCO2, SALL4 and TBX5 genes in the susceptibility to thalidomide teratogenesis. *Sci. Rep.* **2019**, *9*, 11413. [CrossRef]
53. Falero-Perez, J.; Song, Y.S.; Sorenson, C.M.; Sheibani, N. CYP1B1: A key regulator of redox homeostasis. *Trends Cell Mol. Biol.* **2018**, *13*, 27–45.
54. Kwon, Y.J.; Baek, H.S.; Ye, D.J.; Shin, S.; Kim, D.; Chun, Y.J. CYP1B1 Enhances Cell Proliferation and Metastasis through Induction of EMT and Activation of Wnt/β-Catenin Signaling via Sp1 Upregulation. *PLoS ONE* **2016**, *11*, e0151598. [CrossRef]
55. Ma, Y.; Zhang, P.; Wang, F.; Yang, J.; Yang, Z.; Qin, H. The relationship between early embryo development and tumourigenesis. *J. Cell. Mol. Med.* **2010**, *14*, 2697–2701. [CrossRef] [PubMed]
56. Markowetz, F. All biology is computational biology. *PLoS Biol.* **2017**, *15*, e2002050. [CrossRef] [PubMed]
57. Jafari, M.; Guan, Y.; Wedge, D.C.; Ansari-Pour, N. Re-evaluating experimental validation in the Big Data Era: A conceptual argument. *Genome Biol.* **2021**, *22*, 71. [CrossRef]
58. Shinde, V.; Perumal Srinivasan, S.; Henry, M.; Rotshteyn, T.; Hescheler, J.; Rahnenführer, J.; Grinberg, M.; Meisig, J.; Blüthgen, N.; Waldmann, T.; et al. Comparison of a teratogenic transcriptome-based predictive test based on human embryonic versus inducible pluripotent stem cells. *Stem Cell Res. Ther.* **2016**, *7*, 190. [CrossRef] [PubMed]

59. Kirschner, K.M. Reduce, replace, refine-Animal experiments. *Acta Physiol.* **2021**, *233*, e13726. [CrossRef] [PubMed]
60. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [CrossRef]
61. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [CrossRef]
62. Wei, C.H.; Allot, A.; Leaman, R.; Lu, Z. PubTator central: Automated concept annotation for biomedical full text articles. *Nucleic Acids Res.* **2019**, *47*, W587–W593. [CrossRef]
63. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; McMorran, R.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Res.* **2019**, *47*, D948–D954. [CrossRef]
64. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef] [PubMed]
65. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
66. Chin, C.H.; Chen, S.H.; Wu, H.H.; Ho, C.W.; Ko, M.T.; Lin, C.Y. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8* (Suppl. S4), S11. [CrossRef]
67. Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004**, *20*, 307–315. [CrossRef] [PubMed]
68. Xia, J.; Gill, E.E.; Hancock, R.E. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* **2015**, *10*, 823–844. [CrossRef]
69. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]
70. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2023. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 31 May 2023).
71. Giardine, B.; Riemer, C.; Hardison, R.C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; et al. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **2005**, *15*, 1451–1455. [CrossRef] [PubMed]
72. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
73. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
74. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]
75. Han, H.; Cho, J.W.; Lee, S.; Yun, A.; Kim, H.; Bae, D.; Yang, S.; Kim, C.Y.; Lee, M.; Kim, E.; et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **2018**, *46*, D380–D386. [CrossRef] [PubMed]
76. Povey, S.; Lovering, R.; Bruford, E.; Wright, M.; Lush, M.; Wain, H. The HUGO Gene Nomenclature Committee (HGNC). *Hum. Genet.* **2001**, *109*, 678–680. [CrossRef] [PubMed]
77. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **2012**, *16*, 284–287. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Review

Mesenchymal Stem Cells and Exosomes: A Novel Therapeutic Approach for Corneal Diseases

Basanta Bhujel, Se-Heon Oh, Chang-Min Kim, Ye-Ji Yoon, Young-Jae Kim, Ho-Seok Chung, Eun-Ah Ye, Hun Lee and Jae-Yong Kim *

Department of Ophthalmology, University of Ulsan College of Medicine, Asan Medical Center, 88, Olympic-Ro, Songpa-Gu, Seoul 05505, Republic of Korea; basantabhujel86@gmail.com (B.B.); 72004218osh@gmail.com (S.-H.O.); kcm8821@naver.com (C.-M.K.); yejii0849@gmail.com (Y.-J.Y.); dudwo991023@naver.com (Y.-J.K.); chunghoseok@gmail.com (H.-S.C.); yeyeyeyeye@gmail.com (E.-A.Y.); yhun777@hanmail.net (H.L.)
* Correspondence: jykim2311@amc.seoul.kr; Tel.:+82-2-3010-3680; Fax: +82-2-470-6440

Abstract: The cornea, with its delicate structure, is vulnerable to damage from physical, chemical, and genetic factors. Corneal transplantation, including penetrating and lamellar keratoplasties, can restore the functions of the cornea in cases of severe damage. However, the process of corneal transplantation presents considerable obstacles, including a shortage of available donors, the risk of severe graft rejection, and potentially life-threatening complications. Over the past few decades, mesenchymal stem cell (MSC) therapy has become a novel alternative approach to corneal regeneration. Numerous studies have demonstrated the potential of MSCs to differentiate into different corneal cell types, such as keratocytes, epithelial cells, and endothelial cells. MSCs are considered a suitable candidate for corneal regeneration because of their promising therapeutic perspective and beneficial properties. MSCs compromise unique immunomodulation, anti-angiogenesis, and anti-inflammatory properties and secrete various growth factors, thus promoting corneal reconstruction. These effects in corneal engineering are mediated by MSCs differentiating into different lineages and paracrine action via exosomes. Early studies have proven the roles of MSC-derived exosomes in corneal regeneration by reducing inflammation, inhibiting neovascularization, and angiogenesis, and by promoting cell proliferation. This review highlights the contribution of MSCs and MSC-derived exosomes, their current usage status to overcome corneal disease, and their potential to restore different corneal layers as novel therapeutic agents. It also discusses feasible future possibilities, applications, challenges, and opportunities for future research in this field.

Keywords: cornea; mesenchymal stem cells; exosomes; corneal diseases; corneal regeneration

Citation: Bhujel, B.; Oh, S.-H.; Kim, C.-M.; Yoon, Y.-J.; Kim, Y.-J.; Chung, H.-S.; Ye, E.-A.; Lee, H.; Kim, J.-Y. Mesenchymal Stem Cells and Exosomes: A Novel Therapeutic Approach for Corneal Diseases. *Int. J. Mol. Sci.* **2023**, *24*, 10917. https://doi.org/10.3390/ijms241310917

Academic Editor: Wajid Zaman

Received: 28 May 2023
Revised: 25 June 2023
Accepted: 28 June 2023
Published: 30 June 2023

1. Introduction

Loss of vision is an expanding global burden that affects not only the individual but also society as a whole. The human eye is a highly complex organ that has a crucial role in our life, but unfortunately, it is susceptible to a broad spectrum of disorders. Despite the challenges, there have been numerous efforts to address a wide range of eye diseases [1]. The current therapeutic approaches for treating corneal disorders include anti-inflammatory drugs, limbal stem cell (LSC) transplantation, and corneal transplantation [2]. However, these modalities are not free from limitations. LSC transplantation has a high risk of immune rejection [3,4]. Anti-inflammatory drugs are not fully capable of suppressing angiogenesis, conjuctivalization, and corneal scarring [5]. Although corneal transplantation is still considered an effective means of restoring vision, graft rejection due to immune responses remains a significant cause for concern. As estimated in the 2010 World Health Organization's report on visual impairment, about 285 million people of all age groups experience some level of visual impairment, and 39 million people are completely blind [6].

The cornea is one of the main components of the visual system, and it is the prime layer of the eye. It is crucial for vision, as it focuses light and plays a refractive role [7].

Structurally, the human cornea consists of five layers, which include the epithelium, stroma, endothelium, Bowman's layer, and Descemet's membrane. Nonetheless, the most significant layers among these are the epithelium, stroma, and endothelium [8]. The pathogenesis of corneal diseases may be attributed to a range of clinical conditions, including traumatic injury, chemical exposure, infections, deterioration due to aging, limbal stem cell deficiency, and various types of corneal dystrophies. This latter phenomenon may lead to defects in the structural and cellular components of the cornea [9]. The corneal functions can be compromised due to the development of corneal scars, haze, opacities, and edema, which may lead to visual deterioration. However, early detection and timely treatment can prevent most cases of corneal blindness.

Corneal transplantation using a healthy donor cornea to replace the damaged cornea has been successfully carried out for 100 years. It is currently considered one of the standard treatment strategies for corneal blindness. The cornea is the most frequently transplanted solid tissue in humans [10]. Although there have been notable improvements in corneal surgery in recent years, there are still challenges concerning the shortage of donor tissue, the short lifespan of allografts, the prolonged use of immunosuppressive drugs, legal and cultural restrictions, and the requirement for specialized surgical skills. A significant proportion of patients cannot afford corneal transplantation due to the high costs associated with the surgery and post-operative care. This is especially problematic given the increasing number of elderly individuals, resulting in significant financial and logistical obstacles, thus creating a global burden [5].

The transplantation of corneas is the most commonly performed type of transplant on a global scale, with roughly 180,000 corneal transplantation surgeries being conducted each year [11,12]. During the year 2012, over 184,000 corneal transplantation procedures were performed across 116 different nations [11]. As stated by the Eye Bank Association of America, the quantity of donated corneas and eye globes has been steadily increasing over the past few years, with a 5.2% increase reported in 2013 relative to 2012 [11]. An estimated 12.7 million individuals around the world are in need of corneal transplants, but the availability of donor tissues varies greatly across different regions. In 2019, in the UK, 4504 corneal transplants were carried out, while in the USA, the number of transplanted corneas was significantly higher, at 85,601 [13]. Nevertheless, the supply of transplantable donor tissue is consistently insufficient, as the demand for it surpasses the availability [11].

In recent decades, cell therapy and tissue engineering approaches have gained in popularity as treatments for some corneal diseases [14]. When it comes to tissue engineering, the cornea is well-suited for regenerative cell therapy using natural or systemic scaffolds, with or without cells, because of its immune-privileged and avascular characteristics [15]. Compared to other organs or tissues, there is a lower probability of rejection of the transplanted cells in the cornea. Today, ophthalmologists and visual scientists are increasingly interested in mesenchymal stem cells (MSCs) due to their competence to regenerate and differentiate, making them a potential alternative treatment option for corneal diseases. MSCs have been suggested to exhibit a therapeutic effect through their paracrine effect, which is mediated by exosomes (Figure 1) [16,17]. Research investigating the mechanism of MSC-based therapies has yielded compelling evidence indicating that exosomes facilitate intercellular communication by transporting various biomolecules, including nucleic acids and proteins, to the recipient cells. Additionally, they are also involved in essential processes such as cell migration and differentiation, and they participate in many physiological and pathological functions [18,19]. The effectiveness and safety of exosomes for curing corneal diseases have garnered attention due to the exosomes' cell-free nature, which enables them to retain the therapeutic advantages of their parent cells while eliminating the risks associated with stem-based therapies. As a result, the role of exosomes in regenerative medicine is expanding. To conduct this review, we performed a thorough search of PubMed (https://pubmed.ncbi.nlm.nih.gov/ (accessed on 1 May 2023) and Google Scholar using the terms "corneal diseases," "exosomes," and "mesenchymal stem cells" as keywords from July 2002 to April 2023.

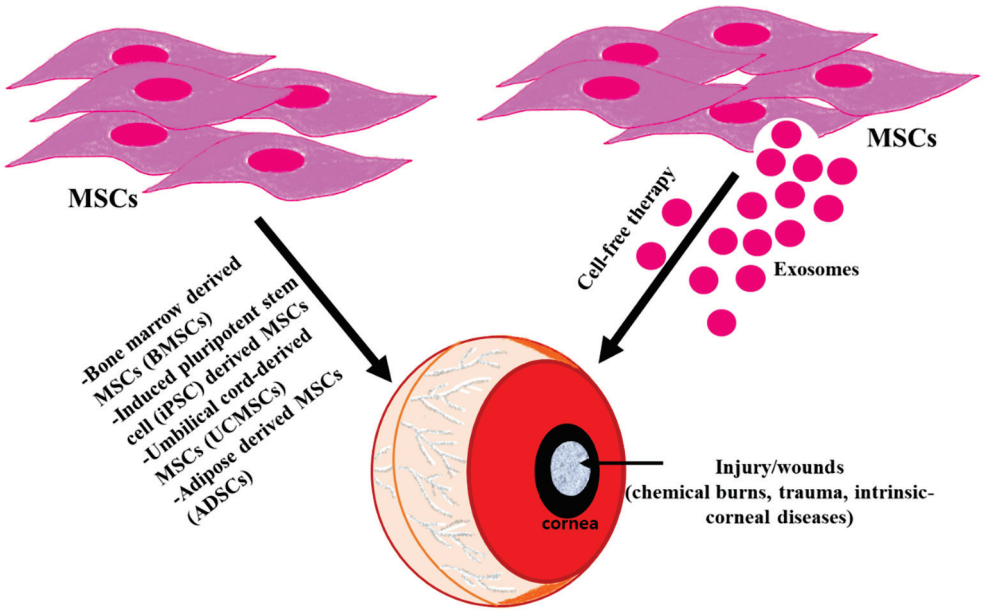


Figure 1. Cornea can be damaged from both external and internal factors (chemical burns, trauma, intrinsic–corneal diseases) leading to impaired vision. Different approaches have been employed to improve the regeneration of the injured cornea. Utilizing MSCs (cell-based therapy) and their secretions (cell-free therapy) can restore the normal corneal function in the diseased/injured cornea.

2. Mesenchymal Stem Cells (MSCs)

Stem cells can be categorized into two types depending on their source: embryonic and adult stem cells. Adult stem cells are present in a multitude of organs and tissues within the body, such as skeletal muscles, brain, bone marrow, dental pulp, liver, spinal cord, cornea, adipose tissue, and more. Among these, MSCs, regarded as multipotent progenitor cells, can arise from either embryonic or adult sources [20]. Wharton's jelly, umbilical cord blood, placenta, and embryo can provide embryonic stem cells, whereas dental pulp, bone marrow, adipose tissue, and other tissues are considered adult sources [21]. Embryonic MSCs contain a large number of primitive phenotypes, more active telomeres, and higher propagation ability compared to stem cells derived from adult tissues [22]. However, to obtain a sufficient amount of embryonic MSCs for therapeutic applications, ex vivo expansion is necessary, which may lead to a decline in their functional activity [23].

Protocols for the in vitro derivation of MSCs from human pluripotent stem cells, such as induced pluripotent stem cells (iPSCs) and embryonic stem cells, have been formulated and implemented [24]. Both in vitro and in vivo, MSCs can differentiate into various mesenchymal lineages, such as osteoblasts, adipocytes, and chondrocytes. Furthermore, they can migrate to injured sites, where they can differentiate and proliferate, secrete various anti-inflammatory and growth factors, promote wound healing, and thus reconstruct the damaged tissues [25].

MSCs play a vital role in the modulation of immune responses via paracrine action and interact with both innate and adaptive immune cells [26]. MSCs derived from different sources have different functions. The phenotypic markers CD13, CD73, CD90, CD105, and STRO-1 are expressed by both bone marrow and adipose MSCs, but they have different expression patterns of CD34, CD49d, CD54, and CD106 [27]. MSCs derived from adult tissues like bone marrow show less proliferation, engraftment ability, and differential potential than MSCs derived from birth-associated tissues (cord blood, umbilical cord, placenta, and amnion) [28]. Different proteomic profiles can be observed in the MSCs derived from the umbilical cord, Wharton's jelly, or cord blood. MSCs derived from the umbilical cord show ameliorated results in musculoskeletal tissue engineering [29]. Similarly, the MSCs derived from adult adipose tissue sources also demonstrate variations in functionality [30]. MSCs derived from subcutaneous fat tissue have quicker tissue

proliferation than those derived from the omental region [31]. MSCs are identified by the existence of specific markers such as CD90, CD73, CD71, CD44, CD105, and CD271, but they do not express hematopoietic markers (CD14, CD34, and CD45) nor do they express stimulant molecules (CD86, CD40, and CD80) [32]. Studies have demonstrated that MSCs produce a diverse array of exosomes, which perform various tasks like immune modulation, repairing damaged tissue, and downregulating inflammation via paracrine action [33,34].

MSCs exhibit potential immunomodulatory effects, which contribute to their therapeutic potential and ability to evade rejection both in vivo and in vitro. MSCs, considered to be immune-privileged cells, diminish the presence of major histocompatibility complex (MHC) class II molecules and co-stimulatory molecules (CD80, CD86, and CD40) on their cell surface [35]. In vitro, MSCs affect the innate immune system by suppressing the maturation and activation of dendritic cells (DCs) as well as the cytotoxicity of natural killer cells. Moreover, they suppress adaptive immune responses by inhibiting the proliferation and secretion of cytokines by T cells, as well as impeding the maturation of B cells [36]. Various soluble factors, such as transforming growth factor-beta (TGF-β), interleukin-6 (IL-6), interleukin-10 (IL-10), matrix metalloproteinases (MMPs), prostaglandin E2 (PGE2), indoleamine-2, 3-dioxygenase (IDO), human leukocyte antigen-G5 (HLA-G) and nitric oxide are involved in the immunosupressive function of MSCs [37]. Moreover, MSCs can suppress the production of interferon-gamma (IFN-γ) by Th1 cells while increasing the production of IL-4 and IL-10 by Th2 cells. This alteration in cytokine expression facilitates the immune response of native CD4+ T cells, promoting a shift toward a Th2-type immune response [38]. In co-culture with native T cells, human MSCs support the differentiation and proliferation of regulatory T cells (Tregs) during mixed-lymphocyte reactions. This effect is mediated through the secretion of prostaglandin E2 (PGE2) and TGF-β. Tregs, which are a specialized subset of T cells, retain their ability to suppress the activation of other T cells and help regulate immune system activity [39]. In addition, microencapsulation within microspheres or hydrogels can also protect MSCs physically from immune recognition and prevent cell aggregation, thereby avoiding any direct contact and reducing the risk of rejection in the transplanted site [40,41].

Even though MSCs have great potential in regenerative medicine, there are concerns about their uncontrolled proliferation or overgrowth, which can lead to unwanted effects. Implementing kill switches in the context of MSCs involves incorporating mechanisms to control and terminate their growth when necessary. One potential approach is to engineer MSCs with genetic circuits that allow for conditional cell death or growth arrest. These circuits can be designed to respond to specific signals or triggers in the cellular environment [42]. Likewise, MSCs can be genetically engineered to control their proliferative capacity and immunomodulatory effects, preventing over-expansion [43]. In addition to this, manipulating cell cycle regulators like $p16^{Ink4a}$ or $p21^{cip1}$ can block cell cycle progression and reduce their proliferation. Additionally, modulating the activity of key regulators, such as retinoblastoma protein (pRB) or p53, can also impact MSCs growth [44]. Hypoxic conditions can slow cell proliferation and promote a quiescent state of MSCs. TGF-β signaling is involved in cell cycle regulation and can affect growth of MSCs. Modulating the activity of TGF-β signaling pathway components, such as SMAD proteins or TGF-β receptors, can help control MSCs proliferation [45]. Incorporating MSCs and growth factors or bioactive molecules within scaffolds can regulate MSC behavior. Furthermore, some scaffold materials, such as collagen and gelatin, have been reported to support controlled MSCs growth and prevent overgrowth [46].

3. MSC Mobilization, Migration, and Homing in Corneal Changes

Injury and inflammation are responsible for inducing the mobilization, migration, and colonization of stem cells [25,34]. The mechanism of MSC homing to the sites of injury remains unclear. When the cornea is injured by trauma and infection, specific chemoattractants stimulate endogenous bone marrow MSCs, causing them to mobilize and enter the peripheral blood. To promote wound healing, these circulating MSCs travel

to the site of injury in the cornea and attach themselves to it [47]. In a study of a murine alkali-burn model of the injured cornea, the researcher demonstrated the vigorous migration and engraftments into the injured cornea of bone marrow MSCs administered intravenously [48]. Chemokines such as stromal cell-derived factor 1 (SDF-1) and substance P have been identified as regulators of MSC mobilization and recruitment to the cornea [47]. Additionally, MSCs' ability to locate and bind to target tissues is facilitated by robust mechanisms similar to leukocyte activity, including adhesion mediated by selectin and integrin, transmigration, and passive entrapment [49]. Efforts have been made to upgrade the targeting of MSCs to ocular tissues. The sub-conjunctival injection of MSCs and their co-transplantation with amnion onto the damaged corneal tissue have both resulted in a substantial enhancement [50–53].

Despite their ability to migrate to the site of injury, it appears that the migration and homing of MSCs are not essential for their therapeutic effects. In one study, the damage caused by inflammation in the cornea was reduced by tumor necrosis factor-stimulated gene/protein-6 (TSG-6) released by MSCs administered systemically, even without MSC engraftment [54]. Furthermore, MSCs have been found to secrete TSG-6, which has been shown to have a positive effect on reducing inflammation and improving cardiac function after a myocardial infarction [55]. A recent study showed that administering human MSCs through the systemic route can diminish inflammatory damage in the cornea by releasing anti-inflammatory substances when prompted by injury signals from the cornea, with no need for engraftment [54]. One potential approach to enhancing the effectiveness of MSCs in treating corneal injuries is to administer them directly to the affected area using methods such as subconjunctival injection, transplantation with an amniotic membrane (AM), or application through a plastic tube. This could result in an increased MSC concentration at the site of the injury, leading to improved outcomes [51,56–58].

4. Corneal Regeneration with MSCs

There is considerable research indicating that MSCs are capable of reducing inflammation and facilitating the restoration of corneal transparency in the aftermath of ocular injuries [48,57,59]. Despite the fact that MSCs promote angiogenesis in certain tissues, they exhibit an opposing effect in the cornea by inhibiting angiogenesis. This distinctive quality of MSCs represents an intriguing and noteworthy characteristic of these cells [60]. Researchers have explored two primary methods for administering MSC treatment for corneal injuries: intravenous injection and topical application. A number of studies have examined the effectiveness of intravenous injection in repairing corneal injuries [61–63]. In a study using adipose-derived stem cells (ADSCs) on an amniotic membrane, the researchers observed a reduction in inflammation, an increase in corneal transparency, and alleviation of the corneal damage [64]. Similarly, the damage to the cornea was partially repaired via the secretion of anti-inflammatory proteins by the injected MSCs [54]. Research has revealed that the culture medium derived from MSCs has the potential to treat corneal ulcers by suppressing inflammation, enhancing cell viability, stimulating proliferation, and modulating the immune response [65]. The therapeutic role of MSCs in different layers of the cornea is shown in Table 1. Figure 2 shows the therapeutic roles of MSCs in regenerating the cornea.

Table 1. The therapeutic role of MSCs in different layers of cornea.

Corneal Tissues	MSC Source	Aim of Experiment	Human or Animal Model	Study Performed	In Vitro Study Results	In Vivo Study Results	References
	Human adipose	-To investigate the impacts of inhibiting glycogen synthase kinase-3 (GSK3) and transforming growth factor β (TGF β) signaling on the epithelial differentiation of ADSCs.	Rat	In vitro In vivo	-Downregulation of the mesenchymal genes and upregulation of epithelial genes (E-cadherin, cytokeratins, and occudin).	-Demonstration of the human E-cadherin CK3, and 12 on rat corneal surface in rat model of total limbal stem cell deficiency.	[66]

Table 1. *Cont.*

Corneal Tissues	MSC Source	Aim of Experiment	Human or Animal Model	Study Performed	In Vitro Study Results	In Vivo Study Results	References
	Human adipose	-To study whether extraocular human ADSCs exhibit some characteristics of corneal epithelial-like cells cultured in vitro.	-	In vitro	-Induction of corneal epithelial-like cells from human adipose-tissue-derived MSCs including the expression of CK3 and CK12 when cultured in corneal epithelium-conditioned media.	-	[65]
	Rabbit bone marrow	-To study the potential of bone-marrow-derived MSCs (BM-MSCs) to differentiate into corneal epithelial cells both in vitro and in vivo.	Rabbit	In vitro In vivo	-Differentiation of rabbit MSCs vigorously into cells with similar morphological and molecular characteristics to corneal epithelial-like cells. -Induction of CK3 expression with co-culture with rabbit limbal stem cells.	-Implantation of the cells incorporated with fibrin gel regenerated the corneal epithelium in alkali-induced corneal deficiency rabbit model and expressed CK3.	[67]
	Rabbit bone marrow	-To inspect the suitability of bone marrow MSCs trans-differentiating into corneal epithelial cells in a rat model with a deficit of LSCs.	Rat	In vitro In vivo	-Observation of CK12 expression and epithelial cell characteristics with co-culture on rat corneal stromal cells.	-Differentiation into epithelial-like cells expressing CK12, improving corneal opacity, and reconstructing the corneal surface in rats with transplantation on amnion in alkali injury rat corneal deficiency model.	[53]
1. Corneal epithelium	Human bone marrow	-To explore the potential of human MSCs to differentiate into corneal epithelial cells and to assess their ability to regenerate damaged corneal tissue.	Rat	In vivo	-	- Restoration of the injured or damaged surface of the cornea in rats. -Inhibition of corneal inflammation (CD45, IL-2, MMP-2) and angiogenesis in the presence of amnion in rat alkali-injured epithelial defect model.	[57]
	Allogenic bone marrow	-To show that MSCs used in transplantation can be safe and effective and help in treating corneal pathology due to limbal stem cell deficiency (LSCD).	Human	In vivo	-	-Improvement of epithelial damage and demonstration of a more corneal epithelial-like phenotype in the central cornea with allogenic BM-MSCs.	[68]
	Rabbit bone marrow, adipose tissue	-To investigate whether MSCs or corneal limbal epithelial cells (LSCs) restore the corneal epithelium and optical properties in an alkali burn rabbit model.	Rabbit	In vivo	-	-Improvement in corneal optical properties, restoration of antioxidant protective mechanism and epithelial regeneration with MSCs.	[69]
	Conjunctiva-derived MSCs (CJMSCs)	-To identify the potential of hybrid polyurethane (PU) and silk nanofibrous scaffold with CJMSCs in treating corneal epithelium.	-	In vitro	-Improvement in the function of corneal epithelium with CJMSCs incorporated with silk fibers and PU fibers. -CJMSCs differentiate into corneal epithelial-like cells.	-	[70]

Table 1. *Cont.*

Corneal Tissues	MSC Source	Aim of Experiment	Human or Animal Model	Study Performed	In Vitro Study Results	In Vivo Study Results	References
	Human bone marrow	-To demonstrate the aptitude of human MSCs derived from bone marrow to differentiate into functional cells with epithelial-like characteristics in vitro.	-	In vitro	-Differentiation of human MSCs derived from bone marrow into functional epithelial cells in the epithelial differentiation medium containing keratinocyte growth factor, epidermal growth factor, hepatocyte growth factor, and insulin-like growth factor.	-	[71]
	Human dental pulp	-To demonstrate the potential of adult dental pulp cells to differentiate into Keratocytes.	Mouse	In vitro In vivo	-Induction of Keratocan and Keratan sulfate proteoglycan (KSPG) in culture with keratocyte differentiation medium.	-Production of stromal components like human type-I collagen and Keratocan with intrastromal injection in mouse's corneal stroma.	[72]
	Human adipose	-To determine whether keratocyte-specific phenotypic markers are expressed by ADSCs when cultured.	-	In vitro	-Induction of expression of stromal matrix components like KSPG, aldehyde-3-dehydrogenase-3A1 (ALDH3A1) in cell culture under reduced serum conditions supplemented with insulin and ascorbate.	-	[73]
2. Corneal stroma	Human bone marrow	-To investigate whether MSCs can differentiate into corneal keratocyte-like cells by using keratocyte-conditioned medium (KCM).	-	In vitro	-Expression of keratocyte markers such as aldehyde-1-dehydrogenase-1A1 (ALDH1A1), Lumican, and Kera in KCM. -Demonstrate MSCs could proliferate and differentiate into cells with similar characteristics to keratocytes when cultured in KCM.	-	[56]
	Rabbit adipose	-To explore if the combination of autologous rabbit adipose-derived stem cells and polylactic-co-glycolic acid (PLGA) scaffold could be used to repair corneal stromal defects in a rabbit.	Rabbit	In vivo	-	-Differentiation of MSCs into functional keratocytes, detection of their presence up to 24 weeks following transplantation. -Differentiation to Kera and ALDH3A1 expressing cells in mechanically induced rabbit stromal defect model via transplantation of cells on a PLGA scaffold.	[74]
	Human periodontal ligament (PDL)	-To show the possibility of using PDL cells as a potential source for regenerative corneal cell therapy to treat corneal disorders.	-	In vivo	-Organ culture shows the presence of CD34, ALDH3A1, Kera and Lumican, collagen type 8 alpha 2 (COL8A2), CHST6 genes and their expressions. -Reduction in fibrosis, neurogenesis, and vaculogenesis gene expression.	-	[75]
3. Corneal endothe-lium	Human umbilical cord MSCs (HUC-MSCs)	-To study the significance of HUC-MSCs in treating corneal endothelial disease in a rabbit model with bullous keratopathy.	Rabbit	In vivo In vitro	-Stimulation of the expression of NA, K-ATPase in a medium containing GSK3β inhibitor.	-Progress in corneal thickness and transparency with cell injection in rabbit bullous keratopathy model.	[76]

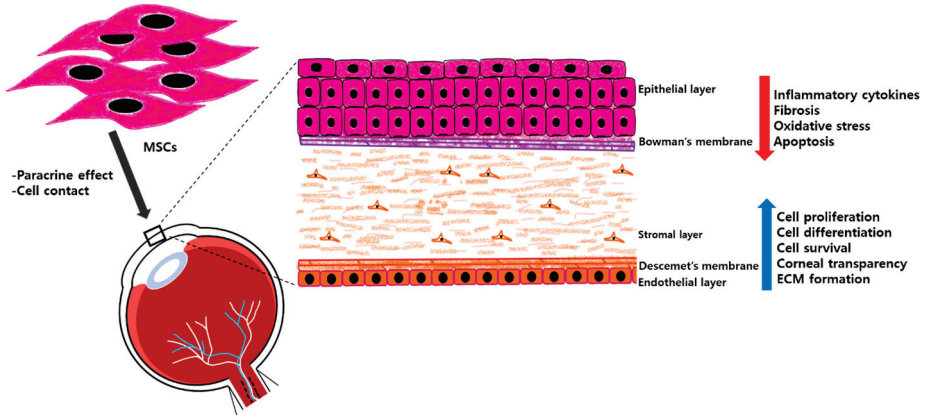


Figure 2. MSCs for treating corneal diseases. MSCs are well known to have a therapeutic effect through paracrine secretions, and by making direct cell contacts with cells in different corneal layers, such as epithelium, stoma, and endothelium. This can help to promote the survival, growth, and specialization of cells, while reducing cell death, inflammation, and fibrosis.

4.1. Corneal Epithelial Regeneration with MSCs

The corneal epithelium is fairly uniformly composed of 5–7 layers of stratified non-keratinized squamous cells, each 50 µm thick, which covers the outermost portion of the cornea [77,78]. The maintenance of a healthy corneal epithelial layer is essential for appropriate visual function and translucency of the cornea [79]. The tear film covers the surface of the epithelium, which is made up of a non-keratinized squamous layer that originates from the superficial ectoderm during embryonic development at around 5–6 weeks [78]. This layer has a crucial function in correcting any abnormalities present on the surface of the cornea [6]. Any disruption of the structural soundness of the corneal epithelium due to physical trauma, infection, or LSC deficiency, could lead to damage to the epithelium, corneal inflammation, neovascularization, and opacities. Eventually, this phenomenon can culminate in corneal blindness [80,81]. MSCs can differentiate into cell lineages derived from the neuroectoderm and epithelial cells [82]. An in vitro study co-cultured rabbit bone marrow MSCs and rabbit LSCs and found that the cells displayed a polygonal and cobblestone morphology similar to epithelial cells. Moreover, these cells expressed cytokeratin 3 (CK3), specific to corneal epithelium [82]. Similarly, a study involving co-culture of rat corneal stromal cells (CSCs) and MSCs found that MSCs underwent trans-differentiation into cells that resembled epithelial cells, confirmed through the observation of CK12 expression. Furthermore, in vivo experiments showed reduced corneal opacity and neovascularization grades, indicating the potential of MSCs to improve corneal integrity [53]. When human ADSCs were grown in a culture medium that had been conditioned by corneal epithelial cells, the result illustrated the upregulation of the expression of CK3 and CK12 with an epithelial-like cell appearance [83,84]. An in vivo study on mesenchymal–epithelial transition (MET) was inconclusive. When human BM-MSCs were tissue engineered onto human amnion and then transplanted into rat corneas damaged by alkali, there was a decline in inflammation (as demonstrated by the markers CD45, IL-2, MMP-2) and angiogenesis markers accompanied by reconstruction of the corneal surface [57].

The utilization of small-molecule chemicals has become increasingly popular since they have shown potential in initiating and modifying the cellular changes involved in the transition between mesenchymal and epithelial phenotypes. Furthermore, these small molecules can also regulate cell fate and facilitate the reprogramming of target genes [85,86]. A protocol has been previously reported for generating corneal epithelial progenitors, referred to as MET-Epi, from human ADSCs. This protocol involves an approach that antagonizes both the GSK and transforming growth factor β (TGF β) pathways [47]. These small molecules can penetrate cell membranes, lack immunogenicity, have modifiable dosages, are cost-effective, and expedite standardization. In an investigation utilizing a

rat corneal surface alkali injury model, it was observed that the application of ADSCs-derived epithelial progenitors engineered on fibrin gel led to an enhancement of corneal transparency and stability of the surface. The damaged corneal surface was restored with the creation of a multilayered epithelium. Moreover, higher levels of human epithelial cell adhesion molecule (EpCAM) and CK3/12 expressions were noted. The sham group had less reduction of corneal haze and no CK3/12 expression [48].

An alternative surgical approach to corneal epithelial failure, which can arise from severe limbal stem cell deficiency (LSCD), involves the use of cultivated limbal epithelial transplantation (CLET) with either an amniotic membrane or fibrin gel acting as a carrier [87]. A new procedure that involves ex vivo propagation of autologous or allogeneic epithelial stem cells from a limbal biopsy can help stabilize the ocular surface, and it has theoretical advantages over conventional limbal transplantation. A recent study demonstrated the potential of MSC therapy to achieve similar therapeutic results as CLET [88]. Another study compared the safety and efficacy of allogeneic BM-MSC transplantation and allogeneic CLET for the restoration of corneal epithelia in LSCD over a one-year follow-up period, and it was revealed that both treatments were effective and safe [68]. In another study of transplantation of human foreskin-derived mesenchymal stem cells (hMSCs) into alkaline damaged rabbit epithelium, differentiation of these cells into corneal cells and their migration into corneal stroma was observed [89]. Based on the findings of these studies, MSCs may serve as a viable alternative option for treating damaged corneal epithelium and promoting the repair of the ocular surface. Although MSC therapy has shown promise in treating corneal epithelial damage, additional research is necessary to provide evidence for its clinical effectiveness, safety, and long-term stability.

4.2. Corneal Stromal Regeneration with MSCs

The corneal stroma, the thickest layer of the cornea, consists of specific extracellular matrix (ECM) elements and collagen fibrils organized into flattened lamellae that run perpendicular to each other. The corneal stromal keratocytes (CSKs) are typically quiescent and are located between collagenous lamellae [90]. The cornea remains transparent, biomechanically strong, and structurally intact due to the precise arrangement and packing of collagen fibrils within the stroma, as well as the unique composition of several substances such as KSPG (lumican, keratocan, mimecan, and decorin), stromal crystalline (transketolase, ALDH3A1, and ALDH1A1) and ECM proteins (collagen type I and V) [90]. When the cornea is damaged or diseased, the CSKs can die. They are responsible for producing proteoglycans and maintaining the collagen fibrils. Their death can lead to a reduction in the production of proteoglycans, the degradation of collagen fibers, and an increase in the glycation of collagen molecules. Nevertheless, some of the surviving CSKs can be activated and turned into repair-type stromal fibroblasts near the damaged area, contributing to the healing process of the cornea. The combined action of serum and cytokines such as PDGF and TGF β can cause certain fibroblasts to transform into highly contractile myofibroblasts. This mechanism can lead to the development of corneal haze, opacity, and scar formation. Such phenomena may disrupt the transmission of light through the cornea, which can result in visual impairment and ultimately lead to blindness.

Research has revealed that MSCs obtained from bone marrow and the lining of the umbilical cord can transform into keratocyte-like cells and potentially restore clarity to the corneal stroma [56,61,63]. A study involving the administration of human umbilical cord lining MSCs directly into the corneal stroma found improvement in the abnormal collagen structure, restoration of the corneal thickness, and enhancement of corneal transparency. In addition, the injected cells downregulated inflammatory cytokines, leading to a low risk of rejection [61]. In the same way, when human MSCs obtained from bone marrow, adipose tissue, and limbal stroma were cultured under conditions that promoted keratocyte differentiation and were supplemented with TGF β 3, basic fibroblast growth factor (bFGF), and ascorbic acid, genes associated with corneal stromal keratocytes were upregulated at the RNA and protein levels [56,73,91]. Similarly, in a study of mechanically induced

corneal stromal defects, transplantation of rabbit adipose MSCs cultivated on a bio-scaffold made of polylactic-co-glycolic acid could repair the defects via the induction of ALDH1A1 and keratocan expression without triggering corneal neovascularization [74].

Another study showed that corneal stroma stem cells (CSSCs) derived from the limbal stroma share many characteristics with MSCs. These cells were found to express Pax6 and MSC markers (CD90 and CD73) [92,93]. CSSCs have the potential to repair and regenerate transparent stromal tissue as well as downregulate inflammation in the cornea and reduce scarring [93,94]. CSSCs can differentiate into CSKs when grown in a serum-free environment that has been enriched with bFGF and TGF β3. This suggests that CSSCs have the potential for stromal regeneration, with the deposition of an ECM similar to that of the native stroma [95,96]. Likewise, in the lumican-null mouse model, the injection of human CSSCs into the cornea repairs the defects in collagen fibrils and restores the stromal thickness, ultimately resulting in complete restoration of corneal transparency [93]. At the L.V. Prasad Eye Institute in India, medical treatment is being performed to address patients with unilateral superficial corneal scars resulting from bacterial/fungal keratitis. This procedure entails the transplantation of cultivated allogeneic limbal stromal cells [97].

Dental MSCs are gaining popularity in regenerative medicine because of their versatility, high adaptability, lack of ethical concerns, and ease of procurement [98]. The developmental pathways of CSKs are similar to those of periodontal ligament stem cells (PDLSCs) and dental pulp stem cells (DPSCs) [99]. Transplanting human DPSCs intrastromally into a mouse cornea resulted in the expression of collagen type I and keratocan, and they exhibited a phenotype resembling that of CSKs while preserving corneal transparency and maintaining the stromal volume [72]. PDLSCs exhibit multilineage potential and differentiate into adipocytes, chondrocytes, and osteoblasts. Additionally, they express markers present on MSCs and embryonic and neural stem cells [100–102]. A preliminary clinical study was conducted using stromal cell therapy, in which patients with advanced keratoconus were treated with autologous ADSCs. The cells were injected with 1 mL of saline into the pocket of the corneal stroma. New collagen was produced, and thus, the use of autologous ADSCs for cellular therapy for human corneal stroma is regarded as safe [103]. In addition, a clinical study involving 11 patients with advanced keratoconus demonstrated positive outcomes after autologous MSC transplantation, with or without decellularized donor corneal stromal lamina sheets. Within three months after the surgery, all patients had fully regained their corneal transparency [104]. However, it is necessary to have a larger group of participants and to follow them up for a longer duration to validate the effectiveness of this treatment.

In another study, the reconstruction of corneal stroma was observed with human-processed lipoaspirate derived (PLA). This outcome proved that human ADSCs retained their morphology up to 10 weeks after transplantation. Similarly, differentiation of these cells into keratocytes was seen in the rabbit cornea 12 weeks after transplantation with the synthesis of collagens type I and VI [105]. Likewise, transplantation of MSCs derived from dental pulp into rat's eyes induced the production of an extracellular stromal matrix consisting of collagen type I and keratocon [72].

4.3. Corneal Endothelium Reconstruction with MSCs

The corneal endothelium is a thin, single-cell layer. It is the innermost layer of the cornea, which forms the boundary between the stromal and anterior chambers. According to one study, around 38% of cases requiring corneal transplantation are attributable to issues with the corneal endothelium [11]. Maintaining clarity is among the primary functions of the corneal endothelium [106]. The corneal endothelium transports fluid from the stromal layer to the anterior chamber, and its cells have a limited ability to undergo cell division [107]. In the event of any obstruction or dysfunction in the corneal endothelium, there is a risk of gradual accumulation of fluid in the stromal and epithelial layers, which can result in corneal edema and ultimately lead to impaired vision and even blindness [108]. One of the important indicators for corneal transplantation is corneal endothelial dysfunc-

tion. Recently, there has been an increase in the popularity of using Descemet's stripping automated endothelial keratoplasty (DSAEK) and Descemet's membrane endothelial keratoplasty (DMEK) for the treatment of endothelial dysfunctions [109]. Researchers are also exploring in-cell therapy as an alternative to corneal transplantation, which is often limited by a shortage of donors, and they are working to overcome the obstacles of growing endothelial cells in culture. They have discovered that Rho-associated protein kinase (ROCK) has the potential to repair and regenerate corneal endothelial cells, making corneal endothelial transplantation a viable option. In the rabbit endothelial dysfunction model, by using a ROCK inhibitor called Y-27,632, corneal transparency was restored by the transplantation of corneal endothelial cells. Even though this study is based on an animal model, there is a high likelihood that this method will yield positive results in clinical studies [110]. A recent investigation transplanted human corneal endothelial cells (HCECs) into the corneas of patients with bullous keratopathy after in vitro culture. The cells were mixed with ROCK inhibitor and transferred into the anterior chamber in a volume of 30 microliters containing 1 million cells. After 24 weeks, the corneas showed an improvement in transparency and an increase in corneal endothelial cell density. This approach was declared a less invasive treatment option for bullous keratopathy [111]. However, many obstacles remain to be overcome before applying this method to large-scale clinical treatment. Acquiring a sufficient number of cells and ensuring their proper proliferation in the laboratory can be difficult due to variations in cell fate and MET [112].

For corneal endothelial replacement, using MSCs and a conditioned medium is a promising approach [113]. One study found that exposing HCECs to the conditioned medium from BM-MSCs resulted in changes in the morphological and phenotypic characteristics of the HCECs [107]. Similarly, in another study, a damaged endothelium was restored by transplanting BM-MSCs, which were grown on a glutaraldehyde crosslinked gelatin scaffold, into the endothelial layers of rabbit corneas [114]. Likewise, in a rabbit model, transplantation of HUC-MSCs cultured on type I collagen sheets into the corneas resulted in a reduction in edema and an increment in corneal transparency [76]. Another study involved co-culturing skin-derived precursors (SKPs) and B4G12 cells in a serum-free medium, which after 4 days resulted in the formation of corneal endothelial-like cells with characteristics similar to human CECs. The markers of endothelium, NA+/K+ ATPase and zonula occludens-1 (ZO-1), were expressed well, as confirmed by immunofluorescence staining. In vivo, these cells were transplanted into rabbits and monkeys after Descemet's membrane was mechanically removed, resulting in an increase in corneal transparency and a sharper appearance of the cornea after just 7 days [115].

Recently, a study demonstrated that hMSCs can differentiate into cells similar to corneal endothelial (CE) cells. The differentiated cells were found to express the markers typically found in corneal epithelial cells, such as ZO-1, Na/K-ATPase, COL-8, and paired-like homeodomain transcription factor 2 (PITX2). The differentiation process was achieved via the use of Descemet membrane biomimetic microphotography [116]. In this study, researchers implanted Warton jelly-derived stem cells in a Descemet membrane that had a collagen-like topography to create an endothelium-like layer. The researchers found an increase in the expression of genes specific to endothelial cells (COL-8, ZO-1, Na/K-ATPase, and PITX2). These results suggest that Wharton jelly-derived stem cells possess the capability to transform into cells with endothelial characteristics. Furthermore, transplantation of these cells in ex vivo rabbit cornea signified the formation of functional endothelium and transparent cornea [117].

5. Fate of MSCs in Corneal Inflammation and Angiogenesis

MSCs are recognized for their ability to regulate angiogenesis and reduce inflammation, making them a promising treatment option for various corneal diseases. Many studies have indicated that topical or sub-conjunctival administration of BM-MSCs can reduce inflammation and angiogenesis in murine models with chemical injuries [48,50,57]. Incorporating MSCs into corneal tissue has been shown to decrease the infiltration of inflammatory

cells and macrophages expressing CD68. This leads to a reduction in pro-inflammatory cytokines, including interleukin-2 (IL-2), IL-1, monocyte chemoattractant protein (MCP-1), and matrix metalloproteinase 2 (MMP2), as well as pro-angiogenic factors. MSC treatment can also increase the expression of molecules with anti-inflammatory effects, such as IL-6, IL-10, TGF-β1, and TSG-6, along with anti-angiogenic mediators, such as thrombospondin-1 (TSP-1) and pentraxin-3 [54,57]. By modifying the pro-inflammatory environment, the corneal epithelium can be restored, leading to the healing of ocular surface injuries in the damaged cornea [51].

MSCs express pro and anti-angiogenic factors, depending on the tissue microenvironment. In order to inhibit angiogenesis, MSCs increase TSP-1 by disrupting the signaling of CD47 and vascular endothelial growth factor (VEGF) receptor 2, and inhibiting the VEGF-Akt-Enos pathway. The release of TSP-1, a potential inflammatory cytokine with pro-angiogenic activity, not only induces endothelial cell apoptosis but also decreases the expression of MMP2 [118]. Recent studies have demonstrated the modulatory actions of the CSSCs derived from the limbal stroma in corneal inflammation and scarring [119]. One study implemented CSCs in an acute corneal wound mouse model, and the inhibition of neutrophils and a decline in the expression of fibrotic markers like tenascin C, alpha-smooth muscle actin (α-SMA), and secreted protein acidic and rich in cysteine (SPARS), mediated through the TSG-6 pathway, was observed [94]. Likewise, in murine corneas, corneal mesenchymal stromal cells also perform an anti-angiogenic role via expression of PEDF and soluble fms-like tyrosine kinase-1 (sFLT) and suppression of macrophage infiltration [59]. Suture-induced corneal neovascularization was inhibited with the intravenous administration of CMSC, which decreased the expression of various genes involved in angiogenesis, such as VEGF-C, VEGF-D, TEK, and mannose receptor C type-1 and 2, within the stromal matrix [120].

6. Corneal Transplantation with MSCs

Numerous studies have explored the functions of MSCs in enhancing the survival of grafts [118,121,122]. The immune-modulating and anti-inflammatory characteristics of MSCs make them a promising choice for the transplantation of corneal allografts. Since the cornea is considered an immune-privileged site, the survival rate of corneal transplants is higher than other types of solid organ transplant. Regardless of this, corneal graft immune rejection can sometimes lead to corneal allograft failure [123]. The maturation of B-cells can be prevented by MSCs, and they can also inhibit the release of cytokines by T-cells. This modulation helps to sustain tolerance of the allograft and promotes its survival by influencing the generation of regulatory T-cells [124,125]. MSCs have a cell surface glycocalyx that contains a high concentration of anti-inflammatory molecules such as TSG-6, versican, and pentraxin-3. These molecules are involved in regulating the inflammatory response of the host [125]. The initial immunomodulatory impacts of recipient-derived MSCs from a pig-to-rat model were examined in the context of penetrating keratoplasty. When allogeneic rat MSCs were applied topically, they caused T-cells to differentiate into Th2 cells. However, even though the researchers induced an increase in Th2-type cytokines using MSCs, they did not observe a notable improvement in the survival of pig corneal tissue transplanted into rats [126].

Similarly, in a study conducted using a rat corneal transplant model, the researchers demonstrated the immunosuppressive capabilities of MSCs. Specifically, they injected MSCs from the donor into the recipient rats at varying intervals and with different doses of cyclosporine A (CsA). Prolonged corneal graft survival and no corneal allograft rejection were associated with a post-operative infusion of MSCs, whereas a pre-operative infusion was ineffective. Furthermore, MSCs in allogeneic keratoplasty inhibited allogeneic T cell responses in both in vitro and in vivo rat models, suggesting MSCs prevent allograft immune rejection and increase the survival rate of allografts by upregulating the number of Tregs [127]. Similarly, in a study performed on a mouse model of corneal allotransplantation, the peri-transplant intravenous infusion of human MSCs led to the

suppression of inflammation and decreased the activation of antigen-presenting cells in the cornea, thereby increasing the survival rate of allografts and decreasing the risk of immune rejection [122]. Despite the potential of MSCs as an alternative method to treat and prevent immune rejection after corneal transplantation, their effectiveness in addressing corneal allograft rejection in various animal models and clinical scenarios is still unclear.

7. Challenges of MSC Therapy

Despite tremendous applications in regenerative medicine, MSCs have some significant risks, and solving this issue is a major challenge [128]. The major problem associated with MSCs is cellular heterogeneity. MSCs sourced from different origins can result in inconsistent outcomes. Furthermore, an additional critical aspect of cell therapy is the various techniques used to obtain, isolate, and cultivate the cells, which can result in differing outcomes [129]. One of the considerable problems associated with insufficient reproducibility of experimental findings is the variation in MSC protocols [128]. Full documentation of the complete process of MSC isolation, sorting, ex vivo expansion, purification, phenotyping, and conducting follow-up examinations should be carried out to ensure reproducible clinical efficacy and outcomes [130]. Prior to the clinical settings, it is necessary to establish standardized protocols for the isolation and ex vivo preparation of MSCs. However, parameters such as age, genetic traits, and the medical history of donors can impede corneal cell therapy. For example, the limited number of MSCs acquired from such cases makes it difficult to utilize autologous transplantation in elderly patients [131]. Furthermore, some studies have demonstrated that treatment with MSCs may not be sufficiently safe, as they are not immune-privileged [132]. However, the cornea itself is immune-privileged, and this is less important when treating corneal conditions. Likewise, senescence and decreases in differential capacity can be observed in MSCs after the sixth passage; this may lead to alterations in gene expression profiles and cell morphology, which can have adverse effects during cell therapy [133,134]. Another critical concern with MSCs is their tendency to transform into different lineages during ex vivo expansion. Thus, controlling and documenting the entire MSC expansion procedure should be made mandatory to ensure reproducibility in the preparation of these cells [135]. Furthermore, ethical challenges should be addressed before undertaking clinical trials. These challenges involve minimizing harm, appropriately selecting and recruiting subjects, ensuring informed decision making through the consent process, and preventing therapeutic misconception [136]. Table 2 illustrates the advantages and disadvantages of MSCs and MSC-derived exosomes.

Table 2. Advantages and disadvantages of MSCs and MSC-derived exosomes in clinical applications [137–139].

Origin	Pros	Cons
MSCs	-Easy to isolate and obtain from accessible sources -High rate of proliferation, multilineal differentiation -Easily cultured in vitro -Low risk of immune-related problems -High stability in various pathological and physiological conditions	-Ethical issues -Risk of potentially transmitting genetic diseases and infections -Low number of cells -After transplantation, risk of teratoma formation is high
MSC-derived exosomes	-Capacity to cross natural barriers like blood–brain barrier -Perfect immune-compatibility and non-cytotoxic -Compared with cells, stable upon freezing and thawing -Ability of natural homing -Capacity for intracellular delivery of cargo by fusion of membranes	-No standard isolation protocol -No excellent mass production protocol -Less and immature research on exosome-based therapies -In vivo, after administration, fast clearance from the blood -Hard to isolate and purify the exosomes

8. Paracrine Effect of MSCs

Within the realm of regenerative medicine, MSCs can have a therapeutic impact by producing soluble factors that aid in the regulation of tissue healing, inflammation, angiogenesis, and immune responses [82,140]. Most MSCs are naturally found in the filtering organs like the lungs, liver, and spleen. In one study, the injection of MSCs beneath

the conjunctiva in corneas injured by alkali exposure facilitated the healing of corneal wounds, even though the MSCs remained localized in the subconjunctival space [51]. In the same way, when MSCs or conditioned media from MSCs were topically administered in a mouse model of corneal epithelial injury, they alleviated corneal inflammation, reduced the formation of new blood vessels, and facilitated wound healing [141]. The fact that the majority of the MSCs were present in the corneal stroma rather than the epithelium implies that the therapeutic impact of MSCs is accomplished via a paracrine mechanism rather than direct cell substitution. This mechanism involves the release of soluble factors from MSCs, which can be delivered through extracellular vesicles or exosomes [140,142,143].

9. Mesenchymal Stem Cell-Derived Exosomes

9.1. Exosome Biogenesis

Exosomes, which are a subset of extracellular vehicles (EVs), are membrane-bound and generated within the endosomal compartment of almost all eukaryotic cells [144]. Exosomes are present in a wide range of biological tissues and fluids, and they are found in almost all cells, tissues, and bodily fluids, including urine, blood, plasma, cerebrospinal fluid, sweat, breast milk, semen, gastrointestinal secretions, saliva, and amniotic fluid [145]. Unlike microvesicles, which are formed during apoptosis, exosomes are produced directly from the plasma membrane and apoptotic bodies, and they originate from endosomes [146]. Once exosomes are secreted into the extracellular space, they bind to multivesicular bodies (MVBs), which have intraluminal vesicles (ILVs), then merge with the plasma membrane. The formation and packaging of exosomes are managed by endosomal sorting complexes required for protein (ESCRT), which are attracted to MVBs. Furthermore, many other associated proteins, such as Tsg101, Alix, and VPS4, are also actively involved in this process [146]. The amount and composition of exosomes released from the same parental MSCs differ due to the influence of external factors that dictate the secretion of exosomes [33]. Exosomes are internalized by recipient cells in the local microenvironment or transported to distant regions via the circulatory system. The uptake of exosomes by target cells can occur through three main mechanisms: (1) interactions between ligands and receptors, (2) endocytosis by the recipient cells, and (3) direct fusion with the cell membrane [147] (Figure 3). Once the exosomes have merged with the recipient cells, their contents are released into the cytoplasmic space.

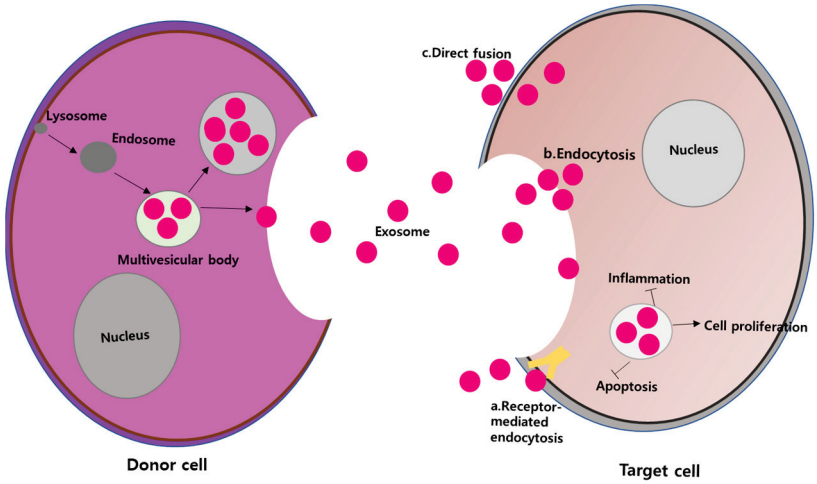


Figure 3. Exosomes are capable of transferring bioactive molecules to recipient cells through three mechanisms: (a) intercellular signaling via receptor–ligand interaction, (b) endocytosis by recipient cells, and (c) direct fusion with the recipient cell membrane, leading to the release of their cargo into the target cells. These modes of transfer are the result of the biogenesis of exosomes [97,148].

9.2. MSC-Derived Exosome Components

The molecular composition of exosomes is subject to considerable variability and is influenced by several factors such as the cell type that produces them, changes made to them, and the pathological environment in which they are formed. The biological roles that exosomes fulfill depend on the types of nucleic acids, proteins, and lipids that they contain. The composition of the exosomes has become well known in recent decades due to massive progress in the fields of biotechnology, including proteomics, lipidomics, transcriptomics, and bioinformatics, which provide a theoretical basis for the use of exosome therapy in treating various diseases [33]. Phosphatidic acid, cholesterol, sphingomyelin, arachidonic acid, prostaglandins, and leukotrienes are exosomal lipids. They are involved in the formation of exosomes and the maintenance of their biological stability [145]. Exosomes contain various multifaceted proteins involved in their production, such as synthenin, ALIX, TSG101, and ESCRT complex. In addition, exosomes carry membrane transporter and fusion proteins, including annexins, heat shock proteins, and Rab GTPase [149]. Again, exosomes are enriched with various nucleic acids, such as genomic DNA, cDNA, mitochondrial DNA (mtDNA), long coding RNAs (IncRNAs), circular RNAs (CirRNAs), microRNAs (miRNAs), and mRNAs [150]. They have an indispensable role in modulating biological processes and the epigenetic remodeling of cells.

MSCs have been demonstrated to produce a greater quantity of exosomes compared to other types of primary cells [151]. Proteomic analysis of MSC exosomes has revealed 1927 distinct proteins that possess various functions required for their characteristics and formation [152]. The exosomes derived from MSCs display commonly occurring exosome surface proteins, such as tetraspanin (CD81, CD63, and CD9), ALIX, and Tsg101. Furthermore, they also contain heat shock proteins (HSP90, HSP70, and HSP60). Similarly, MSC membrane proteins for several adhesion molecules (CD73, CD44, and CD29) are also expressed by exosomes [144]. Likewise, diverse types of nucleic acids including lncRNAs, miRNAs, and mRNAs are found in MSC-derived exosomes. The miRNAs are given a greater interest than others, as it is believed that they are a form of non-coding RNA; they are approximately 22 nucleotides in length and modulate post-transcriptional gene expression [153]. These miRNAs are crucial molecules in MSC exosomes since they are involved in several biological activities, such as cell differentiation, angiogenesis, apoptosis, and inflammatory pathways. Researchers have found that a single miRNA can regulate several messenger RNAs, and conversely, a single messenger RNA can be influenced by several miRNAs [154]. Thus, these complex networks demonstrate the potential of MSC exosomes to alter numerous functional, physiological, and pathological effects.

9.3. Isolation and Storage of Exosomes

Several methods are commonly used for isolating exosomes, such as ultracentrifugation, size exclusion chromatography (SEC), polymer precipitation, immune affinity capture, microfluidics, and ultrafiltration (UF). These techniques are utilized to separate exosomes from other components in biological samples and obtain a pure exosome population for downstream analysis. Among these, for separating exosomes, ultracentrifugation is the most commonly used method and is considered the gold standard approach [155]. Ultracentrifugation separates proteins, vesicles, cell debris, and cells into uniform suspensions based on their differential sedimentation rate. SEC uses the size of small-molecule–protein complexes as the basis for separation. UF separates exosomes according to their size [156]. Immune affinity chromatography separates exosomes based on their interactions with high-specific-affinity antibodies and antigens [157]. Polymer precipitation alters the solubility or dispersion of exosomes in body fluids or cell cultures to precipitate them from samples, usually through the use of polyethylene glycol or agglutinin [158].

To date, there is no standard universally accepted guideline for the proper storage of exosomes. The integrity of exosome lipid membranes and their therapeutic efficacy can be influenced by the buffer composition, storage temperature, and the number of freeze–thaw cycles. One study showed no alterations in the size of MSC exosomes or overall exosomal

membrane integrity after −20 °C freeze–thaw cycles in PBS. However, the size of the vesicles was significantly reduced after being stored at 37 °C for two days and at 4 °C for three days [159]. Likewise, one study found that exosomes could be stored at −20 °C for six months without altering their biochemical activity [97]. Another study found that when neutrophil-derived exosomes were frozen at −20 °C, the vesicle size decreased but it did not at −80 °C [160]. Another study found that adding protease inhibitors to urinary exosomes prior to freezing at −20 °C did not halt the decline in exosomal biochemical activity. However, complete recovery of activity was observed after seven months of storage when the exosomes were frozen at −80 °C [161]. A recent study showed that the stability of the exosomal membrane and biochemical function was further improved by adding 25 mM of trehalose [162]. The findings from these studies have provided evidence that exosomes can maintain their functional stability even when stored for extended periods at relatively mild temperatures.

10. Therapeutic Promise of MSC-Derived Exosomes for Ocular Tissue

Several studies have provided evidence that MSC-derived exosomes have a significant influence on eye tissues. In a rat model of experimental autoimmune uveitis (EAU), injecting exosomes derived from HUC-MSCs around the eye reduced the movement and accumulation of immune cells, such as leukocytes, macrophages, and natural killer cells. This was accomplished by inhibiting the MCP1/C-C motif chemokine ligand 2 (CCL21) and MYD88-dependent pathways. In addition, these exosomes restored retinal function and stimulated the expression of Gr-1, CD4, CD68, CD161, and IL17 [163]. Likewise, in a study of laser-induced retinal injury, injecting exosomes derived from MSCs cultured from either umbilical cord or adipose tissue into the eye improved visual function and modified the pro-inflammatory environment. This was achieved by hindering the production of pro-inflammatory cytokines MCP1, intercellular adhesion molecule-1 (ICAM-1), and TNF-α [164]. Similarly, in a study conducted on diabetic rats with hyperglycemia-induced retinal inflammation, injecting HUC-MSC-derived exosomes directly into the eye (intravitreal injection) improved their visual condition. This was achieved by suppressing the high mobility group Box 1 (HMGB1) signaling pathway, aided by miR-126 [165]. Furthermore, in a study involving rats with retinal damage induced by blue light, injecting umbilical cord MSC-derived exosomes directly into the eye (intravitreal injection) resulted in a dose-dependent reduction in choroidal neovascularization. This suppression was achieved by reducing the production of VEGFA and inhibiting the NFkB pathway through the transfer of miR-16 [166,167]. In addition, in an experiment using a rat model of optic nerve crush, injecting exosomes derived from BM-MSCs directly into the eye (intravitreal injection) restored the growth of retinal ganglion cells. This restoration occurred through the activation of argonaute-2 signaling [168]. Other studies showed that intravenous administration of MSCs can restore retinal function in models of EAU and laser-induced retinal injury [169,170]. A clinical trial was conducted recently to explore the impact of intravitreal injection of HUC-MSCs-derived exosomes on patients with refractory macular holes. The trial involved five patients and showed the intervention resulted in both functional and anatomical recovery. However, one of the patients had a severe inflammatory response [171].

11. MSC-Derived Exosomes and the Cornea

Injuries to the cornea caused by chemical or thermal burns, traumatic injury, immune disorders, and hereditary conditions can lead to inflammation, neovascularization, scarring, and ulceration. Delayed and careless treatment may cause permanent blindness. However, MSC therapy is considered a suitable candidate to provide anti-inflammation, anti-angiogenesis, and immunomodulatory activities during treatment. Several studies have shown that the paracrine action of MSCs can greatly enhance the process of wound healing by regulating inflammation, angiogenesis, and tissue regeneration through various

factors. Moreover, limited studies conducted in vitro and in vivo have shown the healing benefits of the bioactive molecules present in MSC exosomes on corneal injury models.

In an experiment where rabbit CSCs were cultured with ADSC exosomes, there was a robust increase in cell proliferation, reduced cell death, downregulation of MMPs, and a substantial deposition of new ECM molecules, specifically collagen [172]. Furthermore, in a mouse model of a superficial stromal wound, topical application of exosomes derived from CSSCs reduced corneal inflammation and scarring by inhibiting the entry of neutrophils through a TSG-6-dependent pathway and decreasing the activity of genes related to fibrosis such as actin alpha 2 (ACTA2), collagen type 3 alpha 1 (COL3A1), and acidic and rich in cysteine (SPARC) [94]. Likewise, exosomes derived from human corneal mesenchymal stromal cells have the ability to stimulate the recovery of damaged corneal epithelium in mice [173]. In addition, in a murine model of mucopolysaccharidosis, HUC-MSCs exosomes that carried β-glucuronidase were found to reduce the accumulation of glycosaminoglycans, thereby leading to a decrease in corneal haze [34]. Similarly, a study investigated the impact of MSC-derived factors on the function of keratocytes in vitro and found a significant improvement in the cells' performance. Furthermore, the paracrine activity of MSCs enhanced the survival of keratocytes by inhibiting apoptosis [174]. A research study utilized exosomes from ADSCs loaded with miRNA 24-3p and incorporated into a modified hyaluronic acid hydrogel. The resulting miRNA 24-3p-rich exosomes (Exos-miRNA 24-3p) were observed to enhance corneal epithelial defect healing, cell migration, and maturation while inhibiting fibrosis and reducing the levels of inflammatory cytokines (CD163, and MMP9) in both in vitro and in vivo rabbit models of corneal epithelial defects caused by alkali burns [175]. Additionally, in a study with iPSC-MSCs exosomes combined with thermosensitive chitosan-based hydrogels (CHI hydrogel) in a rat cornea damage model using a trephine mold, the iPSC-MSCs exosomes promoted reconstruction of the damaged corneal epithelium and stromal layer. In vivo, this study documented the downregulation of mRNA expression of three major collagen types, collagen type I alpha (COL1A), collagen type V alpha 1 (COL5A1), and collagen type V alpha 2 (COL5A2). Additionally, this led to the inhibition of scar formation, and ECM decomposition was prevented by inhibiting the translocation-associated membrane protein 2 (TRAM2) [176].

In a rat model of corneal allograft rejection, transplanting MSC-exosomes resulted in a significant increase in graft survival time by inhibiting the infiltration of CD4+ and CD25+ T cells, reducing the levels of interferon-gamma (IFN-γ) and C-X-C motif chemokine 11 (CXCL11), and inhibiting the Th1 signaling pathway [177]. In addition, a study using exosomes derived from HUC-MSCs showed significantly enhanced in vitro proliferation and the migration of corneal epithelial cells in a corneal mechanical wound healing model in rats. This was attributed to downregulation of phosphatase and tensin homolog (PTEN) levels and activation of the PI3K/Akt signaling pathway through the transfer of miRNA-21, resulting in better corneal repair and regeneration [178]. In an in vitro study using MSC-EVs in HCECs, it was revealed that MSCs-EVs (5–20 × 103 MSC-EV/cell) were capable of regenerating damaged HCECs by decreasing the number of apoptotic cells [179]. Exosomal microRNA from ADSCs inhibited the expression of homeodomain-interacting protein kinase 2 (HIPK2) in rabbit corneal keratocytes, suppressing the transformation of rabbit corneal keratocytes into myofibroblasts [180]. When comparing MSCs-Exos and iPSCs exosomes, researchers found that iPSC exosomes showed a better effect in vitro by inducing proliferation, migration, and cell cycle progression, and inhibiting apoptosis [181]. This study found that BM-MSCs-derived exosomes could promote the expansion and movement of human corneal epithelial cells in a manner that depended on the dosage used. This effect was related to the activation of the p44/42 MAPK signaling pathway. In an animal experiment using a mouse alkali burn model, the injection of BM-MSCs-derived exosomes was observed to promote the healing of corneal damage by reducing inflammation and mitigating the overproduction of proteins associated with fibrosis (α-SMA) and vascularization (CD31) [182]. In a murine corneal damage model caused by alkali burns, scientists demonstrated promising results by topically administering BM-

MSCs-derived extracellular vesicles (BMSC-EVs) embedded with methylcellulose. The BM-MSCs-EVs were found to regulate cell death, inflammation, and angiogenesis in the damaged tissue, leading to faster corneal regeneration [183]. Similarly, it was demonstrated that exosomes obtained from HUC-MSCs influenced autophagy in both human corneal epithelial cells and a mouse corneal injury model. In vitro, this treatment caused an increase in cell proliferation, migration, and upregulation of proliferating cell nuclear antigen (PCNA), cyclin-dependent kinase 2 (CDK2), cyclin A, and cyclin E expression. In addition, when combined with an autophagy activator, HUC-MSCs ameliorated corneal defects by decreasing apoptotic and inflammatory gene expression in vivo [184]. Likewise, it has been observed that MSC-derived exosomes can mitigate inflammation and cell death in the cornea following injury. This is supported by the decreased expression of proinflammatory molecules like TNF-α, IL-1β, IL-8, and NF-κB, as well as the pro-apoptotic protein Cas-8. Furthermore, these exosomes have shown the potential to reduce corneal angiogenesis by inhibiting the expression of pro-angiogenic factors (VEGF) and angiogenesis-associated genes (MMP-2 and MMP-9) [185]. Moreover, researchers have proven that exosome-mediated targeting of NF-κB c-Rel can effectively accelerate corneal wound healing. Using exosomes loaded with c-Rel-specific siRNA on the corneal surface as a topical treatment reduced the expression of inflammatory cytokines and accelerated corneal wound healing in mice, including in cases of diabetic corneal injury. The results of this study suggested that inhibiting c-Rel might be an effective treatment approach for managing a corneal injury [186].

In a study with a corneal epithelial defect model in diabetic mice, exosomes derived from BM-MSCs labeled with PKH-26 regenerated the corneal epithelium by inducing the proliferation, repair, and migration of corneal epithelial cells. Furthermore, infiltrating inflammatory cells were relatively diminished in the cornea treated with exosomes [187]. A study assessed the effectiveness of ADSCs compared to MSC-derived exosomes for treating corneal injuries induced by alkali in rats. Both ADSCs and exosomes had the potential to enhance corneal healing and prevent complications resulting from alkali burns. This is due to their ability to reduce inflammation and prevent angiogenesis. However, the researcher came to the conclusion that exosomes from MSCs could be an especially auspicious alternative since they were less risky than stem cells, and their diminutive size facilitated their penetration through biological barriers to reach specific organs [188]. Interestingly, in a recent study, when MSC exosomes were treated with a human cornea-on-a-chip (developed using microfluidic technology), they had a favorable outcome on corneal epithelial wound healing by reducing the release of MMP-2 protein and acting as an anti-angiogenetic factor [189]. Furthermore, ADSCs inhibited ECM formation, increased proliferation, and promoted the reverse migration of C-X-C chemokine receptor type 4 (CXCR4) neutrophils, which alleviated neovascularization. These findings suggest that ADSCs have an antiangiogenic effect during corneal wound healing and can be a potential treatment approach that reverses neutrophil migration [190]. In addition, it was observed that exosomes from CSSC downregulated the fibrotic gene (Acta2) expression, blocked neutrophil infiltration, reduced scar formation, and restored the corneal morphology by transferring miRNA to corneal cells [191]. Likewise, in a rat model of corneal scarring, topically administration of MSC exosomes alleviated corneal injuries by inhibiting angiogenesis, modulating the immune response, minimizing scar formation and enhancing the healing process [192]. These studies have demonstrated that MSC exosomes can enhance the treatment of corneal ocular surface diseases and congenital corneal metabolic disorders by promoting ECM formation and cell proliferation, migration, and survival (Figure 4 and Table 3).

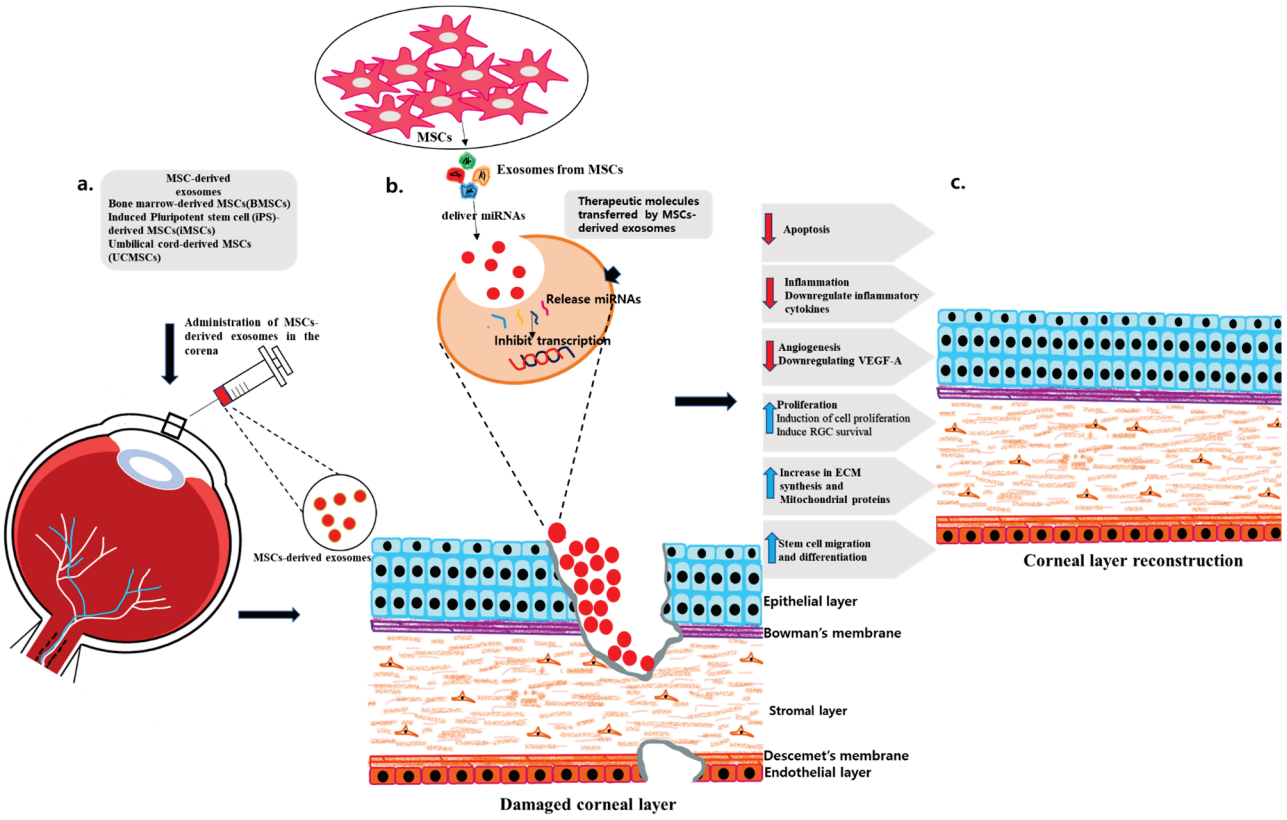


Figure 4. (**a**) A schematic representation of the experimental administration of MSC-derived exosomes in cornea. (**b**) In the cornea, therapeutic molecules are delivered that inhibit angiogenesis, cell proliferation, cell migration, and differentiation. (**c**) Regeneration of corneal layers is observed after treatment with MSC-derived exosomes.

Table 3. Preclinical studies of stem-cell-derived exosomes with their therapeutic potential for treating various corneal diseases.

Origin of Exosome	Aim of Experiment	Animal Model	Study Performed	In Vitro Study Conclusion	In Vivo Study Conclusion	References
ADSC exosomes	-To study how ADSCs exosomes may lead to phenotypic alterations in vitro in CSCs.	-	In vitro	-Inhibition of apoptosis, downregulation of MMPs, upregulation of ECM-related proteins (collagens and fibronectin) and significant proliferation of CSCs by ADSCs- exosomes.	-	[172]
Human corneal mesenchymal stromal cell-derived exosomes	-To investigate the impact of exosomes derived from human corneal MSCs on the healing of corneal epithelial wounds.	Mouse	In vitro In vivo	-Observation of corneal epithelial cell migration, proliferation, and adhesion. - Modulation of expression of genes related to cell signaling, inflammation, ECM remodeling.	-Improvement in corneal epithelial wound healing by augmenting cell proliferation.	[173]
BM-MSCs	-To examine how soluble factors derived from MSCs affect the functions of keratocytes (activation, migration, proliferation and synthesis of ECM).	Mouse	In vitro In vivo	-Enhancement of keratocye survival by inhibiting apoptosis, upregulation of ECM genes, increasing cell viability, and migration.	-Demonstration that various wound healing mediators like vascularly endothelial growth factor (VEGF), platelet-derived growth factor (PDGF), hepatocyte growth factor (HGF) accelerate corneal re-epithelization.	[174]

Table 3. *Cont.*

Origin of Exosome	Aim of Experiment	Animal Model	Study Performed	In Vitro Study Conclusion	In Vivo Study Conclusion	References
ADSCs exosomes	-To explore the potential of ocu-micro-RNA 24-3p to facilitate the migration and repair of rabbit corneal epithelial cells.	Rabbit	In vivo In vitro	-Increase in migration and proliferation of corneal epithelial cells.	-Promotion of rabbit corneal epithelial cell migration and repair by inhibiting fibrosis and keratitis, decreasing inflammatory reactions.	[175]
iPSC-MSCs exosomes	-To determine the effectiveness of exosomes obtained from iPSC-MSCs to repair damaged corneal epithelium and stromal layer, by decreasing the formation of scars and speeding up the healing process.	Rat	In vitro In vivo	-Downregulation of mRNA expression of COL1A, and COL5A2 in anterior lamellar stroma damage model in rats. -Reducing scar formation and regenerate corneal epithelium.	-Suppression of translocation-associated membrane protein 2 (TRAM2) by mi R-432-5p to prevent ECM decomposition.	[176]
MSC exosomes	-To study the effects of MSC-exosomes in corneal allograft rejection model.	Rat	In vivo	-	-Subconjunctival injection of 10 ug exosomes can effectively prolong the graft survival time. -Inhibit infiltration of CD4+ and CD25+T cells, downregulation of IFN-γ and CXCL11 levels in grafts.	[177]
HUC-MSCs-derived small extracellular vesicles (HUMSC-s EVs)	-To explore the mechanism through which HUMSCs-sEVs impact the healing process of corneal epithelial wounds.	Rat	In vitro In vivo	-Promotion of cell proliferation and migration via upregulating the P13k/Akt signaling pathway, achieved by restricting PTEN with transfer of miR-21.	-Corneal fluorescein staining and histological staining showed the healing of corneal wound in corneal mechanical wound rat model.	[178]
Stem cell-derived extracellular vesicles	-To explore the contribution of extracellular vesicles derived from stem cells in stem cell-induced regeneration, by reprogramming injured cells and triggering pro-degenerative pathways.	-	In vitro	-Decline in the quantity of apoptotic cells and faster wound repair in human corneal endothelial cells treatment with MSC-EVs.	-	[179]
ADSCs exosomes	-To examine the impact of exosomal miRNAs obtained from ADSCs on the differentiation process of rabbit corneal keratocytes.	-	In vitro	-Inhibition of HIPK2 expression suppresses the differentiation of corneal keratocytes into myofibroblasts. -Reduction in the expression of markers promoting pro-fibrosis and ECM components.	-	[180]
iPSCs exos and MSC exos	-To compare the efficacy of iPSCs-exos and MSC-exos in the treatment of corneal epithelial defects.	Rat	In vitro In vivo	-More promising result for iPSCs-exos than MSCs-exos. -Greater proliferation, migration, cell cycle progression, and inhibition of apoptosis in human corneal epithelial cells.	-Demonstration of stronger effects of iPSC-exos in healing corneal epithelial defect model than MSC-exos.	[181]
BM-MSCs exos	-To explore how BM-MSCs-exos promote corneal wound healing by activating the p44/42 MAPK signaling pathway.	Mouse	In vitro In vivo	-Enhancing the growth and migration of human corneal epithelial cells by BM-MSCs-exos.	-Downregulation of inflammation, fibrosis (α-SMA) fibrosis and vascularization (CD31) in corneal tissues of mice with alkali burn injury.	[182]
HUC-MSCs exos	-To explore the molecular mechanisms of HUC-MSCs-exos affect autophagy in vitro and corneal injury (CI) models in vivo.	Mouse	In vitro In vivo	-Combination of HUC- and MSCs-exos and an autophagy activator enhances cell proliferation, increases migration capacity, and boosts the expression of PCNA, Cyclin A, Cyclin E, and CDK2.	-Reduction in expression of apoptotic genes (Bax and Caspase 3), decrease in the inflammatory markers (TNF-α, IL-1β, IL-6 and CXCL-2) increment in BCL-2 in CI mice model.	[184]

Table 3. *Cont.*

Origin of Exosome	Aim of Experiment	Animal Model	Study Performed	In Vitro Study Conclusion	In Vivo Study Conclusion	References
HUC-MSCs	-To study whether the use of UMSC transplantation into corneal stroma has the potential to contribute in breakdown of glycosaminoglycans (GAGs), offering a viable approach for cell-based therapy for mucopolysaccharidoses (MPS)	Mouse	In vitro In vivo	-Release of neutral vesicles by HUC-MSCs which are taken by fibroblasts in a coculture assay	-Restoration of dendritic and hexagonal morphology of host keratocytes and endothelial cells, respectively, reduction in corneal haze. -Participation in breaking down extracellular GAGs and facilitating the host keratocytes to metabolize accumulated GAG products.	[52]
hMSCs exosomes	-To demonstrate the effectiveness of c-Rel-specific siRNA delivered through exosomes in accelerating corneal wound healing.	Mouse	In vivo	-	-Nano-polymers or exosomes containing c-Rel-specific siRNA on the corneal surface as a topical treatment help speed up the healing of corneal wounds in both regular and diabetic cases.	[186]
BM-MSCs-derived exosomes	-To examine how exosomes obtained from mouse BM-MSCs impact the regeneration of corneal epithelium in mice with diabetes.	Mouse	In vivo	-	-Restoration of corneal epithelial injury in diabetic mice by exosomes labelled with PKH-26 by downregulating the infiltration of cytokines and proliferation of corneal cells. -Detection of exosomes in the corneal stroma and nourishing it.	[187]
ADSCs and MSCs exosomes	-To assess and compare the effectiveness of ADSCs versus MSCs-exosome in treating corneal injuries induced by alkali in rats.	Rat	In vivo	-	-Improvement of corneal layers with decrease in inflammation and anti-angiogenic effects by MSCs-exosome-treated group in alkali burn injury model.	[188]
MSC exosomes	-To study the wound-healing and immunomodulatory effects of MSC exosomes in a rat corneal scarring model	Rat	In vitro In vivo	-Exerted immunomodulatory effect by modulating the expression and secretion of chemo-attractants.	-Improved the corneal epithelial wound healing by reducing corneal haze, supressing corneal neovascularization, downregulating inflamatory cytokines, and inhibiting angiogenesis in an excimer laser-induced rat corneal injury model.	[192]

12. Challenges and Future Perspectives

MSC therapy is considered a promising strategy in different pre-clinical studies for treating different types of corneal disorders, either by directly repairing the damaged tissue or communicating with other cells through soluble factors. The immune-privileged nature of corneas results in a lower risk of rejection and tumorigenesis.

Exosomes play a critical role as mediators in exchanging cellular information in various non-ocular models. The recipient cells take up the proteins and miRNAs present in the exosome cargoes, resulting in a reduction in inflammation, immune modulation, induction of angiogenesis, wound repair, and overall improvements in functional and biological recovery. The use of autologous serum eye drops has been commonly employed to manage various conditions like severe dry eyes, persistent epithelial defects, chemical injuries, and neurotrophic keratopathy [193]. MSC therapy can be a potential option for treating corneal injuries or diseases, and the topical application of MSC exosomes could be used as an additional therapy to address corneal inflammation and neovascularization. Exosome eye drops derived from MSCs may provide a safer alternative compared to cell-based therapies since they have a lower risk of immunological rejection, uncontrolled proliferation, tumor formation, toxicity, and higher bioavailability.

The commercially available autologous serum eye drops are non-specific, typically focus on lubrication and symptom relief, and hardly restore the immune homeostasis in the diseased cornea without actively promoting tissue repair. In a prospective clinical trial, after topical administration of MSC-derived exosomes to patients with a diseased cornea, there was a reduction in fluorescein scores, a longer rear film breakup time, and an increment in tear secretion [194]. In addition, to obtain the potential therapeutic and biological effects of MSCs, coupling MSC therapy with topical MSC exosomes would be a better option. Thus, with this increasing experimental evidence with attractive advantages over MSCs, MSC-derived exosomes would be the perfect candidate for cell-free therapy, having low immunogenicity and less risk of tumor formation. More than 850 unique proteins and 150 different miRNAs are delivered in the cargoes of MSC-derived exosomes to affect various pathways in the target cells [115].

Due to their cell-free nature, MSC-derived exosomes can potentially reduce issues associated with immunological rejection and the first-pass effect often observed with systemic MSC treatment. This makes MSC-derived exosomes a promising candidate for use as biological carriers of therapeutic agents. Despite the therapeutic potential demonstrated by MSC-derived exosomes in numerous studies, it is crucial to address various ethical concerns before their application in clinical settings. The fact that exosomes are derived from stem cells but are not actual cells presents a potential hurdle in defining legal classifications and obtaining approval from regulatory bodies in different countries for their clinical use [195].

Other challenges remain to be overcome. First, the absence of standardization of the methods used for the purification, separation, and profiling of MSC-derived exosomes on a global scale may result in contentious issues arising from varied laboratory investigations. The transportation, preservation, and storage of exosomes are vital aspects that should be seriously considered. One study showed that thawed freeze-drying could effectively purify exosomes and facilitate their transportation and long-term storage. However, it is necessary to conduct additional research to determine whether this process has any impact on the characteristics of the exosomes. Second, research has shown the exosomes derived from MSCs to be safe and effective for the treatment of corneal diseases. Regardless, the pharmacological characteristics like bioavailability, targeting, pharmacokinetics, and bio-distribution of exosomes should be studied for proper therapeutic applications of exosomes. It is essential to closely monitor and understand the precise mechanism of action of exosomes prior to their use in clinical applications. Third, another important consideration is determining the optimal route of exosome administration, whether local or systemic. It is also essential to establish the appropriate dosage and dosing intervals and to conduct potency assays to evaluate potential toxicity at different doses. Fourth, optimization of the culture methods and isolation techniques is vital. Finally, the healing efficacy of exosomes is mainly attributed to their inherent cargoes, such as RNA, DNA, lipids, and proteins, which should be investigated more thoroughly. Additional research is necessary to clarify the exact mechanism responsible for the regeneration process. Recent research has focused on factors that regulate miRNA sorting into exosomes. Still, further research is necessary to understand the significance of other components, such as IncRNA, cirRNA, lipids, and proteins, and their effects on the physiological functions of exosomes. Interestingly, recent research has proposed the use of exosome-mimicking nanovesicles (NVs) as a substitute for natural MSC-derived exosomes, as they can be produced in larger quantities. However, a suitable method for producing them in a highly efficient manner is still lacking [196]. Thus, a comprehensive examination should be carried out before applying MSC-derived exosomes in clinical trials to guarantee their safety.

13. Conclusions

In conclusion, both MSCs and MSC-derived exosomes have shown diverse functions in the treatment of corneal diseases. Utilizing exosomes as a cell-free treatment option may reduce the risk associated with cell therapy. Nevertheless, our comprehension of the

mechanisms and operations of MSC-derived exosomes is insufficient and requires further study. Additional research is necessary to determine the optimal dosage, administration route, interval, and mechanism of action of these exosomes before they can be used in clinical trials. The low risk of tumor formation and their ability to regulate cell fate make MSC-derived exosomes a promising approach in regenerative medicine.

Author Contributions: Conceptualization and methodology, J.-Y.K., H.L., E.-A.Y., H.-S.C. and B.B.; writing, B.B.; data acquisition B.B., S.-H.O., C.-M.K., Y.-J.Y. and Y.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (MEST) (NRF-2022R1F1A1073895), and by a grant (2023IL0023, 2023IP0089) from the Asan Institute for Life Sciences, Seoul, Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

MSCs	Mesenchymal stem cells
LSC	Limbal stem cell
iPSCs	Induced pluripotent stem cells
TSG-6	Tumor necrosis factor-stimulated gene/protein-6
ECM	Extracellular matrix
SPARS	Secreted protein acidic and rich in cysteine
PLGA	Polylactic-co-glycolic acid
NVs	Nanovesicles
TNF-α	Tumor necrosis factor-alpha
IL-1β	Interleukin 1 β
HMGB1	High mobility group Box 1
PDLSCs	Periodontal ligament stem cells
DPSCs	Dental pulp stem cells
SEC	Size exclusion chromatography
MCP-1	Monocyte chemoattractant protein-1
MMP2	Matrix metallopeptidase 2
ADSCs	Adipose derived stem cells
TRAM2	Translocation-associated membrane protein 2
CK3	Cytokeratin 3
GSK	Glycogen synthase kinase 3
TGF β	Transforming growth factor β
ROCK	Rho-associated protein kinase
MVBs	Multicellular bodies
ILVs	Intraluminal vesicles
ESCRT	Endosomal sorting complexes required for protein ESCRT
EAU	Experimental autoimmune uveitis
KCM	Keratocyte-conditioned medium
NVs	Nanovesicles
VEGF	Vascularly endothelial growth factor
PDGF	Platelet-derived growth factor
HGF	Hepatocyte growth factor
HCEC	Human corneal endothelial cell
DSAEK	Descemet's stripping automated endothelial keratoplasty
DMEK	Descemet's membrane endothelial keratoplasty

References

1. Aghamollaei, H.; Pirhadi, S.; Shafiee, S.; Sehri, M.; Goodarzi, V.; Jadidi, K. Application of polymethylmethacrylate, acrylic, and silicone in ophthalmology. In *Materials for Biomedical Engineering*; Elsevier: Berlin/Heidelberg, Germany, 2019; pp. 507–554.
2. Eslani, M.; Baradaran-Rafii, A.; Movahedan, A.; Djalilian, A.R. The ocular surface chemical burns. *J. Ophthalmol.* **2014**, *2014*, 196827. [CrossRef] [PubMed]
3. Reinhard, T.; Spelsberg, H.; Henke, L.; Kontopoulos, T.; Enczmann, J.; Wernet, P.; Berschick, P.; Sundmacher, R.; Böhringer, D. Long-term results of allogeneic penetrating limbo-keratoplasty in total limbal stem cell deficiency. *Ophthalmology* **2004**, *111*, 775–782. [CrossRef] [PubMed]
4. Eslani, M.; Haq, Z.; Movahedan, A.; Moss, A.; Baradaran-Rafii, A.; Mogilishetty, G.; Holland, E.J.; Djalilian, A.R. Late acute rejection after allograft limbal stem cell transplantation: Evidence for long-term donor survival. *Cornea* **2017**, *36*, 26. [CrossRef] [PubMed]
5. Tan, D.T.; Dart, J.K.; Holland, E.J.; Kinoshita, S. Corneal transplantation. *Lancet* **2012**, *379*, 1749–1761. [CrossRef]
6. Flaxman, S.R.; Bourne, R.R.; Resnikoff, S.; Ackland, P.; Braithwaite, T.; Cicinelli, M.V.; Das, A.; Jonas, J.B.; Keeffe, J.; Kempen, J.H. Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis. *Lancet Glob. Health* **2017**, *5*, e1221–e1234. [CrossRef]
7. Avadhanam, V.S.; Liu, C.S. A brief review of Boston type-1 and osteo-odonto keratoprostheses. *Br. J. Ophthalmol.* **2015**, *99*, 878–887. [CrossRef]
8. DelMonte, D.W.; Kim, T. Anatomy and physiology of the cornea. *J. Cataract Refract. Surg.* **2011**, *37*, 588–598. [CrossRef]
9. Meek, K.M.; Knupp, C. Corneal structure and transparency. *Prog. Retin. Eye Res.* **2015**, *49*, 1–16. [CrossRef]
10. Chong, E.-M.; Dana, M.R. Graft failure IV. Immunologic mechanisms of corneal transplant rejection. *Int. Ophthalmol.* **2008**, *28*, 209–222. [CrossRef]
11. Gain, P.; Jullienne, R.; He, Z.; Aldossary, M.; Acquart, S.; Cognasse, F.; Thuret, G. Global survey of corneal transplantation and eye banking. *JAMA Ophthalmol.* **2016**, *134*, 167–173. [CrossRef]
12. Fouladi, N.; Parker, M.; Kennedy, V.; Binley, K.; McCloskey, L.; Loader, J.; Kelleher, M.; Mitrophanous, K.A.; Stout, J.T.; Ellis, S. Safety and efficacy of OXB-202, a genetically engineered tissue therapy for the prevention of rejection in high-risk corneal transplant patients. *Hum. Gene Ther.* **2018**, *29*, 687–698. [CrossRef]
13. Liu, S.; Wong, Y.L.; Walkden, A. Current perspectives on corneal transplantation. *Clin. Ophthalmol.* **2022**, *16*, 631–646. [CrossRef]
14. Shafiq, M.A.; Gemeinhart, R.A.; Yue, B.Y.; Djalilian, A.R. Decellularized human cornea for reconstructing the corneal epithelium and anterior stroma. *Tissue Eng. Part C Methods* **2012**, *18*, 340–348. [CrossRef]
15. Taylor, A.W. Ocular immune privilege and transplantation. *Front. Immunol.* **2016**, *7*, 37. [CrossRef]
16. Sikora, B.; Skubis-Sikora, A.; Prusek, A.; Gola, J. Paracrine activity of adipose derived stem cells on limbal epithelial stem cells. *Sci. Rep.* **2021**, *11*, 19956. [CrossRef]
17. Alvites, R.; Branquinho, M.; Sousa, A.C.; Lopes, B.; Sousa, P.; Maurício, A.C. Mesenchymal stem/stromal cells and their paracrine activity—Immunomodulation mechanisms and how to influence the therapeutic potential. *Pharmaceutics* **2022**, *14*, 381. [CrossRef]
18. Bhujel, B.; Shin, H.-E.; Choi, D.-J.; Han, I. Mesenchymal stem cell-derived exosomes and intervertebral disc regeneration. *Int. J. Mol. Sci.* **2022**, *23*, 7306. [CrossRef]
19. Hefley, B.S.; Deighan, C.; Vasini, B.; Khan, A.; Hjortdal, J.; Riaz, K.M.; Liu, Y.; Karamichos, D. Revealing the presence of tear extracellular vesicles in Keratoconus. *Exp. Eye Res.* **2022**, *224*, 109242. [CrossRef]
20. Du, Y.; Funderburgh, M.L.; Mann, M.M.; SundarRaj, N.; Funderburgh, J.L. Multipotent stem cells in human corneal stroma. *Stem Cells* **2005**, *23*, 1266–1275. [CrossRef]
21. Bajada, S.; Mazakova, I.; Richardson, J.B.; Ashammakhi, N. Updates on stem cells and their applications in regenerative medicine. *J. Tissue Eng. Regen. Med.* **2008**, *2*, 169–183. [CrossRef]
22. Guillot, P.V.; Gotherstrom, C.; Chan, J.; Kurata, H.; Fisk, N.M. Human first-trimester fetal MSC express pluripotency markers and grow faster and have longer telomeres than adult MSC. *Stem Cells* **2007**, *25*, 646–654. [PubMed]
23. Ho, P.-J.; Yen, M.-L.; Tang, B.-C.; Chen, C.-T.; Yen, B.L. H_2O_2 accumulation mediates differentiation capacity alteration, but not proliferative decline, in senescent human fetal mesenchymal stem cells. *Antioxid. Redox Signal.* **2013**, *18*, 1895–1905. [CrossRef] [PubMed]
24. Yen, M.-L.; Hou, C.-H.; Peng, K.-Y.; Tseng, P.-C.; Jiang, S.-S.; Shun, C.-T.; Chen, Y.-C.; Kuo, M.-L. Efficient derivation and concise gene expression profiling of human embryonic stem cell-derived mesenchymal progenitors (EMPs). *Cell Transplant.* **2011**, *20*, 1529–1545. [CrossRef]
25. Kang, S.K.; Shin, I.S.; Ko, M.S.; Jo, J.Y.; Ra, J.C. Journey of mesenchymal stem cells for homing: Strategies to enhance efficacy and safety of stem cell therapy. *Stem Cells Int.* **2012**, *2012*, 342968. [CrossRef]
26. Wang, M.; Yuan, Q.; Xie, L. Mesenchymal stem cell-based immunomodulation: Properties and clinical application. *Stem Cells Int.* **2018**, *2018*, 3057624. [CrossRef] [PubMed]
27. De Ugarte, D.A.; Alfonso, Z.; Zuk, P.A.; Elbarbary, A.; Zhu, M.; Ashjian, P.; Benhaim, P.; Hedrick, M.H.; Fraser, J.K. Differential expression of stem cell mobilization-associated molecules on multi-lineage cells from adipose tissue and bone marrow. *Immunol. Lett.* **2003**, *89*, 267–270. [CrossRef]

28. Barlow, S.; Brooke, G.; Chatterjee, K.; Price, G.; Pelekanos, R.; Rossetti, T.; Doody, M.; Venter, D.; Pain, S.; Gilshenan, K. Comparison of human placenta-and bone marrow–derived multipotent mesenchymal stem cells. *Stem Cells Dev.* **2008**, *17*, 1095–1108. [CrossRef]
29. Wang, L.; Ott, L.; Seshareddy, K.; Weiss, M.L.; Detamore, M.S. Musculoskeletal tissue engineering with human umbilical cord mesenchymal stromal cells. *Regen. Med.* **2011**, *6*, 95–109. [CrossRef]
30. Skubis, A.; Gola, J.; Sikora, B.; Hybiak, J.; Paul-Samojedny, M.; Mazurek, U.; Łos, M.J. Impact of antibiotics on the proliferation and differentiation of human adipose-derived mesenchymal stem cells. *Int. J. Mol. Sci.* **2017**, *18*, 2522. [CrossRef]
31. Van Harmelen, V.; Röhrig, K.; Hauner, H. Comparison of proliferation and differentiation capacity of human adipocyte precursor cells from the omental and subcutaneous adipose tissue depot of obese subjects. *Metabolism* **2004**, *53*, 632–637.
32. Radtke, C.L.; Nino-Fong, R.; Gonzalez, B.P.E.; Stryhn, H.; McDuffee, L.A. Characterization and osteogenic potential of equine muscle tissue–and periosteal tissue–derived mesenchymal stem cells in comparison with bone marrow–and adipose tissue–derived mesenchymal stem cells. *Am. J. Vet. Res.* **2013**, *74*, 790–800. [CrossRef]
33. Yu, B.; Zhang, X.; Li, X. Exosomes derived from mesenchymal stem cells. *Int. J. Mol. Sci.* **2014**, *15*, 4142–4157. [CrossRef]
34. Baradaran-Rafii, A.; Eslani, M.; Haq, Z.; Shirzadeh, E.; Huvard, M.J.; Djalilian, A.R. Current and upcoming therapies for ocular surface chemical injuries. *Ocul. Surf.* **2017**, *15*, 48–64. [CrossRef]
35. Wong, R.S. Mesenchymal stem cells: Angels or demons? *J. Biomed. Biotechnol.* **2011**, *2011*, 459510. [CrossRef]
36. De Miguel, M.; Fuentes-Julián, S.; Blázquez-Martínez, A.; Pascual, C.; Aller, M.; Arias, J. Immunosuppressive properties of mesenchymal stem cells: Advances and 477 applications. *Curr. Mol. Med.* **2012**, *12*, 574–591. [CrossRef]
37. Bassi, Ê.J.; de Almeida, D.C.; Moraes-Vieira, P.M.M.; Câmara, N.O.S. Exploring the role of soluble factors associated with immune regulatory properties of mesenchymal stem cells. *Stem Cell Rev. Rep.* **2012**, *8*, 329–342. [CrossRef]
38. Fiorina, P.; Jurewicz, M.; Augello, A.; Vergani, A.; Dada, S.; La Rosa, S.; Selig, M.; Godwin, J.; Law, K.; Placidi, C. Immunomodulatory function of bone marrow-derived mesenchymal stem cells in experimental autoimmune type 1 diabetes. *J. Immunol.* **2009**, *183*, 993–1004. [CrossRef]
39. English, K.; Ryan, J.; Tobin, L.; Murphy, M.; Barry, F.P.; Mahon, B.P. Cell contact, prostaglandin E2 and transforming growth factor beta 1 play non-redundant roles in human mesenchymal stem cell induction of CD4+ CD25Highforkhead box P3+ regulatory T cells. *Clin. Exp. Immunol.* **2009**, *156*, 149–160. [CrossRef]
40. Bussche, L.; Harman, R.M.; Syracuse, B.A.; Plante, E.L.; Lu, Y.-C.; Curtis, T.M.; Ma, M.; Van de Walle, G.R. Microencapsulated equine mesenchymal stromal cells promote cutaneous wound healing in vitro. *Stem Cell Res. Ther.* **2015**, *6*, 66. [CrossRef]
41. Xu, K.; Cantu, D.A.; Fu, Y.; Kim, J.; Zheng, X.; Hematti, P.; Kao, W.J. Thiol-ene Michael-type formation of gelatin/poly (ethylene glycol) biomatrices for three-dimensional mesenchymal stromal/stem cell administration to cutaneous wounds. *Acta Biomater.* **2013**, *9*, 8802–8814. [CrossRef]
42. Sardesai, V.S.; Shafiee, A.; Fisk, N.M.; Pelekanos, R.A. Avoidance of maternal cell contamination and overgrowth in isolating fetal chorionic villi mesenchymal stem cells from human term placenta. *Stem Cells Transl. Med.* **2017**, *6*, 1070–1084. [CrossRef] [PubMed]
43. Freitas, J.; Santos, S.G.; Gonçalves, R.M.; Teixeira, J.H.; Barbosa, M.A.; Almeida, M.I. Genetically engineered-MSC therapies for non-unions, delayed unions and critical-size bone defects. *Int. J. Mol. Sci.* **2019**, *20*, 3430. [CrossRef] [PubMed]
44. Chikenji, T.S.; Saito, Y.; Konari, N.; Nakano, M.; Mizue, Y.; Otani, M.; Fujimiya, M. p16INK4A-expressing mesenchymal stromal cells restore the senescence–clearance–regeneration sequence that is impaired in chronic muscle inflammation. *EBioMedicine* **2019**, *44*, 86–97. [CrossRef] [PubMed]
45. Grafe, I.; Alexander, S.; Peterson, J.R.; Snider, T.N.; Levi, B.; Lee, B.; Mishina, Y. TGF-β family signaling in mesenchymal differentiation. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a022202. [CrossRef]
46. Willerth, S.M.; Sakiyama-Elbert, S.E. Combining stem cells and biomaterial scaffolds for constructing tissues and cell delivery. *StemJournal* **2019**, *1*, 1–25. [CrossRef]
47. Lan, Y.; Kodati, S.; Lee, H.S.; Omoto, M.; Jin, Y.; Chauhan, S.K. Kinetics and function of mesenchymal stem cells in corneal injury. *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 3638–3644. [CrossRef]
48. Ye, J.; Yao, K.; Kim, J. Mesenchymal stem cell transplantation in a rabbit corneal alkali burn model: Engraftment and involvement in wound healing. *Eye* **2006**, *20*, 482–490. [CrossRef]
49. Karp, J.M.; Teo, G.S.L. Mesenchymal stem cell homing: The devil is in the details. *Cell Stem Cell* **2009**, *4*, 206–216. [CrossRef]
50. Reinshagen, H.; Auw-Haedrich, C.; Sorg, R.V.; Boehringer, D.; Eberwein, P.; Schwartzkopff, J.; Sundmacher, R.; Reinhard, T. Corneal surface reconstruction using adult mesenchymal stem cells in experimental limbal stem cell deficiency in rabbits. *Acta Ophthalmol.* **2011**, *89*, 741–748. [CrossRef]
51. Yao, L.; Li, Z.-R.; Su, W.-R.; Li, Y.-P.; Lin, M.-L.; Zhang, W.-X.; Liu, Y.; Wan, Q.; Liang, D. Role of mesenchymal stem cells on cornea wound healing induced by acute alkali burn. *PLoS ONE* **2012**, *7*, e30842. [CrossRef]
52. Coulson-Thomas, V.J.; Caterson, B.; Kao, W.W.-Y. Transplantation of human umbilical mesenchymal stem cells cures the corneal defects of mucopolysaccharidosis VII mice. *Stem Cells* **2013**, *31*, 2116–2126.
53. Jiang, T.-S.; Cai, L.; Ji, W.-Y.; Hui, Y.-N.; Wang, Y.-S.; Hu, D.; Zhu, J. Reconstruction of the corneal epithelium with induced marrow mesenchymal stem cells in rats. *Mol. Vis.* **2010**, *16*, 1304.

54. Roddy, G.W.; Oh, J.Y.; Lee, R.H.; Bartosh, T.J.; Ylostalo, J.; Coble, K.; Rosa Jr, R.H.; Prockop, D.J. Action at a distance: Systemically administered adult stem/progenitor cells (MSCs) reduce inflammatory damage to the cornea without engraftment and primarily by secretion of TNF-α stimulated gene/protein 6. *Stem Cells* **2011**, *29*, 1572–1579.
55. Lee, R.H.; Pulin, A.A.; Seo, M.J.; Kota, D.J.; Ylostalo, J.; Larson, B.L.; Semprun-Prieto, L.; Delafontaine, P.; Prockop, D.J. Intravenous hMSCs improve myocardial infarction in mice because cells embolized in lung are activated to secrete the anti-inflammatory protein TSG-6. *Cell Stem Cell* **2009**, *5*, 54–63. [CrossRef]
56. Park, S.H.; Kim, K.W.; Chun, Y.S.; Kim, J.C. Human mesenchymal stem cells differentiate into keratocyte-like cells in keratocyte-conditioned medium. *Exp. Eye Res.* **2012**, *101*, 16–26. [CrossRef]
57. Ma, Y.; Xu, Y.; Xiao, Z.; Yang, W.; Zhang, C.; Song, E.; Du, Y.; Li, L. Reconstruction of chemically burned rat corneal surface by bone marrow–derived human mesenchymal stem cells. *Stem Cells* **2006**, *24*, 315–321. [CrossRef]
58. Askenasy, N.; Zorina, T.; Farkas, D.L.; Shalit, I. Transplanted hematopoietic cells seed in clusters in recipient bone marrow in vivo. *Stem Cells* **2002**, *20*, 301–310. [CrossRef]
59. Eslani, M.; Putra, I.; Shen, X.; Hamouie, J.; Tadepalli, A.; Anwar, K.N.; Kink, J.A.; Ghassemi, S.; Agnihotri, G.; Reshetylo, S. Cornea-derived mesenchymal stromal cells therapeutically modulate macrophage immunophenotype and angiogenic function. *Stem Cells* **2018**, *36*, 775–784. [CrossRef]
60. Yao, L.; Bai, H. Mesenchymal stem cells and corneal reconstruction. *Mol. Vis.* **2013**, *19*, 2237.
61. Liu, H.; Zhang, J.; Liu, C.-Y.; Wang, I.-J.; Sieber, M.; Chang, J.; Jester, J.V.; Kao, W.W. Cell therapy of congenital corneal diseases with umbilical mesenchymal stem cells: Lumican null mice. *PLoS ONE* **2010**, *5*, e10707. [CrossRef]
62. Mittal, S.K.; Omoto, M.; Amouzegar, A.; Sahu, A.; Rezazadeh, A.; Katikireddy, K.R.; Shah, D.I.; Sahu, S.K.; Chauhan, S.K. Restoration of corneal transparency by mesenchymal stem cells. *Stem Cell Rep.* **2016**, *7*, 583–590. [CrossRef] [PubMed]
63. Liu, H.; Zhang, J.; Liu, C.Y.; Hayashi, Y.; Kao, W.W.Y. Bone marrow mesenchymal stem cells can differentiate and assume corneal keratocyte phenotype. *J. Cell. Mol. Med.* **2012**, *16*, 1114–1124. [CrossRef] [PubMed]
64. Galindo, S.; Herreras, J.M.; López-Paniagua, M.; Rey, E.; de la Mata, A.; Plata-Cordero, M.; Calonge, M.; Nieto-Miguel, T. Therapeutic effect of human adipose tissue-derived mesenchymal stem cells in experimental corneal failure due to limbal stem cell niche damage. *Stem Cells* **2017**, *35*, 2160–2174. [CrossRef] [PubMed]
65. Nieto-Miguel, T.; Galindo, S.; Reinoso, R.; Corell, A.; Martino, M.; Perez-Simon, J.A.; Calonge, M. In vitro simulation of corneal epithelium microenvironment induces a corneal epithelial-like cell phenotype from human adipose tissue mesenchymal stem cells. *Curr. Eye Res.* **2013**, *38*, 933–944. [CrossRef] [PubMed]
66. Setiawan, M.; Tan, X.-W.; Goh, T.-W.; Yam, G.H.-F.; Mehta, J.S. Inhibiting glycogen synthase kinase-3 and transforming growth factor-β signaling to promote epithelial transition of human adipose mesenchymal stem cells. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 1381–1388. [CrossRef]
67. Gu, S.; Xing, C.; Han, J.; Tso, M.O.; Hong, J. Differentiation of rabbit bone marrow mesenchymal stem cells into corneal epithelial cells in vivo and ex vivo. *Mol. Vis.* **2009**, *15*, 99.
68. Calonge, M.; Pérez, I.; Galindo, S.; Nieto-Miguel, T.; López-Paniagua, M.; Fernández, I.; Alberca, M.; García-Sancho, J.; Sánchez, A.; Herreras, J.M. A proof-of-concept clinical trial using mesenchymal stem cells for the treatment of corneal epithelial stem cell deficiency. *Transl. Res.* **2019**, *206*, 18–40. [CrossRef]
69. Cejka, C.; Holan, V.; Trosan, P.; Zajicova, A.; Javorkova, E.; Cejkova, J. The favorable effect of mesenchymal stem cell treatment on the antioxidant protective mechanism in the corneal epithelium and renewal of corneal optical properties changed after alkali burns. *Oxidative Med. Cell. Longev.* **2016**, *2016*, 5843809. [CrossRef]
70. Soleimanifar, F.; Mortazavi, Y.; Nadri, S.; Soleimani, M. Conjunctiva derived mesenchymal stem cell (CJMSCs) as a potential platform for differentiation into corneal epithelial cells on bioengineered electrospun scaffolds. *J. Biomed. Mater. Res. Part A* **2017**, *105*, 2703–2711. [CrossRef]
71. Păunescu, V.; Deak, E.; Herman, D.; Siska, I.R.; Tănasie, G.; Bunu, C.; Anghel, S.; Tatu, C.A.; Oprea, T.I.; Henschler, R. In vitro differentiation of human mesenchymal stem cells to epithelial lineage. *J. Cell. Mol. Med.* **2007**, *11*, 502–508. [CrossRef]
72. Syed-Picard, F.N.; Du, Y.; Lathrop, K.L.; Mann, M.M.; Funderburgh, M.L.; Funderburgh, J.L. Dental pulp stem cells: A new cellular resource for corneal stromal regeneration. *Stem Cells Transl. Med.* **2015**, *4*, 276–285. [CrossRef] [PubMed]
73. Du, Y.; Roh, D.S.; Funderburgh, M.L.; Mann, M.M.; Marra, K.G.; Rubin, J.P.; Li, X.; Funderburgh, J.L. Adipose-derived stem cells differentiate to keratocytes in vitro. *Mol. Vis.* **2010**, *16*, 2680.
74. Ma, X.-Y.; Bao, H.-J.; Cui, L.; Zou, J. The graft of autologous adipose-derived stem cells in the corneal stromal after mechanic damage. *PLoS ONE* **2013**, *8*, e76103. [CrossRef]
75. Yam, G.H.F.; Teo, E.P.W.; Setiawan, M.; Lovatt, M.J.; Yusoff, N.Z.B.M.; Fuest, M.; Goh, B.T.; Mehta, J.S. Postnatal periodontal ligament as a novel adult stem cell source for regenerative corneal cell therapy. *J. Cell. Mol. Med.* **2018**, *22*, 3119–3132. [CrossRef]
76. Yamashita, K.; Inagaki, E.; Hatou, S.; Higa, K.; Ogawa, A.; Miyashita, H.; Tsubota, K.; Shimmura, S. Corneal endothelial regeneration using mesenchymal stem cells derived from human umbilical cord. *Stem Cells Dev.* **2018**, *27*, 1097–1108. [CrossRef]
77. Li, W.; Hayashida, Y.; Chen, Y.-T.; Tseng, S.C. Niche regulation of corneal epithelial stem cells at the limbus. *Cell Res.* **2007**, *17*, 26–36. [CrossRef]
78. Farjo, A.; McDermott, M.; Soong, H.K. Corneal anatomy, physiology, and wound healing. *Ophthalmology* **2009**, *44*, 203–208.

79. Movahedan, A.; Majdi, M.; Afsharkhamseh, N.; Sagha, H.M.; Saadat, N.S.; Shalileh, K.; Milani, B.Y.; Ying, H.; Djalilian, A.R. Notch inhibition during corneal epithelial wound healing promotes migration. *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 7476–7483. [CrossRef]
80. Afsharkhamseh, N.; Movahedan, A.; Gidfar, S.; Huvard, M.; Wasielewski, L.; Milani, B.Y.; Eslani, M.; Djalilian, A.R. Stability of limbal stem cell deficiency after mechanical and thermal injuries in mice. *Exp. Eye Res.* **2016**, *145*, 88–92. [CrossRef]
81. Dua, H.S.; Shanmuganathan, V.; Powell-Richards, A.; Tighe, P.; Joseph, A. Limbal epithelial crypts: A novel anatomical structure and a putative limbal stem cell niche. *Br. J. Ophthalmol.* **2005**, *89*, 529–532. [CrossRef]
82. Phinney, D.G.; Prockop, D.J. Concise review: Mesenchymal stem/multipotent stromal cells: The state of transdifferentiation and modes of tissue repair—Current views. *Stem Cells* **2007**, *25*, 2896–2902. [CrossRef] [PubMed]
83. Sikora, B.; Skubis-Sikora, A.; Kimsa-Furdzik, M.; Ciszek, W.; Kostrzewski, M.; Stojko, J.; Mazurek, U.; Gola, J. Adipose-derived stem cells undergo differentiation after co-culture with porcine limbal epithelial stem cells. *Stem Cell Res.* **2019**, *41*, 101609. [CrossRef] [PubMed]
84. Nieto-Nicolau, N.; Martínez-Conesa, E.M.; Fuentes-Julián, S.; Arnalich-Montiel, F.; García-Tuñón, I.; De Miguel, M.P.; Casaroli-Marano, R.P. Priming human adipose-derived mesenchymal stem cells for corneal surface regeneration. *J. Cell. Mol. Med.* **2021**, *25*, 5124–5137. [CrossRef] [PubMed]
85. Teng, N.-Y.; Liu, Y.-S.; Wu, H.-H.; Liu, Y.-A.; Ho, J.H.; Lee, O.K.-S. Promotion of mesenchymal-to-epithelial transition by Rac1 inhibition with small molecules accelerates hepatic differentiation of mesenchymal stromal cells. *Tissue Eng. Part A* **2015**, *21*, 1444–1454. [CrossRef] [PubMed]
86. Ghosh, A.; Degyatoreva, N.; Kukielski, C.; Story, S.; Bhaduri, S.; Maiti, K.; Nahar, S.; Ray, A.; Arya, D.P.; Maiti, S. Targeting miRNA by tunable small molecule binders: Peptidic aminosugar mediated interference in miR-21 biogenesis reverts epithelial to mesenchymal transition. *MedChemComm* **2018**, *9*, 1147–1154. [CrossRef]
87. Rama, P.; Ferrari, G.; Pellegrini, G. Cultivated limbal epithelial transplantation. *Curr. Opin. Ophthalmol.* **2017**, *28*, 387–389. [CrossRef]
88. Ramachandran, C.; Basu, S.; Sangwan, V.S.; Balasubramanian, D. Concise review: The coming of age of stem cell treatment for corneal surface damage. *Stem Cells Transl. Med.* **2014**, *3*, 1160–1168. [CrossRef]
89. Guo, T.; Wang, W.; Zhang, J.; Chen, X.; Li, B.; Li, L. Experimental study on repairing damage of corneal surface by mesenchymal stem cells transplantation. *[Zhonghua Yan Ke Za Zhi] Chin. J. Ophthalmol.* **2006**, *42*, 246–250.
90. West-Mays, J.A.; Dwivedi, D.J. The keratocyte: Corneal stromal cell with variable repair phenotypes. *Int. J. Biochem. Cell Biol.* **2006**, *38*, 1625–1631. [CrossRef]
91. Basu, S.; Hertsenberg, A.J.; Funderburgh, M.L.; Burrow, M.K.; Mann, M.M.; Du, Y.; Lathrop, K.L.; Syed-Picard, F.N.; Adams, S.M.; Birk, D.E. Human limbal biopsy–derived stromal stem cells prevent corneal scarring. *Sci. Transl. Med.* **2014**, *6*, 266ra172. [CrossRef]
92. Kureshi, A.K.; Funderburgh, J.L.; Daniels, J.T. Human Corneal Stromal Stem Cells Exhibit Survival Capacity Following Isolation From Stored Organ–Culture Corneas. *Investig. Ophthalmol. Vis. Sci.* **2014**, *55*, 7583–7588. [CrossRef]
93. Du, Y.; Carlson, E.C.; Funderburgh, M.L.; Birk, D.E.; Pearlman, E.; Guo, N.; Kao, W.W.-Y.; Funderburgh, J.L. Stem cell therapy restores transparency to defective murine corneas. *Stem Cells* **2009**, *27*, 1635–1642. [CrossRef]
94. Hertsenberg, A.J.; Shojaati, G.; Funderburgh, M.L.; Mann, M.M.; Du, Y.; Funderburgh, J.L. Corneal stromal stem cells reduce corneal scarring by mediating neutrophil infiltration after wounding. *PLoS ONE* **2017**, *12*, e0171712. [CrossRef] [PubMed]
95. Funderburgh, J.L.; Funderburgh, M.L.; Du, Y. Stem cells in the limbal stroma. *Ocul. Surf.* **2016**, *14*, 113–120. [CrossRef]
96. Wu, J.; Du, Y.; Watkins, S.C.; Funderburgh, J.L.; Wagner, W.R. The engineering of organized human corneal tissue through the spatial guidance of corneal stromal stem cells. *Biomaterials* **2012**, *33*, 1343–1352. [CrossRef] [PubMed]
97. Mansoor, H.; Ong, H.S.; Riau, A.K.; Stanzel, T.P.; Mehta, J.S.; Yam, G.H.-F. Current trends and future perspective of mesenchymal stem cells and exosomes in corneal diseases. *Int. J. Mol. Sci.* **2019**, *20*, 2853. [CrossRef] [PubMed]
98. Yam, G.H.-F.; Peh, G.S.-L.; Singhal, S.; Goh, B.-T.; Mehta, J.S. Dental stem cells: A future asset of ocular cell therapy. *Expert Rev. Mol. Med.* **2015**, *17*, e20. [CrossRef]
99. Seo, B.-M.; Miura, M.; Gronthos, S.; Bartold, P.M.; Batouli, S.; Brahim, J.; Young, M.; Robey, P.G.; Wang, C.Y.; Shi, S. Investigation of multipotent postnatal stem cells from human periodontal ligament. *Lancet* **2004**, *364*, 149–155. [CrossRef]
100. Trubiani, O.; Zalzal, S.F.; Paganelli, R.; Marchisio, M.; Giancola, R.; Pizzicannella, J.; Bühring, H.J.; Piattelli, M.; Caputi, S.; Nanci, A. Expression profile of the embryonic markers nanog, OCT-4, SSEA-1, SSEA-4, and frizzled-9 receptor in human periodontal ligament mesenchymal stem cells. *J. Cell. Physiol.* **2010**, *225*, 123–131. [CrossRef]
101. Tomokiyo, A.; Maeda, H.; Fujii, S.; Monnouchi, S.; Wada, N.; Kono, K.; Yamamoto, N.; Koori, K.; Teramatsu, Y.; Akamine, A. A multipotent clonal human periodontal ligament cell line with neural crest cell phenotypes promotes neurocytic differentiation, migration, and survival. *J. Cell. Physiol.* **2012**, *227*, 2040–2050. [CrossRef]
102. Kawanabe, N.; Murata, S.; Murakami, K.; Ishihara, Y.; Hayano, S.; Kurosaka, H.; Kamioka, H.; Takano-Yamamoto, T.; Yamashiro, T. Isolation of multipotent stem cells in human periodontal ligament using stage-specific embryonic antigen-4. *Differentiation* **2010**, *79*, 74–83. [CrossRef]
103. Del Barrio, J.L.A.; El Zarif, M.; de Miguel, M.P.; Azaar, A.; Makdissy, N.; Harb, W.; El Achkar, I.; Arnalich-Montiel, F.; Alió, J.L. Cellular therapy with human autologous adipose-derived adult stem cells for advanced keratoconus. *Cornea* **2017**, *36*, 952–960. [CrossRef]

104. Alió, J.L.; Del Barrio, J.L.A.; El Zarif, M.; Azaar, A.; Makdissy, N.; Khalil, C.; Harb, W.; El Achkar, I.; Jawad, Z.A.; De Miguel, M.P. Regenerative surgery of the corneal stroma for advanced keratoconus: 1-year outcomes. *Am. J. Ophthalmol.* **2019**, *203*, 53–68. [CrossRef]
105. Arnalich-Montiel, F.; Pastor, S.; Blazquez-Martinez, A.; Fernandez-Delgado, J.; Nistal, M.; Alio, J.L.; De Miguel, M.P. Adipose-derived stem cells are a source for cell therapy of the corneal stroma. *Stem Cells* **2008**, *26*, 570–579. [CrossRef]
106. Goldstein, A.S.; Janson, B.J.; Skeie, J.M.; Ling, J.J.; Greiner, M.A. The effects of diabetes mellitus on the corneal endothelium: A review. *Surv. Ophthalmol.* **2020**, *65*, 438–450. [CrossRef]
107. Nakahara, M.; Okumura, N.; Kay, E.P.; Hagiya, M.; Imagawa, K.; Hosoda, Y.; Kinoshita, S.; Koizumi, N. Corneal endothelial expansion promoted by human bone marrow mesenchymal stem cell-derived conditioned medium. *PLoS ONE* **2013**, *8*, e69009. [CrossRef]
108. Afshari, N.A.; Pittard, A.B.; Siddiqui, A.; Klintworth, G.K. Clinical study of Fuchs corneal endothelial dystrophy leading to penetrating keratoplasty: A 30-year experience. *Arch. Ophthalmol.* **2006**, *124*, 777–780. [CrossRef]
109. Melles, G.R.; San Ong, T.; Ververs, B.; Van der Wees, J. Preliminary clinical results of Descemet membrane endothelial keratoplasty. *Am. J. Ophthalmol.* **2008**, *145*, 222–227 e221. [CrossRef]
110. Okumura, N.; Koizumi, N.; Ueno, M.; Sakamoto, Y.; Takahashi, H.; Tsuchiya, H.; Hamuro, J.; Kinoshita, S. ROCK inhibitor converts corneal endothelial cells into a phenotype capable of regenerating in vivo endothelial tissue. *Am. J. Pathol.* **2012**, *181*, 268–277. [CrossRef]
111. Kinoshita, S.; Koizumi, N.; Ueno, M.; Okumura, N.; Imai, K.; Tanaka, H.; Yamamoto, Y.; Nakamura, T.; Inatomi, T.; Bush, J. Injection of cultured cells with a ROCK inhibitor for bullous keratopathy. *N. Engl. J. Med.* **2018**, *378*, 995–1003. [CrossRef]
112. Chen, S.; Zhu, Q.; Sun, H.; Zhang, Y.; Tighe, S.; Xu, L.; Zhu, Y. Advances in culture, expansion and mechanistic studies of corneal endothelial cells: A systematic review. *J. Biomed. Sci.* **2019**, *26*, 2. [CrossRef] [PubMed]
113. Yamagami, S.; Mimura, T.; Yokoo, S.; Takato, T.; Amano, S. Isolation of human corneal endothelial cell precursors and construction of cell sheets by precursors. *Cornea* **2006**, *25*, S90–S92. [CrossRef] [PubMed]
114. Liu, X.-W.; Zhao, J.-L. Transplantation of autologous bone marrow mesenchymal stem cells for the treatment of corneal endothelium damages in rabbits. *[Zhonghua Yan Ke Za Zhi] Chin. J. Ophthalmol.* **2007**, *43*, 540–545.
115. Shen, L.; Sun, P.; Zhang, C.; Yang, L.; Du, L.; Wu, X. Therapy of corneal endothelial dysfunction with corneal endothelial cell-like cells derived from skin-derived precursors. *Sci. Rep.* **2017**, *7*, 13400. [CrossRef]
116. Gutermuth, A.; Maassen, J.; Harnisch, E.; Kuhlen, D.; Sauer-Budge, A.; Skazik-Voogt, C.; Engelmann, K. Descemet's membrane biomimetic microtopography differentiates human mesenchymal stem cells into corneal endothelial-like cells. *Cornea* **2019**, *38*, 110. [CrossRef]
117. Feiertag, E.; Maassen, J.; Mueller, A.; Harnisch, E.; Skazik-Voogt, C.; Engelmann, K.; Schrage, N.; Gutermuth, A. From Cord to Eye: Wharton Jelly-Derived Stem Cells Differentiate Into Corneal Endothelial–Like Cells. *Cornea* **2020**, *39*, 877–885. [CrossRef]
118. Oh, J.Y.; Kim, M.K.; Shin, M.S.; Lee, H.J.; Ko, J.H.; Wee, W.R.; Lee, J.H. The anti-inflammatory and anti-angiogenic role of mesenchymal stem cells in corneal wound healing following chemical injury. *Stem Cells* **2008**, *26*, 1047–1055. [CrossRef]
119. Pinnamaneni, N.; Funderburgh, J.L. Concise review: Stem cells in the corneal stroma. *Stem Cells* **2012**, *30*, 1059–1063. [CrossRef]
120. Song, H.B.; Park, S.Y.; Ko, J.H.; Park, J.W.; Yoon, C.H.; Kim, D.H.; Kim, J.H.; Kim, M.K.; Lee, R.H.; Prockop, D.J. Mesenchymal stromal cells inhibit inflammatory lymphangiogenesis in the cornea by suppressing macrophage in a TSG-6-dependent manner. *Mol. Ther.* **2018**, *26*, 162–172. [CrossRef]
121. Casiraghi, F.; Azzollini, N.; Todeschini, M.; Cavinato, R.; Cassis, P.; Solini, S.; Rota, C.; Morigi, M.; Introna, M.; Maranta, R. Localization of mesenchymal stromal cells dictates their immune or proinflammatory effects in kidney transplantation. *Am. J. Transplant.* **2012**, *12*, 2373–2383. [CrossRef]
122. Oh, J.Y.; Lee, R.H.; Yu, J.M.; Ko, J.H.; Lee, H.J.; Ko, A.Y.; Roddy, G.W.; Prockop, D.J. Intravenous mesenchymal stem cells prevented rejection of allogeneic corneal transplants by aborting the early inflammatory response. *Mol. Ther.* **2012**, *20*, 2143–2152.
123. Niederkorn, J.Y.; Stein-Streilein, J. History and physiology of immune privilege. *Ocul. Immunol. Inflamm.* **2010**, *18*, 19–23. [CrossRef]
124. Casiraghi, F.; Remuzzi, G.; Perico, N. Mesenchymal stromal cells to promote kidney transplantation tolerance. *Curr. Opin. Organ Transplant.* **2014**, *19*, 47–53. [CrossRef]
125. Coulson-Thomas, V.J.; Gesteira, T.F.; Hascall, V.; Kao, W. Umbilical cord mesenchymal stem cells suppress host rejection: The role of the glycocalyx. *J. Biol. Chem.* **2014**, *289*, 23465–23481. [CrossRef]
126. Oh, J.Y.; Kim, M.K.; Ko, J.H.; Lee, H.J.; Lee, J.H.; Wee, W.R. Rat allogeneic mesenchymal stem cells did not prolong the survival of corneal xenograft in a pig-to-rat model. *Vet. Ophthalmol.* **2009**, *12*, 35–40. [CrossRef]
127. Jia, Z.; Jiao, C.; Zhao, S.; Li, X.; Ren, X.; Zhang, L.; Han, Z.C.; Zhang, X. Immunomodulatory effects of mesenchymal stem cells in a rat corneal allograft rejection model. *Exp. Eye Res.* **2012**, *102*, 44–49. [CrossRef]
128. Lukomska, B.; Stanaszek, L.; Zuba-Surma, E.; Legosz, P.; Sarzynska, S.; Drela, K. Challenges and controversies in human mesenchymal stem cell therapy. *Stem Cells Int.* **2019**, *2019*, 9628536. [CrossRef]
129. Mo, M.; Wang, S.; Zhou, Y.; Li, H.; Wu, Y. Mesenchymal stem cell subpopulations: Phenotype, property and therapeutic potential. *Cell. Mol. Life Sci.* **2016**, *73*, 3311–3321. [CrossRef]
130. Anitua, E.; Prado, R. Addressing reproducibility in stem cell and PRP therapies. *Trends Biotechnol.* **2019**, *37*, 340–344. [CrossRef]

131. Dufrane, D. Impact of age on human adipose stem cells for bone tissue engineering. *Cell Transplant.* **2017**, *26*, 1496–1504. [CrossRef]
132. Lohan, P.; Treacy, O.; Griffin, M.D.; Ritter, T.; Ryan, A.E. Anti-donor immune responses elicited by allogeneic mesenchymal stem cells and their extracellular vesicles: Are we still learning? *Front. Immunol.* **2017**, *8*, 1626. [CrossRef] [PubMed]
133. Yang, Y.-H.K. Aging of mesenchymal stem cells: Implication in regenerative medicine. *Regen. Ther.* **2018**, *9*, 120–122. [CrossRef] [PubMed]
134. Wagner, W.; Horn, P.; Castoldi, M.; Diehlmann, A.; Bork, S.; Saffrich, R.; Benes, V.; Blake, J.; Pfister, S.; Eckstein, V. Replicative senescence of mesenchymal stem cells: A continuous and organized process. *PLoS ONE* **2008**, *3*, e2213. [CrossRef] [PubMed]
135. Pamies, D.; Bal-Price, A.; Simeonov, A.; Tagle, D.; Allen, D.; Gerhold, D.; Yin, D.; Pistollato, F.; Inutsuka, T.; Sullivan, K. Good cell culture practice for stem cells and stem-cell-derived models. *Altern. Anim. Exp. ALTEX* **2017**, *34*, 95–132. [CrossRef]
136. King, N.M.; Perrin, J. Ethical issues in stem cell research and therapy. *Stem Cell Res. Ther.* **2014**, *5*, 85. [CrossRef]
137. Lai, R.C.; Tan, S.S.; Teh, B.J.; Sze, S.K.; Arslan, F.; De Kleijn, D.P.; Choo, A.; Lim, S.K. Proteolytic potential of the MSC exosome proteome: Implications for an exosome-mediated delivery of therapeutic proteasome. *Int. J. Proteom.* **2012**, *2012*, 971907. [CrossRef]
138. Chen, T.S.; Lai, R.C.; Lee, M.M.; Choo, A.B.H.; Lee, C.N.; Lim, S.K. Mesenchymal stem cell secretes microparticles enriched in pre-microRNAs. *Nucleic Acids Res.* **2010**, *38*, 215–224.
139. Yin, K.; Wang, S.; Zhao, R.C. Exosomes from mesenchymal stem/stromal cells: A new therapeutic paradigm. *Biomark. Res.* **2019**, *7*, 8.
140. Burrello, J.; Monticone, S.; Gai, C.; Gomez, Y.; Kholia, S.; Camussi, G. Stem cell-derived extracellular vesicles and immune-modulation. *Front. Cell Dev. Biol.* **2016**, *4*, 83. [CrossRef]
141. Lin, K.-J.; Loi, M.-X.; Lien, G.-S.; Cheng, C.-F.; Pao, H.-Y.; Chang, Y.-C.; Ji, A.T.-Q.; Ho, J.H.-C. Topical administration of orbital fat-derived stem cells promotes corneal tissue regeneration. *Stem Cell Res. Ther.* **2013**, *4*, 72. [CrossRef]
142. Trounson, A.; McDonald, C. Stem cell therapies in clinical trials: Progress and challenges. *Cell Stem Cell* **2015**, *17*, 11–22. [CrossRef]
143. Phinney, D.G.; Pittenger, M.F. Concise review: MSC-derived exosomes for cell-free therapy. *Stem Cells* **2017**, *35*, 851–858. [CrossRef]
144. Colombo, M.; Raposo, G.; Théry, C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 255–289. [CrossRef]
145. Pegtel, D.M.; Gould, S.J. Exosomes. *Annu. Rev. Biochem.* **2019**, *88*, 487–514. [CrossRef]
146. Birtwistle, L.; Chen, X.-M.; Pollock, C. Mesenchymal stem cell-derived extracellular vesicles to the rescue of renal injury. *Int. J. Mol. Sci.* **2021**, *22*, 6596. [CrossRef]
147. Sun, S.-J.; Wei, R.; Li, F.; Liao, S.-Y.; Tse, H.-F. Mesenchymal stromal cell-derived exosomes in cardiac regeneration and repair. *Stem Cell Rep.* **2021**, *16*, 1662–1673. [CrossRef]
148. Tiwari, A.; Singh, A.; Verma, S.; Stephenson, S.; Bhowmick, T.; Sangwan, V.S. Mini review: Current trends and understanding of exosome therapeutic potential in corneal diseases. *Front. Pharmacol.* **2021**, *12*, 684712. [CrossRef]
149. Xunian, Z.; Kalluri, R. Biology and therapeutic potential of mesenchymal stem cell-derived exosomes. *Cancer Sci.* **2020**, *111*, 3100–3110. [CrossRef]
150. Nakamura, Y. Biomarkers for immune checkpoint inhibitor-mediated tumor response and adverse events. *Front. Med.* **2019**, *6*, 119. [CrossRef]
151. Yu, L.-L.; Zhu, J.; Liu, J.-X.; Jiang, F.; Ni, W.-K.; Qu, L.-S.; Ni, R.-Z.; Lu, C.-H.; Xiao, M.-B. A comparison of traditional and novel methods for the separation of exosomes from human samples. *BioMed Res. Int.* **2018**, *2018*, 3634563. [CrossRef]
152. Li, M.; Zhang, A.; Li, J.; Zhou, J.; Zheng, Y.; Zhang, C.; Xia, D.; Mao, H.; Zhao, J. Osteoblast/fibroblast coculture derived bioactive ECM with unique matrisome profile facilitates bone regeneration. *Bioact. Mater.* **2020**, *5*, 938–948. [CrossRef] [PubMed]
153. Xie, L.; Chen, Z.; Liu, M.; Huang, W.; Zou, F.; Ma, X.; Tao, J.; Guo, J.; Xia, X.; Lyu, F. MSC-derived exosomes protect vertebral endplate chondrocytes against apoptosis and calcification via the miR-31-5p/ATF6 axis. *Mol. Ther. -Nucleic Acids* **2020**, *22*, 601–614. [CrossRef] [PubMed]
154. Kong, X.; Ding, X.; Li, X.; Gao, S.; Yang, Q. 53 BP 1 suppresses epithelial–mesenchymal transition by downregulating ZEB 1 through micro RNA-200b/429 in breast cancer. *Cancer Sci.* **2015**, *106*, 982–989. [CrossRef] [PubMed]
155. Ailuno, G.; Baldassari, S.; Lai, F.; Florio, T.; Caviglioli, G. Exosomes and extracellular vesicles as emerging theranostic platforms in cancer research. *Cells* **2020**, *9*, 2569. [CrossRef] [PubMed]
156. Sidhom, K.; Obi, P.O.; Saleem, A. A review of exosomal isolation methods: Is size exclusion chromatography the best option? *Int. J. Mol. Sci.* **2020**, *21*, 6466. [CrossRef]
157. Schiel, J.E.; Hage, D.S. Kinetic studies of biological interactions by affinity chromatography. *J. Sep. Sci.* **2009**, *32*, 1507–1522. [CrossRef]
158. Wei, H.; Qian, X.; Xie, F.; Cui, D. Isolation of exosomes from serum of patients with lung cancer: A comparison of the ultra-high speed centrifugation and precipitation methods. *Ann. Transl. Med.* **2021**, *9*, 882. [CrossRef]
159. Sokolova, V.; Ludwig, A.; Hornung, S.; Rotan, O.; Horn, P.; Epple, M.; Giebel, B. Colloids and surfaces B: Biointerfaces Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. *Colloids Surf. B Biointerfaces* **2011**, *87*, 146–150. [CrossRef]
160. Lorincz, Á.; Timár, C.; Marosvári, K.; Veres, D.; Otrokocsi, L.; Kittel, Á.; Ligeti, E. Effect of storage on physical and functional properties of extracellular vesicles derived from neutrophilic granulocytes. *J. Extracell. Vesicles* **2014**, *3*, 25465. [CrossRef]

161. Zhou, H.; Yuen, P.S.; Pisitkun, T.; Gonzales, P.A.; Yasuda, H.; Dear, J.W.; Gross, P.; Knepper, M.A.; Star, R.A. Collection, storage, preservation, and normalization of human urinary exosomes for biomarker discovery. *Kidney Int.* **2006**, *69*, 1471–1476. [CrossRef]
162. Bosch, S.; de Beaurepaire, L.; Allard, M.; Mosser, M.; Heichette, C.; Chrétien, D.; Jegou, D.; Bach, J.-M. Trehalose prevents aggregation of exosomes and cryodamage. *Sci. Rep.* **2016**, *6*, 36162. [CrossRef]
163. Bai, L.; Shao, H.; Wang, H.; Zhang, Z.; Su, C.; Dong, L.; Yu, B.; Chen, X.; Li, X.; Zhang, X. Effects of mesenchymal stem cell-derived exosomes on experimental autoimmune uveitis. *Sci. Rep.* **2017**, *7*, 4323. [CrossRef]
164. Yu, B.; Shao, H.; Su, C.; Jiang, Y.; Chen, X.; Bai, L.; Zhang, Y.; Li, Q.; Zhang, X.; Li, X. Exosomes derived from MSCs ameliorate retinal laser injury partially by inhibition of MCP-1. *Sci. Rep.* **2016**, *6*, 34562. [CrossRef]
165. Zhang, W.; Wang, Y.; Kong, Y. Exosomes derived from mesenchymal stem cells modulate miR-126 to ameliorate hyperglycemia-induced retinal inflammation via targeting HMGB1. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, 294–303. [CrossRef]
166. Lee, J.-K.; Park, S.-R.; Jung, B.-K.; Jeon, Y.-K.; Lee, Y.-S.; Kim, M.-K.; Kim, Y.-G.; Jang, J.-Y.; Kim, C.-W. Exosomes derived from mesenchymal stem cells suppress angiogenesis by down-regulating VEGF expression in breast cancer cells. *PLoS ONE* **2013**, *8*, e84256. [CrossRef]
167. He, C.; Zheng, S.; Luo, Y.; Wang, B. Exosome theranostics: Biology and translational medicine. *Theranostics* **2018**, *8*, 237. [CrossRef]
168. Mead, B.; Tomarev, S. Bone marrow-derived mesenchymal stem cells-derived exosomes promote survival of retinal ganglion cells through miRNA-dependent mechanisms. *Stem Cells Transl. Med.* **2017**, *6*, 1273–1285. [CrossRef]
169. Zhang, X.; Ren, X.; Li, G.; Jiao, C.; Zhang, L.; Zhao, S.; Wang, J.; Han, Z.C.; Li, X. Mesenchymal stem cells ameliorate experimental autoimmune uveoretinitis by comprehensive modulation of systemic autoimmunity. *Investig. Ophthalmol. Vis. Sci.* **2011**, *52*, 3143–3152. [CrossRef]
170. Mead, B.; Berry, M.; Logan, A.; Scott, R.A.; Leadbeater, W.; Scheven, B.A. Stem cell treatment of degenerative eye disease. *Stem Cell Res.* **2015**, *14*, 243–257. [CrossRef]
171. Zhang, X.; Liu, J.; Yu, B.; Ma, F.; Ren, X.; Li, X. Effects of mesenchymal stem cells and their exosomes on the healing of large and refractory macular holes. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2018**, *256*, 2041–2052. [CrossRef]
172. Shen, T.; Zheng, Q.-Q.; Shen, J.; Li, Q.-S.; Song, X.-H.; Luo, H.-B.; Hong, C.-Y.; Yao, K. Effects of adipose-derived mesenchymal stem cell exosomes on corneal stromal fibroblast viability and extracellular matrix synthesis. *Chin. Med. J.* **2018**, *131*, 704–712. [CrossRef]
173. Samaeekia, R.; Rabiee, B.; Putra, I.; Shen, X.; Park, Y.J.; Hematti, P.; Eslani, M.; Djalilian, A.R. Effect of human corneal mesenchymal stromal cell-derived exosomes on corneal epithelial wound healing. *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 5194–5200. [CrossRef] [PubMed]
174. Jiang, Z.; Liu, G.; Meng, F.; Wang, W.; Hao, P.; Xiang, Y.; Wang, Y.; Han, R.; Li, F.; Wang, L. Paracrine effects of mesenchymal stem cells on the activation of keratocytes. *Br. J. Ophthalmol.* **2017**, *101*, 1583–1590. [CrossRef] [PubMed]
175. Sun, X.; Song, W.; Teng, L.; Huang, Y.; Liu, J.; Peng, Y.; Lu, X.; Yuan, J.; Zhao, X.; Zhao, Q. MiRNA 24-3p-rich exosomes functionalized DEGMA-modified hyaluronic acid hydrogels for corneal epithelial healing. *Bioact. Mater.* **2022**, *25*, 640–656. [CrossRef] [PubMed]
176. Tang, Q.; Lu, B.; He, J.; Chen, X.; Fu, Q.; Han, H.; Luo, C.; Yin, H.; Qin, Z.; Lyu, D. Exosomes-loaded thermosensitive hydrogels for corneal epithelium and stroma regeneration. *Biomaterials* **2022**, *280*, 121320. [CrossRef] [PubMed]
177. Jia, Z.; Lv, Y.; Zhang, W.; Zhang, X.; Li, F.; Lu, X.; Zhao, S. Mesenchymal stem cell derived exosomes-based immunological signature in a rat model of corneal allograft rejection therapy. *Front. Biosci.-Landmark* **2022**, *27*, 86. [CrossRef]
178. Liu, X.; Li, X.; Wu, G.; Qi, P.; Zhang, Y.; Liu, Z.; Li, X.; Yu, Y.; Ye, X.; Li, Y. Umbilical cord mesenchymal stem cell-derived small extracellular vesicles deliver miR-21 to promote corneal epithelial wound healing through PTEN/PI3K/Akt pathway. *Stem Cells Int.* **2022**, *2022*, 1252557. [CrossRef]
179. Nuzzi, R.; Buono, L.; Scalabrin, S.; De Iuliis, M.; Bussolati, B. Effect of stem cell-derived extracellular vesicles on damaged human corneal endothelial cells. *Stem Cells Int.* **2021**, *2021*, 6644463. [CrossRef]
180. Shen, T.; Zheng, Q.; Luo, H.; Li, X.; Chen, Z.; Song, Z.; Zhou, G.; Hong, C. Exosomal miR-19a from adipose-derived stem cells suppresses differentiation of corneal keratocytes into myofibroblasts. *Aging* **2020**, *12*, 4093. [CrossRef]
181. Wang, S.; Hou, Y.; Li, X.; Song, Z.; Sun, B.; Li, X.; Zhang, H. Comparison of exosomes derived from induced pluripotent stem cells and mesenchymal stem cells as therapeutic nanoparticles for treatment of corneal epithelial defects. *Aging* **2020**, *12*, 19546. [CrossRef]
182. Zhou, J.; Ding, Y.; Zhang, Y.; Zheng, D.; Yan, L.; Guo, M.; Mao, Y.; Yang, L. Exosomes from bone marrow-derived mesenchymal stem cells facilitate corneal wound healing via regulating the p44/42 MAPK pathway. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2023**, *261*, 723–734. [CrossRef]
183. Saccu, G.; Menchise, V.; Gai, C.; Bertolin, M.; Ferrari, S.; Giordano, C.; Manco, M.; Dastrù, W.; Tolosano, E.; Bussolati, B. Bone Marrow Mesenchymal Stromal/Stem Cell-Derived Extracellular Vesicles Promote Corneal Wound Repair by Regulating Inflammation and Angiogenesis. *Cells* **2022**, *11*, 3892. [CrossRef]
184. Ma, S.; Yin, J.; Hao, L.; Liu, X.; Shi, Q.; Diao, Y.; Yu, G.; Liu, L.; Chen, J.; Zhong, J. Exosomes from human umbilical cord mesenchymal stem cells treat corneal injury via autophagy activation. *Front. Bioeng. Biotechnol.* **2022**, *10*, 879192. [CrossRef]
185. Tao, H.; Chen, X.; Cao, H.; Zheng, L.; Li, Q.; Zhang, K.; Han, Z.; Han, Z.-C.; Guo, Z.; Li, Z. Mesenchymal stem cell-derived extracellular vesicles for corneal wound repair. *Stem Cells Int.* **2019**, *2019*, 5738510. [CrossRef]

186. Zhao, W.; He, X.; Liu, R.; Ruan, Q. Accelerating corneal wound healing using exosome-mediated targeting of NF-κB c-Rel. *Inflamm. Regen.* **2023**, *43*, 6. [CrossRef]
187. Wu, W.; Zhou, J.; Zhu, D.; Ma, S. Effect of PKH-26-labeled exosomes derived from bone marrow mesenchymal stem cells on corneal epithelium regeneration in diabetic mice. *Ann. Transl. Med.* **2023**, *11*, 167. [CrossRef]
188. El-Mahalaway, A.M.; El-Azab, N.E.-E.; Abdrabbo, M.; Said, O.M.; Sabry, D.; El Sibaie, M.M. Comparative light and electron microscopic study on the therapeutic efficacy of adipose derived stem cells versus exosomes for experimentally induced acute corneal injuries in rats. *Stem Cell Res. Ther.* **2018**, *8*, 2. [CrossRef]
189. Yu, Z.; Hao, R.; Du, J.; Wu, X.; Chen, X.; Zhang, Y.; Li, W.; Gu, Z.; Yang, H. A human cornea-on-a-chip for the study of epithelial wound healing by extracellular vesicles. *Iscience* **2022**, *25*, 104200. [CrossRef]
190. Shang, Q.; Chu, Y.; Li, Y.; Han, Y.; Yu, D.; Liu, R.; Zheng, Z.; Song, L.; Fang, J.; Li, X. Adipose-derived mesenchymal stromal cells promote corneal wound healing by accelerating the clearance of neutrophils in cornea. *Cell Death Dis.* **2020**, *11*, 707. [CrossRef]
191. Shojaati, G.; Khandaker, I.; Funderburgh, M.L.; Mann, M.M.; Basu, R.; Stolz, D.B.; Geary, M.L.; Dos Santos, A.; Deng, S.X.; Funderburgh, J.L. Mesenchymal stem cells reduce corneal fibrosis and inflammation via extracellular vesicle-mediated delivery of miRNA. *Stem Cells Transl. Med.* **2019**, *8*, 1192–1201. [CrossRef]
192. Ong, H.S.; Riau, A.K.; Yam, G.H.-F.; Yusoff, N.Z.B.M.; Han, E.J.; Goh, T.-W.; Lai, R.C.; Lim, S.K.; Mehta, J.S. Mesenchymal Stem Cell Exosomes as Immunomodulatory Therapy for Corneal Scarring. *Int. J. Mol. Sci.* **2023**, *24*, 7456. [CrossRef] [PubMed]
193. Dalmon, C.A.; Chandra, N.S.; Jeng, B.H. Use of autologous serum eyedrops for the treatment of ocular surface disease: First US experience in a large population as an insurance-covered benefit. *Arch. Ophthalmol.* **2012**, *130*, 1612–1613. [CrossRef] [PubMed]
194. Zhou, T.; He, C.; Lai, P.; Yang, Z.; Liu, Y.; Xu, H.; Lin, X.; Ni, B.; Ju, R.; Yi, W. miR-204–containing exosomes ameliorate GVHD-associated dry eye disease. *Sci. Adv.* **2022**, *8*, eabj9617. [CrossRef] [PubMed]
195. Liao, F.-l.; Tan, L.; Liu, H.; Wang, J.-j.; Ma, X.-t.; Zhao, B.; Chen, Y.; Bihl, J.; Yang, Y.; Chen, R.-l. Hematopoietic stem cell-derived exosomes promote hematopoietic differentiation of mouse embryonic stem cells in vitro via inhibiting the miR126/Notch1 pathway. *Acta Pharmacol. Sin.* **2018**, *39*, 552–560. [CrossRef]
196. Ko, K.-W.; Yoo, Y.-I.; Kim, J.Y.; Choi, B.; Park, S.-B.; Park, W.; Rhim, W.-K.; Han, D.K. Attenuation of tumor necrosis factor-α induced inflammation by umbilical cord-mesenchymal stem cell derived exosome-mimetic nanovesicles in endothelial cells. *Tissue Eng. Regen. Med.* **2020**, *17*, 155–163. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

Proteome Landscape during Ripening of Solid Endosperm from Two Different Coconut Cultivars Reveals Contrasting Carbohydrate and Fatty Acid Metabolic Pathway Modulation

Jean Wildort Félix [1], María Inés Granados-Alegría [1], Rufino Gómez-Tah [2], Miguel Tzec-Simá [1], Eliel Ruíz-May [3], Blondy Canto-Canché [2], Jesús Alejandro Zamora-Briseño [3], Esaú Bojórquez-Velázquez [3], Carlos Oropeza-Salín [2] and Ignacio Islas-Flores [1,*]

[1] Unidad de Bioquímica y Biología Molecular de Plantas, Centro de Investigación Científica de Yucatán, A.C., Calle 43 No. 130 x 32 y 34, Chuburná de Hidalgo, Mérida C.P. 97205, Yucatán, Mexico; felixjeandewildort@yahoo.fr (J.W.F.); granados.alegria@gmail.com (M.I.G.-A.); tzecmyr@cicy.mx (M.T.-S.)
[2] Unidad de Biotecnología, Centro de Investigación Científica de Yucatán, A.C., Calle 43 No. 130 x 32 y 34, Chuburná de Hidalgo, Mérida C.P. 97205, Yucatán, Mexico; rufino26gt@gmail.com (R.G.-T.); cantocanche@cicy.mx (B.C.-C.); cos@cicy.mx (C.O.-S.)
[3] Red de Estudios Moleculares Avanzados, Instituto de Ecología, A.C., Carretera antigua a Coatepec 351, Colonia El Haya, Xalapa C.P. 91073, Veracruz, Mexico; eliel.ruiz@inecol.mx (E.R.-M.); alejandro.zamora@inecol.mx (J.A.Z.-B.); esau.bojorquez@inecol.mx (E.B.-V.)
* Correspondence: islasign@cicy.mx; Tel.: +52-(999)-9-42-83-30

Abstract: *Cocos nucifera* L. is a crop grown in the humid tropics. It is grouped into two classes of varieties: dwarf and tall; regardless of the variety, the endosperm of the coconut accumulates carbohydrates in the early stages of maturation and fatty acids in the later stages, although the biochemical factors that determine such behavior remain unknown. We used tandem mass tagging with synchronous precursor selection (TMT-SPS-MS3) to analyze the proteomes of solid endosperms from Yucatan green dwarf (YGD) and Mexican pacific tall (MPT) coconut cultivars. The analysis was conducted at immature, intermediate, and mature development stages to better understand the regulation of carbohydrate and lipid metabolisms. Proteomic analyses showed 244 proteins in YGD and 347 in MPT; from these, 155 proteins were shared between both cultivars. Furthermore, the proteomes related to glycolysis, photosynthesis, and gluconeogenesis, and those associated with the biosynthesis and elongation of fatty acids, were up-accumulated in the solid endosperm of MPT, while in YGD, they were down-accumulated. These results support that carbohydrate and fatty acid metabolisms differ among the developmental stages of the solid endosperm and between the dwarf and tall cultivars. This is the first proteomics study comparing different stages of maturity in two contrasting coconut cultivars and may help in understanding the maturity process in other palms.

Keywords: proteomics; coconut cultivars; solid endosperm; maturation stages; carbohydrate and lipid metabolisms; TMT-SPS-MS3

Citation: Félix, J.W.; Granados-Alegría, M.I.; Gómez-Tah, R.; Tzec-Simá, M.; Ruíz-May, E.; Canto-Canché, B.; Zamora-Briseño, J.A.; Bojórquez-Velázquez, E.; Oropeza-Salín, C.; Islas-Flores, I. Proteome Landscape during Ripening of Solid Endosperm from Two Different Coconut Cultivars Reveals Contrasting Carbohydrate and Fatty Acid Metabolic Pathway Modulation. *Int. J. Mol. Sci.* **2023**, *24*, 10431. https://doi.org/10.3390/ijms241310431

Academic Editor: Wajid Zaman

Received: 23 May 2023
Revised: 16 June 2023
Accepted: 16 June 2023
Published: 21 June 2023

1. Introduction

The coconut tree (*Cocos nucifera* L.) is an ancient palm species of great importance in tropical and subtropical areas [1]. Coconut belongs to the Arecaceae family, which also includes the oil palm (*Elaeis guineensis*) and the date palm (*Phoenix dactylifera* L.); these palms are sources of oil (the former) and carbohydrates (the latter) for the food and pharmaceutical industries [2,3].

The coconut tree shows many morphological and genetic variations, e.g., in tree and fruit sizes, as well as flower pollination, but coconuts are classified into two groups based on morphology and growth habits: the "tall" and the "dwarf" varieties [1,4]. Additionally, hybrids exist, which result from cross-pollination between a tall and a dwarf variety [5,6]. Tall varieties begin to flower 5–7 years after planting and continue to produce inflorescences

and fruits until they are 80 to 100 years old. As adults, the tall palms reach 20–30 m in height, producing medium- to large-sized fruits with abundant solid endosperm and high oil content [7]. In the case of dwarf palm varieties, they begin to flower and fruit 3–4 years after planting and continue to produce until they are 50 years old. As adults, they are 8–10 m tall and produce small to medium-sized fruits with moderate amounts of solid endosperm, but lower oil content than tall varieties [8]. The seed of the coconut tree is a drupe; its shape varies from spherical to oval and includes the solid endosperm, the embryo embedded in the solid endosperm, and the slightly sweet liquid endosperm (coconut water) [9].

The solid endosperm, also called "coconut meat", and when dry, "copra", is a source of oil composed of triacylglycerol (TAG) with long-, medium- and short-chain fatty acids. In the solid endosperm, medium-chain saturated fatty acids (MCSFA) predominate, e.g., lauric acid (C12:0), myristic acid (C14:0) and palmitic acid (C16:0), among others, and the total fatty acid content is higher in the tall than in the dwarf variety [10,11]. Coconut oil contains high levels of lauric acid and exhibits characteristics such as increased oxidative stability, low melting points and stable emulsion formation, all highly appreciated in the food and chemical industries [12,13].

The growth and maturation of coconut fruit are intrinsically related to the development of the different components of the seed, i.e., endosperms, pericarp and embryo. However, in contrast to the liquid endosperm, the solid endosperm of the coconut fruit does not accumulate synchronically with fruit maturation because it begins to accumulate around the internal periphery of the embryo sac sixth months after pollination; early solid endosperm cells have an appearance of a cream to transparent jelly-like matrix and contain many kinds of structural carbohydrates [14]. When fruits are ripening, the solid endosperm becomes thicker, white in color, with a more compact structure; then, endosperm cells stop their deposition and begin to function as a reservoir of proteins and triacylglycerols [14].

Bourgis et al. [15] analyzed the transcriptomes of oil and date palms, two Arecaceae species with contrasting metabolisms. They found that the high oil content in *E. guineensis* was associated with high transcriptional expression levels of fatty acid biosynthesis genes, while transcriptional levels of glycolysis enzymes were comparable in both species. Naganeeswaran et al. [16] performed transcriptomics analysis of *P. dactylifera* embryogenic calli, identifying 53,251 transcripts. The KEGG-enriched pathways were predominantly represented by metabolic pathways that include carbohydrate and energy, consistent with the almost exclusive accumulation of carbohydrates in date palm. In the case of coconut palm, the seed accumulates carbohydrates in the early stages of development, while in the intermediate and mature stages, it accumulates fatty acids [5]. Transcriptomics analysis is highly valuable, but transcript expression does not necessarily correlate with protein translation; thus, proteomics analysis is also necessary.

Regarding proteomics studies carried out on coconut solid endosperm, D'Amato et al. [17] reported a proteome of 307 proteins associated with coconut milk, a derivate of the solid endosperm. Huang et al. [18], using two-dimensional electrophoresis (2-DE), generated a proteomic map of common proteins in the solid endosperms of 14 coconut samples collected in different provinces of China, Thailand and Vietnam. They found at least 200 protein spots, and sequencing some of these proteins by MALDI-TOF-TOF-MS revealed homology to 7S globulin, glutelin and a putative receptor-like protein kinase. Recently, Ma et al. [19], using shotgun protein sequencing of coconut solid endosperm, identified 1686 proteins. The in silico digestion of globulins and antioxidant proteins with pepsin and trypsin helped them to predict peptides with emulsifying and antioxidant activities. Unfortunately, the database for this research is not publicly available, preventing further analysis. Moreover, all these investigations seem incomplete because they omit information about the varieties and developmental stages of the fruit used, since proteomic changes are influenced, at least in part, by these factors.

Coconut is a non-climacteric fruit with only a few studies focused on biochemical maturation processes. The Yucatan green dwarf (YGD) and the Mexican pacific tall (MPT)

are two coconut cultivars with contrasting characteristics; the fruit of the former is used mainly for water consumption, while the second is mainly used for copra production and oil extraction [6]. However, the molecular and biochemical basis for these differences have not been revealed yet.

This study aims to investigate the regulation of carbohydrate and fatty acid metabolisms during coconut fruit ripening and uncover the proteins involved in these metabolic processes in coconut fruit. Using tandem mass tag–synchronous precursor selection (TMT-SPS-MS3), proteomic analysis was carried out on the solid coconut endosperm, comparing the meat of immature, intermediate and mature fruits of the Yucatan green dwarf (YGD) and the Mexican pacific tall (MPT) cultivars.

A total of 244 proteins were identified in YGD and 347 in MPT cultivars; enzymes involved in lipid and carbohydrate metabolism were found in both cultivars. Proteins involved in glycolysis and methylglyoxal pathways were up-accumulated in MPT but down-accumulated in YGD, consistent with their roles in lipid accumulation in other plants. Regarding lipid metabolism, the proteins showed higher accumulation in MPT than in YGD as maturation progressed, which corroborates that the MPT cultivar is best suited for oil extraction because of their high TAG content.

To the best of our knowledge, this is the first comprehensive proteomics report focusing on the maturity process of solid coconut endosperms of two contrasting varieties, thus contributing to a better understanding of lipids and carbohydrate metabolisms in coconut, but also in other oilseeds, which are poorly studied in terms of their maturation processes.

2. Results

2.1. Physiological and Physicochemical Parameters of the Coconut Fruit

Our study was centered on three coconut ripening stages: immature, intermediate and mature fruits of the Yucatan green dwarf (YGD) and Mexican pacific tall (MPT) cultivars (Figure 1A,B). In both cultivars, the main morphological characteristics of coconut fruits during ripening were associated with changes in color and weight. At the immature stage, the fruits are green, while at the intermediate, they are yellowish-green, and finally, they are pale brown at the mature stage. The color modifications were accompanied by the drying of the pericarp, a decrease in coconut water volume, and a loss of weight of mature fruits (Figures 1A,B and S1).

Both cultivars exhibited similar morphological features during solid endosperm ripening (Figure 1A,B). The immature solid endosperm was gelatinous and slightly transparent to whitish in color. In the intermediate stage, the endosperm was thicker and harder than in the immature fruit. In the mature stage, the solid endosperm was white, thicker and harder than in the previous stages (Figure 1A,B).

Complex protein profiles were observed in a 12% SDS-PAGE for each of the different stages of maturity and in both cultivars, with observed molecular sizes between 17 to 199 kDa. Four major polypeptides with molecular weights of 46, 33, 24.8 and 17 kDa were present in all stages of both cultivars, but particular differences were also observed. A 30.7 kDa polypeptide was present in all three stages of maturation in the YGD cultivar, whereas it was only observed in the immature stage of the MPT cultivar. A 20.7 kDa polypeptide was more abundant in the mature stage in the YGD cultivar (Figure 1C). The comparative-proteomics-based TMT-SPS-MS3 allowed the identification of 244 and 347 proteins in YGD and MPT cultivars, respectively (Figure 1D,E and Table S1). Both cultivars shared 155 proteins, indicating that these proteins were the core of the proteome, whereas 89 proteins were unique to YGD and 192 were unique to MPT (Figure 1E).

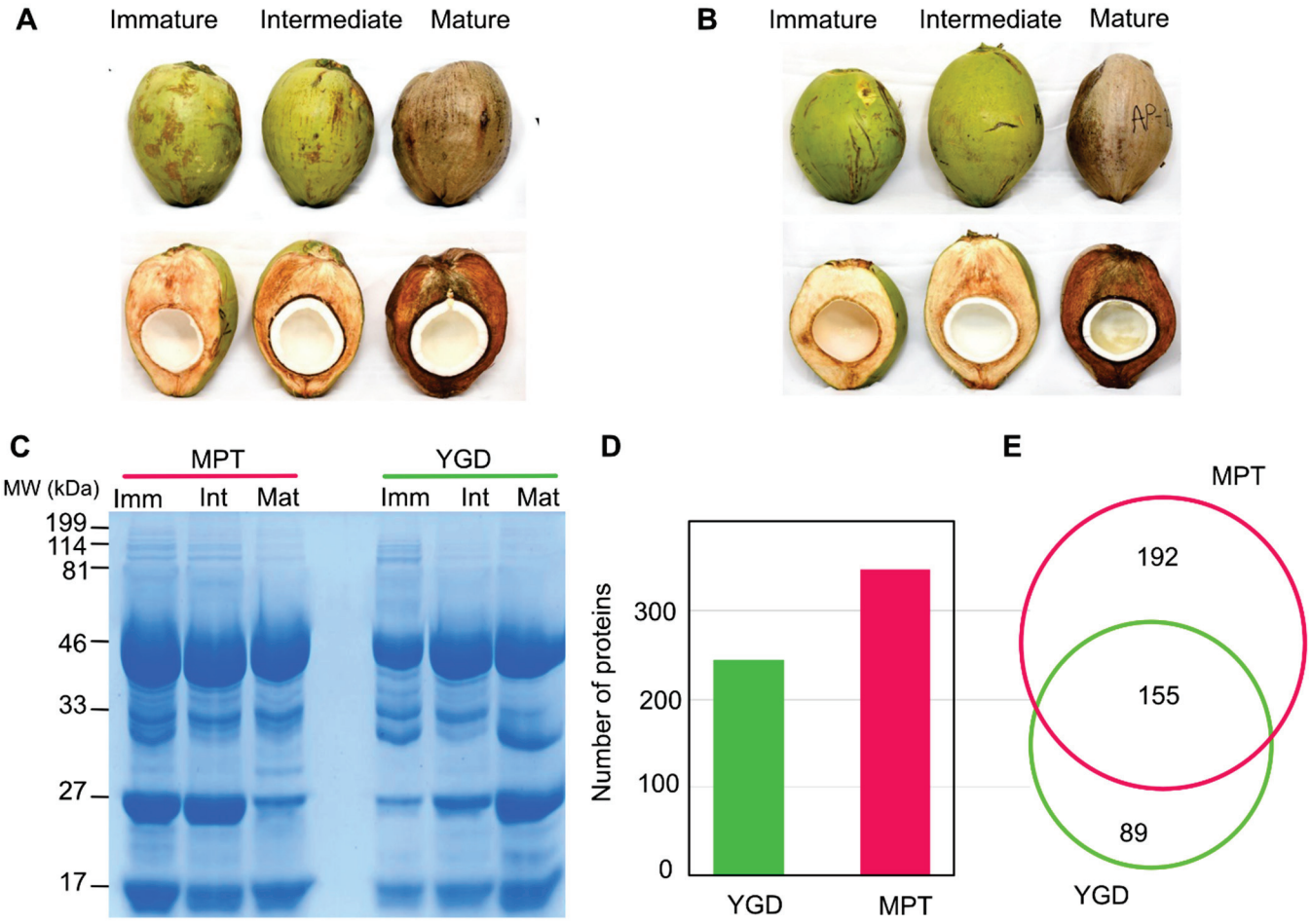


Figure 1. Coconut ripening and proteins identified in Yucatan green dwarf (YGD) and Mexican pacific tall (MPT) cultivars. Visual changes in fruit ripening of YGD (**A**) and MTP (**B**). SDS-PAGE protein profile of solid endosperm during ripening (**C**), the protein molecular weight standard is indicated on the left side of the figure. Number of proteins identified in YGD and MTP by TMT-SPS-MS3 (**D**). Proportional Venn diagram representation of proteins identified in each cultivar (**E**).

2.2. Proteome Landscape of Coconut Solid Endosperm during Ripening

The core clusters of the MPT and YGD cultivars were visualized in TreeMap to analyze the enriched metabolic pathways. The glycolytic process, including malate metabolism, protein folding, proteasomal ubiquitin independent protein catabolism, together with the stress response, were the main biological processes enriched in both cultivars (Figure 2A–C and Table S2). In addition, YGD exhibited an enrichment of fatty acid biosynthesis, protein refolding, and embryo development ending in seed dormancy (Figure 2B). The MPT super cluster was particularly enriched with proteins associated with biological processes such as the tricarboxylic acid cycle, the glyoxylate cycle, citrate and malate metabolism, tetrahydrofolate interconversion, chlorophyll biosynthesis, photosynthesis, cytoskeleton organization, and protein transport (Figure 2C).

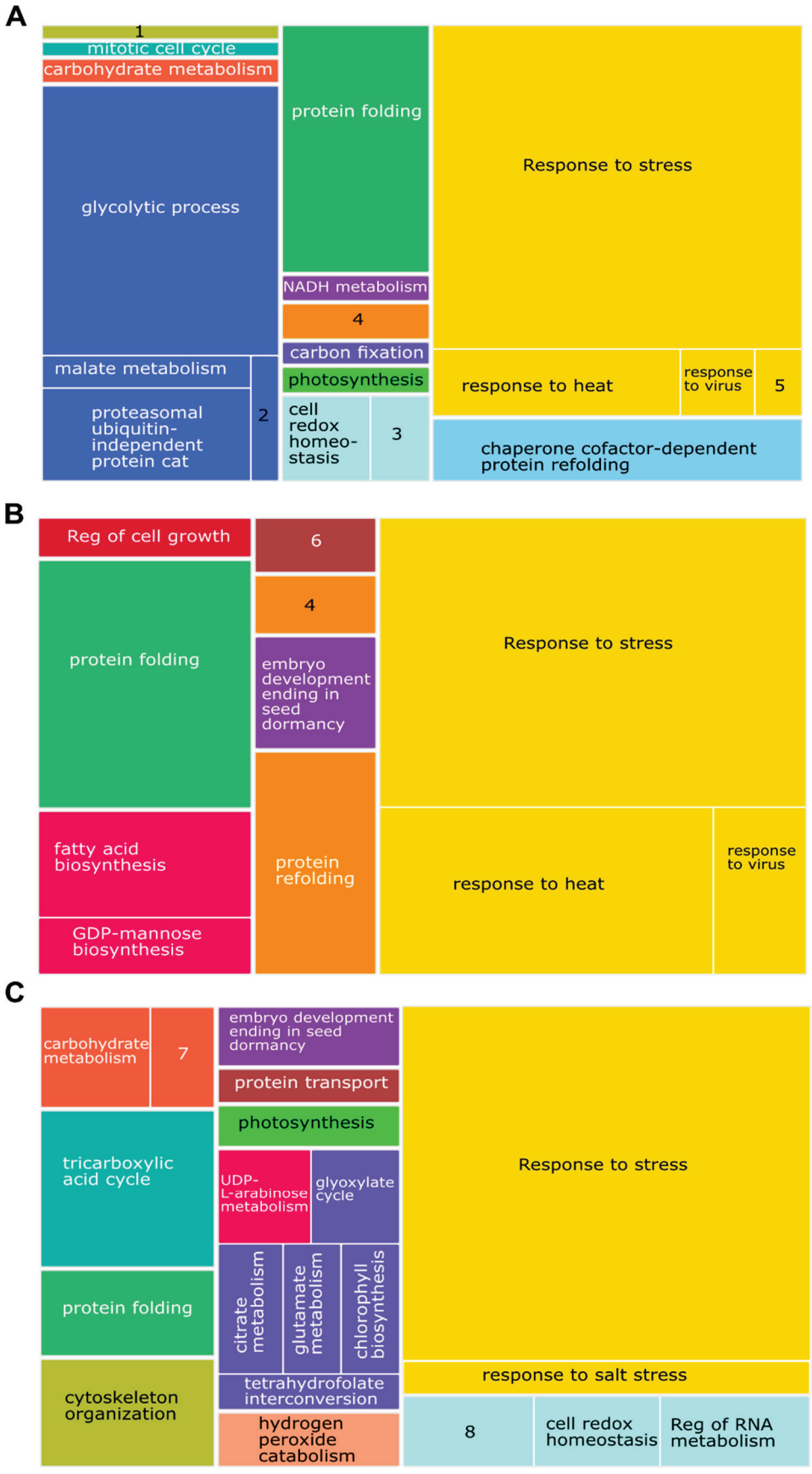


Figure 2. Enrichment of the biological processes found in solid endosperm of coconut visualized as a TreeMap. Proteins found in both cultivars (core proteome, **A**). Proteins identified particularly in

YGD (**B**) and MPT (**C**) cultivars. Each box represents the *p*-value weight within the treemap. Components (boxes) of the same color represent semantic similarity (medium allowed similarity, 0.7). The most enriched biological processes are shown as larger components within the map. 1, microtubule cytoskeleton org; 2, Fructose 1,6-bisphosphate metabolisms; 3, regulation of vacuole fusion, non-autophagic; 4, microtubule-based process; 5, response to arsenic-containing substance; 6, retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum; 7, carbohydrate derivative metabolism; 8, negative regulation of response to water deprivation.

2.3. Dynamics of the Coconut Solid Endosperm Proteome during Ripening

The normalized protein abundance of the biological replicates originating from the immature, intermediate and mature stages was analyzed by principal component analysis (PCA). The biological replicates under study were grouped together according to each stage of maturation and the cultivar (Figure 3A,E). In the YGD cultivar, the accumulated variance of PC1 (59.8%) and PC2 (25.7%) explained up to 85.5% of the variation observed between the ripening stages analyzed (Figure 3A); while in MPT, the accumulated variance of PC1 (72.1%) and PC2 (18.4%) explained up to 90.5% of the variation observed there (Figure 3E). An additional PCA was carried out to compare the two cultivars and their stages of maturity between them; the PCA shows differences between cultivars and among their different stages of maturation (Figure S2).

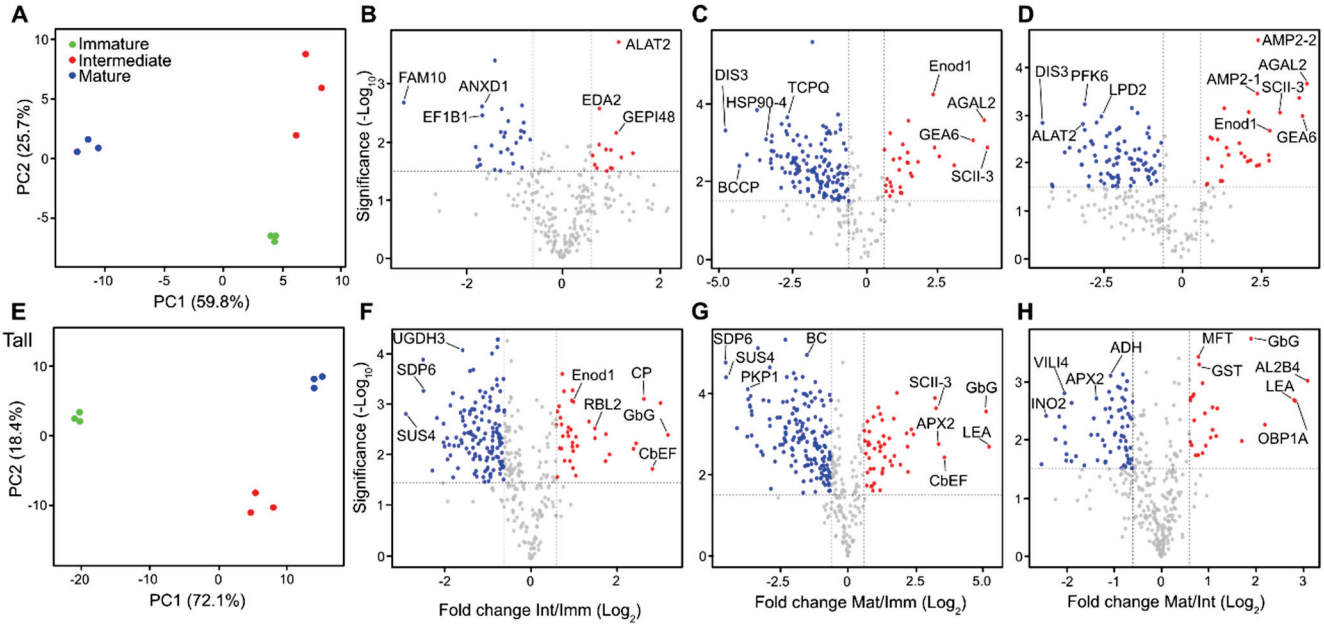


Figure 3. Comparative proteomics analyses of solid endosperm during ripening. Principal component analysis (PCA) of protein abundances of YGD (**A**). Significant differential accumulated proteins represented in volcano plots in the Log 2 ratios intermediate/immature (**B**), mature/immature (**C**) and mature/intermediate samples (**D**). PCA of protein abundances of MPT (**E**). Significant differential accumulated proteins are represented in (**F**–**H**). The *x*-axis corresponds to the log2 fold change value, and the *y*-axis displays the −log10 *p*-value. The points in blue represent the down-accumulated proteins (<1/1.5-fold change, FC), those in red represent the over-accumulated proteins (≥1.5 FC) and those in grey represent proteins with not changes in log2 fold change. Alanine aminotransferase 2 (ALAT2), L-ascorbate peroxidase 2 (APX2), Alpha-galactosidase 2 (AGAL2), Annexin D1 (ANXD1), Aldehyde dehydrogenase family 2 member B4 (AL2B4), Biotin carboxylase (BC), Calcium-binding EF-hand family protein (CbEF), Carboxypeptidase (CP), Elongation factor 1-beta 1 (EF1B1), putative

Elongation factor 1-alpha (EF1A4), early nodulin-like protein 1 (Enod1), Em-like protein (GEA6), Dihydrolipoyl dehydrogenase 2 (LPD2), Inositol-3-phosphate synthase isozyme 2 (INO2), Probable fructokinase-6 (PFK6), Glycerol-3-phosphate dehydrogenase (SDP6), Heat shock protein 90-4 (HSP90-4), Glutathione S-transferase family protein (GST), putative DIS3-like exonuclease 2, UDP-glucose 4-epimerase (GEPI48), putative FAM10 family protein, glucan endo-1,3-beta-glucosidase-like (GbG), 11 kDa late embryogenesis abundant protein (LEA), RHOMBOID-like protein 2 (RBL2), 11 kDa late embryogenesis abundant protein (LEA), putative DIS3-like exonuclease 2, Protein MOTHER of FT and TFL1 (MFT), Oil-body-associated protein 1A (OBP1A), T-complex protein 1 subunit theta (TCPQ), Plastidial pyruvate kinase 2 (PKP1), Plastidial pyruvate kinase 2 (PKP1), Sucrose synthase 4 (SUS4), serine carboxypeptidase II-3 (SCII-3), putative serine protease (EDA2), UDP-glucose 6-dehydrogenase 3 (UGDH3), vicilin-like antimicrobial peptides 2-2 (AMP2-2), and Villin-4 (VILI4).

Subsequently, the trend in protein accumulation was analyzed during coconut ripening, focusing on the ratios intermediate/immature (Int/Imm), mature/immature (Mat/Imm), and mature/intermediate (Mat/Int). According to the log2-fold change (log2FC), a total of 225 differentially accumulated proteins (DAPs) were found in YGD, whereas 251 DAPs were found in MPT (Table S1). With respect to the Mat/Int in YGD, proteins like putative FAM10 family, ANXD1 and EF1B1 were down-accumulated, while ALAT2, EDA2 and GEP148 were up-accumulated. When comparing the Mat/Imm and Mat/Int of YGD coconut meat, HSP90-4, BCCP, TCPQ, DIS3, LPD2, PFK6 and ALAT2 were down-accumulated in the mature stage, while proteins like Enod1, AGAL2, GEA6 and SCII-3 were up-accumulated (Figure 3B–D). In the MPT coconut cultivar, SUS4 and SPD6 were down-accumulated in the intermediate and mature stages, while the GbG protein increased during maturation and, finally, LEA was up-accumulated in the later stages (Figure 3F–H).

2.4. Detoxification of ROS and Methylglyoxal Pathways in MPT and YGD Cultivars

ROS detoxification was more active in MPT than in YGD. Proteins such as superoxide dismutase (P24704, O81235) and catalase (P25819) were significantly up-accumulated in the intermediate and late stages of MPT maturation (Figure 4). Furthermore, the glutathione S-transferase DHAR2 (Q9FRL8), belonging to the glutathione ascorbate cycle, followed a similar trend. Proteins such as methylglyoxal reductase (Q0PGJ6) and lactoylglutathione lyase (Q9M8R4, O65398, Q04760), as well as L-ascorbate peroxidase (Q1PER6, Q05431), involved in the glutathione ascorbate cycle were also up-accumulated in the last two stages of ripening in MPT compared to YGD, evidencing the activation of methylglyoxal degradation pathways I and III in the MPT cultivar (Figure 4).

2.5. DAPs Involved in Carbohydrate Metabolism in the YGD and MPT Cultivars

2.5.1. Glycolysis and Gluconeogenesis

Proteins involved in acetaldehyde synthesis were down-accumulated in the immature and intermediate compared with mature YGD fruits, while proteins involved in gluconeogenesis, and glycan, starch, sucrose and sugar-nucleotide synthesis were slightly up-accumulated in the immature YGD; however, these proteins were down-accumulated in the intermediate and mature fruits compared to immature YGD fruits (Figure 5A). In the case of the MPT cultivar, the proteins involved in gluconeogenesis and the synthesis of glycan, glycogen, starch and sucrose were down-accumulated in the immature stage compared to intermediate and mature fruits; only the proteins involved in the synthesis of acetaldehyde showed a slight up-accumulation in all the analyzed stages of MPT fruits, in comparison with YGD fruits. Proteins involved in gluconeogenesis and the synthesis of glycan, starch, sucrose and sugar-nucleotide synthesis were up-accumulated in the intermediate and mature stages of MPT cultivar (Figure 5A).

The analysis of proteins involved in the gluconeogenesis pathway showed that malate dehydrogenase (P57106, Q9SN86), independent 2,3-bisphosphoglycerate (O04499), phosphoglycerate kinase (Q9LD57, Q9SAJ4), glyceraldehyde 3-phosphate dehydrogenase (Q9FX54, Q5E924), fructose bisphosphate aldolase (Q9ZU52, Q9SJQ9), and glucose 6 phosphate isomerase (P34795, Q8H103) were down-accumulated in the intermediate and mature

stage of YGD coconut (Figure 5B). In the case of the MPT cultivar, all these proteins started with low amounts in the immature stage (log2FC −1.07) and increased in the intermediate and mature stages, reaching log2FC +0.36 and +1.07, respectively (Figure 5B). Phosphoglycerate kinase (Q9LD57, Q9SAJ4) and glyceraldehyde 3-phosphate dehydrogenase (Q9FX54, Q5E924) showed the highest accumulation among the proteins involved in the gluconeogenesis pathway in MPT coconut, the former in the intermediate and mature stages and the latter in the intermediate stage (Figure 5B).

The analysis of glycolysis pathway determined the down-accumulation of 6-phosphofructokinase (Q94AA4), diphosphate-fructose-6-phosphate 1-phosphotransferase (Q8W4M5, Q9C9K3), Glucose 6-phosphate isomerase (P34795, Q8H103), triose-phosphate isomerase (Q9SJQ9), glyceraldehyde 3-phosphate dehydrogenase (Q9FX54, Q9LD57, Q9SAJ4, and Q5E924), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (O04499), phosphoglycerate kinase (Q9LD57, Q9SAJ4, and D7L0K4), enolase (P25696, Q9C9C4), fructose bisphosphate aldolase (Q9ZU52, Q9SJQ9) and pyruvate kinase (Q9FM97, Q9FNN1, Q94KE3, Q9LIK0, and Q9FLW9) in the meat of mature YGD coconut; all proteins showed a decrease in the mature stage in glycolysis I, II and IV (Figures 5C and S3). In contrast, triose-phosphate isomerase (P48491) doubled its amount in the mature stage (log2FC +1.07) (Figures 5B,C and S3).

Regarding glycolysis in the MPT cultivar, most glycolytic enzymes increased as maturity progressed. The protein 6-phosphofructokinase (Q94AA4), increased moderately in the intermediate stage compared to the immature stage, and doubled its amount in the mature stage (Figures 5C and S3). In the case of diphosphate:fructose-6-phosphate 1-phosphotransferase (Q8W4M5, Q9C9K3), involved in glycolysis IV, it started at a low level in the immature stage, but then increased during intermediate and mature stages (Figure S3). Likewise, in YGD, the proteins that increased during the progression of maturity were triose-phosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, enolase, pyruvate kinase, phosphoglycerate kinase, and the fructose bisphosphate aldolase, all of them observed with low amounts in the immature stage which then increased in the intermediate and mature stages, eventually doubling the amount of initial protein (Figure 5C). In summary, glycolysis I, II and IV were more active in the intermediate and mature stages in the MPT cultivar than in YGD.

2.5.2. The Tricarboxylic Acid Cycle (TCA)

In YGD, proteins involved in the TCA cycle (variation V), i.e., the citrate synthase (P20115), citrate hydro-lyase (cis-aconitate-forming) (Q9SIB9), the cis-aconitate hydratase (Q9SIB9), succinate: quinone oxidoreductase (O82663), and malate dehydrogenase (Q9ZP06, mMDH1, and PMDH1), did not show changes in protein abundance during the maturation of the solid endosperm; the levels were between −0.444 to +0.444 (Figure 6A,B). In contrast, succinyl-CoA synthetase (O82662) decreased from the immature to the mature stage, regardless of whether the TCA type V or II was present. Regarding malate dehydrogenase (P57106, Q9SN86, c-NAD-MDH2, pNAD-MDH), it showed down-accumulation in the intermediate and mature stages in comparison with the immature stage (Figure 6A,B).

When proteins in the TCA cycle were analyzed in MPT endosperm, it was found that citrate synthase, citrate hydro-lyase (cis-aconitate-forming), cis-aconitate hydratase, PMDH1, pNAD-MDH, c-NAD-MDH2 and mMDH1 were initially low in the immature stage, and then increased in the intermediate and mature stages (Figure 6A,B), while the amount of succinate: quinone oxidoreductase (O82663) was stable throughout the maturity process, as was similarly observed in YGD. Some enzymes, e.g., malate dehydrogenase, were only active in the immature stage of YGD. In the case of citrate synthase, it was up-accumulated in the intermediate and mature stages of the endosperm of the MPT cultivar.

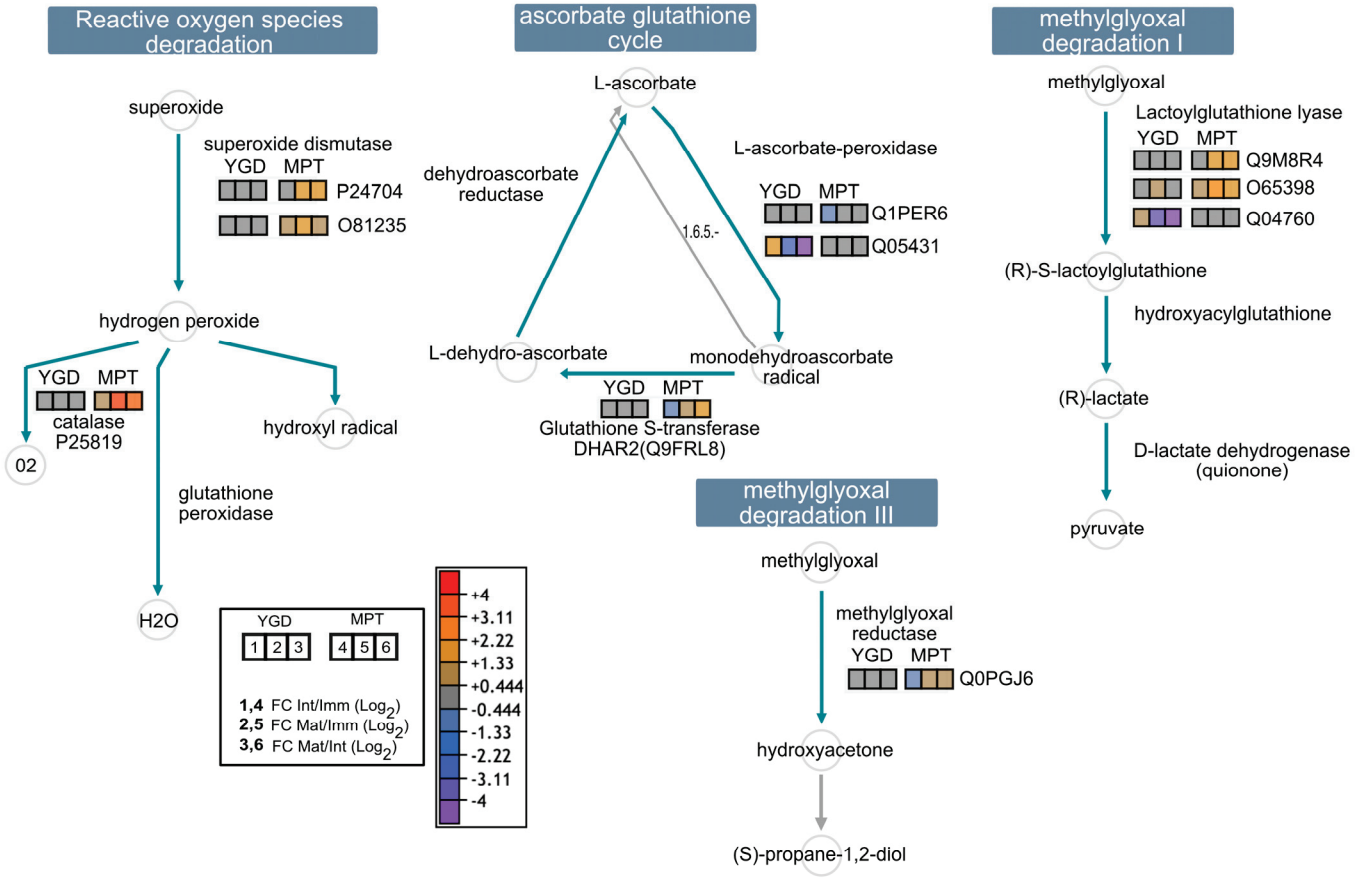


Figure 4. Visual representation of differential proteins associated with enzymes related to reactive oxygen species (ROS) and methylglyoxal detoxification pathways in YGD and MPT cultivars. Pathway reconstruction was carried out using the "cellular overview" tool on the Plant Metabolic Network (PMN) database platform. Numbers in squares specify, 1,4: log2 FC intermediate/immature (Int/Imm); 2,5: log2 FC mature/immature (Mat/Imm); 3,6: log2 FC mature/intermediate (Mat/Int). Numbers 1,2,3 indicate YGD variety, while 4,5,6 indicate MPT variety. The same applies to the following figures.

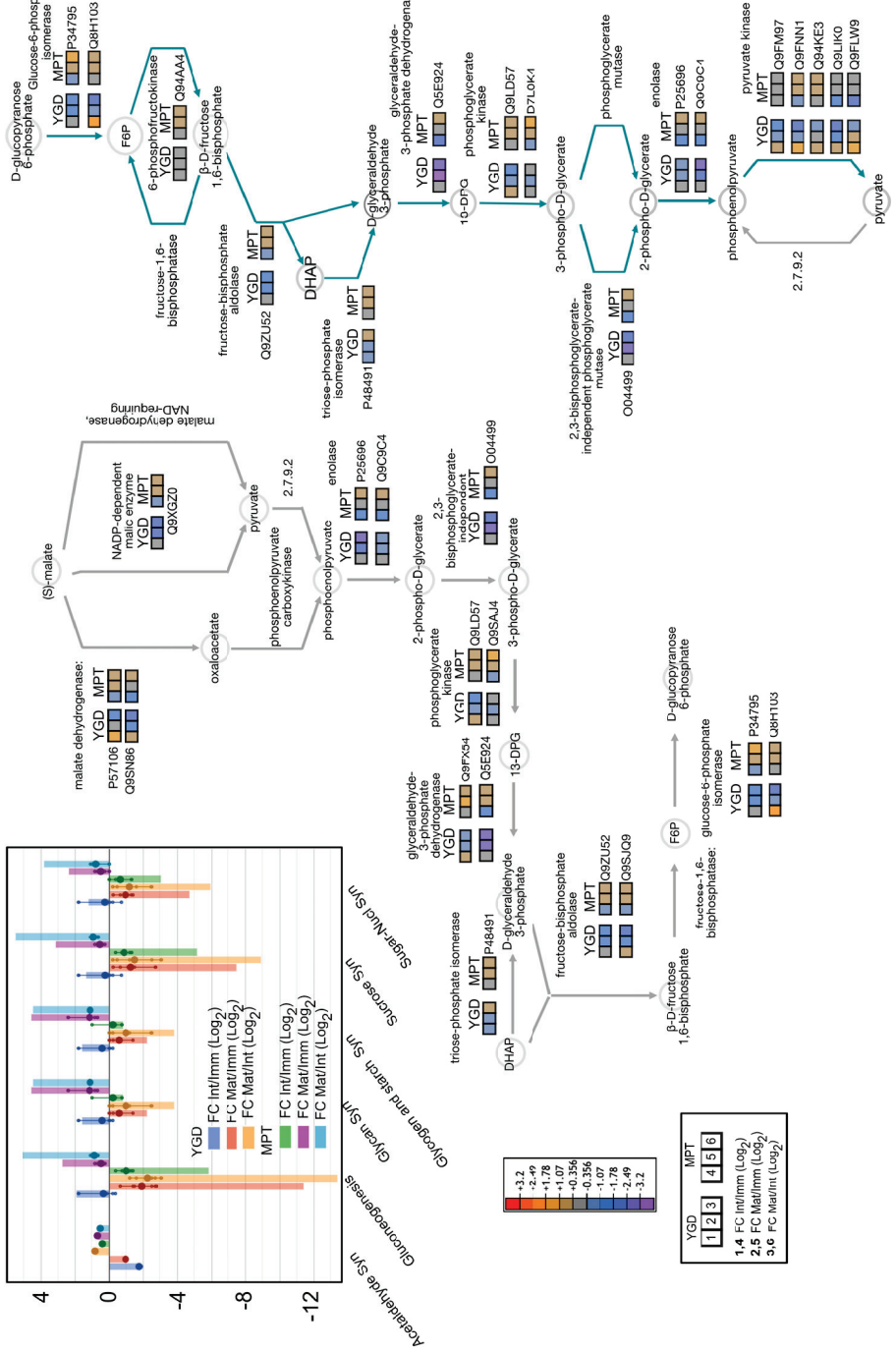


Figure 5. Visual representation of differential proteins associated with the carbohydrate biosynthetic main pathways. Pattern of accumulation of carbohydrate-related proteins determined by TMT-SPS-MS3. Small circles specify the log2FC values of each protein; the big circles are the average of the log2FC values, and columns indicate the sum of the log2FC (**A**). Dynamic of identified differential enzymes in the gluconeogenesis I (**B**) and glycolysis I (from glucose 6-phosphate) (**C**). Glycolysis II and IV are presented in Figure S3. The pathway reconstruction was carried out by using the "cellular overview" tool in the Plant Metabolic Network (PMN) database platform.

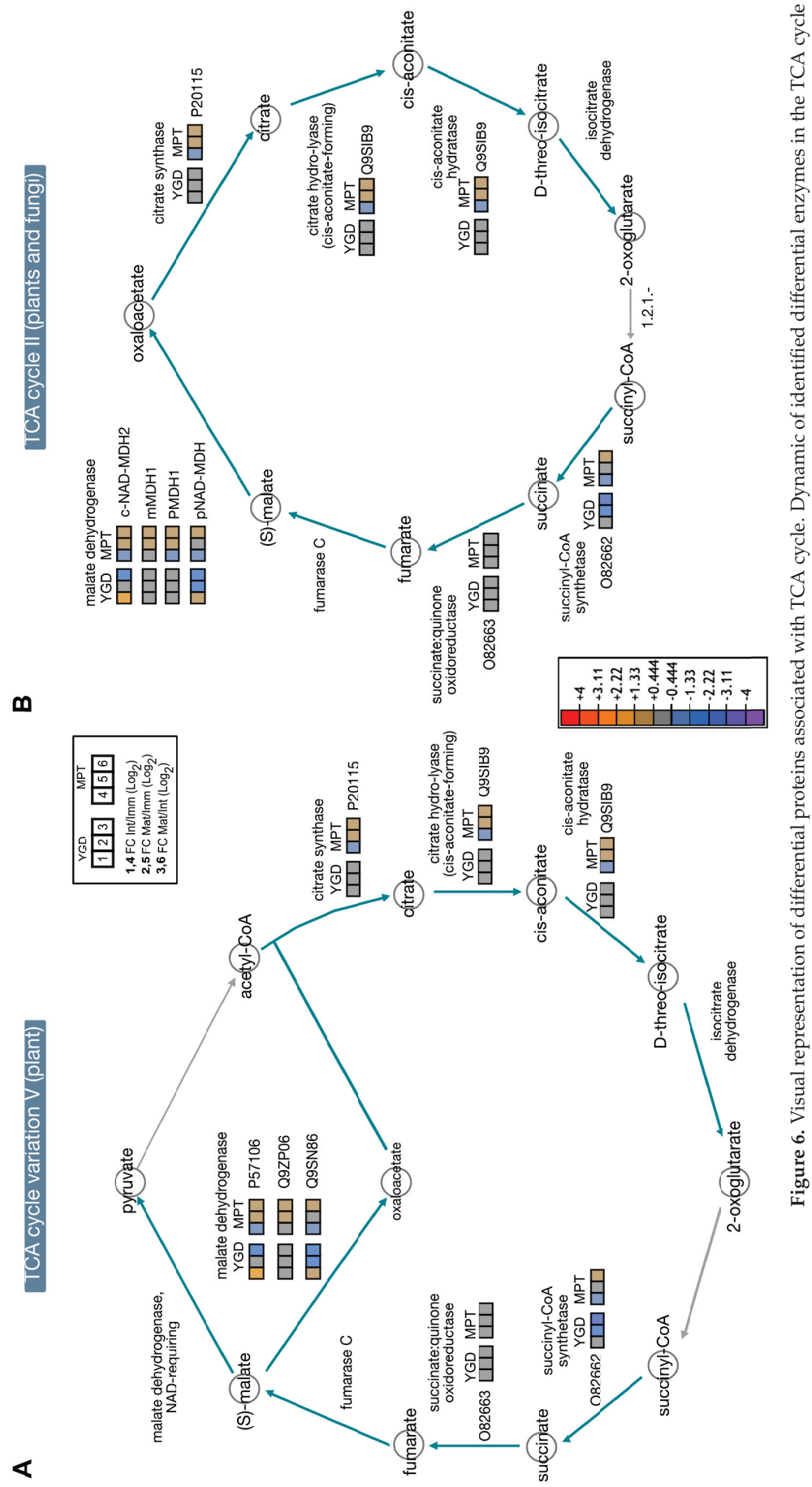


Figure 6. Visual representation of differential proteins associated with TCA cycle. Dynamic of identified differential enzymes in the TCA cycle V (plants, **A**) and II (plants and fungi, **B**).

2.6. DAPs Involved in Lipid Metabolism in the YGD and MPT Cultivars

The proteins involved in fatty acid initiation were up-accumulated in the immature stage of YGD coconut, while they were down-accumulated in intermediate and mature stages; proteins involved in fatty acid elongation and palmitate and stearate synthesis were down-accumulated in the intermediate and mature stages, compared to the immature stage. They were also down-accumulated in the mature stage compared to the intermediate stage (Figure 7A). In the MPT cultivar, the proteins involved in fatty acid initiation showed constant accumulation, remaining at the basal level. Proteins involved in fatty acid elongation and palmitate and stearate synthesis were up-accumulated in the intermediate and mature stages compared to the immature stage (Figure 7A).

The analysis of palmitate biosynthesis showed that 3-oxoacyl-[ACP] reductase (P33207) and 2,3,4-saturated fatty acyl-[ACP]: NAD+ oxidoreductase (Q9SLA8) were down-accumulated in the YGD, while the 3-hydroxyacyl-[ACP] dehydratase (Q9LX13) was up-accumulated, starting from log2FC −1.33 in the immature stage and increasing to log2FC +1.33 in the mature stage (Figure 7B). Regarding the MPT coconut, 3-hydroxyacyl-[ACP] dehydratase and 2,3,4-saturated fatty acyl-[ACP]NAD+ oxidoreductase (Q9SLA8) showed an up-accumulation trend during the maturity stages, with a log2FC from −0.444 to +4 (Figure 7B). The analysis of the protein dynamics involved in fatty acid elongation-saturation showed that three enzymes were down-accumulated in the mature stage of YGD, while the same enzymes were moderately up-accumulated in the mature stage of the MPT coconut (Figure 7C).

2.7. Other Proteins

Some proteins were shared between the two cultivars and the three maturation stages showing similar behaviors, such as the embryonic proteins DC-8 (A0A8B8ZT31), vicilin-like antimicrobial peptides 2-2 (A0A8B7BEE0), vicilin-like seed storage protein (A0A8B7BPL4), elongation factor 2 (AT1G56070), tubulin alpha chain (AT4G14960), tubulin beta-chain (AT5G12250), heat shock protein 83 (AT5G52640), and late embryogenesis abundant protein of 31-like (A0A8B7CDB9), whose accumulation increased throughout maturation of both cultivars (Table S1).

The exclusive DAPs found in the YGD cultivar included a putative mitochondrial cytochrome c1-1 heme protein (AT5G40810), the ribonuclease TUDOR 1 (AT5G07350), the coatomer subunit alpha-1 (AT1G62020), the ERBB-3 binding protein 1 (AT3G51800), the adenylate kinase 4 (AT5G50370), and the hypothetical protein COCNU_10G003830 (AT3G20050), which were down-accumulated along the maturation event (Table S1). In contrast, DAPs, such as actin-101 (AT5G09810), putative mitochondrial heat shock 70 kDa (AT5G09590), proteasome subunit alpha type-5 (AT3G14290), putative mitochondrial formate dehydrogenase (AT5G14780), aminotransferase ALD1 (AT4G33680), serine/threonine-protein kinase SAPK7 (AT1G10940), and glutathione S-transferase 3 (A0A8B7C1W8), increased in the intermediate stage but decreased in the mature stage (Table S1).

In the MPT coconut, DAPs included polyadenylate-binding protein 5 (A0A8B7MTK7), acetylornithine deacetylase (AT4G17830), putative isopentenyl-diphosphate delta-isomerase I (AT5G16440), mitochondrial gamma-aminobutyrate transaminase 1 (AT3G22200), actin-101 (AT5G09810), two-component response regulator 24 (A0A8B7D411), which were down-accumulated throughout ripening (Table S1). On the contrary, DAPs such as hypothetical protein COCNU_08G003720 (AT3G01650), putative low-temperature-induced 65 kDa protein (A0A8B7C9J1), putative coatomer beta'-1 (AT1G79990), FT-interacting protein 7 (AT3G57880), putative seed maturation protein (AT3G10960), monocopper oxidase-like protein SKU5 (AT4G12420), and cinnamoyl-CoA reductase 1 (AT2G33590) were up-accumulated throughout the maturation process (Table S1).

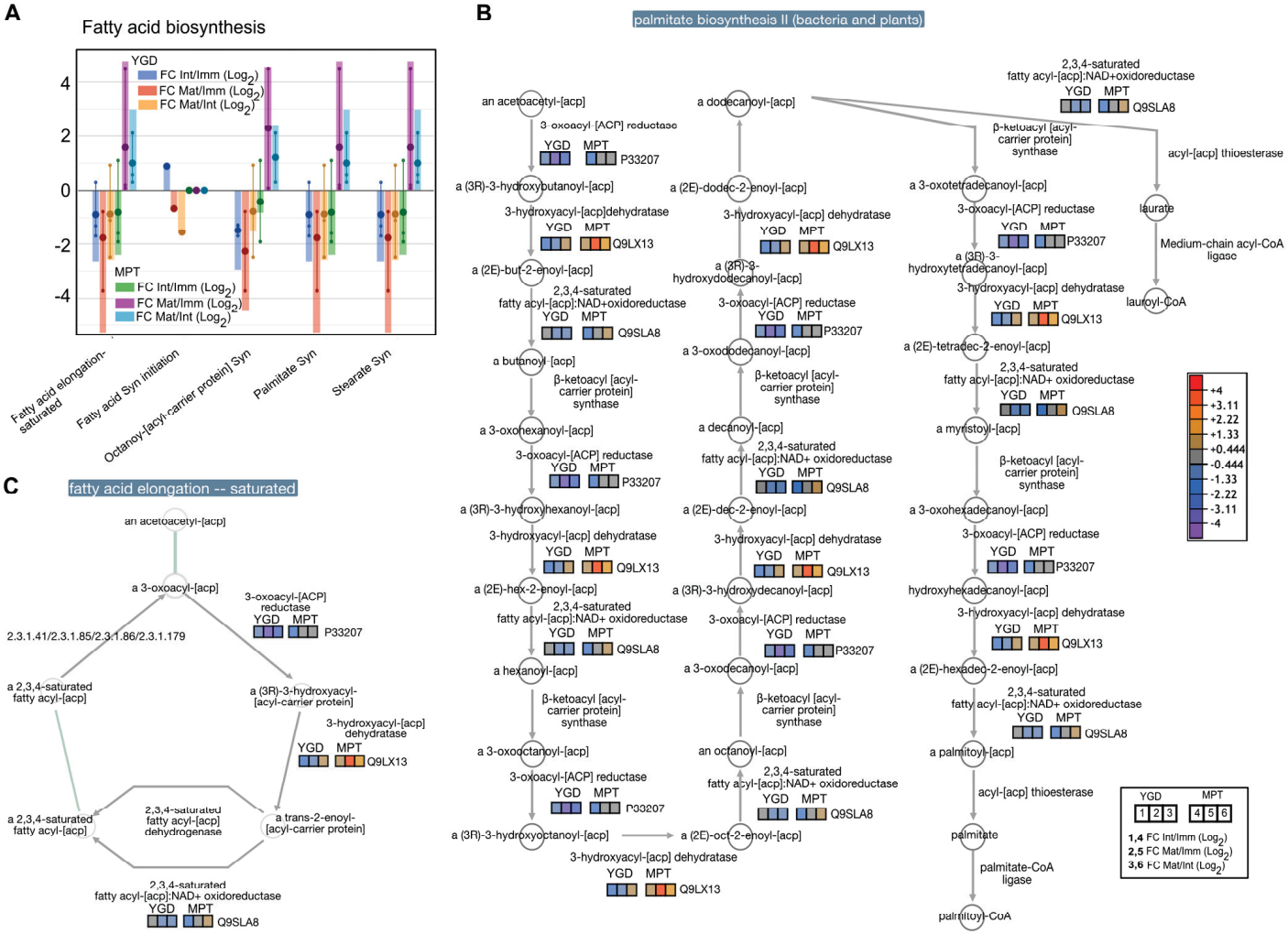


Figure 7. Visual representation of differential proteins associated with the fatty acid biosynthetic pathways. Pattern of accumulation of lipid-related proteins determined using TMT-SPS-MS3. Small circles specify the log2FC values of each protein; the big circles are the average of the log2FC values, and columns indicate the sum of the log 2FC (**A**). Dynamic of identified differential enzymes in the palmitate biosynthetic (**B**) and saturated fatty acid elongation pathways (**C**). The pathway reconstruction was carried out by using the "cellular overview" tool in the Plant Metabolic Network (PMN) database platform.

3. Discussion

Metabolic changes linked to physiological, biochemical and molecular processes associated with seed and fruit maturity are influenced by endogenous and external signals [20]. Changes include modifications of the cell wall, the accumulation and degradation of starch and triacylglycerol (TAG), among others [21]. On the other hand, color, hardness/softening and sugar content are used to monitor fruit development and maturity [22]. Most studies classify the maturity of the fruits into three main stages: unripe, breaker and ripe, as in *Persea americana* Mill. [23], or endosperm development, embryo development, seed dehydration and maturation like in *Brassica napus* L. [24].

Coconut seed ripening occurs in three stages: immature, intermediate, and mature [10]. However, the maturation of the coconut fruit does not follow the same pattern as other fruits; coconut meat begins to develop at the opposite end to the stalk and the embryo; it grows during maturation, and the amount of different proteins in the solid endosperm increases [25]. In the immature stage, the endosperm of the coconut has a soft consistency and hardens as it matures, and the fruit loses water. In this report, coconut cultivars YGD and MPT were compared; the MPT coconut was heavier and contained a thicker solid endosperm than YGD (Figures 1 and S1). According to Farooq et al. [26], the development of the endosperm requires large amounts of energy, which is fulfilled mainly by glycolysis.

The regulation of fatty acid and carbohydrate metabolisms in the developing endosperm of oil seeds is less understood than in legumes [13,27]. In this work, an enrichment analysis of biological processes identified proteins associated with the response to stress, particularly to heat and viruses, as the most represented core proteome in the YGD and MPT cultivars (Figure 2A–C). The identification of stress response proteins among enriched biological processes is consistent with the fact that coconut palm grows near coastal areas, where solar radiation and heat are high [28].

Proteomics showed metabolic differences between both cultivars (Figure 2B,C). Regarding carbohydrate metabolism (photosynthesis and glycolytic processes), the coconut is a C3 plant, i.e., initial CO_2 fixation is in ribulose 1,5-bisphosphate that immediately breaks down into 1,3-bisphophoglycerate. After three rounds of CO_2 fixation, the first 3-phosphoglycerate is biosynthesized and channeled to produce sucrose or starch, or to triose catabolism to produce TCA cycle intermediaries and NADH, FADH2 and ATP for cell redox balance and energy homeostasis [29]. Glycolysis was higher in MPT than in YGD (Figure 2B,C), and the former accumulated greater amounts of lipids in the solid endosperm. This was expected since lipid biosynthesis requires carbon and energy sources, which are provided by glycolysis and the TCA cycle [30].

Increased glycolytic flux results in the accumulation of methylglyoxal, a toxic byproduct of glycolysis [31]. This explains why lactoylglutathione lyase and methylglyoxal reductase, proteins involved in methylglyoxal detoxification, were over-accumulated as ripening progressed in the MPT coconut; meanwhile, these enzymes were stably produced in YGD (Figure 4). The accumulation of methylglyoxal is evident in many plants with intracellular oxidative stress [32,33]. Similarly, the proteins involved in coping with oxidative stress were also up-accumulated with the progress of ripening in MPT. Superoxide dismutase, which detoxifies superoxide and catalase, an efficient hydrogen peroxide scavenger, was up-accumulated in MPT coconut.

The PCA evidently grouped the replicates of immature, intermediate and mature fruits from each coconut cultivar, which supported the reproducibility of the sample collection, processing and analysis (Figure 3A,E). The volcano plots allowed for an analysis of the fold change vs. the significance of protein up-accumulation in the Int/Imm, Mat/Imm and Mat/Int in the YGD and MPT cultivars. The ALATA 2 protein, detected in the Int/Imm of the YGD, is involved in the transamination reactions between L-alanine and 2-oxoglutarate, and the reverse reaction between L-glutamate and pyruvate, an essential intermediate for the metabolism of carbohydrates and lipids. On the other hand, the UDP-glucose 4 epimerase (GEPI48) catalyzes the interconversion of UDP-galactose (UDP-Gal)

to UDP-glucose (UDP-Glc), thus channeling glucose towards more complex carbohydrate biosynthesis or carbohydrate catabolism to produce pyruvate [34,35].

In the intermediate stage of YGD, there were down-accumulated proteins involved in chaperone refolding activity (FAM10), as well as annexin D1 (ANXD1), a membrane-bound protein involved in the reduction in oxidative stress [36] (Figure 3B), suggesting that the transition from the immature to intermediate stage in this cultivar is a step with low proteolytic activity and oxidative stress. Likewise, in the mature stage compared to the immature stage, TCPQ, a subunit of chaperons that assists in the correct folding of nascent proteins [37], and HSP904, another molecular chaperone involved in the response to various stresses [38], were down-accumulated (Figure 3C), as well as the BCCP complex protein that regulates the first step in fatty acid biosynthesis [39], and proteins involved in fatty acid synthesis initiation and fatty acid elongation saturation (Figure 7A). Therefore, the transition to the mature stage in YGD occurred with low oxidative stress and, more importantly, it was strongly associated with a decrease in proteins involved in lipid biosynthesis (Figure 3C). In the mature stage, compared to the intermediate stage in cultivar YGD, maturity was also accompanied by a decrease in proteins related to carbohydrate metabolism, for example, phospho-fructokinase-6 (PFK6) and dihydrolipoyl dehydrogenase 2 (LPD2) (Figure 3D), in agreement with the decrease in proteins involved in the synthesis of glycan gluconeogenesis, sucrose and starch, among others, observed in our study (Figure 5A).

In the case of the MPT cultivar, the transition from the immature to intermediate stage (Figure 3F,G) was associated with an accumulation of proteins involved in development, for example, the calcium-binding EF-hand protein family (CbEF), a biological switch that regulates signal transduction by Ca^{2+} and controls cell growth and plant development [40]. Similarly, there was an accumulation of proteases, like carboxypeptidases, that supply nitrogen. Other up-accumulated proteins included glucan endo-1,3-β-glucosidase (GbG), which is responsible for β-glucan cleavage and releases glucan oligosaccharides, and the early nodulin-like protein 1 (Enod1), which transports glucose, fructose or sucrose for plant development [41] (Figure 3F). These proteins increased throughout maturation (Figure 3G,H), consistent with the cell expansion and development of the seed and fruit.

The proteins most down-accumulated in the intermediate stage compared to the immature stage in MPT were UDP-glucose 6-dehydrogenase 3 (UGDH3) and sucrose synthase 4 (SUS4) (Figure 3F), with the former being involved in the synthesis of hemicellulose for the plant cell wall [42] and the latter catalyzing the reversible cleavage of sucrose into fructose and UDP-glucose or ADP-glucose, which could then be channeled into energy production or for cell wall synthesis [43]. In MPT, the differentiation from the immature to intermediate solid endosperm was accompanied by a decrease in hemicellulose biosynthesis, possibly preventing the reinforcement of the cell wall, therefore facilitating cell expansion.

In the mature stage, compared to the immature stage in MPT, the down-accumulated proteins included a biotin carboxylase (BC) enzyme that carboxylates the N1′ of biotin [44]; glycerol-3-phosphate dehydrogenase (SDP6) that reduces dihydroxyacetone phosphate to glycerol-3-phosphate, a key step in glycerolipid metabolism [45]; and the phospho-fructokinase-6 (PFK6), a key enzyme in the carbohydrate metabolism [46] (Figure 3G). This is consistent with the decrease in proteins involved in the biosynthesis of carbohydrates (Figure 5A) and fatty acids (Figure 7A) observed in the mature MPT.

Comparing the proteomics of the mature stage and the intermediate stage in MPT, a down-accumulation of aldehyde dehydrogenase (ALDH) was observed in the mature stage (Figure 3H). ALDH has been described in the biosynthesis of aromatic volatiles during maturation in postharvest banana [47] and in the degradation of fatty acids [48]; its down-accumulation in the MPT coconut is understood since solid endosperm from the mature fruits of MPT is used for copra production due to its high fatty acid content.

Regarding APX2, it detoxifies the H_2O_2 during biotic and abiotic stresses [49]. APX2 was down-accumulated in the mature stage, while glutathione S-transferase (GST) was up-accumulated (Figure 3H), showing that antioxidant proteins are well coordinated in

the MPT endosperm (Figure 4) to cope with different kinds of stresses. A primary supply of myo-inositol is more necessary in the intermediate stage compared to the mature stage, since the INO2 protein was down-accumulated in the latter stage. This coincides with reports in *Arabidopsis thaliana* and *Glycine max*, where the INO2 enzyme is necessary for the normal development of endosperm and embryo [50].

YGD and MPT coconut cultivars also showed differences in gluconeogenesis (Figure 5B). During gluconeogenesis and the Calvin–Benson cycle, the photosynthetic activities of the phosphoglycerate kinase isoenzyme 1 and 3 were down-accumulated in the YGD cultivar; the decrease in photosynthetic activity and glycolysis may be a strategy for metabolic adjustment to optimize plant or fruit growth [51]. Yang et al. [52] reported that most DAPs are associated with carbohydrate metabolism in other varieties of coconut in response to low temperature. In *Camellia oleifera*, Ye et al. [48] found that the metabolism of carbohydrates and lipids was among the DAPs, but each metabolism integrated the results from all stages, and, thus, the dynamics of DAPs between these two plants cannot be compared stage by stage.

Enolase catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate (PEP), an important step in all classes of glycolysis (I, II, and IV), (Figures 5C and S3). Hassan et al. [53] found that enolase was up-accumulated during the development of the mesocarp of oil palm, related to the production of oil in the mesocarp. In other fruits, such as Chinese bayberry [54] and apricot [55], the production of enolase protein also increased throughout ripening. In our work, it was found that enolase protein was more accumulated in MPT than in YGD during seed maturation. Again, this is in agreement with the use of MPT solid endosperm for oil production. As a product of glycolysis, pyruvate is converted to acetyl-CoA by the pyruvate dehydrogenase (PDH) complex, and acetyl-CoA has several fates: as a substrate for the glyoxylate cycle, the TCA cycle, or a precursor for fatty acid elongation and lipid biosynthesis [15,56]. Glycolysis was more active in MPT than in YGD (Figure 5C), and pyruvate may be channeled to support lipid biosynthesis in this case.

Regarding malate dehydrogenase (MDH), a key enzyme involved in carbohydrate and lipid metabolisms, it was found to be up-accumulated in the TCA cycle in the MPT cultivar (Figure 6), similar to seeds of castor bean, where glycolysis and TCA were among the most active metabolic processes [57]. Likewise, in *Brassica napus*, MDH was up-accumulated in the cytosol and mitochondria of seeds having a higher content of oil [58], supporting a key role of MDH in oilseeds.

In plants, the synthesis of fatty acids occurs in the plastids while their breakdown occurs in the peroxisomes, the glyoxysomes, and to a lesser degree in the mitochondria, and elongation occurs in the endoplasmic reticulum [59,60]. However, although the compartment regulation of plant lipid biosynthesis is clear, the temporal and spatial regulation diverge among plants, and oil content in seeds greatly varies in quantitative and qualitative terms. Studies show that lauric acid is 65% in *C. nucifera* and it is 85% in *E. guineensis* [61,62], while oleic acid predominates in *Arachis hypogaea* (45–56%), *P. americana* (60%), and *Styrax tonkinensis* (60%) [63–65]. According to the previous findings, it is necessary to further understand the metabolic pathways of carbohydrates and lipids of oilseeds such as coconut, since there may be differences among the plant families regarding the regulation of these metabolic pathways [15]. In the case of the coconut tree, Maskromo et al. [66] and Tenda et al. [11] described significant differences in the fatty acid content of tall and dwarf varieties, which allows us to hypothesize that there is a correlation between the fatty acid content and the accumulation of proteins related to that metabolism.

There was a higher accumulation of proteins involved in the biosynthesis of fatty acids in the intermediate and mature stages of the MPT cultivar than in YGD (Figure 7A), suggesting a high production of acetyl-CoA. These results are in concordance with coconut seed uses: MPT for copra and oil extraction and YGD for water consumption (Prades et al., 2012). Acetyl-CoA carboxylase (ACCase) regulates the most essential step in fatty acid biosynthesis by converting acetyl CoA to malonyl CoA [67]. In oilseeds, acetyl-CoA combines with oxaloacetate in the glyoxylate cycle; then, succinate, one of the products

of this cycle, enters mitochondrial TCA and gluconeogenesis, where it is a precursor to hexoses, necessary for postgermination growth [68]. Likewise, the enzyme 3-hydroxyacyl-ACP dehydratase, which transforms 3-hydroxyacyl-ACP to enoyl-ACP and enoyl-ACP reductase to acyl-ACP in palmitate biosynthesis and fatty acid elongation, showed the greatest up-accumulation in the MPT cultivar (Figure 7B,C), similar to that reported in *B. napus*, *R. communis*, *P. americana*, and the mesocarp of *E. guineensis* [69].

It can be said that during the development and germination of the seed, the proteome is greatly represented by proteins with enzymatic activities, such as those involved in the metabolism of lipids and carbohydrates, while the proteome associated with structural or storage functions is represented by globulins, vicilin, and oleosins, among others, which play a crucial role in supporting cell shape and structure or as energy reserves for the development of embryos and seedlings [70].

In our work, during the protein profiling of coconut solid endosperm at different stages of development of YGD and MPT, it was observed that vicilin, glutelin type A-1, 11S globulin, cocosin 1, late embryogenesis abundant protein 3 (LEA), and heat shock 70 kDa proteins were detected in the "core proteome" (Table S1). Our results were in line with those of Garcia et al. [71], who reported 11S globulin and 7S vicilin in the solid coconut endosperm; the former, called cocosin, is known as the main coconut seed storage protein. Similarly, Patil and Benjakul [72] reported glutelin, and Dave [73] found oleosins involved in the stabilization of oil bodies in the solid endosperm. Our results evidenced that coconut storage proteins and heat shock proteins were distributed from the immature to mature stages in YGD and MPT cultivars, suggesting that their presence is crucial for endosperm development, accumulation and structural formation. Biochemical evidence in *Arabidopsis* suggests that storage proteins contribute to seed germination vigor, support early seedling growth, and could have a role in seed longevity [74]. Regarding the heat-shock proteins, these are important in fruit maturation, protein stabilization, protein refolding, and protection against environmental stresses [30,33].

4. Materials and Methods

4.1. Plant Material Collection

Yucatan green dwarf and Mexican pacific tall coconuts were harvested from a coconut plantation located at 20°24′03.95″ N and 89°30′56.86″ W in San Crisanto, Yucatan, Mexico. The youngest inflorescence is at the top of the palm, closest to the apical meristem, and shows a closed spathe (0 stage, not one ovule fertilized in the ovaries). Coconut palms produce one inflorescence per month, with the first fertilized ovules occurring in inflorescence at stage 1, so from inflorescence stage 0, the age of inflorescences and fruits increases each month in the direction to the ground [75,76]. The stages of fruit maturity were classified as immature for fruits from 6 to 8 months, intermediate for fruits from 9 to 10 months, and mature for fruits from 11 to 14 months. The fruits were disinfested with sterile distilled water and manually dehusked, the endocarp was cut with a sterile steel knife, and 20 g of meat was stored at −80 °C until further processing.

4.2. Proteome Extraction from Coconut Solid Endosperm

Proteins were extracted according to Islas-Flores et al. [77], with modifications. Briefly, one gram (1 g) of solid endosperm was ground to a fine powder with a sterile mortar and pestle in liquid nitrogen, and then 1 mL of buffer was added, which contained 50 mM Tris-HCl, pH 7.4, 50 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM EGTA, 5 mM ß-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), and $\frac{1}{4}$ of pill protease inhibitor cocktail (Roche, Basel, Switzerland). After homogenization, it was centrifuged for 30 min at 13,000 rpm and 4 °C, and the supernatant was stored at −20 °C until use. Each sample from each developmental stage was extracted individually; biological replicates were performed in triplicate.

4.3. Protein Quantification and Analysis by SDS-PAGE

Protein concentration in each supernatant was determined according to the Bradford method [78]. Protein analysis (30 µg) was performed on 12% SDS-PAGE gel and electrophoresed at 80 V for 2 h at room temperature. After electrophoresis, the gels were incubated overnight in Coomassie blue stain and distained in a solution containing 10% acetic acid, 25% methanol, and 65% sterile distilled water.

4.4. Sample Preparation for Proteomics

For each sample, 1200 µg of proteins was precipitated with chloroform–methanol according to Burgos-Canul et al. [79]. The proteins (100 µg per sample) were reduced for 45 min at 60 °C with 10 mM Tris (2-carboxyethyl) phosphine (TCEP), alkylated with 30 mM iodoacetamide (IA) for 1 h at room temperature in the dark, then quenched with 30 mM dithiothreitol (DTT). The proteins were then precipitated overnight at −20 °C in cold acetone and centrifuged at 10,000× *g* for 15 min at 4 °C. The pellets were dried for 15 min at room temperature, and the proteins were suspended in 100 µL of digestion buffer containing 50 mM triethylammonium bicarbonate (TEAM) and 0.1% sodium dodecyl sulfate (SDS) and digested overnight at 37 °C with trypsin 1:30 (*w*/*w*) (Promega, Madison, WI, USA, Trypsin Gold, Mass Spectrometry Grade). After this period, freshly prepared trypsin was added 1:60 w/w and incubation occurred for 4 h at 37 °C. The digestion reaction was stopped at 80 °C [80]. The trypsin-digested peptides were desalted and dried at room temperature.

4.5. TMT Labeling and Peptide Fractionation

Tandem mass tags (TMT), in this case TMT-9plex reagent, were added with acetonitrile and gently agitated. Peptide samples were resuspended in 100 mM triethylammonium bicarbonate (TEAB) buffer for 5 min and conjugated with TMT at room temperature for 1 h, following the manufacturer's instructions. The Yucatan green dwarf (YGD) and Mexican pacific tall (MPT) samples were labeled and analyzed separately. For each cultivar, the peptides of coconut in the immature stage were labeled with the reagents 126, 127N, and 128N; the intermediate stage proteins were labeled with 128C, 129N, and 129C; and the mature stage proteins were labeled with 130N, 130C, and 131. Peptide labeling was stopped with 5% hydroxylamine, incubating at room temperature for 15 min. All samples from the same variety were pooled and fractionated using high-pH reversed-phase C18 cartridges (Thermo Scientific, San Jose, CA, USA). The pooled samples were desalted with C18 cartridges, cleaned with ZipTip, and dried using a CentriVap (Labconco, Kansas City, MO, USA).

4.6. Nano-LC-MS/MS and Synchronous Precursor Selection (SPS)-MS3

The dried samples were dissolved with 0.1% formic acid in LC-MS-grade water (solvent A) and 20 µL of this solution was injected onto a nano-LC platform (UltiMate 3000 RSLC system, Dionex, Sunnyvale, CA, USA) through a nanoviper C18 trap column (3 µm, 75 µm × 2 cm, Dionex), and fractionated on an EASY spray C-18 RSLC column (2 µm, 75 µm × 25 cm) at a flow rate of 300 nL/min.

The mobile-phase gradient was as follows: a 10 min gradient was set using Solvent A and 0.1% formic acid in 90% acetonitrile (Solvent B), 10 min Solvent A, 7–20% Solvent B for 25 min, 20% Solvent B for 15 min, 20–25% Solvent B for 15 min, 25–95% Solvent B for 20 min, and 8 min Solvent A. The nano-LC was hyphenated to an Orbitrap Fusion Tribrid (Thermo-Fisher Scientific, San Jose, CA, USA) mass spectrometer equipped with an "EASY Spray" nano-ion source (Thermo-Fisher Scientific, San Jose, CA, USA). The mass spectrometer was operated in positive ionization mode with the nanospray voltage set at 3.5 kV and a source temperature of 280 °C. Caffeine, Met-Arg-Phe-Ala (MRFA) and Ultramark 1621 were used for external calibration.

Full MS scans were performed on the Orbitrap analyzer at a resolution of 120,000 (FWHM), scan range of 350–1500 m/z, AGC of 2.0×10^5, maximum injection time of 50 ms,

intensity threshold of 5.0e3, dynamic exclusion 1 at 70 s, and mass tolerance of 10 ppm. For the MS2 analysis, the 20 most abundant MS1s were isolated with charge states set to 27. The fragmentation parameters included CID with 35% collision energy and activation Q of 0.25, an AGC of 1.0×10^4, a maximum injection time of 50 ms, a precursor selection mass range of 400–1200 m/z, a precursor ion exclusion width low of 18 m/z and high of 5 m/z, isobaric tag loss TMT and in-trap detection ion; MS3 spectra were acquired using a 10-notch SPS isolation. MS3 precursors were fragmented using HCD with 65% collision energy and analyzed using the Orbitrap at 60,000 resolution with a 120–500 m/z scan range, 2 m/z isolation window, 1.0×10^5 AGC and maximum injection time of 120 ms with 1 microscan.

The spectra acquired using MS/MS and (SPS)-MS3 were analyzed with the program Proteome Discoverer 2.4 (Thermo-Fisher Scientific Inc.). Data were processed using the search engines Amanda and Mascot server (version 2.4.1, Matrix Science, Boston, MA, USA) and the Sequest HT algorithm, in which searches were conducted against *Arabidopsis* and the sequences of those of the Arecaceae family downloaded from the NCBI protein database. The sequences were concatenated into one single file and were filtered to remove the undetermined amino acids. To reduce the protein redundancy, sequences were collapsed with the CDHIT [81] program using 0.98, 5, and 0 for -c, -n, and -t parameters, respectively. Full-tryptic protease specificity, two missed cleavages, carbamidomethylation of cysteine (+57.021 Da) and TMT 9-plex N-terminal/lysine residues (+229.163 Da) were considered. Methionine oxidation (+15.995 Da) and deamidation in asparagine/glutamine (+0.984 Da) as dynamic modifications were considered as well.

Protein identification was carried out at a lower resolution in the linear ion trap with tolerances of $\pm$10 ppm and $\pm$0.6 Da. Peptide hits were filtered for a maximum of 1% false discovery rate (FDR) using the Percolator algorithm [82]. In addition, TMT quantification was performed at the MS3 level in the Orbitrap analyzer with the precursor co-isolation filter set to 45%. Data are available via ProteomeXchange with identifier PXD036949 (https://www.ebi.ac.uk/pride/, accessed on 22 September 2022).

4.7. Bioinformatics Analyses

Protein and peptide abundance were generated on the Proteome Discoverer 2.4 platform. Normalization was carried out based on the total peptide amount. The fold change in protein abundance was determined in the ratio Int/Imm, Mat/Imm, and Mat/Int. Differential proteins (fold change >1.5 and <0.66 (log2 fold change >0.59 and <$-$0.59)) were analyzed based on biological process gene ontology (GO) enrichment using the platform DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/summary.jsp, accessed on 22 September 2022), and the resulting GO terms were then summarized by finding a representative subset with a simple clustering algorithm that relies on semantic similarity measures using the REVIGO software (http://revigo.irb.hr, accessed on 22 September 2022). We used protein homologues to *Arabidopsis*, *E. guineensis,* and *P. dactylifera* as reference databases. The output data were presented using Treemaps, displaying a two-level hierarchy of GO terms—each rectangle is a single cluster representative, and the representatives were joined into "superclusters" of loosely related terms. PCA-based normalized protein abundances were calculated to visualize the differences in protein abundances between conditions. Log2 ratios Int/Imm, Mat/Imm, and Mat/Int; significant values (p-value $-$log10), and a Manhattan distance as a criterion for ranking hits in the platform VolcaNoseR (https://huygens.science.uva.nl/VolcaNoseR2/, accessed on 22 September 2022) were calculated. The plant metabolic database (https://plantcyc.org/, accessed on 22 September 2022) was used and samples were computed using the cell overview/omics viewer tool to generate pathway diagrams and protein accumulation intensities based on the Int/Imm, Mat/Imm, and Mat/Int ratios.

5. Conclusions

The present work adds new information to the proteomic landscape associated with coconut solid endosperm, focusing on the metabolisms of carbohydrates and lipids in the

immature, intermediate, and mature stages in two contrasting coconut cultivars, MPT and YGD. The former is a high lipid and fatty acid producer, while the latter is a less efficient producer of lipids but is highly valued as young (immature) fruits because of the taste of its coconut water. Interestingly, proteins involved in carbohydrate metabolism, e.g., in glycolysis, were more greatly accumulated in the mature stage of MPT than in YGD; this most likely occurs in order to synthesize precursors for fatty acids through enzymes such as MDH and enolase, whose products may be channeled for fatty acid production. Carbohydrate metabolism was higher in the immature stage of YGD, but was reduced in the intermediate and mature stages. Future studies combining proteomics analysis with metabolomics may help to elucidate the fate of those carbohydrates during the intermediate and mature stages of solid endosperm development. In agreement with our hypothesis, the proteins of solid endosperm involved in fatty acid metabolism were most accumulated in the intermediate and mature stages of MPT compared to YGD.

By investigating dynamic proteomics in carbohydrate and lipid metabolisms during solid endosperm maturation, our results, for first time, provide findings that contribute to a better understanding of the biochemical events related to maturation of the coconut. Although proteomics revealed differences in carbohydrate and lipid metabolism between MPT and YGD, the regulation of these metabolisms was found to be conserved with other fruits, suggesting that the regulation of carbohydrate and lipid metabolisms is ancient and preserved, regardless of the taxonomy or physiology of the fruits. Further analyses, for example focusing on the coconut endocarp, mesocarp, and pericarp, can help to unravel metabolic differences between coconut and other fruits, e.g., climacteric fruits. The present work greatly expands the current knowledge surrounding the fundamental biology of this non-climacteric species, which, in general terms, is a group still poorly studied.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241310431/s1.

Author Contributions: Conceptualization: I.I.-F., B.C.-C., E.R.-M., C.O.-S. and J.W.F.; software: E.R.-M., E.B.-V. and J.A.Z.-B.; data curation and formal analysis: J.W.F., M.I.G.-A., R.G.-T. and E.R.-M.; original draft preparation: I.I.-F., B.C.-C., E.R.-M. and J.W.F.; writing—review and editing: B.C.-C., E.R.-M., I.I.-F., C.O.-S., J.W.F., M.T.-S., E.B.-V. and J.A.Z.-B.; visualization: E.R.-M. and M.T.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), México, grant number CB-2017-2018-A1-S-10398.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Maxim V. Berezovski for support with the proteomic facilities which carried out the first coconut solid endosperm analysis, and José Miguel Elizalde Contreras for their technical assistance and sample processing. The authors also thank CONAHCyT for funding the project CB-2017-2018-A1-S-10398 and for the scholarships granted to Jean Wildort Félix (775223), María Inés Granados Alegría (774063), and Rufino Gómez Tah (774047) for their doctoral studies.

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

References

1. Niral, V.; Jerard, B.A.; Rajesh, M.K. Germplasm Resources: Diversity and Conservation. In *The Coconut Genome*; Rajesh, M.K., Ramesh, S.V., Perera, L., Kole, C., Eds.; Compendium of Plant Genomes; Springer International Publishing: Cham, Switzerland, 2021; pp. 27–46. ISBN 978-3-030-76648-1.
2. Burns, D.T.; Johnston, E.-L.; Walker, M.J. Authenticity and the Potability of Coconut Water—A Critical Review. *J. AOAC Int.* **2020**, *103*, 800–806. [CrossRef] [PubMed]
3. Kong, E.Y.Y.; Biddle, J.; Foale, M.; Panis, B.; Adkins, S.W. The Potential to Propagate Coconut Clones through Direct Shoot Organogenesis: A Review. *Sci. Hortic.* **2021**, *289*, 110400. [CrossRef]

4. Nartvaranant, P. Genetic Variations for "Nam Hom" Coconut (*Cocos nucifera* L.) Grownin the Western Region of Thailand Using AFLP Markers. *J. Thai Interdiscip. Res.* **2019**, *14*, 12. [CrossRef]
5. Ignacio, I.-F.; Miguel, T.-S. Research Opportunities on the Coconut (*Cocos nucifera* L.) Using New Technologies. *S. Afr. J. Bot.* **2021**, *141*, 414–420. [CrossRef]
6. Sudha, R.; Niral, V.; Samsudeen, K. Botanical Study and Cytology. In *The Coconut Genome*; Rajesh, M.K., Ramesh, S.V., Perera, L., Kole, C., Eds.; Compendium of Plant Genomes; Springer International Publishing: Cham, Switzerland, 2021; pp. 13–25. ISBN 978-3-030-76648-1.
7. Lantican, D.V.; Strickler, S.R.; Canama, A.O.; Gardoce, R.R.; Mueller, L.A.; Galvez, H.F. De Novo Genome Sequence Assembly of Dwarf Coconut (*Cocos nucifera* L. 'Catigan Green Dwarf') Provides Insights into Genomic Variation between Coconut Types and Related Palm Species. *G3 Genes Genomes Genet.* **2019**, *9*, 2377–2393. [CrossRef] [PubMed]
8. Boonkaew, T.; Mongkolsiriwatana, C.; Vongvanrungruang, A.; Srikulnath, K.; Peyachoknagul, S. Characterization of GA20ox Genes in Tall and Dwarf Types Coconut (*Cocos nucifera* L.). *Genes Genom.* **2018**, *40*, 735–745. [CrossRef]
9. Naik, M.; Sunil, C.M.; Rawson, A.; Venkatachalapathy, N. Tender Coconut Water: A Review on Recent Advances in Processing and Preservation. *Food Rev. Int.* **2022**, *38*, 1215–1236. [CrossRef]
10. Zheng, Y.; Jin, Y.; Yuan, Y.; Feng, D.; Chen, L.; Li, D.; Zhou, P. Identification and Function Analysis of a Type 2 Diacylglycerol Acyltransferase (DGAT2) from the Endosperm of Coconut (*Cocos nucifera* L.). *Gene* **2019**, *702*, 75–82. [CrossRef]
11. Tenda, E.; Miftahorrachman; Kumaunang, J. Profile of Amino Acids and Fatty Acids of Some Indonesia Tall Coconut Varieties. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *974*, 012133. [CrossRef]
12. Kumar, S.N. Variability in Coconut (*Cocos nucifera* L.) Germplasm and Hybrids for Fatty Acid Profile of Oil. *J. Agric. Food Chem.* **2011**, *59*, 13050–13058. [CrossRef]
13. Reynolds, K.B.; Cullerne, D.P.; El Tahchy, A.; Rolland, V.; Blanchard, C.L.; Wood, C.C.; Singh, S.P.; Petrie, J.R. Identification of Genes Involved in Lipid Biosynthesis through de Novo Transcriptome Assembly from *Cocos nucifera* Developing Endosperm. *Plant Cell Physiol.* **2019**, *60*, 945–960. [CrossRef] [PubMed]
14. Angeles, J.; Lado, J.; Pascual, E.; Cueto, C.; Laurena, A.; Laude, R. Towards the Understanding of Important Coconut Endosperm Phenotypes: Is There an Epigenetic Control? *Agronomy* **2018**, *8*, 225. [CrossRef]
15. Bourgis, F.; Kilaru, A.; Cao, X.; Ngando-Ebongue, G.-F.; Drira, N.; Ohlrogge, J.B.; Arondel, V. Comparative Transcriptome and Metabolite Analysis of Oil Palm and Date Palm Mesocarp That Differ Dramatically in Carbon Partitioning. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12527–12532. [CrossRef]
16. Naganeeswaran, S.; Fayas, T.P.; Rajesh, M.K. Dataset of Transcriptome Assembly of Date Palm Embryogenic Calli and Functional Annotation. *Data Brief* **2020**, *31*, 105760. [CrossRef]
17. D'Amato, A.; Fasoli, E.; Righetti, P.G. Harry Belafonte and the Secret Proteome of Coconut Milk. *J. Proteom.* **2012**, *75*, 914–920. [CrossRef]
18. Huang, J.; Liu, X.; Lan, Q.; Lai, X.; Luo, Z.; Yang, G. Proteomic Profile of Coconuts. *Eur. Food Res. Technol.* **2016**, *242*, 449–455. [CrossRef]
19. Ma, J.; Pan, C.; Chen, H.; Chen, W.; Chen, W.; Zhang, M.; Zhong, Q. Insight of the Functional and Biological Activities of Coconut (*Cocos nucifera* L.) Protein by Proteomics Analysis and Protein-Based Bioinformatics. *Molecules* **2022**, *27*, 2987. [CrossRef] [PubMed]
20. Corpas, F.J.; Palma, J.M. Nitric Oxide on/off in Fruit Ripening. *Plant Biol.* **2018**, *20*, 805–807. [CrossRef] [PubMed]
21. Zhang, W.; Li, X.; Li, L.; Tang, Y.; Qi, W.; Liu, X.; Qiao, L.; Wang, W.; Jia, X. A Label-Free Quantitative Proteomic Investigation Reveals Stage-Responsive Ripening Genes in Apricot Fruits. *J. Hortic. Sci. Biotechnol.* **2017**, *92*, 261–269. [CrossRef]
22. Williams, R.S.; Benkeblia, N. Biochemical and Physiological Changes of Star Apple Fruit (*Chrysophyllum cainito*) during Different "on Plant" Maturation and Ripening Stages. *Sci. Hortic.* **2018**, *236*, 36–42. [CrossRef]
23. Juarez-Escobar, J.; Guerrero-Analco, J.A.; Zamora-Briseño, J.A.; Elizalde-Contreras, J.M.; Bautista-Valle, M.V.; Bojórquez-Velázquez, E.; Loyola-Vargas, V.M.; Mata-Rosas, M.; Ruíz-May, E. Tissue-Specific Proteome Characterization of Avocado Seed during Postharvest Shelf Life. *J. Proteom.* **2021**, *235*, 104112. [CrossRef] [PubMed]
24. Zhou, Z.; Lin, B.; Tan, J.; Hao, P.; Hua, S.; Deng, Z. Tandem Mass Tag-Based Quantitative Proteomics Reveals Implication of a Late Embryogenesis Abundant Protein (BnLEA57) in Seed Oil Accumulation in *Brassica napus* L. *Front. Plant Sci.* **2022**, *13*, 907244. [CrossRef] [PubMed]
25. Jackson, J.C.; Gordon, A.; Wizzard, G.; McCook, K.; Rolle, R. Changes in Chemical Composition of Coconut (*Cocos nucifera*) Water during Maturation of the Fruit. *J. Sci. Food Agric.* **2004**, *84*, 1049–1052. [CrossRef]
26. Farooq, M.A.; Zhang, X.; Zafar, M.M.; Ma, W.; Zhao, J. Roles of Reactive Oxygen Species and Mitochondria in Seed Germination. *Front. Plant Sci.* **2021**, *12*, 781734. [CrossRef] [PubMed]
27. Miray, R.; Kazaz, S.; To, A.; Baud, S. Molecular Control of Oil Metabolism in the Endosperm of Seeds. *Int. J. Mol. Sci.* **2021**, *22*, 1621. [CrossRef]
28. Nair, K.P. The Coconut Palm (*Cocos nucifera* L.). In *Tree Crops*; Springer International Publishing: Cham, Switzerland, 2021; pp. 79–128. ISBN 978-3-030-62139-1.

29. Baslam, M.; Mitsui, T.; Sueyoshi, K.; Ohyama, T. Recent Advances in Carbon and Nitrogen Metabolism in C3 Plants. *Int. J. Mol. Sci.* **2020**, *22*, 318. [CrossRef]
30. Ramzan, A.; Shah, M.; Ullah, N.; Sheheryar; Nascimento, J.R.S.; Campos, F.A.P.; Domont, G.B.; Nogueira, F.C.S.; Abdellattif, M.H. Proteomic Analysis of Embryo Isolated from Mature *Jatropha curcas* L. Seeds. *Front. Plant Sci.* **2022**, *13*, 843764. [CrossRef]
31. Allaman, I.; Bélanger, M.; Magistretti, P.J. Methylglyoxal, the Dark Side of Glycolysis. *Front. Neurosci.* **2015**, *9*, 23. [CrossRef]
32. Hoque, T.S.; Hossain, M.A.; Mostofa, M.G.; Burritt, D.J.; Fujita, M.; Tran, L.-S.P. Methylglyoxal: An Emerging Signaling Molecule in Plant Abiotic Stress Responses and Tolerance. *Front. Plant Sci.* **2016**, *7*, 1341. [CrossRef]
33. Kalapos, M.P. The Tandem of Free Radicals and Methylglyoxal. *Chem. Biol. Interact.* **2008**, *171*, 251–271. [CrossRef]
34. Xu, Z.; Ma, J.; Qu, C.; Hu, Y.; Hao, B.; Sun, Y.; Liu, Z.; Yang, H.; Yang, C.; Wang, H.; et al. Identification and Expression Analyses of the Alanine Aminotransferase (AlaAT) Gene Family in Poplar Seedlings. *Sci. Rep.* **2017**, *7*, 45933. [CrossRef]
35. Fushinobu, S. Molecular Evolution and Functional Divergence of UDP-Hexose 4-Epimerases. *Curr. Opin. Chem. Biol.* **2021**, *61*, 53–62. [CrossRef]
36. Mortimer, J.C.; Laohavisit, A.; Macpherson, N.; Webb, A.; Brownlee, C.; Battey, N.H.; Davies, J.M. Annexins: Multifunctional Components of Growth and Adaptation. *J. Exp. Bot.* **2008**, *59*, 533–544. [CrossRef]
37. Zhao, X.; Xiu, J.; Li, Y.; Ma, H.; Wu, J.; Wang, B.; Guo, G. Characterization and Expression Pattern Analysis of the T-Complex Protein-1 Zeta Subunit in *Musca domestica* L. (Diptera). *J. Insect Sci.* **2017**, *17*, 90. [CrossRef]
38. Lin, X.; Wu, X.; Liu, X. Temperature Stress Response of Heat Shock Protein 90 (Hsp90) in the Clam Paphia Undulata. *Aquac. Fish.* **2018**, *3*, 106–114. [CrossRef]
39. Keereetaweep, J.; Liu, H.; Zhai, Z.; Shanklin, J. Biotin Attachment Domain-Containing Proteins Irreversibly Inhibit Acetyl CoA Carboxylase. *Plant Physiol.* **2018**, *177*, 208–215. [CrossRef]
40. Kundu, P.; Nehra, A.; Gill, R.; Tuteja, N.; Gill, S.S. Unraveling the Importance of EF-Hand-Mediated Calcium Signaling in Plants. *S. Afr. J. Bot.* **2022**, *148*, 615–633. [CrossRef]
41. Denancé, N.; Szurek, B.; Noël, L.D. Emerging Functions of Nodulin-Like Proteins in Non-Nodulating Plant Species. *Plant Cell Physiol.* **2014**, *55*, 469–474. [CrossRef] [PubMed]
42. Jia, T.; Ge, Q.; Zhang, S.; Zhang, Z.; Liu, A.; Fan, S.; Jiang, X.; Feng, Y.; Zhang, L.; Niu, D.; et al. UDP-Glucose Dehydrogenases: Identification, Expression, and Function Analyses in Upland Cotton (*Gossypium hirsutum*). *Front. Genet.* **2021**, *11*, 597890. [CrossRef] [PubMed]
43. Stein, O.; Granot, D. An Overview of Sucrose Synthases in Plants. *Front. Plant Sci.* **2019**, *10*, 95. [CrossRef] [PubMed]
44. Tong, L. Structure and Function of Biotin-Dependent Carboxylases. *Cell. Mol. Life Sci.* **2013**, *70*, 863–891. [CrossRef]
45. Zhao, Y.; Liu, M.; He, L.; Li, X.; Wang, F.; Yan, B.; Wei, J.; Zhao, C.; Li, Z.; Xu, J. A Cytosolic NAD+-Dependent GPDH from Maize (ZmGPDH1) Is Involved in Conferring Salt and Osmotic Stress Tolerance. *BMC Plant Biol.* **2019**, *19*, 16. [CrossRef]
46. Yao, Y.; Geng, M.-T.; Wu, X.-H.; Sun, C.; Wang, Y.-L.; Chen, X.; Shang, L.; Lu, X.-H.; Li, Z.; Li, R.-M.; et al. Identification, Expression, and Functional Analysis of the Fructokinase Gene Family in Cassava. *Int. J. Mol. Sci.* **2017**, *18*, 2398. [CrossRef]
47. Ueda, Y.; Zhao, W.; Ihara, H.; Imahori, Y.; Tsantili, E.; Wendakoon, S.; Chambers, A.; Bai, J. Functional Characteristics of Aldehyde Dehydrogenase and Its Involvement in Aromatic Volatile Biosynthesis in Postharvest Banana Ripening. *Foods* **2022**, *11*, 347. [CrossRef] [PubMed]
48. Ye, Z.; Yu, J.; Yan, W.; Zhang, J.; Yang, D.; Yao, G.; Liu, Z.; Wu, Y.; Hou, X. Integrative ITRAQ-Based Proteomic and Transcriptomic Analysis Reveals the Accumulation Patterns of Key Metabolites Associated with Oil Quality during Seed Ripening of *Camellia oleifera*. *Hortic. Res.* **2021**, *8*, 157. [CrossRef] [PubMed]
49. Wu, B.; Wang, B. Comparative Analysis of Ascorbate Peroxidases (APXs) from Selected Plants with a Special Focus on *Oryza sativa* Employing Public Databases. *PLoS ONE* **2019**, *14*, e0226543. [CrossRef] [PubMed]
50. Abid, G.; Silue, S.; Muhovski, Y.; Jacquemin, J.-M.; Toussaint, A.; Baudoin, J.-P. Role of Myo-Inositol Phosphate Synthase and Sucrose Synthase Genes in Plant Seed Development. *Gene* **2009**, *439*, 1–10. [CrossRef] [PubMed]
51. Rosa-Téllez, S.; Anoman, A.D.; Flores-Tornero, M.; Toujani, W.; Alseek, S.; Fernie, A.R.; Nebauer, S.G.; Muñoz-Bertomeu, J.; Segura, J.; Ros, R. Phosphoglycerate Kinases Are Co-Regulated to Adjust Metabolism and to Optimize Growth. *Plant Physiol.* **2018**, *176*, 1182–1198. [CrossRef]
52. Yang, Y.; Saand, M.A.; Abdelaal, W.B.; Zhang, J.; Wu, Y.; Li, J.; Fan, H.; Wang, F. ITRAQ-Based Comparative Proteomic Analysis of Two Coconut Varieties Reveals Aromatic Coconut Cold-Sensitive in Response to Low Temperature. *J. Proteom.* **2020**, *220*, 103766. [CrossRef]
53. Hassan, H.; Amiruddin, M.D.; Weckwerth, W.; Ramli, U.S. Deciphering Key Proteins of Oil Palm (*Elaeis Guineensis* Jacq.) Fruit Mesocarp Development by Proteomics and Chemometrics. *Electrophoresis* **2019**, *40*, 254–265. [CrossRef]
54. Chen, Y.-Y.; Zhang, Z.-H.; Zhong, C.-Y.; Song, X.-M.; Lin, Q.-H.; Huang, C.-M.; Huang, R.-H.; Chen, W. Functional Analysis of Differentially Expressed Proteins in Chinese Bayberry (*Myrica rubra* Sieb. et Zucc.) Fruits during Ripening. *Food Chem.* **2016**, *190*, 763–770. [CrossRef]
55. D'Ambrosio, C.; Arena, S.; Rocco, M.; Verrillo, F.; Novi, G.; Viscosi, V.; Marra, M.; Scaloni, A. Proteomic Analysis of Apricot Fruit during Ripening. *J. Proteom.* **2013**, *78*, 39–57. [CrossRef]

56. Correa, S.M.; Fernie, A.R.; Nikoloski, Z.; Brotman, Y. Towards Model-Driven Characterization and Manipulation of Plant Lipid Metabolism. *Prog. Lipid Res.* **2020**, *80*, 101051. [CrossRef] [PubMed]
57. Gu, J.; Hou, D.; Li, Y.; Chao, H.; Zhang, K.; Wang, H.; Xiang, J.; Raboanatahiry, N.; Wang, B.; Li, M. Integration of Proteomic and Genomic Approaches to Dissect Seed Germination Vigor in *Brassica napus* Seeds Differing in Oil Content. *BMC Plant Biol.* **2019**, *19*, 21. [CrossRef] [PubMed]
58. Yu, A.; Li, F.; Liu, A. Comparative Proteomic and Transcriptomic Analyses Provide New Insight into the Formation of Seed Size in Castor Bean. *BMC Plant Biol.* **2020**, *20*, 48. [CrossRef] [PubMed]
59. Fan, J.; Yu, L.; Xu, C. A Central Role for Triacylglycerol in Membrane Lipid Breakdown, Fatty Acid β-Oxidation, and Plant Survival under Extended Darkness. *Plant Physiol.* **2017**, *174*, 1517–1530. [CrossRef] [PubMed]
60. Zulu, N.N.; Zienkiewicz, K.; Vollheyde, K.; Feussner, I. Current Trends to Comprehend Lipid Metabolism in Diatoms. *Prog. Lipid Res.* **2018**, *70*, 1–16. [CrossRef] [PubMed]
61. Teng, M.; Zhao, Y.J.; Khoo, A.L.; Yeo, T.C.; Yong, Q.W.; Lim, B.P. Impact of Coconut Oil Consumption on Cardiovascular Health: A Systematic Review and Meta-Analysis. *Nutr. Rev.* **2020**, *78*, 249–259. [CrossRef] [PubMed]
62. Cui, J.; Lamade, E.; Tcherkez, G. Seed Germination in Oil Palm (*Elaeis guineensis* Jacq.): A Review of Metabolic Pathways and Control Mechanisms. *Int. J. Mol. Sci.* **2020**, *21*, 4227. [CrossRef]
63. Liu, H.; Li, H.; Gu, J.; Deng, L.; Ren, L.; Hong, Y.; Lu, Q.; Chen, X.; Liang, X. Identification of the Candidate Proteins Related to Oleic Acid Accumulation during Peanut (*Arachis hypogaea* L.) Seed Development through Comparative Proteome Analysis. *Int. J. Mol. Sci.* **2018**, *19*, 1235. [CrossRef]
64. Pedreschi, R.; Uarrota, V.; Fuentealba, C.; Alvaro, J.E.; Olmedo, P.; Defilippi, B.G.; Meneses, C.; Campos-Vargas, R. Primary Metabolism in Avocado Fruit. *Front. Plant Sci.* **2019**, *10*, 795. [CrossRef]
65. Wu, Q.; Chen, C.; Wang, X.; Zhang, Z.; Yu, F.; Guy, R.D. Proteomic Analysis of Metabolic Mechanisms Associated with Fatty Acid Biosynthesis during *Styrax tonkinensis* Kernel Development. *J. Sci. Food Agric.* **2021**, *101*, 6053–6063. [CrossRef]
66. Maskromo, I.; Karouw, S.; Pandin, D.S.; Mahayu, W.M.; Santosa, B.; Alouw, J.C. Physichochemical Properties of Kebumen Entog Dwarf Coconut. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *418*, 012037. [CrossRef]
67. Yu, L.; Fan, J.; Zhou, C.; Xu, C. Chloroplast Lipid Biosynthesis Is Fine-Tuned to Thylakoid Membrane Remodeling during Light Acclimation. *Plant Physiol.* **2021**, *185*, 94–107. [CrossRef] [PubMed]
68. Yang, Y.; Benning, C. Functions of Triacylglycerols during Plant Development and Stress. *Curr. Opin. Biotechnol.* **2018**, *49*, 191–198. [CrossRef] [PubMed]
69. Kilaru, A.; Cao, X.; Dabbs, P.B.; Sung, H.-J.; Rahman, M.M.; Thrower, N.; Zynda, G.; Podicheti, R.; Ibarra-Laclette, E.; Herrera-Estrella, L.; et al. Oil Biosynthesis in a Basal Angiosperm: Transcriptome Analysis of *Persea americana* Mesocarp. *BMC Plant Biol.* **2015**, *15*, 203. [CrossRef]
70. Antonets, K.S.; Belousov, M.V.; Sulatskaya, A.I.; Belousova, M.E.; Kosolapova, A.O.; Sulatsky, M.I.; Andreeva, E.A.; Zykin, P.A.; Malovichko, Y.V.; Shtark, O.Y.; et al. Accumulation of Storage Proteins in Plant Seeds Is Mediated by Amyloid Formation. *PLoS Biol.* **2020**, *18*, e3000564. [CrossRef] [PubMed]
71. Garcia, R.N.; Arocena, R.V.; Laurena, A.C.; Tecson-Mendoza, E.M. 11S and 7S Globulins of Coconut (*Cocos nucifera* L.): Purification and Characterization. *J. Agric. Food Chem.* **2005**, *53*, 1734–1739. [CrossRef] [PubMed]
72. Patil, U.; Benjakul, S. Coconut Milk and Coconut Oil: Their Manufacture Associated with Protein Functionality. *J. Food Sci.* **2018**, *83*, 2019–2027. [CrossRef] [PubMed]
73. Dave, A.C.; Ye, A.; Singh, H. Structural and Interfacial Characteristics of Oil Bodies in Coconuts (*Cocos nucifera* L.). *Food Chem.* **2019**, *276*, 129–139. [CrossRef]
74. Nguyen, T.-P.; Cueff, G.; Hegedus, D.D.; Rajjou, L.; Bentsink, L. A Role for Seed Storage Proteins in *Arabidopsis* Seed Longevity. *J. Exp. Bot.* **2015**, *66*, 6399–6413. [CrossRef] [PubMed]
75. Oropeza, C.; Cordova, I.; Chumba, A.; Narváez, M.; Sáenz, L.; Ashburner, R.; Harrison, N. Phytoplasma Distribution in Coconut Palms Affected by Lethal Yellowing Disease. *Ann. Appl. Biol.* **2011**, *159*, 109–117. [CrossRef]
76. Perera, P.I.P.; Hocher, V.; Weerakoon, L.K.; Yakandawala, D.M.D.; Fernando, S.C.; Verdeil, J.-L. Early Inflorescence and Floral Development in *Cocos nucifera* L. (Arecaceae: Arecoideae). *S. Afr. J. Bot.* **2010**, *76*, 482–492. [CrossRef]
77. Islas-Flores, I.; Oropeza, C.; Hernández-Sotomayor, S.M.T. Protein Phosphorylation during Coconut Zygotic Embryo Development1. *Plant Physiol.* **1998**, *118*, 257–263. [CrossRef] [PubMed]
78. Bradford, M.M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254. [CrossRef]
79. Burgos-Canul, Y.Y.; Canto-Canché, B.; Berezovski, M.V.; Mironov, G.; Loyola-Vargas, V.M.; Barba de Rosa, A.P.; Tzec-Simá, M.; Brito-Argáez, L.; Carrillo-Pech, M.; Grijalva-Arango, R.; et al. The Cell Wall Proteome from Two Strains of *Pseudocercospora fijiensis* with Differences in Virulence. *World J. Microbiol. Biotechnol.* **2019**, *35*, 105. [CrossRef] [PubMed]
80. Hernández-Pérez, A.; Zamora-Briseño, J.A.; Ruiz-May, E.; Pereira-Santana, A.; Elizalde-Contreras, J.M.; Pozos-González, S.; Torres-Irineo, E.; Hernández-López, J.; Gaxiola-Cortés, M.G.; Rodríguez-Canul, R. Proteomic Profiling of the White Shrimp *Litopenaeus vannamei* (Boone, 1931) Hemocytes Infected with White Spot Syndrome Virus Reveals the Induction of Allergy-Related Proteins. *Dev. Comp. Immunol.* **2019**, *91*, 37–49. [CrossRef] [PubMed]

81. Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
82. Käll, L.; Canterbury, J.D.; Weston, J.; Noble, W.S.; MacCoss, M.J. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. *Nat. Methods* **2007**, *4*, 923–925. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Review

The Role of Increased Expression of Sirtuin 6 in the Prevention of Premature Aging Pathomechanisms

Adrianna Dzidek [1], Olga Czerwińska-Ledwig [2], Małgorzata Żychowska [3], Wanda Pilch [2] and Anna Piotrowska [2,*]

[1] Doctoral School of Physical Culture Science, University of Physical Education, 31-571 Krakow, Poland
[2] Institute for Basic Sciences, Faculty of Physiotherapy, University of Physical Education, 31-571 Krakow, Poland
[3] Faculty of Health Sciences and Physical Culture, Biological Fundation of Physical Culture, Kazimierz Wielki University in Bydgoszcz, 85-064 Bydgoszcz, Poland
* Correspondence: anna.piotrowska@awf.krakow.pl

Abstract: Sirtuins, in mammals, are a group of seven enzymes (SIRT1–SIRT7) involved in the post-translational modification of proteins—they are considered longevity proteins. SIRT6, classified as class IV, is located on the cell nucleus; however, its action is also connected with other regions, e.g., mitochondria and cytoplasm. It affects many molecular pathways involved in aging: telomere maintenance, DNA repair, inflammatory processes or glycolysis. A literature search for keywords or phrases was carried out in PubMed and further searches were carried out on the ClinicalTrials.gov website. The role of SIRT6 in both premature and chronological aging has been pointed out. SIRT6 is involved in the regulation of homeostasis—an increase in the protein's activity has been noted in calorie-restriction diets and with significant weight loss, among others. Expression of this protein is also elevated in people who regularly exercise. SIRT6 has been shown to have different effects on inflammation, depending on the cells involved. The protein is considered a factor in phenotypic attachment and the migratory responses of macrophages, thus accelerating the process of wound healing. Furthermore, exogenous substances will affect the expression level of *SIRT6*: resveratrol, sirtinol, flavonoids, cyanidin, quercetin and others. This study discusses the importance of the role of SIRT6 in aging, metabolic activity, inflammation, the wound healing process and physical activity.

Keywords: sirtuins; SIRT6; gene expression; aging; inflammation; metabolism; aging pathomechanism; physical activity

Citation: Dzidek, A.; Czerwińska-Ledwig, O.; Żychowska, M.; Pilch, W.; Piotrowska, A. The Role of Increased Expression of Sirtuin 6 in the Prevention of Premature Aging Pathomechanisms. *Int. J. Mol. Sci.* **2023**, *24*, 9655. https://doi.org/10.3390/ijms24119655

Academic Editor: Wajid Zaman

Received: 19 April 2023
Revised: 23 May 2023
Accepted: 29 May 2023
Published: 2 June 2023

1. Introduction

Silent information regulator (Sir) proteins belong to NAD^+-dependent deacetylases, enzymes catalyzing the deacetylation reaction. Sirtuins were first discovered in the yeast *Saccharomyces cerevisiae* by virtue of their role in the establishment of transcriptional silencing of mating-type loci. Further studies have shown that Sir2 is also crucial in the partitioning of carbonylated proteins between mother and daughter cells, as well as for silencing at yeast telomeres and in the rDNA. The conserved enzymes called sirtuins, originally discovered in yeast, are produced by almost all organisms, from non-nucleated prokaryotes, to unicellular archaea and bacteria, to mammals [1,2].

Sirtuins play an important role in maintaining health and affect many pathways that increase the lifespan of organisms [1]. In mammals, they constitute a group of seven proteins (SIRT 1–7) belonging to class III histone deacetylases (HDACs). Sirtuins have a common catalytic domain of NAD^+, consisting of about 260 amino acid residues [2]. Individual sirtuin isoforms differ in sequence and length in the N- and C-terminal domains, which affects their enzymatic activity, cellular localization and substrate specificity [3].

Initially, sirtuins were studied in the context of organism aging and gene silencing, but many other biological functions of proteins belonging to this group have been revealed

in mammalian cells [4]. Studies on yeast aging showed a relationship between the *Silent Information Regulator 2 (Sir2)* gene and the viability of budding yeast [5].

Sirtuin 6 (SIRT6), belonging to class IV, is in the cell nucleus. SIRT6 promotes the repair of double-stranded DNA breaks by forming a complex with DNA-dependent protein kinase (DNA-PK) [6]. It has a hydrophobic pocket where the myristoylated group is located before being cleaved from the modified protein [7,8]. It has been suggested that this enzyme exhibits higher deacylase activity than deacetylase activity [4].

In a healthy organism or with minor damage, this protein promotes cell proliferation and activation of repair processes, while in the case of serious damage, it supports the apoptosis of damaged cells. It is in this mechanism that SIRT6 prevents the proliferation of damaged cancer cells [9,10]. SIRT6 slows down the course of gluconeogenesis by inhibiting the action of the PGC-1α factor. The activity of PGC-1α depends on the degree of its acetylation. This process is controlled by the acetyltransferase GCN5. SIRT6 causes deacetylation and phosphorylation of GCN5, which increases its acetylase activity. Over-acetylating PGC-1α, the acetyltransferase leads to a decrease in its activity and, consequently, to the inhibition of gluconeogenesis [11,12].

The association of sirtuins with aging is a well-known topic. Imai et al. [13] pointed to several mechanisms, including: effects on metabolism and regulation of circadian clock mechanisms, and emphasized that sirtuin activity decreases with age, which should be associated with a decrease in NAD^+. The most important sirtuin described by the paper's authors was SIRT1. Furthermore, it has been the most extensively studied protein of the mammalian sirtuin family. However, it has been shown that not only a loss of SIRT1 activity enhances DNA damage, but also SIRT6 [14]. Therefore, the authors of this review considered it interesting to examine the significance of SIRT6 in the aging process. The mechanisms of SIRT6 action indicate that this protein should be considered as an important element coordinating processes related to maintaining homeostasis and thereby inhibiting the premature aging of the organism. In this respect, knowledge of the mechanisms of SIRT6 action and factors modifying its expression becomes an interesting issue for a number of specialists. Therefore, the aim of this study was to indicate how this protein acts in a healthy organism, what changes are observed in pathological processes and what internal and external stimuli modify the process of *SIRT6* expression.

2. Structure

SIRT6 is a widely expressed protein [15]. Mahlknecht et al. [15] demonstrated that the human *SIRT6* genomic sequence has one single genomic locus, which spans a region of 8427 bp. The human *SIRT6* gene is located on chromosome 19p13.3 and consists of eight exons ranging in size from 60 bp (exon 4) to 838 bp (exon 8). The mRNA for human SIRT6 encodes a protein of 355 amino acids, with a predicted molecular weight of 39.1 kDa and an isoelectric point of 9.12 [9]. SIRT6 is composed of two main domains—a large domain containing a nucleotide binding and Rossmann fold (responsible for binding NAD^+) and a small domain containing a Zn^{2+} binding loop. Zinc ions play a role in maintaining the integrity of the catalytic domain and stabilizing the enzyme structure [16].

Activation and Reactivation of Gene Expression

The DNA molecule, which is a long strand, is wound around histone proteins that constitute a structural element of chromatin. The modification reaction of histone proteins is histone deacetylation, which involves the detachment of an acetyl group from the N-terminus of the histone. This reaction is carried out by histone deacetylases [17].

SIRT6, as a histone deacetylase, inhibits the transcriptional activity of several transcription factors and deacetylates specific histone lysine residues, such as H3K9 and H3K56, depending on NAD^+ [18].

The N-terminal extension of SIRT6 is important for chromatin association and the internal catalytic activity of the core domain [16]. As a non-histone protein deacetylase, SIRT6 deacetylates forkhead box O1 protein, pyruvate kinase M2, C-terminal binding

protein and GATA-binding protein 3 [18]. The C-terminal extension is essential for proper nuclear localization [16].

Although SIRT6-mediated histone deacetylation generally correlates with chromatin condensation and gene silencing [19,20], there is also evidence that SIRT6 can activate certain genes by mediating histone deacetylation. For example, SIRT6 acts as a transcriptional coactivator of erythroid 2-related factor 2 (Nrf2) to protect against oxidative stress in human mesenchymal stem cells, and it has been found that SIRT6 is in a protein complex with Nrf2 and RNAP II [21,22]. Nrf2 is one of the master regulators of antioxidant responses—binds to the antioxidant response elements (AREs) and activates antioxidant genes, for example, heme oxygenase 1 (HMOX1), which is known for counteracting reactive oxygen species (ROS). In aging cells, a decrease in Nrf2-ARE activity is observed, which may be caused by oxidative stress-related tissue degeneration [21]. Nrf2 activation occurs during evasion from Keap1-mediated proteasomal degradation in the cytosol [23]. SIRT6 activates Nrf2 by inhibiting Keap1 transcription and directly interacting with Nrf2 [24]. It is indicated that SIRT6 is required for HOMX1 induction in response to oxidative stress and SIRT6 deficiency increases basal cellular ROS [21] (Figure 1).

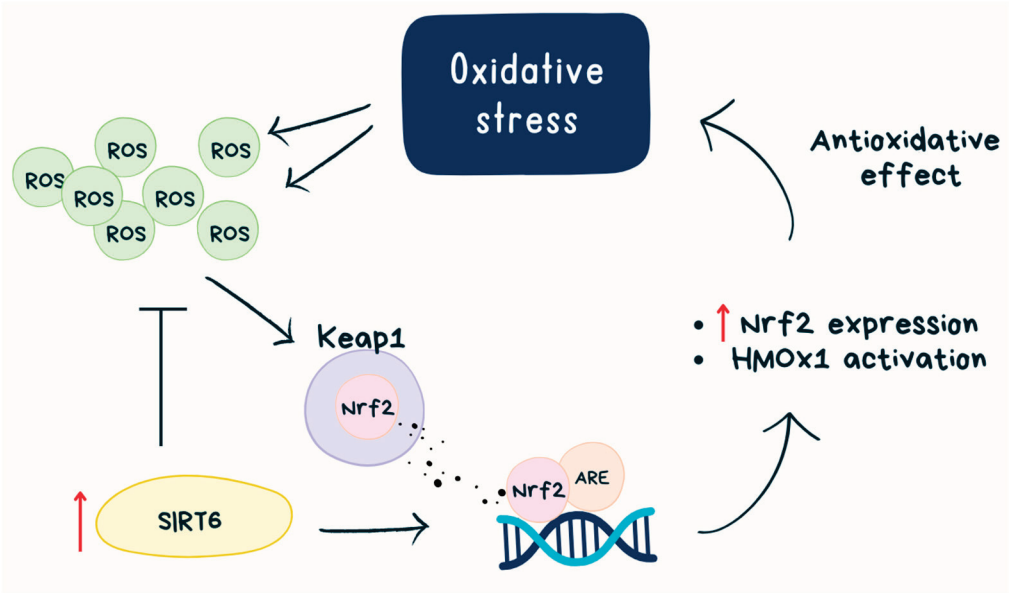


Figure 1. SIRT6 and Nrf2 path. ROS—reactive oxygen species; SIRT6—sirtuin 6, Keap1—Kelch-like ECH-associated protein 1, Nrf2—erythroid 2-related factor 2; ARE—antioxidant response elements; HMOX1—heme oxygenase 1.

3. Aging

Aging of the body is a physiological, dynamic and inevitable process. It is believed that the aging process begins at different times in different systems (the skin is one of the first organs to age). However, aging usually begins in the fourth decade of life and ends at the end of biological life. Important factors influencing aging are extrinsic, biological and psychosocial factors. However, the basic determinant of organism aging is the individual's genotype [25]. Lopez-Otin et al. [26] proposed nine so-called hallmarks of aging: telomere attrition, genomic instability, loss of proteostasis, epigenetic alterations, deregulated nutrient sensing, mitochondrial dysfunction, stem cell exhaustion, cellular senescence and altered intercellular communication. There are some symptoms that occur in psychological aging, but at a younger age [27]. It is associated with accelerated or premature aging, where interrelated molecular and cellular phenomena are intensified.

Nowadays, it is hypothesized that SIRT6 regulates lifespan by influencing a series of processes that control aging, such as genome stability, transcriptional processes of

DNA repair elements, the ability to regulate telomere length and metabolic homeostasis (including carbohydrate metabolism) [9,28,29] (Figure 2).

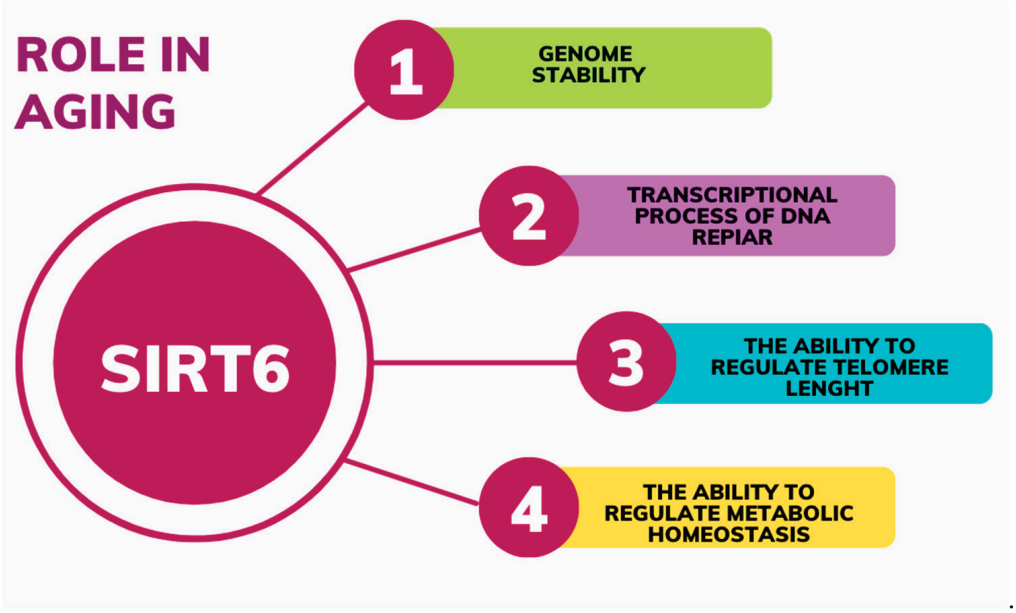


Figure 2. The role of the SIRT6 in aging.

Mostoslavsky et al. [30] demonstrated that mice deficient in SIRT6 are small and develop dysfunctions related to the loss of subcutaneous fat, lordokyphosis and severe metabolic defects by 2–3 weeks of age. These symptoms can be considered similar to those associated with premature aging. These mice eventually died after about 4 weeks. On the other hand, Kanfi et al. [8] observed an extended lifespan of transgenic mice with overexpression of SIRT6 compared to wild-type mice. Interestingly, a significant difference was only observed in males. The authors suggested that this may be related to lower levels of insulin-like growth factor 1 (IGF1) in serum, altered levels of phosphorylation of major IGF1 signaling components and higher levels of IGF1 binding protein, which affects lifespan regulation [31].

Betry et al. [32] conducted a study on 14 men and 12 women in their sixth and seventh decades of life with similar anthropometric and metabolic parameters. The participants underwent six tests: mini-mental state examination—a test that allows for the detection of cognitive dysfunction; Doubis' five-word test—a screening memory task specifically built to discriminate hippocampal memory deficits; Chair Stand Test—allowing for the detection of fall risk in senior patients; Time Up&Go, walking speed and balance tests—allowing for the quantifiable assessment of motor function and serving as an important observation of functional capacity, Center for Epidemiologic Studies Depression scale—CES-D; and Cognitive Difficulties Scale—CDS. Based on the results obtained in the above tests, patients were assigned to the frail group (first tercile of the index score) or the robust group (individuals from the third tercile). The researchers [32] indicated that skeletal muscle biopsies taken from individuals in the robust group showed significantly lower expression of *SIRT6*. However, no differences were observed in the level of expression of this protein in leukocytes isolated from the patients' blood. In the Betry et al. [32] study, an interesting correlation was also noted. The levels of *CPT1b* mRNA are strongly correlated with SIRT6 levels in skeletal muscle. The authors suggested that the overexpression of *CPT1b* in the robust group may at least partially explain their better metabolic profile, and the significant inverse correlation with the level of *SIRT6* expression in muscle could be associated.

SIRT6 is a central regulator of mitochondrial activity in the brain. Deficiency of this protein in the central nervous system leads to mitochondrial deficiency with marked changes

in metabolite content and global downregulation of mitochondria-related genes [33]. The metabolomic changes observed in SIRT6 deficiency lead to an increase in ROS generation. Smirnov et al. [33] indicated that the deficit of SIRT6 in the brain occurs during the aging process of the human brain, especially in patients with Parkinson's, Alzheimer's, Huntington's and Amyotrophic lateral sclerosis disease.

Telomeres shorten with increased turnover and chronological age in somatic stem cells. In early adult life, telomeres help to stabilize the nuclear genome, whereas, in post-reproductive age, they decrease the intensity of age-related changes [34]. Michishita et al. [35] indicated that SIRT6 is associated with specific telomeres, and cells with reduced SIRT6 activity show telomere structures resembling defects observed in Werner syndrome. The Werner syndrome protein (WRN) plays an important role in telomere metabolism and during DNA replication. The connection between WRN and chromatin is closely related to SIRT6-dependent deacetylation of telomeric residues H3K9 and H3K59. Genomic instability with loss of SIRT6 may thus be related to loss of WRN-chromatin connection and influence the redirection of the cell towards premature aging [36]. The cited researchers [35] concluded that SIRT6 contributes to maintaining the specialized chromatin state in mammalian telomeres, which in turn is required for proper telomere metabolism and function. Michishita et al.'s study links SIRT6's chromatin regulation with telomere maintenance and premature aging syndrome [35].

Another mechanism in which *SIRT6* expression will interfere with the aging process is its influence on retrotransposon activity. Transposable elements are DNA sequences capable of integration into other genomic locations [12]. Class I retrotransposons transport via intermediate RNA. Balestrieri et al. [37] identified L1 retrotransposons as a class of transcriptional elements that strongly mediate genomic instability and play a role in age-related pathologies. SIRT6 is a strong repressor of L1 activity, binding to the 5′ UTR of L1 loci. In response to DNA damage and aging, SIRT6 has a reduced ability to bind to L1 loci, contributing to the activation of previously silenced retroelements [38].

Skin aging is the first and most visible sign of aging in the body. Technical advancements in recent years have allowed for a broader examination of the aging process in various types of skin cells. Fibroblasts, as the most numerous types of cells in the dermis, are directly or indirectly responsible for most skin aging characteristics [39]. Sharma et al. [6] demonstrated significantly decreased *SIRT6* expression in fibroblasts of older individuals (above 50 years old) compared to younger individuals (below 18 years old). Furthermore, it was shown that fibroblasts of older individuals were more resistant to reprogramming into induced pluripotent stem cells (iPSCs) compared to fibroblasts of younger individuals. The inclusion of SIRT6 with classical Yamanaka factors in older fibroblasts improved their reprogramming efficiency, which was then similar to that of younger individuals [6].

MicroRNAs (miRNAs) are a class of non-coding RNAs that play a significant role in gene regulation; modulating numerous biological processes; including aging. The influence of miRNAs on the post-transcriptional regulation of SIRT6 is still being studied. Sharma et al. [6] confirmed that SIRT6 regulates the transcription of miR-766, and blocking miR-766 significantly improves the efficiency of reprogramming aging cells. Higher levels of miR-766 in older fibroblasts were inversely correlated with SIRT6 levels [6]. Increasing SIRT6 activity in fibroblasts may therefore slow down skin aging.

An important factor accelerating the aging of the organism is an unhealthy diet and low levels of physical activity and their far-reaching consequences in the form of metabolic syndrome. SIRT6 is indicated as a potential therapeutic target against metabolic syndrome. This protein is associated with calorie reduction. Activating SIRT6 is also suggested to have potentially beneficial effects on age-related metabolic diseases. The development of small molecule *SIRT6* activators could therefore have great therapeutic potential [40].

4. Metabolic Activity

SIRT6 is involved in regulating glucose homeostasis in the body. It has been shown that SIRT6 levels are increased during fasting. Additionally, by increasing the expression of genes involved in gluconeogenesis, it controls this process in the liver. Zhong et al. [41] discovered that the expression of gluconeogenic genes was increased in livers with SIRT6 deficiency. Researchers also identified the role of SIRT6 as a corepressor of Hif1α (Hypoxia-inducible factor 1α), a critical regulator of the response to nutritional stress. Cells deficient in SIRT6 exhibit increased Hif1α activity, thus showing increased glucose uptake. SIRT6, therefore, acts as a corepressor of the transcription factor Hif1α, reducing glycolysis during normal nutrition and stimulating mitochondrial fatty acid oxidation [41].

Kim et al. [42] pointed out that liver-specific deletion of *SIRT6* in mice leads to profound changes in gene expression, causing increased glycolysis, intensified triglyceride synthesis, decreased β-oxidation and liver steatosis. Other authors suggest that SIRT6 may therefore be a potential target in the treatment of liver diseases characterized by lipid accumulation [43].

Dominy et al. [11] indicated the usefulness of activating liver *SIRT6* in the therapeutic treatment of insulin-resistant diabetes. Xiong et al. [44] demonstrated a significant decrease (50%) in glucose-stimulated insulin secretion in mice subjected to *SIRT6* knockdown in pancreatic beta cells. The mice also had lower levels of ATP in the studied cells compared to the wild-type control group. An increased number of damaged mitochondria was also identified. Based on the obtained results, the authors suggested that SIRT6 regulates proper insulin secretion through the regulation of mitochondrial glucose oxidation. Therefore, activating the protein may be helpful in improving insulin secretion in diabetic states. The process described here takes place in the mitochondria. This underscores that SIRT6 acts in many cellular organelles, so it can be said that it acts throughout the cell.

Further experiments have indicated that mice with *SIRT6* knockout were more susceptible to high-fat diet-induced obesity, attributed to adipocyte hypertrophy. Moreover, increased macrophage infiltration into the examined adipose tissue was observed, indicating an intensified inflammatory process in these mice. It was found that SIRT6 regulates energy homeostasis by modulating the activity and expression of lipase in adipose tissue [45]. Kanfi et al. [40] noticed increased glucose tolerance and insulin secretion stimulated by glucose in mice with *SIRT6* overexpression. Their study indicated that *SIRT6* overexpression increases triglyceride clearance in the blood and reduces triglyceride production in adipose tissue. Mice with *SIRT6* overexpression subjected to a high-fat diet accumulated significantly less visceral fat, LDL cholesterol and triglycerides compared to wild-type mice. This suggests a protective role of SIRT6 against metabolic consequences of obesity caused by an improper diet [40]. Tang et al. [46] came to similar conclusions—*SIRT6* overexpression in the arcuate nucleus of the hypothalamus in mice reduced their body weight induced by a high-fat diet. A decrease in the weight of eWAT (epididymal white adipose tissue) and iWAT (inguinal white adipose tissue) and adipocyte size was also observed. Furthermore, the impairment of leptin activity in the POMC neurons of mice subjected to *SIRT6* neuron-specific knockout was demonstrated. These mice exhibited a predisposition to obesity and increased food consumption.

Increased SIRT6 levels were observed in in vivo models—in mice after fasting, rats after caloric restriction and in vitro—in cell cultures in a medium deprived of nutrients. The authors indicated that the increase in SIRT6 levels is caused by the stabilization of the SIRT6 protein, not an increase in *SIRT6* transcription. Moreover, p53 positively regulates SIRT6 protein levels under standard growth conditions but does not play a role in regulating *SIRT6* under caloric restriction [47]. It is widely known that calorie restriction diets (CR) slow down aging processes and may contribute to extending life [48]. In light of the above research, it can be suggested that the beneficial effects of a calorie-restricted diet are strongly related to increased SIRT6 stability, and the increased expression of the protein generated by factors other than starvation may exhibit similar effects to CR [47]. The role of SIRT6 in metabolism has been summarized in Table 1.

Table 1. Role of the SIRT6 in metabolic activity regulation.

	SIRT6 and Metabolic Activity
Obesity	SIRT6 plays a protective role against the metabolic consequences of diet-induced obesity, which suggests a potentially beneficial effect of SIRT6 activation on age-related metabolic diseases [40]. In obese patients, the expression of SIRT6 is reduced. It suggests that SIRT6 is an attractive therapeutic target for treating obesity and obesity-related metabolic disorders [45].
Fat metabolism	SIRT6 plays a critical role in fat metabolism, and may therefore be a potential target in the treatment of liver diseases characterized by lipid accumulation [42].
Carbohydrate metabolism	Activation of hepatic by SIRT6 may be therapeutically useful for treating insulin-resistant diabetes [11]. SIRT6 may be useful to improve insulin secretion in diabetes [44].
Energy balance	SIRT6 is an important molecular regulator for POMC neurons to promote negative energy balance [46].
Other	SIRT6 appears to function as a corepressor of the Hif1α—a critical regulator of nutrient stress responses [41]. Expression of SIRT6 increased upon nutrient deprivation in cultured cells. The increase in SIRT6 levels is due to the stabilization of the SIRT6 protein, and not via an increase in SIRT6 transcription [47].

The "sirtfoods" diet, combing sirtuin-activating foods belonging to both Mediterranean and Asian diets, may be a promising dietary strategy in preventing chronic diseases, thereby ensuring healthy aging [49]. "Sirtfoods" can be found in: olive oil [50]; red wine [51]; grapes [52,53]; apple, strawberries, onion and cabbage [53]; soybeans [54]; tofu [55]; licorice [56]; shallot [56]. Pallauf et al. [49] suggested that omega-3 fatty acids, vitamins and antioxidants do not work in isolation. They should synergistically work to prevent chronic diseases.

5. Inflammation

The role of SIRT6′s involvement in inflammation is complex. This protein can have both pro- and anti-inflammatory properties, depending on the type of cells involved [36]. It has been suggested that SIRT6 acts as a pro-inflammatory agent by functioning as a lysine deacetylase, removing fatty acyl modifications from K19 and K20 of TNF-α and thereby promoting the secretion of TNF-α from the cell [57].

A characteristic feature of inflammatory diseases is endothelial dysfunction. Factors associated with this phenomenon include the increased release of pro-inflammatory cytokines, adhesion molecules and tissue matrix-degrading enzymes (collagenases and others). At the transcriptional level, this is regulated by the histone deacetylase SIRT6 through its action on the pro-inflammatory transcription factor nuclear factor-κB (*NF-κB*) [58]. In a study by Lappas et al. [59], the role of SIRT6 in regulating inflammation in endothelial cells was determined. They evaluated markers of inflammation in human umbilical vein endothelial cells (HUVECs) in the presence of lipopolysaccharide (LPS) as a model agent for initiating inflammation. LPS decreased *SIRT6* expression in HUVEC cells, which agrees with previous observations by other authors [60,61]. In contrast, *SIRT6* knockdown increased the expression of a number of pro-inflammatory cytokines (IL-1β, IL-6, IL-8), the COX-prostaglandin system, extracellular matrix remodeling enzymes (MMP-2, MMP-9 and PAI-1), the adhesion molecule ICAM-1 and angiogenesis-related growth factors (VEGF and FGF-2). Knockdown of *SIRT6* increased the expression of *NF-κB*. In contrast, *SIRT6* overexpression was associated with decreased *NF-κB* transcriptional activity. Thus, they indicated that the loss of *SIRT6* in endothelial cells is associated with the up-regulation of genes involved in the progression of inflammation and associated vascular remodeling.

Hence, the authors conclude that the up-regulation of *SIRT6* expression is a potential pharmacological target for drugs that reduce the inflammatory destruction of blood vessels. Such action will have the potential to nullify the negative manifestations of cardiovascular disease, diabetes and a range of neurodegenerative diseases. An extensive discussion of this topic was carried out in the work of Guo et al. [62].

Earlier, the role of LPS was identified as a factor that decreases the expression of *SIRT6*. It is suggested that this mechanism is based on the weakening of NF-κB signaling through H3K39 deacetylation on chromatin. NF-κB is a key factor involved in regulating cell apoptosis, aging, inflammation and immune system function [63]. Hyperactive NF-κB signaling can therefore contribute to both normal and premature aging. Hence, the role of SIRT6 in reducing the activity of this factor is significant [64].

Balestrieri et al. [37] demonstrated reduced *SIRT6* expression and lower interstitial collagen content in cells obtained from homogenates of atherosclerotic plaques from the carotid arteries of people with diabetes compared to plaques obtained from people without diabetes. In addition, plaques obtained from patients with diabetes were characterized by greater inflammation and oxidative stress. Subsequent trials showed that atherosclerotic plaques obtained from diabetic patients treated with GLP-1 (glucagon-like peptide-1 receptor) drugs for 26 ± 8 months indicated higher expression of *SIRT6* and collagen, as well as a lower intensity of inflammation and oxidative stress than cells obtained from untreated patients. This indicates the involvement of SIRT6 in the inflammatory changes of atherosclerotic lesions in diabetes and the role of GLP-1 in modulating it. These results were further supported by observations in an in vitro model. Endothelial progenitor cells (EPCs) and endothelial cells (ECs) were used as model cells. Both types of cells treated with high glucose medium (25 mmol/L) in the presence of GLP-1 (100 nmol/liraglutide) showed higher expression of *SIRT6* and lower expression of nuclear factor-κB compared to control cells (only treated with high glucose). The results obtained by Balestrieri et al. [37] confirmed the controlling involvement of SIRT6 in the inflammatory pathways of diabetic atherosclerotic lesions. From the perspective of skin wellbeing, this is an important observation. The incidence of diabetic complications involving the skin and its appendages is high, and identifying treatment options to nullify the underlying pathomechanism, i.e., vascular destruction, is an important topic here.

The promising role of *SIRT6* regulation is indicated to support the treatment of: atherosclerosis [65]; diabetic atherosclerosis [66]; diabetes-associated inflammatory diseases [67]; diabetic nephropathy [68]; adipose tissue inflammation [69]; oxidative stress in the heart triggered by high-fat diet-induced obesity [70] and osteoarthritis [61,71].

6. Role in the Wound Healing Process

With age, the loss of homeostasis leads to a series of clinical consequences, including impaired wound healing. Macrophages play a key role in wound healing, as they are responsible for suppressing excessive inflammatory states, removing cell debris and initiating and coordinating tissue remodeling and regeneration [72,73]. Activated macrophages can be divided into two phenotypes: M1 (mainly involved in pro-inflammatory responses) and M2 (responsible for anti-inflammatory actions) [19,73].

SIRT6 is considered a key factor in switching macrophage phenotypes and migratory responses, and therefore, it is believed to play a role in the wound healing process [18]. To illustrate the effect of SIRT6 on wound healing, Koo et al. [18] created a mouse model with *SIRT6* deletion (mS6KO). Wounds were excised on the dorsal skin of female mS6KO and control group mice (WT). Slower wound healing was observed in mS6KO mice. Reduced collagen content in the granulation tissue of mS6KO mice was demonstrated, as well as suppressed angiogenesis gene activity in mS6KO mice. Moreover, increased M1 macrophage infiltration with decreased M2 macrophage numbers, as well as more pro-inflammatory cytokines TNF-α, IL-1β and IL-6 were noted. This suggests that the specific deficiency of SIRT6 in bone marrow cells has an impact on the temporal response of the wound healing process by inhibiting epithelial regeneration, angiogenesis and collagen deposition.

Thandavarayan et al. [22] arrived at similar conclusions. In a mouse model with diabetes, they showed that SIRT6 deficiency reduced *VEGF* expression, increased the expression of pro-inflammatory markers (including TNF-α, IL-1β) and intensified oxidative stress when siRNA against *SIRT6* was administered. Yang et al. [74] described the molecular mechanism that connects SIRT6 and angiogenesis.

Furthermore, exogenous substances that are associated with the modification of the intensity of sirtuin expression can affect changes in the rate of wound healing. This issue has been the subject of several research projects [75–79].

Resveratrol, a sirtuin activator, by increasing the proliferation of keratinocytes, accelerates wound healing, while sirtinol, a sirtuin inhibitor, will delay their closure. This suggests the role of administering natural or synthetic *SIRT6* activators as factors accelerating wound healing processes [77].

Flavonoids are polyphenolic secondary metabolites synthesized by fungi and plants; possessing various pharmacological activities. It has been shown that compounds from this group modulate the activity of SIRT6 by altering its structure.

SIRT6 activators presumably bind near the acetylated peptide substrate binding site. Inhibitors, on the other hand, are likely to bind in a way that disrupts NAD^+ binding. Cyanidin is indicated as the strongest SIRT6 activator [80]. Other activators include: N-acetylamines [81], icariin [82], ergothioneine [83]. Significant inhibitory power for SIRT6 activity has been shown for galloylated catechins [80]. Compounds that will exhibit inhibitory activity on SIRT6 activity include quercetin, vitexin [84], peptides containing netioacetylated lysine [85] and EX-527 [86]. The role of SIRT6 in the wound healing process is shown in Figure 3.

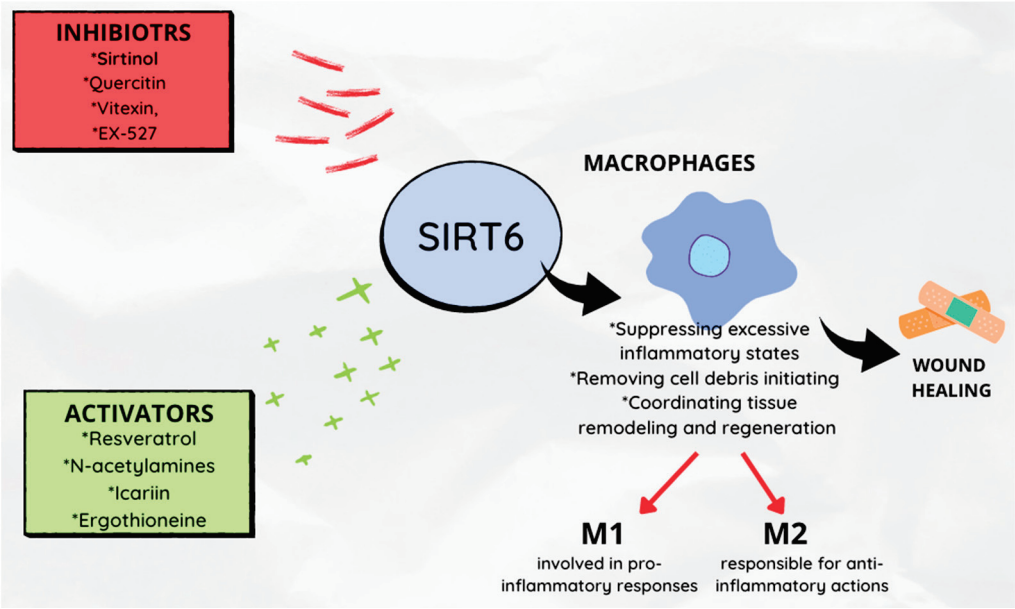


Figure 3. The role of SIRT6 in the wound healing process.

7. Role of Physical Activity

Appropriate levels of physical activity are an important element in preventing premature disability, the progression of various lifestyle diseases and aging. Increasing the endurance capacity of skeletal muscles is a new strategy in the fight against metabolic diseases associated with obesity, and this can be achieved by shifting skeletal muscle fibers towards slow-twitch, oxidative fibers. Aging leads to a gradual decrease in physical activity and disturbances in energy homeostasis. Numerous studies have focused on the correlation between changes in activity levels and *SIRT6* expression [87–92] (Figure 4).

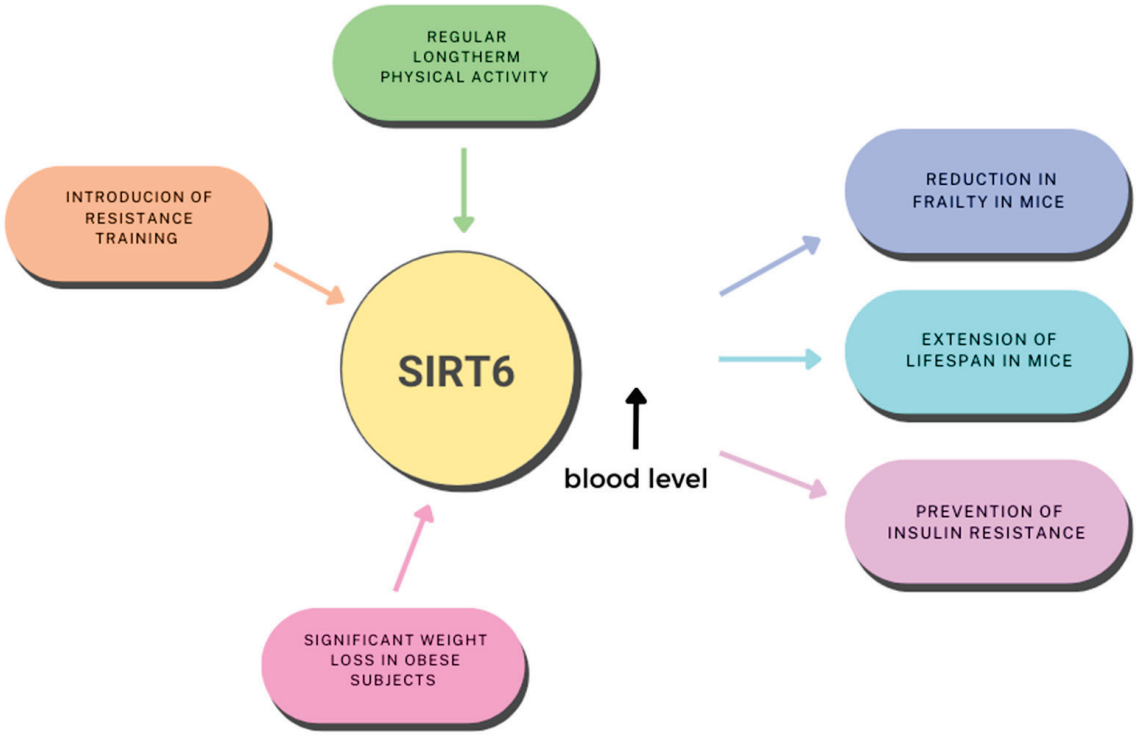


Figure 4. Modifying role of physical activity and the influence of SIRT6 on metabolism based on: [85,88–90].

The expression of *SIRT6* is increased in chronically exercising individuals. Introducing resistance training as a modification of current physical activity allows for an increase in SIRT6 concentration in the blood of older men [65]. A factor that increases *SIRT6* expression is significant weight loss achieved through diet therapy and exercise in morbidly obese individuals [88].

SIRT6 regulates aging and a series of metabolic processes. Roichman et al. [89] showed that *SIRT6* overexpression leads to a reduction in frailty and an extension of lifespan in both male and female mice. Older transgenic *SIRT6* mice maintained normal glucose production in the liver and glucose homeostasis by improving the utilization of two major gluconeogenic precursors, lactate and glycerol. Therefore, *SIRT6* overexpression protected older individuals from impaired carbohydrate homeostasis. The authors indicated that the mechanism of this beneficial phenomenon is associated with increased gene expression for proteins involved in gluconeogenesis in the liver, de novo synthesis of NAD^+ (which is also the basis for the beneficial effects of physical training) and systemically increases the release of glycerol from adipose tissue. SIRT6 optimizes energy homeostasis in older age, which is associated with the carbohydrate theory of aging. It was also noted that it is precisely SIRT6 that mediates the beneficial effect of regular physical activity in preventing insulin resistance [90].

The study by Song et al. [91] indicated that SIRT6 is crucial in regulating the configuration of muscle fibers toward the oxidative type. They also indicated that a SIRT6 activator (MDL801) could produce the same effects as physical training. Genetic inactivation of SIRT6 in skeletal muscle decreased, while its transgenic overexpression increased mitochondrial oxidative capacity and exercise performance in the model animals studied. The authors indicated that the ablation of *SIRT6* in skeletal muscle led to reduced exercise performance. The authors tested the hypothesis that *SIRT6* induction may be mediated by endocrine/paracrine signaling (Il-6 enhanced *SIRT6* expression) and/or a stimulus effect (stimulating the muscle with electrodes also enhanced *SIRT6* expression). *SIRT6* ablation also reduces the number of mitochondria; knockout muscles also had significantly more mitochondria with abnormal structures. Reduced mitochondrial DNA content and

suppressed expression of genes related to mitochondrial biogenesis were observed. The consequence was the suppression of basal, ATP-related and maximal respiration in *SIRT6* knockout mice observed in this study. *SIRT6* ablation impairs the ability to meet the cell's energy needs through oxidative phosphorylation [91].

The mechanism linked *SIRT6* to the downregulation of *Sox6*, a key repressor of the free-fiber-specific gene, by increasing *CREB* transcription. This mechanism demonstrates how suppressed *SIRT6* expression links to the mitochondrial theory of aging in how the beneficial effects of regular physical activity on life extension and quality of life binding can be explained.

It has also been indicated, in studies conducted on young men, that even a single bout of aerobic exercise (at 80% of peak oxygen uptake) will affect the expression of *SIRT6* and telomerase, the enzyme responsible for maintaining the proper length of telomeres protecting the ends of DNA strands [92]. This observation will again allow for linking SIRT6 to the telomeric theory of aging.

8. Conclusions

Research suggests that the activation of *SIRT6* and its role in various biochemical processes may have a positive impact on health and longevity. Several pathomechanisms associated with accelerated aging of the body as a whole, as well as individual systems, could be inhibited by an increased expression of this protein. Modulating SIRT6 activity can therefore be a valuable tool for supporting the treatment of age-related diseases and improving quality of life, thereby contributing to longevity. In addition to well-known lifestyle modifications, such as caloric restriction and intermittent fasting, research is being conducted to identify new pharmacologically active molecules and dietary components with recognized roles in promoting longevity. Furthermore, physical activity might be a valuable tool for modifying and/or modulating SIRT6 for health and longevity. As seen in relation to the presented results in this review, not all aspects related to the activity of SIRT6 are fully understood. Hence, there is a necessity and need for systematic, controlled research into this continually emerging area of investigation.

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

Funding: This research received no external funding, the APC fee was funded by a grant 318/BS/INP/2023 University of Physical Education in Krakow, Poland.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Haigis, M.C.; Sinclair, D.A. Mammalian Sirtuins: Biological Insights and Disease Relevance. *Annu. Rev. Pathol. Mech. Dis.* **2010**, *5*, 253–295. [CrossRef]
2. Carafa, V.; Rotili, D.; Forgione, M.; Cuomo, F.; Serretiello, E.; Hailu, G.S.; Jarho, E.; Lahtela-Kakkonen, M.; Mai, A.; Altucci, L. Sirtuin Functions and Modulation: From Chemistry to the Clinic. *Clin. Epigenetics* **2016**, *8*, 61. [CrossRef]
3. Frydzińska, Z.; Owczarek, A.; Winiarska, K. Sirtuiny i Ich Rola w Regulacji Metabolizmu. *Postep. Biochem.* **2019**, *65*, 31–40. [CrossRef]
4. Jing, H.; Lin, H. Sirtuins in Epigenetic Regulation. *Chem. Rev.* **2015**, *115*, 2350–2375. [CrossRef]
5. Sinclair, D.A.; Guarente, L. Extrachromosomal RDNA Circles-A Cause of Aging in Yeast. *Cell* **1997**, *91*, 1033–1042. [CrossRef]
6. Sharma, A.; Diecke, S.; Zhang, W.Y.; Lan, F.; He, C.; Mordwinkin, N.M.; Chua, K.F.; Wu, J.C. The Role of SIRT6 Protein in Aging and Reprogramming of Human Induced Pluripotent Stem Cells. *J. Biol. Chem.* **2013**, *288*, 18439–18447. [CrossRef]

7. Madsen, A.S.; Andersen, C.; Daoud, M.; Anderson, K.A.; Laursen, J.S.; Chakladar, S.; Huynh, F.K.; Colaço, A.R.; Backos, D.S.; Fristrup, P.; et al. Investigating the Sensitivity of NAD+-Dependent Sirtuin Deacylation Activities to NADH. *J. Biol. Chem.* **2016**, *291*, 7128–7141. [CrossRef]
8. Kanfi, Y.; Naiman, S.; Amir, G.; Peshti, V.; Zinman, G.; Nahum, L.; Bar-Joseph, Z.; Cohen, H.Y. The Sirtuin SIRT6 Regulates Lifespan in Male Mice. *Nature* **2012**, *483*, 218–221. [CrossRef]
9. Vitiello, M.; Zullo, A.; Servillo, L.; Mancini, F.P.; Borriello, A.; Giovane, A.; Della Ragione, F.; D'Onofrio, N.; Balestrieri, M.L. Multiple Pathways of SIRT6 at the Crossroads in the Control of Longevity, Cancer, and Cardiovascular Diseases. *Ageing Res. Rev.* **2017**, *35*, 301–311. [CrossRef]
10. Ran, L.K.; Chen, Y.; Zhang, Z.Z.; Tao, N.N.; Ren, J.H.; Zhou, L.; Tang, H.; Chen, X.; Chen, K.; Li, W.Y.; et al. SIRT6 Overexpression Potentiates Apoptosis Evasion in Hepatocellular Carcinoma via BCL2-Associated X Protein-Dependent Apoptotic Pathway. *Clin. Cancer Res.* **2016**, *22*, 3372–3382. [CrossRef]
11. Dominy, J.E.; Lee, Y.; Jedrychowski, M.P.; Chim, H.; Jurczak, M.J.; Camporez, J.P.; Ruan, H.B.; Feldman, J.; Pierce, K.; Mostoslavsky, R.; et al. The Deacetylase Sirt6 Activates the Acetyltransferase GCN5 and Suppresses Hepatic Gluconeogenesis. *Mol. Cell* **2012**, *48*, 900–913. [CrossRef]
12. Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as Regulators of Gene Expression. *Science* **2016**, *351*, 6274. [CrossRef]
13. Imai, S.i.; Guarente, L. NAD+ and Sirtuins in Aging and Disease. *Trends Cell Biol.* **2014**, *24*, 464–471. [CrossRef]
14. Tennen, R.I.; Chua, K.F. Chromatin Regulation and Genome Maintenance by Mammalian SIRT6. *Trends Biochem. Sci.* **2011**, *36*, 39–46. [CrossRef]
15. Mahlknecht, U.; Ho, A.; Volter-Mahlknecht, S. Chromosomal Organization and Fluorescence in Situhybridization of the Human Sirtuin 6 Gene. *Int. J. Oncol.* **2006**, *28*, 447–456.
16. Gertler, A.A.; Cohen, H.Y. SIRT6, a Protein with Many Faces. *Biogerontology* **2013**, *14*, 629–639. [CrossRef]
17. Bielawski, A.; Nalepa, I. Sirtuins—Intriguing multi-tasking "keepers" of life processes. *Wszechświat* **2019**, *20*, 4–6.
18. Koo, J.H.; Jang, H.Y.; Lee, Y.; Moon, Y.J.; Bae, E.J.; Yun, S.K.; Park, B.H. Myeloid Cell-Specific Sirtuin 6 Deficiency Delays Wound Healing in Mice by Modulating Inflammation and Macrophage Phenotypes. *Exp. Mol. Med.* **2019**, *51*, 1–10. [CrossRef]
19. Yunna, C.; Mengru, H.; Lei, W.; Weidong, C. Macrophage M1/M2 Polarization. *Eur. J. Pharmacol.* **2020**, *877*, 173090. [CrossRef]
20. Chang, A.R.; Ferrer, C.M.; Mostoslavsky, R. SIRT6, a Mammalian Deacylase with Multitasking Abilities. *Physiol. Rev.* **2020**, *100*, 145–169. [CrossRef]
21. Pan, H.; Guan, D.; Liu, X.; Li, J.; Wang, L.; Wu, J.; Zhou, J.; Zhang, W.; Ren, R.; Zhang, W.; et al. SIRT6 Safeguards Human Mesenchymal Stem Cells from Oxidative Stress by Coactivating NRF2. *Cell Res.* **2016**, *26*, 190–205. [CrossRef]
22. Thandavarayan, R.A.; Garikipati, V.N.S.; Joladarashi, D.; Suresh Babu, S.; Jeyabal, P.; Verma, S.K.; Mackie, A.R.; Khan, M.; Arumugam, S.; Watanabe, K.; et al. Sirtuin-6 Deficiency Exacerbates Diabetes-Induced Impairment of Wound Healing. *Exp. Dermatol.* **2015**, *24*, 773–778. [CrossRef]
23. Itoh, K.; Tong, K.I.; Yamamoto, M. Molecular Mechanism Activating Nrf2-Keap1 Pathway in Regulation of Adaptive Response to Electrophiles. *Free. Radic. Biol. Med.* **2004**, *36*, 1208–1213. [CrossRef]
24. Kanwal, A.; Pillai, V.B.; Gupta, M.P. The Nuclear and Mitochondrial Sirtuins, Sirt6 and Sirt3, Regulate each others' Activity and Protect the Heart from Developing Obesity-Mediated Diabeticcardiomyopathy. *J. Fed. Am. Soc. Exp. Biol.* **2020**, *34*, 14057.
25. Dziechciaż, M.; Filip, R. Biological Psychological and Social Determinants of Old Age: Bio-Psycho-Social Aspects of Human Aging. *Ann. Agric. Environ. Med.* **2014**, *21*, 835–838. [CrossRef]
26. López-Otín, C.; Blasco, M.A.; Partridge, L.; Serrano, M.; Kroemer, G. The Hallmarks of Aging. *Cell* **2013**, *153*, 1194. [CrossRef]
27. Franco, A.C.; Aveleira, C.; Cavadas, C. Skin Senescence: Mechanisms and Impact on Whole-Body Aging. *Trends Mol. Med.* **2022**, *28*, 97–109. [CrossRef]
28. Mao, Z.; Hine, C.; Tian, X.; Van Meter, M.; Au, M.; Vaidya, A.; Seluanov, A.; Gorbunova, V. SIRT6 Promotes DNA Repair under Stress by Activating PARP1. *Science* **2011**, *332*, 1443–1446. [CrossRef]
29. Cardus, A.; Uryga, A.K.; Walters, G.; Erusalimsky, J.D. SIRT6 Protects Human Endothelial Cells from DNA Damage, Telomere Dysfunction, and Senescence. *Cardiovasc. Res.* **2013**, *97*, 571–579. [CrossRef]
30. Mostoslavsky, R.; Chua, K.F.; Lombard, D.B.; Pang, W.W.; Fischer, M.R.; Gellon, L.; Liu, P.; Mostoslavsky, G.; Franco, S.; Murphy, M.M.; et al. Genomic Instability and Aging-like Phenotype in the Absence of Mammalian SIRT6. *Cell* **2006**, *124*, 315–329. [CrossRef]
31. Tower, J. Sex-Specific Gene Expression and Life Span Regulation. *Trends Endocrinol. Metab.* **2017**, *28*, 735–747. [CrossRef]
32. Bétry, C.; Meugnier, E.; Pflieger, M.; Grenet, G.; Hercberg, S.; Galan, P.; Kesse-Guyot, E.; Vidal, H.; Laville, M. High Expression of CPT1b in Skeletal Muscle in Metabolically Healthy Older Subjects. *Diabetes Metab.* **2019**, *45*, 152–159. [CrossRef]
33. Smirnov, D.; Eremenko, E.; Stein, D.; Kaluski, S.; Jasinska, W.; Cosentino, C.; Martinez-Pastor, B.; Brotman, Y.; Mostoslavsky, R.; Khrameeva, E.; et al. SIRT6 Is a Key Regulator of Mitochondrial Function in the Brain. *Cell Death Dis.* **2023**, *14*, 35. [CrossRef]
34. Shay, J.W. Telomeres and Aging. *Curr. Opin. Cell Biol.* **2018**, *52*, 1–7. [CrossRef]
35. Michishita, E.; McCord, R.A.; Berber, E.; Kioi, M.; Padilla-Nash, H.; Damian, M.; Cheung, P.; Kusumoto, R.; Kawahara, T.L.A.; Barrett, J.C.; et al. SIRT6 Is a Histone H3 Lysine 9 Deacetylase That Modulates Telomeric Chromatin. *Nature* **2008**, *452*, 492–496. [CrossRef]
36. Kugel, S.; Mostoslavsky, R. Chromatin and beyond: The Multitasking Roles for SIRT6. *Trends Biochem. Sci.* **2014**, *39*, 72–81. [CrossRef]

37. Balestrieri, M.L.; Rizzo, M.R.; Barbieri, M.; Paolisso, P.; D'onofrio, N.; Giovane, A.; Siniscalchi, M.; Minicucci, F.; Sardu, C.; D'andrea, D.; et al. Sirtuin 6 Expression and Inflammatory 7 Activity in Diabetic Atherosclerotic Plaques: Effects of Incretin Treatment. *Diabetes* **2015**, *64*, 1395–1406. [CrossRef]
38. Van Meter, M.; Kashyap, M.; Rezazadeh, S.; Geneva, A.J.; Morello, T.D.; Seluanov, A.; Gorbunova, V. SIRT6 Represses LINE1 Retrotransposons by Ribosylating KAP1 but This Repression Fails with Stress and Age. *Nat. Commun.* **2014**, *5*, 5011. [CrossRef]
39. Gruber, F.; Kremslehner, C.; Eckhart, L.; Tschachler, E. Cell Aging and Cellular Senescence in Skin Aging—Recent Advances in Fibroblast and Keratinocyte Biology. *Exp. Gerontol.* **2020**, *130*, 110780. [CrossRef]
40. Kanfi, Y.; Peshti, V.; Gil, R.; Naiman, S.; Nahum, L.; Levin, E.; Kronfeld-Schor, N.; Cohen, H.Y. SIRT6 Protects against Pathological Damage Caused by Diet-Induced Obesity. *Aging Cell* **2010**, *9*, 162–173. [CrossRef]
41. Zhong, L.; D'Urso, A.; Toiber, D.; Sebastian, C.; Henry, R.E.; Vadysirisack, D.D.; Guimaraes, A.; Marinelli, B.; Wikstrom, J.D.; Nir, T.; et al. The Histone Deacetylase Sirt6 Regulates Glucose Homeostasis via Hif1α. *Cell* **2010**, *140*, 280–293. [CrossRef]
42. Kim, H.S.; Xiao, C.; Wang, R.H.; Lahusen, T.; Xu, X.; Vassilopoulos, A.; Vazquez-Ortiz, G.; Jeong, W.I.; Park, O.; Ki, S.H.; et al. Hepatic-Specific Disruption of SIRT6 in Mice Results in Fatty Liver Formation Due to Enhanced Glycolysis and Triglyceride Synthesis. *Cell Metab.* **2010**, *12*, 224–236. [CrossRef]
43. Houtkooper, R.H.; Pirinen, E.; Auwerx, J. Sirtuins as Regulators of Metabolism and Healthspan. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 225–238. [CrossRef]
44. Xiong, X.; Wang, G.; Tao, R.; Wu, P.; Kono, T.; Li, K.; Ding, W.X.; Tong, X.; Tersey, S.A.; Harris, R.A.; et al. Sirtuin 6 Regulates Glucose-Stimulated Insulin Secretion in Mouse Pancreatic Beta Cells. *Diabetologia* **2016**, *59*, 151–160. [CrossRef]
45. Kuang, J.; Zhang, Y.; Liu, Q.; Shen, J.; Pu, S.; Cheng, S.; Chen, L.; Li, H.; Wu, T.; Li, R.; et al. Fat-Specific Sirt6 Ablation Sensitizes Mice to High-Fat Diet-Induced Obesity and Insulin Resistance by Inhibiting Lipolysis. *Diabetes* **2017**, *66*, 1159–1171. [CrossRef]
46. Tang, Q.; Gao, Y.; Liu, Q.; Yang, X.; Wu, T.; Huang, C.; Huang, Y.; Zhang, J.; Zhang, Z.; Li, R.; et al. Sirt6 in Pro-Opiomelanocortin Neurons Controls Energy Metabolism by Modulating Leptin Signaling. *Mol. Metab.* **2020**, *37*, 100994. [CrossRef]
47. Kanfi, Y.; Shalman, R.; Peshti, V.; Pilosof, S.N.; Gozlan, Y.M.; Pearson, K.J.; Lerrer, B.; Moazed, D.; Marine, J.C.; de Cabo, R.; et al. Regulation of SIRT6 Protein Levels by Nutrient Availability. *FEBS Lett.* **2008**, *582*, 543–548. [CrossRef]
48. Mccay, C.M.; Crowell, M.F.; Maynard, L.A. The Effect of Retarded Growth upon the Length of Life Span and upon the Ultimate Body Size. 1935. *Nutrition* **1989**, *5*, 155–171.
49. Pallauf, K.; Giller, K.; Huebbe, P.; Rimbach, G. Nutrition and Healthy Ageing: Calorie Restriction or Polyphenol-Rich "MediterrAsian" Diet? *Oxid. Med. Cell Longev.* **2013**, *2013*, 707421. [CrossRef]
50. Menendez, J.A.; Joven, J.; Aragonès, G.; Barrajón-Catalán, E.; Beltrán-Debón, R.; Borrás-Linares, I.; Camps, J.; Corominas-Faja, B.; Cufí, S.; Fernández-Arroyo, S.; et al. Xenohormetic and Anti-Aging Activity of Secoiridoid Polyphenols Present in Extra Virgin Olive Oil: A New Family of Gerosuppressant Agents. *Cell Cycle* **2013**, *12*, 555–578. [CrossRef]
51. Howitz, K.T.; Bitterman, K.J.; Cohen, H.Y.; Lamming, D.W.; Lavu, S.; Wood, J.G.; Zipkin, R.E.; Chung, P.; Kisielewski, A.; Zhang, L.-L.; et al. Small Molecule Activators of Sirtuins Extend Saccharomyces Cerevisiae Lifespan. *Nature* **2003**, *425*, 191–196. [CrossRef]
52. Guerrero, R.F.; Puertas, B.; Fernández, M.I.; Palma, M.; Cantos-Villar, E. Induction of Stilbenes in Grapes by UV-C: Comparison of Different Subspecies of Vitis. *Innov. Food Sci. Emerg. Technol.* **2010**, *11*, 231–238. [CrossRef]
53. Arai, Y.; Watanabe, S.; Kimira, M.; Shimoi, K.; Mochizuki, R.; Kinae, N. Human Nutrition and Metabolism Dietary Intakes of Flavonols, Flavones and Isoflavones by Japanese Women and the Inverse Correlation between Quercetin Intake and Plasma LDL Cholesterol Concentration 1. *J. Nutr.* **2000**, *130*, 2243–2250. [CrossRef]
54. Rasbach, K.A.; Schnellmann, R.G. Isoflavones Promote Mitochondrial Biogenesis. *J. Pharmacol. Exp. Ther.* **2008**, *325*, 536–543. [CrossRef]
55. Song, T.; Barua, K.; Buseman, G.; Murphy, P. Soy Isoflavone Analysis: Quality Control and a New Internalstandard. *Am. J. Clin. Nutr.* **1998**, *68*, 1474–1479. [CrossRef]
56. Hsu, Y.L.; Chia, C.C.; Chen, P.J.; Huang, S.E.; Huang, S.C.; Kuo, P.L. Shallot and Licorice Constituent Isoliquiritigenin Arrests Cell Cycle Progression and Induces Apoptosis through the Induction of ATM/P53 and Initiation of the Mitochondrial System in Human Cervical Carcinoma HeLa Cells. *Mol. Nutr. Food Res.* **2009**, *53*, 826–835. [CrossRef]
57. Jiang, H.; Khan, S.; Wang, Y.; Charron, G.; He, B.; Sebastian, C.; Du, J.; Kim, R.; Ge, E.; Mostoslavsky, R.; et al. SIRT6 Regulates TNF-α Secretion through Hydrolysis of Long-Chain Fatty Acyl Lysine. *Nature* **2013**, *496*, 110–113. [CrossRef]
58. Mendes, K.L.; Lelis, D.d.F.; Santos, S.H.S. Nuclear Sirtuins and Inflammatory Signaling Pathways. *Cytokine Growth Factor Rev.* **2017**, *38*, 98–105. [CrossRef]
59. Lappas, M. Anti-Inflammatory Properties of Sirtuin 6 in Human Umbilical Vein Endothelial Cells. *Mediat. Inflamm.* **2012**, *2012*, 597514. [CrossRef]
60. Zhang, Y.; Wang, L.; Meng, L.; Cao, G.; Wu, Y. Sirtuin 6 Overexpression Relieves Sepsis-Induced Acute Kidney Injury by Promoting Autophagy. *Cell Cycle* **2019**, *18*, 425–436. [CrossRef]
61. Chen, J.; Chen, S.; Cai, D.; Wang, Q.; Qin, J. The Role of Sirt6 in Osteoarthritis and Its Effect on Macrophage Polarization. *Bioengineered* **2022**, *13*, 9677–9689. [CrossRef]
62. Guo, Z.; Li, P.; Ge, J.; Li, H. SIRT6 in Aging, Metabolism, Inflammation and Cardiovascular Diseases. *Aging Dis.* **2022**, *13*, 1787–1822. [CrossRef]

63. Yu, H.; Lin, L.; Zhang, Z.; Zhang, H.; Hu, H. Targeting NF-KB Pathway for the Therapy of Diseases: Mechanism and Clinical Study. *Signal Transduct Target Ther.* **2020**, *5*, 209. [CrossRef]
64. Kawahara, T.L.A.; Michishita, E.; Adler, A.S.; Damian, M.; Berber, E.; Lin, M.; McCord, R.A.; Ongaigui, K.C.L.; Boxer, L.D.; Chang, H.Y.; et al. SIRT6 Links Histone H3 Lysine 9 Deacetylation to NF-KB-Dependent Gene Expression and Organismal Life Span. *Cell* **2009**, *136*, 62–74. [CrossRef]
65. Xia, W.; Xiao, J.; Tong, C.L.; Lu, J.; Tu, Y.; Li, S.; Ni, L.; Shi, Y.; Luo, P.; Zhang, X.; et al. Orientin Inhibits Inflammation in Chondrocytes and Attenuates Osteoarthritis through Nrf2/NF-KB and SIRT6/NF-KB Pathway. *J. Orthop. Res.* **2023**, 1–13. [CrossRef]
66. Zhao, Y.; Jia, X.; Yang, X.; Bai, X.; Lu, Y.; Zhu, L.; Cheng, W.; Shu, M.; Zhu, Y.; Du, X.; et al. Deacetylation of Caveolin-1 by Sirt6 Induces Autophagy and Retards High Glucose-Stimulated LDL Transcytosis and Atherosclerosis Formation. *Metabolism* **2022**, *131*, 155162. [CrossRef]
67. Li, B.; Xin, Z.; Gao, S.; Li, Y.; Guo, S.; Fu, Y.; Xu, R.; Wang, D.; Cheng, J.; Liu, L.; et al. SIRT6-Regulated Macrophage Efferocytosis Epigenetically Controls Inflammation Resolution of Diabetic Periodontitis. *Theranostics* **2023**, *13*, 231–249. [CrossRef]
68. Wang, Z.; Wu, Q.; Wang, H.; Gao, Y.; Nie, K.; Tang, Y.; Su, H.; Hu, M.; Gong, J.; Fang, K.; et al. Diosgenin Protects against Podocyte Injury in Early Phase of Diabetic Nephropathy through Regulating SIRT6. *Phytomedicine* **2022**, *104*, 154276. [CrossRef]
69. Bae, E.J.; Park, B.-H. Multiple Roles of Sirtuin 6 in Adipose Tissue Inflammation. *Diabetes Metab. J.* **2023**, *47*, 164–172. [CrossRef]
70. Gao, S.; Yang, Q.; Peng, Y.; Kong, W.; Liu, Z.; Li, Z.; Chen, J.; Bao, M.; Li, X.; Zhang, Y.; et al. SIRT6 Regulates Obesity-Induced Oxidative Stress via ENDOG/SOD2 Signaling in the Heart. *Cell Biol. Toxicol.* **2022**. [CrossRef]
71. Ji, M.L.; Jiang, H.; Li, Z.; Geng, R.; Hu, J.Z.; Lin, Y.C.; Lu, J. Sirt6 Attenuates Chondrocyte Senescence and Osteoarthritis Progression. *Nat. Commun.* **2022**, *13*, 7658. [CrossRef] [PubMed]
72. Kim, S.Y.; Nair, M.G. Macrophages in Wound Healing: Activation and Plasticity. *Immunol. Cell Biol.* **2019**, *97*, 258–267. [CrossRef] [PubMed]
73. Delavary, B.M.; van der Veer, W.M.; van Egmond, M.; Niessen, F.B.; Beelen, R.H.J. Macrophages in Skin Injury and Repair. *Immunobiology* **2011**, *216*, 753–762. [CrossRef] [PubMed]
74. Yang, Z.; Huang, Y.; Zhu, L.; Yang, K.; Liang, K.; Tan, J.; Yu, B. SIRT6 Promotes Angiogenesis and Hemorrhage of Carotid Plaque via Regulating HIF-1α and Reactive Oxygen Species. *Cell Death Dis.* **2021**, *12*, 77. [CrossRef]
75. Zhang, P.; He, L.; Zhang, J.; Mei, X.; Zhang, Y.; Tian, H.; Chen, Z. Preparation of Novel Berberine Nano-Colloids for Improving Wound Healing of Diabetic Rats by Acting Sirt1/NF-KB Pathway. *Colloids Surf. B Biointerfaces* **2020**, *187*, 110647. [CrossRef]
76. Jiang, X.; Yao, Z.; Wang, K.; Lou, L.; Xue, K.; Chen, J.; Zhang, G.; Zhang, Y.; Du, J.; Lin, C.; et al. MDL-800, the SIRT6 Activator, Suppresses Inflammation via the NF- κ B Pathway and Promotes Angiogenesis to Accelerate Cutaneous Wound Healing in Mice. Oxid. *Med. Cell Longev.* **2022**, *2022*, 1619651. [CrossRef]
77. Spallotta, F.; Cencioni, C.; Straino, S.; Nanni, S.; Rosati, J.; Artuso, S.; Manni, I.; Colussi, C.; Piaggio, G.; Martelli, F.; et al. A Nitric Oxide-Dependent Cross-Talk between Class i and III Histone Deacetylases Accelerates Skin Repair. *J. Biol. Chem.* **2013**, *288*, 11004–11012. [CrossRef]
78. O'Connor, M.; Wang, J.V.; Saedi, N. Whole- and Partial-Body Cryotherapy in Aesthetic Dermatology: Evaluating a Trendy Treatment. *J. Cosmet. Dermatol.* **2019**, *18*, 1435–1437. [CrossRef]
79. Huang, X.; Sun, J.; Chen, G.; Niu, C.; Wang, Y.; Zhao, C.; Sun, J.; Huang, H.; Huang, S.; Liang, Y.; et al. Resveratrol Promotes Diabetic Wound Healing via SIRT1-FoxO1-c-Myc Signaling Pathway-Mediated Angiogenesis. *Front. Pharmacol.* **2019**, *10*, 421. [CrossRef]
80. Rahnasto-Rilla, M.; Tyni, J.; Huovinen, M.; Jarho, E.; Kulikowicz, T.; Ravichandran, S.; Bohr, V.A.; Ferrucci, L.; Lahtela-Kakkonen, M.; Moaddel, R. Natural Polyphenols as Sirtuin 6 Modulators. *Sci. Rep.* **2018**, *8*, 4163. [CrossRef]
81. Rahnasto-Rilla, M.; Kokkola, T.; Jarho, E.; Lahtela-Kakkonen, M.; Moaddel, R. N-Acylethanolamines Bind to SIRT6. *ChemBioChem* **2016**, *17*, 77–81. [CrossRef] [PubMed]
82. Chen, Y.; Sun, T.; Wu, J.; Kalionis, B.; Zhang, C.; Yuan, D.; Huang, J.; Cai, W.; Fang, H.; Xia, S. Icariin Intervenes in Cardiac Inflammaging through Upregulation of Sirt6 Enzyme Activity and Inhibition of the NF-Kappa B Pathway. *Biomed. Res. Int.* **2015**, *2015*, 895976. [CrossRef]
83. D'Onofrio, N.; Servillo, L.; Giovane, A.; Casale, R.; Vitiello, M.; Marfella, R.; Paolisso, G.; Balestrieri, M.L. Ergothioneine Oxidation in the Protection against High-Glucose Induced Endothelial Senescence: Involvement of SIRT1 and SIRT6. *Free. Radic. Biol. Med.* **2016**, *96*, 211–222. [CrossRef]
84. Yasuda, M.; Wilson, D.R.; Fugmann, S.D.; Moaddel, R. Synthesis and Characterization of SIRT6 Protein Coated Magnetic Beads: Identification of a Novel Inhibitor of SIRT6 Deacetylase from Medicinal Plant Extracts. *Anal. Chem.* **2011**, *83*, 7400–7407. [CrossRef]
85. Kokkonen, P.; Rahnasto-Rilla, M.; Kiviranta, P.H.; Huhtiniemi, T.; Laitinen, T.; Poso, A.; Jarho, E.; Lahtela-Kakkonen, M. Peptides and Pseudopeptides as SIRT6 Deacetylation Inhibitors. *ACS Med. Chem. Lett.* **2012**, *3*, 969–974. [CrossRef] [PubMed]
86. Kokkonen, P.; Rahnasto-Rilla, M.; Mellini, P.; Jarho, E.; Lahtela-Kakkonen, M.; Kokkola, T. Studying SIRT6 Regulation Using H3K56 Based Substrate and Small Molecules. *Eur. J. Pharm. Sci.* **2014**, *63*, 71–76. [CrossRef] [PubMed]
87. Hooshmand-Moghadam, B.; Eskandari, M.; Golestani, F.; Rezae, S.; Mahmoudi, N.; Gaeini, A.A. The Effect of 12-Week Resistance Exercise Training on Serum Levels of Cellular Aging Process Parameters in Elderly Men. *Exp. Gerontol.* **2020**, *141*, 111090. [CrossRef]

88. Moschen, A.R.; Wieser, V.; Gerner, R.R.; Bichler, A.; Enrich, B.; Moser, P.; Ebenbichler, C.F.; Kaser, S.; Tilg, H. Adipose Tissue and Liver Expression of SIRT1, 3, and 6 Increase after Extensive Weight Loss in Morbid Obesity. *J. Hepatol.* **2013**, *59*, 1315–1322. [CrossRef]
89. Roichman, A.; Elhanati, S.; Aon, M.A.; Abramovich, I.; Di Francesco, A.; Shahar, Y.; Avivi, M.Y.; Shurgi, M.; Rubinstein, A.; Wiesner, Y.; et al. Restoration of Energy Homeostasis by SIRT6 Extends Healthy Lifespan. *Nat. Commun.* **2021**, *12*, 3208. [CrossRef]
90. Zhang, Y.; Yang, Z.; Xu, Z.; Wan, J.; Hua, T.; Sun, Q. Exercise Ameliorates Insulin Resistance and Improves Sirt6-Mediated Insulin Signaling Transduction in Liver of Obese Rats. *Can. J. Physiol. Pharmacol.* **2021**, *99*, 506–511. [CrossRef]
91. Song, M.Y.; Han, C.Y.; Moon, Y.J.; Lee, J.H.; Bae, E.J.; Park, B.H. Sirt6 Reprograms Myofibers to Oxidative Type through CREB-Dependent Sox6 Suppression. *Nat. Commun.* **2022**, *13*, 1808. [CrossRef] [PubMed]
92. Chilton, W.L.; Marques, F.Z.; West, J.; Kannourakis, G.; Berzins, S.P.; O'Brien, B.J.; Charchar, F.J. Acute Exercise Leads to Regulation of Telomere-Associated Genes and Microrna Expression in Immune Cells. *PLoS ONE* **2014**, *9*, e92088. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Review

Multi-Faceted Role of Luteolin in Cancer Metastasis: EMT, Angiogenesis, ECM Degradation and Apoptosis

Maria Teresa Rocchetti [1,†], Francesco Bellanti [2,†], Mariia Zadorozhna [3], Daniela Fiocco [1] and Domenica Mangieri [1,*]

[1] Department of Clinical and Experimental Medicine, University of Foggia, Via Pinto 1, 71122 Foggia, Italy; mariateresa.rocchetti@unifg.it (M.T.R.); daniela.fiocco@unifg.it (D.F.)
[2] Department of Medical and Surgical Sciences, University of Foggia, Via Pinto 1, 71122 Foggia, Italy; francesco.bellanti@unifg.it
[3] Medical Genetics Unit, Department of Molecular Medicine, University of Pavia, Via Forlanini 14, 27100 Pavia, Italy; mariia.zadorozhna@unipv.it
* Correspondence: domenica.mangieri@unifg.it; Tel.: +39-340-251-6009
† These authors contributed equally to this work.

Abstract: Luteolin (3′,4′,5,7-tetrahydroxyflavone), a member of the flavonoid family derived from plants and fruits, shows a wide range of biomedical applications. In fact, due to its anti-inflammatory, antioxidant and immunomodulatory activities, Asian medicine has been using luteolin for centuries to treat several human diseases, including arthritis, rheumatism, hypertension, neurodegenerative disorders and various infections. Of note, luteolin displays many anti-cancer/anti-metastatic properties. Thus, the purpose of this review consists in highlighting the relevant mechanisms by which luteolin inhibits tumor progression in metastasis, i.e., affecting epithelial-mesenchymal transition (EMT), repressing angiogenesis and lysis of extracellular matrix (ECM), as well as inducing apoptosis.

Keywords: phytochemicals; flavonoids; metastasis prevention; signaling pathways

Citation: Rocchetti, M.T.; Bellanti, F.; Zadorozhna, M.; Fiocco, D.; Mangieri, D. Multi-Faceted Role of Luteolin in Cancer Metastasis: EMT, Angiogenesis, ECM Degradation and Apoptosis. *Int. J. Mol. Sci.* **2023**, *24*, 8824. https://doi.org/10.3390/ijms24108824

Academic Editor: Shun-Fa Yang

Received: 18 April 2023
Revised: 9 May 2023
Accepted: 13 May 2023
Published: 16 May 2023

1. Introduction

Cancer is a complex disease, representing the second leading cause of death in the world [1]. In fact, despite the advances in diagnostic and therapeutic protocols—which include surgery, radiotherapy and chemotherapy, as well as target and gene therapy—the survival rate for patients suffering of cancer remains very poor [2–6]. Probably, the real problem of high cancer mortality is its relapse after months to years, through the occurrence of metastases, an event that consists in the spread of cancer cells from a primary lesion to distant sites [7–9]. Currently available cancer/metastatic treatments often induce high toxicity and are associated with several types of adverse events, which are frequently unpredictable and unexplained [10–15]. Thus, finding novel and more efficacious strategies to treat cancer and prevent metastasis is strongly encouraged [8,16–19]. Because of their low toxicity, and due to their ability to target multiple cell signalings, current therapeutic compounds for cancer include phytochemicals [20,21].

Luteolin is one of the most widespread food-derived flavonoids showing preventive and therapeutic effects against several cancer types [22–24]. The main purpose of this review is to summarize the most relevant knowledge about the antimetastatic properties of luteolin, emphasizing its interference with key events underlying tumor progression and expansion, namely (i) epithelial-to-mesenchymal transition (EMT), (ii) angiogenesis, (iii) degradation of the extracellular matrix (ECM) and iv) induction of apoptosis (Figure 1). In detail, we report and discuss the most relevant findings from in vitro and in vivo studies (including animal models) that describe the effects of luteolin at both the cellular and molecular level, focusing on major signaling pathways involved in tumor metastasis (Table 1).

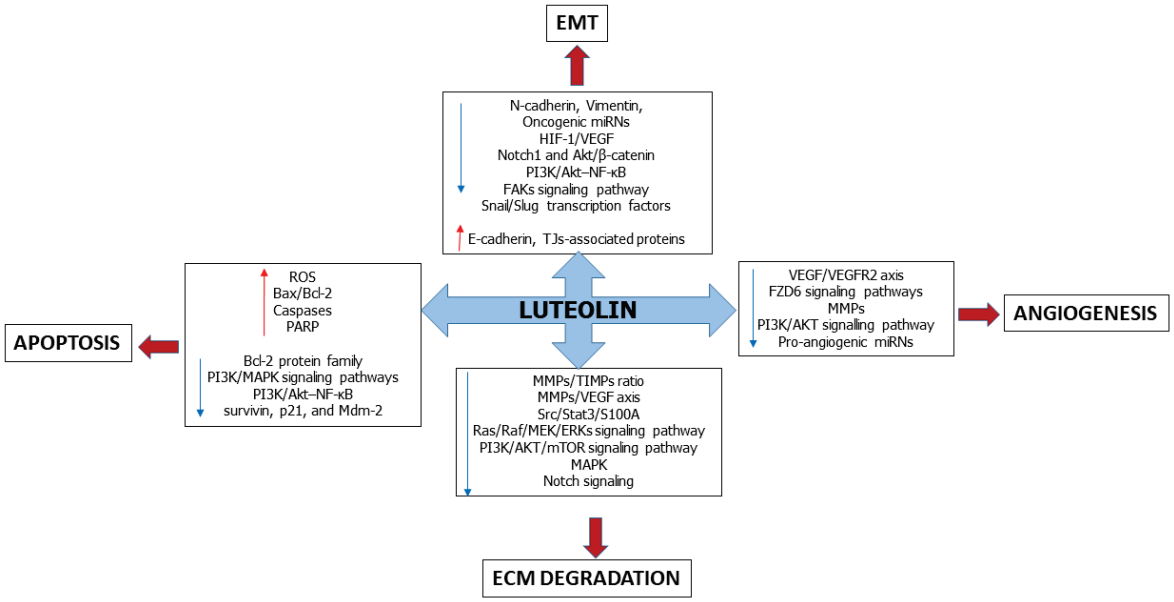


Figure 1. Schematic diagram showing the main molecular targets of luteolin during the initial steps of the metastatic cascade. The flavone interferes with the epithelial-mesenchymal transition (EMT) by inhibiting the expression of N-cadherin, vimentin and several molecular regulators of mesenchymal cell phenotype (e.g., miRNs, HIF-1/VEGF axis, etc.), whereas it enhances the epithelial features. Luteolin exerts anti-angiogenic effects, mainly by blocking the VEGF/VEGFRs signaling pathway and by inhibiting MMPs. Moreover, this compound hampers cell invasiveness by slowing down the extracellular matrix (ECM) degradation through the blockade of MMPs activation and a set of related intracellular pathways. Luteolin induces apoptosis (both the intrinsic and extrinsic routes) by acting on the expression and activities of correlated effectors.

Table 1. Molecular effects of luteolin in different cancer types.

Tumor Entity	EMT	Angiogenesis	ECM Degradation	Apoptosis
Breast Cancer	Reversal of EMT by suppressing β-catenin signaling; inhibition of cancer cell invasion and metastatic potential [25]. Inhibition of the pro-invasive Ras/Raf/MEK/ERK signaling; increase of miR-203 expression [26]. Increase of E-cadherin expression and reduction of protein levels of fibronectin, N-cadherin and vimentin; decrease of transcriptional activity of YAP/TAZ [27].	Blockade of VEGF secretion within breast cancer cells responsive to natural and synthetic progestins, both in vitro and in a xenograft model [28]. Repression of VEGF secretion by TNBC cells and suppression of their metastatic potential in vitro and in vivo [29]. Repression of Notch signaling and its downstream targets Notch-1, Hes-1, VEGF and gelatinases by regulating the level of oncogenic miRs [30].	Modulation of the biogenesis of specific miRs, inhibition of gelatinases secretion and VEGF/Notch signaling pathway [30]. Epigenetically downregulation of gelatinases expression and activation of AKT/mTOR signaling pathway [31].	Reduction of telomerase expression by targeting NF-κB/c-Myc; increase of Bax/Bcl-2 ratio and caspase-3 [32]. By modulating miR-21 and miR-16, upregulation of Bax/Bcl-2 ratio; triggering of both the intrinsic and the extrinsic pathways of apoptosis [33]. Repression of PI3K/Akt pathway; induction of FOXO3a expression and increase of p21 and p27; induction of PARP cleavage and release of cytochrome *c* [34].

Table 1. *Cont.*

Tumor Entity	EMT	Angiogenesis	ECM Degradation	Apoptosis
Gastric Carcinoma	Reversal of EMT by inhibiting Akt/β-catenin and Notch signaling pathways [35]. Decreased migration and invasion by regulating Notch1/ PI3K/ AKT/ mTOR/ERK/STAT3 and P38 signaling pathways; regulation of several oncogenic miRs expression in vitro and in vivo [36].	Suppression of VEGF secretion by acting on Notch1 expression and inhibition of the formation of tube-like structures of HUVECs seeded in a Matrigel layer [37].	, Reduction of gelatinases expression via inhibition of cMet/Akt/ERK signaling [38].	Suppression of PI3K/MAPK signaling with increase of Bax/Bcl-2 ratio and cytochrome c release [39]. Decrease of Bcl-2 expression through upregulation of miR-34a [40].
Lung Cancer	Reversal of TGF-β1 induced-EMT by slowing down the activation of PI3K/Akt/IκBa/NF-κB/Snail pathway [41]. Inhibition of hypoxia-induced EMT by blocking integrin β1 expression and FAK-signaling pathway [42].	Repression of VEGF and gelatinases by upregulating miR-133a-p69, and by regulating MAPK/PI3K/Akt signaling pathways [43].	Downregulation of the pro-metastatic markers CXCR4, gelatinases in vitro and in vivo [44].	Induction of ROS accumulation via suppression of SOD activity; suppression of NF-kB potentiating JNK to sensitize cancer cells to TNF [45]. Upregulation of miR-34a-5p via targeting MDM4 oncogene, increase of p53 and p21expression; increase of Bax/Bcl-2 ratio, followed by activation of caspase-3 and -9 [46].
Pancreatic cancer	Deactivation of STAT3 signaling with consequent reversal of Il-6-induced EMT [47].	Decrease of VEGF secretion and VEGF mRNA expression via NF-κB inhibition [48].	-	Attenuation of EGFR signaling pathway and induction of PARP degradation followed by DNA fragmentation [49].
Melanoma	Upregulation of E-cadherin/N-cadherin ratio through inhibition of HIF-1α/VEGF axis [50]. Enhancement of E-cadherin expression via inhibition of β3 integrin/FAK signal pathway in vitro and in vivo [51].	Inhibition of different pathways of tumor neovascularization, including angiogenesis, vasculogenesis and vasculogenic mimicry, by suppressing PI3K/AKT signaling pathway [52].	Downregulation of the pro-metastatic markers, gelatinases and CXCR4 [53].	-
Choroidal melanoma	-	Reduction of VEGF secretion, in a concentration-dependent manner, followed by induction of cell death [54].	Decrease of gelatinase secretion in vitro via inhibition of PI3K/Akt signaling pathway [55].	Increase of Bax/Bcl-2 ratio [52].

Table 1. *Cont.*

Tumor Entity	EMT	Angiogenesis	ECM Degradation	Apoptosis
Prostate Cancer	-	Inhibition of VEGFR2/AKT/ERK/mTOR/P70S6K signaling pathway and of neovascularization in ex vivo chicken chorioallantoic membrane (CAM) assay, and in a Matrigel plug assay, as well as in a xenograft model [56].	-	Downregulation of miR-301 that promotes the expression of pro-death DNA-binding effector domain-containing protein 2 (DEDD2) [57].
Haemangioma	-	Suppression of VEGF-A expression with consequential inhibition of microvessel density and vasculogenesis in vivo targeting FZD6 signaling pathway [58].	-	-
Glioblastoma	-	-	Inactivation of the p-IGF-1R/PI3K/AKT/mTOR signaling pathway and alteration of the gelatinase/TIMPs ratio [59].	Stimulation of PARP cleavage, DNA degradation and caspases activation [60].
Colon cancer	-	-	-	Activation of antioxidant enzymes and MAPK signaling; by unbalancing ROS, acting on cytochrome c release and caspase-9 and -3 activation [61]. Activation of caspases 3, 7, 9 and PARP cleavage; downregulation of p21, survivin, Mcl-1, Bcl-x(L) and Mdm-2 [62]. By involving Wnt/β-catenin/GSK-3β signaling, increase of Bax/Bcl-2 ratio and activation of caspase-3 [63].
Colorectal Cancer	-	-	Inactivation of gelatinases by suppressing Raf/PI3K signaling pathways [64]. Downregulation of MMP-2, -9, -3 and -16 expression coupled with an enhancement of miR-384 biogenesis and with suppression of PTN [65].	-

Table 1. *Cont.*

Tumor Entity	EMT	Angiogenesis	ECM Degradation	Apoptosis
Cholangiocarcinoma	-	-	-	Inhibition of Nrf2 with consequent downregulation of the antioxidant genes γ-glutamylcysteine ligase and heme oxygenase-1, and increase of mitochondrial membrane potential dissipation and caspases -3 and 9 activation [61].
Cervical cancer	-	-	-	Disruption of pro-apoptotic/anti-apoptotic genes equilibrium interfering with the RAS-RAF/MAPK/AKT/PI3K signaling pathway; triggering collapse of the mitochondrial membrane and DNA fragmentation [33].
Leukemia	-	-	-	Induction of histone H3 hyper-acetylation by activating the ERK /JNKs pathways; increase of Fas and FasL expression culminating in caspases-8/-3 activation [66].

2. Luteolin

Luteolin is a flavonoid belonging to the flavone family; it is isolated from several vegetables and edible herbs, including radicchio, broccoli, raw brussels sprouts, onion leaves, parsley, carrots, peppers and rosemary, where it can occur either as aglycone or bound to one or several carbohydrates as glycoside [67] (Figure 2).

Luteolin is extracted as a yellow crystalline compound and its chemical structure presents a classic flavone C6-C3-C6 skeleton, consisting in two benzene rings with 4 hydroxyl groups located at positions 3, 4, 5 and 7, and one oxygen-containing ring which presents a C2-C3 double bond [67]. All these functional groups account for the biological/biochemical properties of luteolin, some of them being specifically involved in maintaining redox balance in various pathological processes [68,69] (Figure 3).

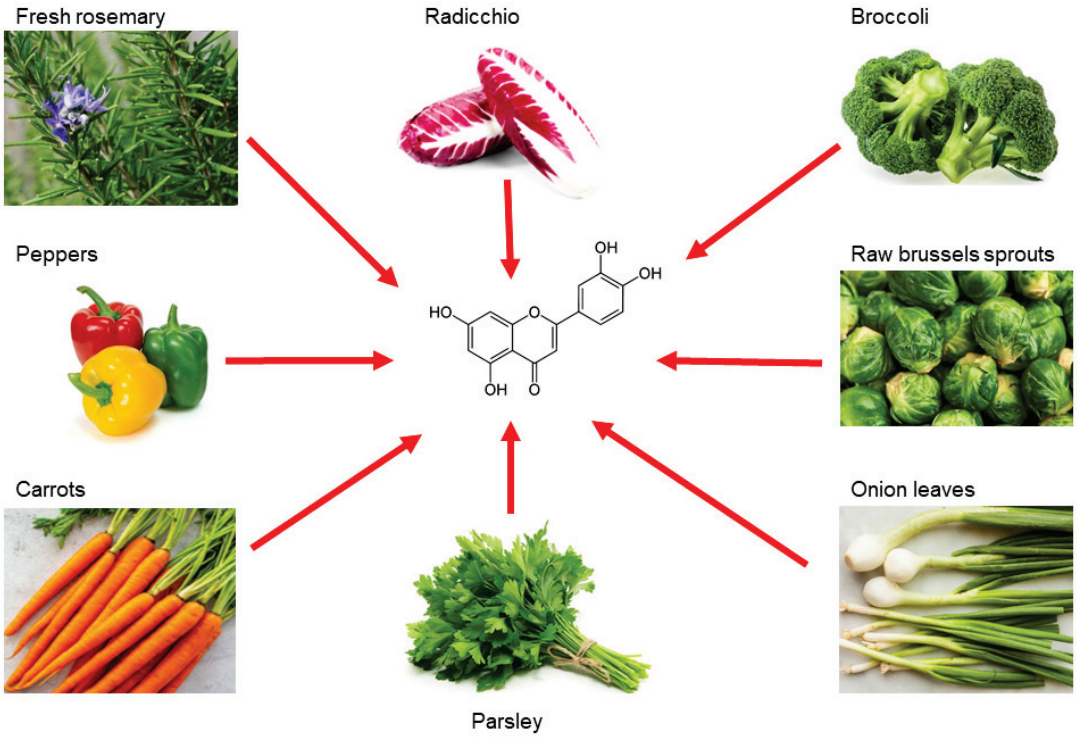


Figure 2. Main sources of luteolin. Edible vegetables, such as radicchio, broccoli, raw brussels sprouts, onion leaves, parsley, carrots, peppers and rosemary are rich in luteolin.

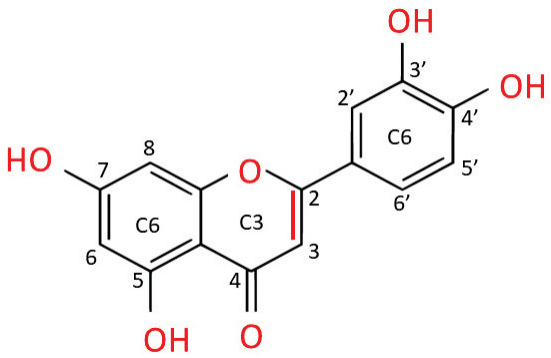


Figure 3. Chemical structure of luteolin (3,4,5,7-tetrahydroxy flavone). Luteolin is a flavone belonging to a group of hydrophobic, naturally occurring compounds, named flavonoids. The functional groups involved in oxido-reductive properties are marked in red.

In fact, preclinical studies ascribed to luteolin several pharmacological properties, including anti-inflammatory, neuroprotective, antimicrobial/antiviral, cardioprotective, antidiabetic and pro-/antioxidant effects [70]. Interestingly, since this flavone can interact with various signaling pathways, experimental evidence attributes to luteolin important chemopreventive effects, indicating its ability to interfere with almost all cellular processes underlying cancer development, including metastasis formation [23,24,71]. The available data related to the pharmacokinetics of the free or glycosylated forms of luteolin derive mainly from studies on rat models [72–74]. Despite the numerous beneficial effects already mentioned, studies on absorption, metabolism and bioavailability of luteolin in humans are very difficult, partly due to its hydrophobicity, which affects bioavailability and limits the yields of the bioactive flavonoid [70].

3. Luteolin Affects the Epithelial-Mesenchymal Transition

EMT is the differentiation process of epithelial cells toward mesenchymal ones; several events are observed during this process, including a disorganization of epithelial cell

polarity, the dissolution of cellular junctions, as well as a reorganization of the cytoskeleton [75–77]. Thus, EMT involves the downregulation of epithelial markers expression, such as E-cadherin, claudins, zonula occludens-1 (ZO-1), and the acquisition of many facets of the mesenchymal cells, including the expression of N-cadherin and vimentin, coupled with a high propensity to cell motility, invasiveness, and resistance to anoikis [76–78]. Several signaling pathways have been identified in EMT induction, such as the transforming growth factor-beta (TGF-β), Notch, Wnt/β-catenin, as well as the Hippo-YAP/TAZ pathways [79,80]. Moreover, several transcription factors (and their target genes) promote EMT, including Snail1/2, Twist1/2, ZEB1/2 and hypoxia-inducible factors 1/2 (HIF1/2) [81]. Physiologically, EMT occurs during embryonic development and in tissue remodeling. In cancer, this process assumes a crucial role in promoting metastasis [76,77]. Therefore, blocking or reversing EMT may represent an attractive approach to prevent cancer spreading [77].

As aforementioned, luteolin inhibits or prevents cell invasion and metastasis in several cancer types, due to a modulation of EMT [82]. In this regard, both in vitro and in vivo studies demonstrated that treatment of highly metastatic triple-negative breast cancer (TNBC) with luteolin reduced β-catenin expression and downregulated mesenchymal markers, such as N-cadherin, vimentin, Snail and Slug [25]. Moreover, cancer cells regained their epithelial features by overexpressing cell-cell junctional proteins, such as E-cadherin and claudins [25].

MicroRNAs (miRs) are small, endogenous, noncoding RNAs that can post-transcriptionally regulate gene expression and play an important role in maintaining normal cellular functions [83]. On the other hand, growing evidence shows that some miRs participate in the initiation and progression of cancer, taking part in the EMT process [84,85]. Luteolin was able to reverse EMT in breast cancer cells (MDA-MB-453 and MCF-7) by overexpressing miR-230, which, in turn, inhibited the Ras/Raf/MEK/ERK signaling pathway, known as a marker of cancer invasiveness [26]. Furthermore, luteolin administration significantly inhibited gastric carcinoma (GC) upregulating miR-139, miR-34a, miR-422a, miR-107 levels, while suppressing the oncogenic expression of miR-21, miR-155, miR-224, miR-340 [36]. Regulation of this panel of miRs was accompanied by reduced cell proliferation, cell cycle arrest, and induction of apoptosis [36]. In a different study, the flavonoid inhibited the oncogenic properties of YAP/TAZ in highly metastatic breast cancer, both in vitro and in xenograft models [27]. Moreover, Zang et al. demonstrated that, by interfering with Notch1 and Akt/β-catenin signaling in GC, luteolin reversed the EMT process and, consequently, inhibited tumor progression and invasion, both in vitro and in vivo [35]. By using cultures of human lung adenocarcinoma cells (A549), Chen et al. showed that luteolin inhibited cell proliferation and migration through an attenuation of TGF-β1-induced EMT, by activating the PI3K/Akt–NF-κB–Snail signaling pathway [41]. Furthermore, luteolin (in a time- and dose-dependent manner) reversed IL-6-induced EMT acting on STAT3 signaling and, consequently, reduced the invasiveness of cancer cells by inhibiting the release of metalloproteases (MMPs) in in vitro models of human pancreatic cancer (i.e., Panc-1 and SW1990 cells) [47].

As abovementioned, HIF-1 and HIF-2 are implicated in cancer-associated EMT [81]. Based on these premises, Li et al. observed that luteolin, by inhibiting the HIF-1α/VEGF signaling pathway, decreased the expression of N-cadherin and vimentin mesenchymal markers, while augmenting the level of epithelial cadherin isoform in human and murine melanoma cells (i.e., A375 and B16-F10 cell lines, respectively) [50]. The inversion of hypoxia-induced EMT was also observed in a murine melanoma model, where luteolin inhibited lung metastasis formation by stopping the β3 integrin/FAK signaling pathway [51]. Similarly, luteolin prevented hypoxia-induced EMT of human non-small cell lung carcinoma cells (NSCLC; A549 and NCI-H1975 cell lines), as evidenced by a downregulation of mesenchymal specific markers, such as vimentin and N-cadherin; contextually, luteolin treatment repressed cancer cell motility and adhesion, interfering with integrin β1 expression and with the FAK-signaling pathway [42].

The ability of luteolin to reverse EMT in cancer cells is often enhanced by adding other flavonoids, such as quercetin [86,87]. Indeed, a mixture of luteolin and quercetin was able to attenuate (in a time- and dose-dependent manner) EMT-correlated events, migration and invasiveness of human squamous carcinoma, both in vivo and in vitro, by suppressing the Src/Stat3/S100A signaling pathway [86]. Similarly, the addition of luteolin to quercetin inhibited metastasis of skin squamous carcinoma by blocking the Akt/mTOR/c-Myc signaling pathway to suppress RPS19-activated EMT signaling [87].

In summary, luteolin appears capable to block or reverse EMT by acting on multiple molecular targets, hindering the first steps of cancer spreading (Figure 4).

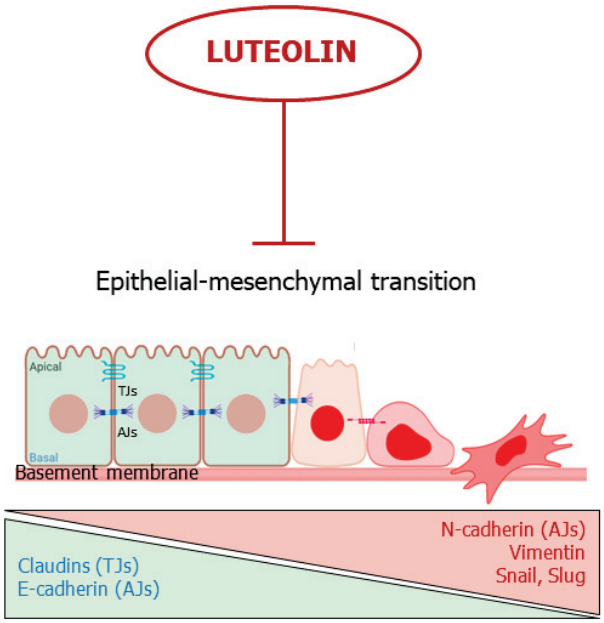


Figure 4. Luteolin interferes with the epithelial-mesenchymal transition. The flavone inhibits or reverses EMT. TJs: Tights Junctions; AJs; Adherens Junctions.

4. Luteolin Suppresses Angiogenesis

Angiogenesis is the process by which new blood vessels form, starting from pre-existing vasculature; this event depends on pro-angiogenic mediators, such as vascular endothelial growth factor (VEGF), basic-fibroblast growth factor (bFGF), metalloproteases, etc., and on negative regulators of angiogenesis, including thrombospondin and endostatin [88]. Angiogenesis plays a crucial role during a variety of physiological processes, such as wound healing, embryonic development and pregnancy [88]. On the other hand, neovascularization is a key event for pathological processes, such as tumor progression, invasion and metastatic cascade [89]. This last aspect is pharmacologically relevant: in fact, many authorized cancer therapies are directed against the tumor-associated vessels [89–92].

Several investigations demonstrate that various flavonoids—including luteolin—act as negative regulators of VEGF and other pro-angiogenic factors [93–95]. For instance, Cai and co-workers demonstrated that luteolin decreased VEGF secretion and *VEGF* mRNA expression in pancreatic carcinoma cells (PANC-1, CoLo-357 and BxPC-3 cell lines), via inhibition of the NF-κB transcriptional factor activity [48]. Furthermore, treating human choroidal melanoma cells (C918 and OCM-1) with luteolin not only inhibited VEGF expression, but also increased cancer cell death [54]. Additionally, an in vitro study on breast cancer showed that luteolin, by interfering with Notch1 expression, inhibited VEGF secretion from tumor cells, decreased endothelial cell migration, proliferation, and their propension to form tube-like structures on a Matrigel layer [37]. Cook and co-coworkers demonstrated the ability of luteolin to block the production of VEGF in human breast cancer cells (T47-D and BT-474) responsive to (natural and synthetic) progestins, both in vitro and in a xenograft model [28]. In a similar study, the authors showed that luteolin

(in a dose- and time-dependent fashion) significantly reduced VEGF secretion in human TNBC cells (i.e., MDA- MB-435 and highly aggressive MDA-MB-231 (4175) LM2), coupled with a significantly decreased cell viability and reduced migration and invasion in vitro; at the same time, the authors showed that the flavonoid inhibited lung metastasis formation in a dedicated xenograft model [29].

The anti-proliferative/anti-mitotic effect on endothelial cells was also tested in infantile haemangioma in vitro and in vivo, by using haemangioma-derived stem cells (HemSC) [58]. In detail, the results of this study demonstrated that luteolin suppressed VEGF expression and inhibited HemSC growth in a dose-dependent manner, and, at the same time, the flavonoid inhibited both angiogenesis and vasculogenesis, in a murine model, by acting on the frizzled6 (FZD6) signaling pathway [58]. In prostate cancer, luteolin strongly suppressed neovascularization in different experimental settings by inhibiting the activation of VEGFR-2/AKT/ERK/mTOR/P70S6K signaling pathways [56]. In fact, it abolished angiogenesis in an ex vivo chicken chorioallantoic membrane (CAM) assay and in a Matrigel plug assay; in addition, the flavone suppressed both vascularization and growth of the tumor in a xenograft model [56]. Moreover, luteolin inhibited different pathways of vascularization in an in vitro model of uveal melanoma, including angiogenesis, vasculogenesis and vasculogenic mimicry, by suppressing the PI3K/AKT signaling pathway [52].

As previously cited, miRs have a role in cancer evolution, including angiogenesis [85]. In an experimental model of NSCLC, luteolin repressed the expression of the rich element binding protein B (PURB) and crucial proangiogenic factors, including VEGF and MMP-2/-9, by overexpressing the miR-133a-p69 and by regulating MAPK and PI3K/Akt signaling pathways [43]. In breast cancer, the flavonoid upregulated the biogenesis of miR-34a, miR-181a, miR-139-5p, miR-224 and miR-246, while it decreased the level of miR-155, coupled with a significant inhibition of VEGF/Notch signaling and MMPs downregulation [30].

Tumor growth and its progression also depend on proangiogenic factors secreted by the surrounding microenvironment cells (e.g., fibroblasts, tumor associated macrophages (TAMs), etc.) [8]. Fang et al. demonstrated that luteolin inhibited the ability of TAMs to induce angiogenesis, thereby inhibiting tumor growth and its spreading, in both normoxic and hypoxic conditions [96].

Taken together, these studies show promise for luteolin as a potent anti-angiogenic agent, evading tumor evolution and metastatic cascade (Figure 5).

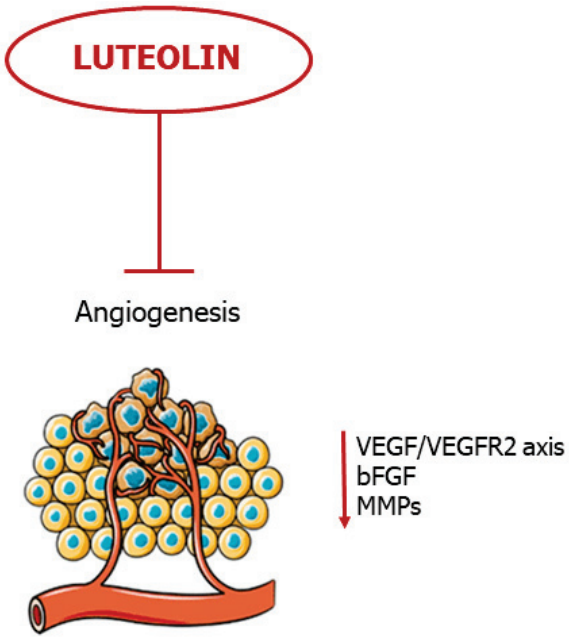


Figure 5. Luteolin blocks neovascularization. The flavonoid blocks angiogenesis by negative regulation of pro-angiogenic mediators.

5. Luteolin Slows down Extracellular Matrix Degradation

Lysis of extracellular matrix (ECM) molecules occurs through the proteolytic action of a family of zinc-dependent metalloproteases (MMPs), facilitating angiogenesis and promoting cancer cell invasion and dissemination [97–100]. Luteolin significantly inhibited MMP-2 and MMP-9 (also known as gelatinases) and VEGF expression by suppressing Notch signaling and by modulating specific miRs with a crucial role in breast cancer progression [30]. Another study showed that the flavonoid repressed the invasiveness of androgen receptor-positive TNBC cells by epigenetically downregulating MMP-9 expression (due to a hypoacetylation of the histone H3) via the induction of the AKT/mTOR signaling pathway [31]. A substantial downregulation of migratory propensity was also reported in an in vitro model of human glioblastoma, where an unbalanced MMPs/TIMPs ratio was attributed to inhibition of the pro-invasive p-IGF-1R/PI3K/AKT/mTOR signaling pathway [59]. Furthermore, the flavonoid abolished MMP-2 and MMP-9 activity in a concentration-dependent effect by suppressing pro-invasive Raf/PI3K signaling in murine colorectal carcinoma (CRC) [64]. Recent studies have further reinforced the negative effect of luteolin in CRC dissemination [65]. In fact, in in vitro and in vivo models of this tumor, luteolin induced a downregulation of gelatinases and MMP -3 and -16 expression, coupled with enhanced miR-384 biogenesis, as well as suppression of pleiotrophin (PTN), a small cytokine with a polyhedral role in tumor evolution (PTN promotes cancer cell migration and invasion, and stimulates angiogenesis) [65]. An independent research showed that the flavonoid inhibited the spreading of GC, both in vitro and in a xenograft model, due to a downregulation of MMP-9 associated to a suppression of cMet/Akt/ERK signaling [38]. Chemopreventive properties of the abovementioned compound were further supported by a B16F10 mouse xenograft model, in which the levels of pro-metastatic markers, including MMP-9, MMP-2 and CXCR4, were significantly decreased in the lung tissues isolated from tumor-bearing nude mice after luteolin treatment [53]. Additionally, Shi and co-workers showed that luteolin (in a dose-dependent manner) suppressed the proliferation, migration and invasion of human choroidal melanoma cells in vitro by blocking the secretion of the gelatinases by cancer cells, essentially interfering with the PI3K/Akt signaling pathway [55].

In conclusion, luteolin is able to interfere with different initial steps of the metastatic cascade, including ECM degradation, angiogenesis and cell invasion, by blocking activation pathways of MMPs (Figure 6).

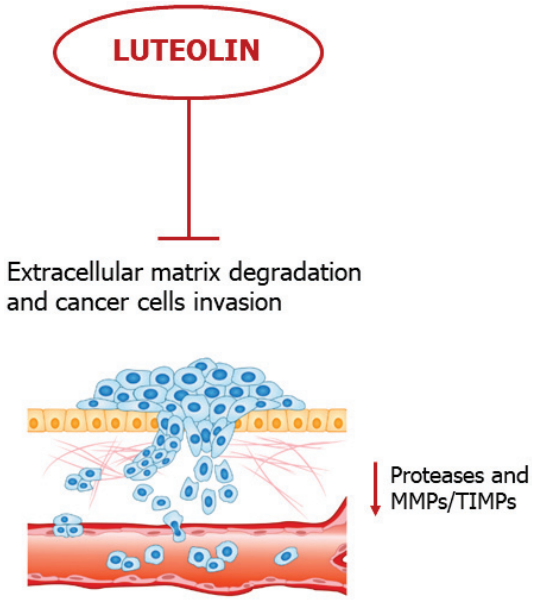


Figure 6. Luteolin slows down ECM lysis. Cancer cell invasion is inhibited by luteolin through suppression of proteases activity.

6. Luteolin Induces Apoptosis

Apoptosis, a programmed cell death, is an essential mechanism involved in several physiological conditions, including organ development and tissue homeostasis, due to its role in controlling cell depletion, somehow opposing to mitosis in the regulation of cell proliferation [101]. In the tumor context, the intricate machinery of apoptosis plays a polyhedral role, ranging from malignant transformation of cells to metastatic cascade or overstepping anticancer drugs' resistance, thus suggesting that the balance of pro-survival and pro-death pathways can be impaired at several steps of the apoptotic process [102]. Apoptosis can occur through either the death receptor-mediated extrinsic pathway and/or the mitochondria-mediated intrinsic route, both converging to caspases (cysteine-aspartate proteases) activation, resulting in characteristic morphological and biochemical cellular changes [103].

Apoptotic cellular death can be triggered by different cellular events, including a perturbation of the intracellular redox status, followed by reactive oxygen species (ROS) overproduction and enhanced cellular oxidative stress [44,104]. In this regard, luteolin exhibits pro-apoptotic effects on various cancer cells, acting both on antioxidant activity and on ROS overproduction [45,61,104,105]. Luteolin- induced accumulation of ROS, suppressed NF-κB and potentiated JNK in an in vitro model of NCSLC, hence sensitizing lung cancer cells to TNF-triggered apoptosis [45]. Furthermore, luteolin caused an overproduction of ROS, coupled with the release of cytochrome c from collapsed mitochondria, and with the activation of caspase-3 and -9 in cholangiocarcinoma cells [61]. These molecular changes were associated with inhibition of the Nrf2 transcription factor, with resulting downregulation of its antioxidant target genes, including those encoding γ-glutamylcysteine ligase and heme oxygenase-1 [61]. Kang and co-workers showed that in vitro treatment of human colon cancer (CC) cells (HT-29 cell line) with luteolin was able to cause apoptosis [105]. The authors described an upregulation of intracellular and mitochondrial ROS and enhanced Bax/Bcl-2 ratio, followed by release of cytochrome c from mitochondria to the cytosol and activation of caspase-9 and -3 [105]. Other studies further support the pro-apoptotic effects of the flavone in cancer; in fact, HT-29 cells treated with 60 μM of luteolin for 48 h showed an increased activation of caspase-3, -7, and -9 and Poly-ADP ribose polymerase (PARP) cleavage, followed by the upregulation of several members of the Bcl-2 protein family (Bcl-xL, Mcl-1, survivin, p21 and Mdm-2) [62]. Similarly, luteolin induced apoptosis of human colon tumor cells (HCT-15 cell line), increasing the Bax/Bcl-2 ratio and activating caspase-3 via Wnt/β-catenin/GSK-3β signaling [63]. The intrinsic pro-apoptotic effect of luteolin was also tested in an in vitro model of human GC [39]. In detail, upon exposure to high concentration of the flavonoid (60 μM), GC cells (BGC-823 cell line) showed a significant increase of the Bax/Bcl-2 ratio, coupled with the release of cytochrome c from mitochondria and consequent caspases activation [39]. These molecular events were correlated with the suppression of PI3K/MAPK signaling pathways [39]. In a very extensive study, luteolin treatment induced apoptosis in HeLa cells, by either the intrinsic or extrinsic pathways [106]. Indeed, increased expression of various pro-apoptotic factors, including BAX, BAD, BID, APAF1, TRADD, FAS, FADD, Caspase-3 and -9, correlated with a reduction of anti-apoptotic mediators, such as MCL-1 and BCL-2. Additionally, in this experimental context, luteolin was found to suppress the AKT/mTOR and MAPK/ERK1/2 pathways, highlighting the pro-apoptotic features of protein p53 [106]. Interestingly, luteolin triggered apoptosis through the downregulation of the human telomerase reverse transcriptase (hTERT) via the NF-κB/c-Myc pathway, which correlated to increased Bax/Bcl-2 ratio and caspase-3 protein expression in human breast cancer cells [32]. Another study, performed in HL-60 leukemia cells, demonstrated that luteolin caused a hyperacetylation of histone H3 by activating the ERK and JNKs pathways, with consequential increase of Fas and FasL expression, followed by caspase-8 and -3 activation [66].

In addition to the modulation of transcription events and/or epigenetic changes, several miRs are integrated into critical pathways of cancer spreading [107]. Accordingly, in vitro studies by Han and collaborators demonstrated that luteolin induced apoptosis

in prostate cancer (PC3 and LNCaP cell lines) through downregulation of miR-301, coupled with the expression of pro-death DNA-binding effector domain-containing protein 2 (DEDD2) [57]. Another study showed that administration of the flavone to human breast cancer cells (MCF-7) modulated the level of miR-21 and miR-16, reducing cell viability in a dose- and time-dependent manner, but also inducing apoptosis through the intrinsic and extrinsic pathways via upregulation of Bax/Bcl-2 expression [33]. Furthermore, luteolin decreased Bcl-2 expression by upregulating miR-34a in human GC (BGC-823 and SGC-7901 cell lines) [40]. Moreover, Jiang and co-workers demonstrated that luteolin triggered intrinsic apoptosis in NSCLC, in both in vitro and in vivo studies, by enhancing the level of miR-34a-5p and targeting the oncogene MDM4, thus indirectly increasing the availability of the pro-apoptotic features of p53 and p21 [46].

Several cancers acquire drug resistance through different mechanisms, including involvement of PI3K/Akt pathway and tyrosine-kinases activation [4,108–110]. Thus, inhibition of PI3K/Akt pathway seems to be crucial to stop cancer progression. Indeed, Lin and co-workers showed that luteolin suppressed PI3K/Akt phosphorylation in an in vitro model of human breast cancer (MCF-7 cells), through the induction of forkhead box O3 (FOXO3a), followed by DNA damage; all these events culminated in mitochondrial apoptotic cascade [34]. The epidermal growth factor receptor (EGFR) is a tyrosine kinase receptor which is commonly upregulated in several types of carcinoma [111]. Various tyrosine kinases inhibitors (TKIs) can suppress EGFR phosphorylation and, consequently, suppress its downstream signaling pathways, including ERK and AKT, with the final effect of inducing apoptosis [112]. Nevertheless, first responses to TKIs (or even to EGFR antibodies) are followed by acquisition of drug resistance for several cancer types. Therefore, studies are consistently searching for new therapeutic alternatives [4,110]. In this scenario, medicinal plants extracts, including luteolin, have been ascribed surprising efficacy as promising anticancer and anti-metastatic compounds [5]. For example, human ductal pancreatic cancer cells (MiaPaCa-2) treated with luteolin showed a markedly decreased activation of EGFR and its downstream protein kinases activity; indeed, treated cells exhibited typical apoptosis features, such as shrinkage of the cell morphology, PARP cleavage and DNA fragmentation [49]. Similarly, luteolin attenuated cell proliferation and growth by inhibiting EGFR/MAP/ERK phosphorylation in an in vitro model of human glioblastoma (U-87 MG and U-251 MG cell lines), inducing apoptosis via PARP cleavage and activation of caspase cascade [60].

To sum up, the anti-metastatic effect of luteolin may depend on its ability to stimulate both intrinsic and extrinsic pathways of apoptosis, by promoting an increase of Bax/Bcl-2 ratio, cytochrome c release and the activation of caspases cascade, by modulating the expression of oncogenic miRs and, in some cases, by overstepping drug resistance (Figure 7).

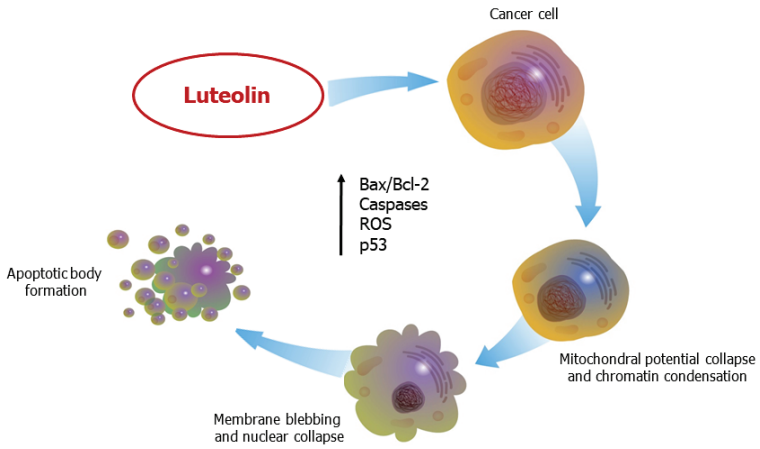


Figure 7. Luteolin induces apoptosis in cancer cells. Both the extrinsic and intrinsic pathways are triggered by luteolin.

7. Conclusions and Perspectives

Luteolin is able to quench the initial steps of the metastatic cascade. In fact, it inhibits EMT, tumor-associated angiogenesis and ECM degradation. Furthermore, by targeting multiple cellular signals, this flavone induces apoptosis in malignant cells through both the intrinsic pathway and receptor-mediated extrinsic route, with minimal side effects and insignificant toxicity on normal cells [71]. Additionally, luteolin is not only able to overstep drug resistance, but also represents a good adjuvant to other flavonoids, including quercetin [86,87]. However, most of the known activities of luteolin have been investigated only in vitro and/or in animal models, except for a very few non-oncology clinical studies in humans [113–118]. Besides, to date, no clinical trials using luteolin as cancer therapy have been performed yet. The main challenges faced by investigators to test this compound in cancer studies include limited funding, absence of product consistency, contamination, and manufacturing difficulties. Indeed, its biological applications are currently limited, due to its hydrophobic nature, although improvement of luteolin bioavailability could be achieved by using nanostructured lipid carriers, microemulsions, and other similar devices [118–121]. Since preclinical experiments strongly advise for the potential efficacy of luteolin in cancer treatment, the development of strategies that improve its bioavailability, increase its efficacy and decrease its toxicity, will support future clinical studies.

Author Contributions: M.T.R. and F.B. wrote the manuscript; M.T.R. prepared the illustrations; M.Z. and D.F. reviewed the paper; D.M. conceived and supervised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

Akt	Protein-kinase B
Bax	BCL2 associated X, apoptosis regulator
Bcl-2	B-cell lymphoma 2
bFGF	Basic-fibroblast growth factor
CAM	Chorioallantoic membrane
Caspases	Cysteine-aspartate proteases
CC	Colon cancer
C-Myc	bHLH transcription factor
CXCR4	C-X-C motif chemokine receptor-4
DNA	Deoxyribonucleic Acid
EGFR	Epidermal growth factor receptor
ECM	Extracellular matrix
EMT	Epithelial-mesenchymal transition
ERK1/2	Extracellular-regulated kinase1/2
FAKs	Focal adhesion kinases
FZD6	Frizzled family receptor 6
GC	Gastric cancer
GSK-3β	Glycogen synthase kinase 3-beta
HIF1/2	Hypoxia-inducible factor 1/2
hTERT	human telomerase reverse transcriptase
HUVECs	Human umbilical vein endothelial cells
IL-6	Interleukin 6

JNK	Jun N-terminal kinase
MAPK	Mitogen-activated protein kinase
MEK	MAP kinase-ERK kinase
miRs	MicroRNA
MMPs	Metalloproteases
mRNA	Messenger ribonucleic acid.
mTOR	Mammalian target of rapamycin
NF-κB	Nuclear factor-κappaB
Notch	Signal transducer and activator of transcription
NSCLC	Non-small cell lung carcinoma
PARP	Poly-ADP ribose polymerase
PI3K	Phosphatidylinositol-3-kinase
Raf	Rapidly accelerated fibrosarcoma
Ras	Small guanosine triphosphatases
Snail	Snail homolog 1/2 of drosophila
ROS	Reactive oxygen species
RPS19	Ribosomial protein S19
Src	Proto-oncogene tyrosine-protein kinase
STAT3	Signal transducer and activator of transcription-3
TGF-β	Transforming growth factor-beta
TIMPs	Tissue inhibitors of metalloproteases
TNBC	Triple-negative breast cancer
TNF	Tumor necrosis factor
TWIST	Twist family bHLH transcription factor
VEGF	Vascular endothelial growth factor
VEGFR2	Vascular endothelial growth factor receptor-2
ZEB	Zinc finger E-box binding homeobox
ZO-1	Zonula Occludens-1

References

1. WHO. 2020. Available online: http://gco.iarc.fr/today/fact-sheets-cancers (accessed on 5 December 2020).
2. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [CrossRef] [PubMed]
3. Scott, A.M.; Wolchok, J.D.; Old, L.J. Antibody therapy of cancer. *Nat. Rev. Cancer* **2012**, *12*, 278–287. [CrossRef] [PubMed]
4. Liang, X.-J.; Chen, C.; Zhao, Y.; Wang, P.C. Circumventing Tumor Resistance to Chemotherapy by Nanotechnology. *Methods Mol. Biol.* **2009**, *596*, 467–488.
5. Seca, A.M.L.; Pinto, D.C.G.A. Plant Secondary Metabolites as Anticancer Agents: Successes in Clinical Trials and Therapeutic Application. *Int. J. Mol. Sci.* **2018**, *19*, 263. [CrossRef]
6. Singh, V.; Khan, N.J.G. Vector engineering, strategies and targets in cancer gene therapy. *Cancer Gene Ther.* **2022**, *29*, 402–417. [CrossRef] [PubMed]
7. Kirkland, J.L. Tumor dormancy and disease recurrence. *Cancer Metastasis Rev.* **2023**, *42*, 9–12. [CrossRef]
8. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70. [CrossRef]
9. Jiang, W.G.; Sanders, A.J.; Katoh, M.; Ungefroren, H.; Gieseler, F.; Prince, M.; Thompson, S.K.; Zollo, M.; Spano, D.; Dhawan, P.; et al. Tissue invasion and metastasis: Molecular, biological and clinical perspectives. *Semin. Cancer Biol.* **2015**, *35*, S244–S275. [CrossRef]
10. Barber, F.D. Adverse Events of Oncologic Immunotherapy and Their Management. *Asia-Pac. J. Oncol. Nurs.* **2019**, *6*, 212–226. [CrossRef]
11. Magee, D.E.; Hird, A.E.; Klaassen, Z.; Sridhar, S.S.; Nam, R.K.; Wallis, C.J.D.; Kulkarni, G.S. Adverse event profile for immunotherapy agents compared with chemotherapy in solid organ tumors: A systematic review and meta-analysis of randomized clinical trial. *Ann. Oncol.* **2020**, *31*, 50–60. [CrossRef]
12. Galateanu, B.; Pușcașu, A.I.; Tircol, S.A.; Tanase, B.C.; Hudita, A.; Negrei, C.; Burcea-Dragomiroiu, G.-T.-A.; Negreanu, L.; Vacaroiu, I.A.; Ginghină, O. Allergy in Cancer Care: Antineoplastic Therapy-Induced Hypersensitivity Reactions. *Int. J. Mol. Sci.* **2023**, *24*, 3886. [CrossRef] [PubMed]
13. Pagani, M.; Bavbek, S.; Alvarez-Cuesta, E.; Berna Dursun, A.; Bonadonna, P.; Castells, M.; Cernadas, J.; Chiriac, A.; Sahar, H.; Madrigal-Burgaleta, R. Hypersensitivity Reactions to Chemotherapy: An EAACI Position Paper. *Allergy* **2022**, *77*, 388–403. [CrossRef] [PubMed]
14. Roselló, S.; Blasco, I.; Fabregat, L.G.; Cervantes, A.; Jordan, K. Management of Infusion Reactions to Systemic Anticancer Therapy: ESMO Clinical Practice Guidelines. *Ann. Oncol.* **2017**, *28*, iv100–iv118. [CrossRef] [PubMed]
15. Iiizumi, M.; Liu, W.; Pai, S.K.; Furuta, E.; Watabe, K. Drug development against metastasis-related genes and their pathways: A rationale for cancer therapy. *Biochim. Biophys. Acta* **2008**, *1786*, 87–104. [CrossRef]

16. Gonzalez-Valdivieso, J.; Girotti, A.; Schneider, J.; Arias, F.J. Advanced nanomedicine and cancer: Challenges and opportunities in clinical translation. *Int. J. Pharm.* **2021**, *599*, 120438. [CrossRef]
17. Shi, J.; Kantoff, P.; Wooster, R.; Farokhzad, O.C. Cancer nanomedicine: Progress, challenges and opportunities. *Nat. Rev. Cancer* **2017**, *17*, 20–37. [CrossRef]
18. Spencer, D.S.; Amey Puranik, S.N.A.P. Intelligent nanoparticles for advanced drug delivery in cancer treatment. *Curr. Opin. Chem. Eng.* **2015**, *7*, 84–92. [CrossRef]
19. Aggarwal, B.B.; Ichikawa, H.; Garodia, P.; Weerasinghe, P.; Sethi, G.; Bhatt, I.D.; Pandey, M.K.; Shishodia, S.; Nair, M.G. From traditional Ayurvedic medicine to modern medicine: Identification of therapeutic targets for suppression of inflammation and cancer. *Expert Opin. Ther. Targets* **2006**, *10*, 87–118. [CrossRef]
20. Ali, M.; Benfante, V.; Stefano, A.; Yezzi, A.; Di Raimondo, D.; Tuttolomondo, A.; Comelli, A. Anti-Arthritic and Anti-Cancer Activities of Polyphenols: A Review of the Most Recent In Vitro Assays. *Life* **2023**, *13*, 361. [CrossRef] [PubMed]
21. Abbaszadeh, H.; Keikhaei, B.M.S. A review of molecular mechanisms involved in anticancer and antiangiogenic effects of natural polyphenolic compounds. *Phytother. Res.* **2019**, *33*, 2002–2014. [CrossRef] [PubMed]
22. Cook, M.T. Mechanism of metastasis suppression by luteolin in breast cancer. *Breast Cancer Targets Ther.* **2018**, *10*, 89–100. [CrossRef] [PubMed]
23. Imran, M.; Rauf, A.; Abu-Izneid, T.; Nadeem, M.; Shariati, M.A.; Khan, I.A.; Imran, A.; Orhan, I.E.; Rizwan, M.; Atif, M.; et al. Luteolin, a flavonoid, as an anticancer agent: A review. *Biomed. Pharmacother.* **2019**, *112*, 108612. [CrossRef] [PubMed]
24. Tuli, H.S.; Rath, P.; Chauhan, A.; Sak, K.; Aggarwal, D.; Choudhary, R.; Sharma, U.; Vashishth, K.; Sharma, S.; Kumar, M.; et al. Luteolin, a Potent Anticancer Compound: From Chemistry to Cellular Interactions and Synergetic Perspectives. *Cancers* **2022**, *14*, 5373. [CrossRef] [PubMed]
25. Lin, D.; Kuang, G.; Wan, J.; Zhang, X.; Li, H.; Gong, X.; Li, H. Luteolin suppresses the metastasis of triple-negative breast cancer by reversing epithelial-to-mesenchymal transition via downregulation of β-catenin expression. *Oncol. Rep.* **2017**, *37*, 895–902. [CrossRef]
26. Gao, G.; Ge, R.; Li, Y.; Liu, S. Luteolin exhibits anti-breast cancer property through up-regulating miR-203. *Artif. Cells Nanomed. Biotechnol.* **2019**, *47*, 3265–3271. [CrossRef] [PubMed]
27. Cao, D.; Zhu, G.Y.; Lu, Y.; Yang, A.; Chen, D.; Huang, H.J.; Peng, S.X.; Chen, L.W.; Li, Y.W. Luteolin suppresses epithelial-mesenchymal transition and migration of triple-negative breast cancer cells by inhibiting YAP/TAZ activity. *Biomed. Pharmacother.* **2020**, *129*, 110462. [CrossRef]
28. Cook, M.T.; Liang, Y.; Besch-Williford, C.; Goyette, S.; Mafuvadze, B.; Hyder, S.M. Luteolin inhibits progestin-dependent angiogenesis, stem cell-like characteristics, and growth of human breast cancer xenografts. *Springerplus* **2015**, *4*, 444. [CrossRef]
29. Cook, M.T.; Liang, Y.; Besch-Williford, C.; Hyder, S.M. Luteolin inhibits lung metastasis, cell migration, and viability of triple-negative breast cancer cells. *Breast Cancer Dove Med. Press* **2016**, *9*, 9–19. [CrossRef]
30. Sun, D.W.; Zhang, H.D.; Mao, L.; Mao, C.F.; Chen, W.; Cui, M.; Ma, R.; Cao, H.X.; Jing, C.W.; Wang, Z.; et al. Luteolin Inhibits Breast Cancer Development and Progression In Vitro and In Vivo by Suppressing Notch Signaling and Regulating MiRNAs. *Cell Physiol. Biochem.* **2015**, *37*, 1693–1711. [CrossRef]
31. Wu, H.T.; Lin, J.; Liu, Y.E.; Chen, H.F.; Hsu, K.W.; Lin, S.H.; Peng, K.Y.; Lin, K.J.; Hsieh, C.C.; Chen, D.R. Luteolin suppresses androgen receptor-positive triple-negative breast cancer cell proliferation and metastasis by epigenetic regulation of MMP9 expression via the AKT/mTOR signaling pathway. *Phytomedicine* **2021**, *81*, 153437. [CrossRef]
32. Huang, L.; Jin, K.; Lan, H. Luteolin inhibits cell cycle progression and induces apoptosis of breast cancer cells through downregulation of human telomerase reverse transcriptase. *Oncol. Lett.* **2019**, *17*, 3842–3850. [CrossRef] [PubMed]
33. Magura, J.; Roshila Moodley, I.M. The effect of hesperidin and luteolin isolated from Eriocephalus africanus on apoptosis, cell cycle and miRNA expression in MCF-7. *J. Biomol. Struct. Dyn.* **2022**, *40*, 1791–1800. [CrossRef] [PubMed]
34. Lin, C.H.; Chang, C.Y.; Lee, K.R.; Lin, H.J.; Chen, T.H.; Wan, L. Flavones inhibit breast cancer proliferation through the Akt/FOXO3a signaling pathway. *BMC Cancer* **2015**, *15*, 958. [CrossRef] [PubMed]
35. Zang, M.D.; Hu, L.; Fan, Z.Y.; Wang, H.X.; Zhu, Z.L.; Cao, S.; Wu, X.Y.; Li, J.F.; Su, L.P.; Li, C.; et al. Luteolin suppresses gastric cancer progression by reversing epithelial-mesenchymal transition via suppression of the Notch signaling pathway. *J. Transl. Med.* **2017**, *15*, 52. [CrossRef]
36. Pu, Y.; Zhang, T.; Wang, J.; Mao, Z.; Duan, B.; Long, Y.; Xue, F.; Liu, D.; Liu, S.G.Z. Luteolin exerts an anticancer effect on gastric cancer cells through multiple signaling pathways and regulating miRNAs. *J. Cancer* **2018**, *9*, 3669–3675. [CrossRef]
37. Zang, M.; Hu, L.; Zhang, B.; Zhu, Z.; Li, J.; Zhu, Z.; Yan, M.; Liu, B. Luteolin suppresses angiogenesis and vasculogenic mimicry formation through inhibiting Notch1-VEGF signaling in gastric cancer. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 913–919. [CrossRef]
38. Lu, J.; Li, G.; He, K.; Jiang, W.; Xu, C.; Li, Z.; Wang, H.; Wang, W.; Wang, H.; Teng, X.; et al. Luteolin exerts a marked antitumor effect in cMet-overexpressing patient-derived tumor xenograft models of gastric cancer. *J. Transl. Med.* **2015**, *13*, 42. [CrossRef]
39. Lu, X.; Li, Y.; Li, X.; Aisa, H.A. Luteolin induces apoptosis in vitro through suppressing the MAPK and PI3K signaling pathways in gastric cancer. *Oncol. Lett.* **2017**, *14*, 1993–2000. [CrossRef]
40. Wu, H.; Huang, M.; Liu, Y.; Shu, Y.; Liu, P. Luteolin induces apoptosis by up-regulating miR-34a in human gastric cancer cells. *Technol. Cancer Res. Treat.* **2015**, *14*, 747–755. [CrossRef]

41. Chen, K.C.; Chen, C.Y.; Lin, C.R.; Yang, T.Y.; Chen, T.H.; Wu, L.C.; Wu, C.C. Luteolin attenuates TGF-β1-induced epithelial-mesenchymal transition of lung cancer cells by interfering in the PI3K/Akt-NF-κB-Snail pathway. *Life Sci.* **2013**, *93*, 924–933. [CrossRef]
42. Ruan, J.; Zhang, L.; Yan, L.; Liu, Y.; Yue, Z.; Chen, L.; Wang, A.Y.; Chen, W.; Zheng, S.; Wang, S.; et al. Inhibition of hypoxia-induced epithelial mesenchymal transition by luteolin in non-small cell lung cancer cells. *Mol. Med. Rep.* **2012**, *6*, 232–238. [CrossRef] [PubMed]
43. Pan, J.; Cai, X.; Zheng, X.; Zhu, X.; Feng, J.; Wang, X. Luteolin inhibits viability, migration, angiogenesis and invasion of non-small cell lung cancer vascular endothelial cells via miR-133a-3p/purine rich element binding protein B-mediated MAPK and PI3K/Akt signaling pathways. *Tissue Cell* **2022**, *75*, 101740. [CrossRef] [PubMed]
44. Matés, J.M.; Segura, J.A.; Alonso, F.J.; Márquez, J. Oxidative stress in apoptosis and cancer: An update. *Arch. Toxicol.* **2012**, *86*, 1649–1665. [CrossRef]
45. Ju, W.; Wang, X.; Shi, H.; Chen, W.; Belinsky, S.A.L.Y. A critical role of luteolin induced reactive oxygen species in blockage of tumor necrosis factor activated nuclear factor-kappab pathway and sensitization of apoptosis in lung cancer cells. *Mol. Pharmacol.* **2007**, *71*, 1381–1388. [CrossRef]
46. Jiang, Z.Q.; Li, M.H.; Qin, Y.M.; Jiang, H.Y.; Zhang, X.; Wu, M.H. Luteolin inhibits tumorigenesis and induces apoptosis of non-small cell lung cancer cells via regulation of microRNA-34a-5p. *Int. J. Mol. Sci.* **2018**, *19*, 447. [CrossRef] [PubMed]
47. Huang, X.; Dai, S.; Dai, J.; Xiao, Y.; Bai, Y.; Chen, B.; Zhou, M. Luteolin decreases invasiveness, deactivates STAT3 signaling, and reverses interleukin-6 induced epithelial-mesenchymal transition and matrix metalloproteinase secretion of pancreatic cancer cells. *OncoTargets Ther.* **2015**, *8*, 2989–3001. [CrossRef] [PubMed]
48. Cai, X.; Lu, W.; Ye, T.; Lu, M.; Wang, J.; Huo, J.; Qian, S.; Wang, X.; Cao, P. The molecular mechanism of luteolin-induced apoptosis is potentially related to inhibition of angiogenesis in human pancreatic carcinoma cells. *Oncol. Rep.* **2012**, *28*, 1353–1361. [CrossRef] [PubMed]
49. Lee, L.T.; Huang, Y.T.; Hwang, J.J.; Lee, P.P.; Nair, M.P.; Kandaswam, C.L.M. Blockade of the epidermal growth factor receptor tyrosine kinase activity by quercetin and luteolin leads to growth inhibition and apoptosis of pancreatic tumor cells. *Anticancer Res.* **2002**, *22*, 1615–1627.
50. Li, C.; Wang, Q.; Shen, S.; Wei, X.; Li, G. HIF-1α/VEGF signaling-mediated epithelial-mesenchymal transition and angiogenesis is critically involved in anti-metastasis effect of luteolin in melanoma cells. *Phyther. Res.* **2019**, *33*, 798–807. [CrossRef]
51. Ruan, J.S.; Liu, Y.P.; Zhang, L.; Yan, L.G.; Fan, F.T.; Shen, C.S.; Wang, A.Y.; Zheng, S.Z.; Wang, S.M.; Lu, Y. Luteolin reduces the invasive potential of malignant melanoma cells by targeting β3 integrin and the epithelial-mesenchymal transition. *Acta Pharmacol. Sin.* **2012**, *33*, 1325–1331. [CrossRef]
52. Chen, Y.F.; Wu, S.; Li, X.; Chen, M.; Liao, H.F. Luteolin Suppresses Three Angiogenesis Modes and Cell Interaction in Uveal Melanoma in Vitro. *Curr. Eye Res.* **2022**, *47*, 1590–1599. [CrossRef] [PubMed]
53. Feng, J.; Zheng, T.; Hou, Z.; Lv, C.; Xue, A.; Han, T.; Han, B.; Sun, X.; Wei, Y. Luteolin, an aryl hydrocarbon receptor ligand, suppresses tumor metastasis in vitro and in vivo. *Oncol. Rep.* **2020**, *44*, 2231–2240. [CrossRef] [PubMed]
54. Shi, M.L.; Chen, Y.F.; Liao, H.F. Effect of luteolin on apoptosis and vascular endothelial growth factor in human choroidal melanoma cells. *Int. J. Ophthalmol.* **2021**, *14*, 186–193. [CrossRef] [PubMed]
55. Shi, M.L.; Chen, Y.F.; Wu, W.Q.; Lai, Y.; Jin, Q.; Qiu, W.L.; Yu, D.L.; Li, Y.Z.; Liao, H.F. Luteolin inhibits the proliferation, adhesion, migration and invasion of choroidal melanoma cells in vitro. *Exp. Eye Res.* **2021**, *210*, 108643. [CrossRef] [PubMed]
56. Pratheeshkumar, P.; Son, Y.O.; Budhraja, A.; Wang, X.; Ding, S.; Wang, L.; Hitron, A.; Lee, J.C.; Kim, D.; Divya, S.P.; et al. Luteolin inhibits human prostate tumor growth by suppressing vascular endothelial growth factor receptor 2-mediated angiogenesis. *PLoS ONE* **2012**, *7*, e52279. [CrossRef] [PubMed]
57. Han, K.; Wei, M.; Zhang, J.J.; Zhou, Y.; Wang, Y.L.; Su, Y.; Lin, S.C.; Gan, Z.H.; Sun, Y.N.; Min, D.L. Luteolin inhibited proliferation and induced apoptosis of prostate cancer cells through miR-301. *OncoTargets Ther.* **2016**, *9*, 3085–3094. [CrossRef]
58. Dai, Y.; Zheng, H.; Liu, Z.; Wang, Y.; Hu, W. The flavonoid luteolin suppresses infantile hemangioma by targeting FZD6 in the Wnt pathway. *Investig. New Drugs* **2021**, *39*, 775–784. [CrossRef]
59. Wang, Q.; Wang, H.; Jia, Y.; Ding, H.; Zhang, L.; Pan, H. Luteolin reduces migration of human glioblastoma cell lines via inhibition of the p-IGF-1R/PI3K/AKT/mTOR signaling pathway. *Oncol. Lett.* **2017**, *14*, 3545–3551. [CrossRef]
60. Anson, D.M.; Wilcox, R.M.; Huseman, E.D.; Stump, T.A.; Paris, R.L.; Darkwah, B.O.; Lin, S.; Adegoke, A.O.; Gryka, R.J.; Jean-Louis, D.S.A.S. Luteolin Decreases Epidermal Growth Factor Receptor-Mediated Cell Proliferation and Induces Apoptosis in Glioblastoma Cell Lines. *Basic Clin. Pharmacol. Toxicol.* **2018**, *123*, 678–686. [CrossRef]
61. Kittiratphatthana, N.; Kukongviriyapan, V.; Prawan, A.; Senggunprai, L. Luteolin induces cholangiocarcinoma cell apoptosis through the mitochondrial-dependent pathway mediated by reactive oxygen species. *J. Pharm. Pharmacol.* **2016**, *68*, 1184–1192. [CrossRef]
62. Lim, D.Y.; Jeong, Y.; Tyner, A.L.; Park, J.H. Induction of cell cycle arrest and apoptosis in HT-29 human colon cancer cells by the dietary compound luteolin. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2007**, *292*, G66–G75. [CrossRef] [PubMed]
63. Pandurangan, A.K.; Dharmalingam, P.; Sadagopan, S.K.; Ramar, M.; Munusamy, A.; Ganapasam, S. Luteolin induces growth arrest in colon cancer cells through involvement of Wnt/β-catenin/GSK-3β signaling. *J. Environ. Pathol. Toxicol. Oncol.* **2013**, *32*, 131–139. [CrossRef] [PubMed]

64. Kim, H.Y.; Jung, S.K.; Byun, S.; Son, J.E.; Oh, M.H.; Lee, J.; Kang, M.J.; Heo, Y.S.; Lee, K.W.; Lee, H.J. Raf and PI3K are the molecular targets for the anti-metastatic effect of luteolin. *Phyther. Res.* **2013**, *27*, 1481–1488. [CrossRef] [PubMed]
65. Yao, Y.; Rao, C.; Zheng, G.; Wang, S. Luteolin suppresses colorectal cancer cell metastasis via regulation of the miR-384/pleiotrophin axis. *Oncol. Rep.* **2019**, *42*, 131–141. [CrossRef] [PubMed]
66. Wang, S.W.; Chen, Y.R.; Chow, J.M.; Chien, M.H.; Yang, S.F.; Wen, Y.C.; Lee, W.J.; Tseng, T.H. Stimulation of Fas/FasL-mediated apoptosis by luteolin through enhancement of histone H3 acetylation and c-Jun activation in HL-60 leukemia cells. *Mol. Carcinog.* **2018**, *57*, 866–877. [CrossRef]
67. López-Lázaro, M. Distribution and biological activities of the flavonoid luteolin. *Mini Rev. Med. Chem.* **2009**, *9*, 31–59. [CrossRef]
68. Ross, J.A.; Kasum, C.M. Dietary flavonoids: Bioavailability, metabolic effects, and safety. *Annu. Rev. Nutr.* **2002**, *22*, 19–34. [CrossRef]
69. Xu, H.; Linn, B.; Zhang, Y.; Ren, J. A review on the antioxidative and prooxidative properties of luteolin. *React. Oxyg. Species* **2019**, *7*, 136–147. [CrossRef]
70. Manzoor, M.F.; Ahmad, N.; Ahmed, Z.; Siddique, R.; Zeng, X.A.; Rahaman, A.; Muhammad Aadil, R.; Wahab, A. Novel extraction techniques and pharmaceutical activities of luteolin and its derivatives. *J. Food Biochem.* **2019**, *43*, e12974. [CrossRef]
71. Prasher, P.; Sharma, M.; Singh, S.K.; Gulati, M.; Chellappan, D.K.; Zacconi, F.; De Rubis, G.; Gupta, G.; Sharifi-Rad, J.; Cho, W.C.; et al. Luteolin: A flavonoid with a multifaceted anticancer potential. *Cancer Cell Int.* **2022**, *22*, 386. [CrossRef]
72. Zhou, P.; Li, L.P.; Luo, S.Q.; Jiang, H.D.; Zeng, S. Intestinal absorption of luteolin from peanut hull extract is more efficient than that from individual pure luteolin. *J. Agric. Food Chem.* **2008**, *56*, 296–300. [CrossRef] [PubMed]
73. Deng, C.; Gao, C.; Tian, X. Pharmacokinetics, tissue distribution and excretion of luteolin and its major metabolites in rats: Metabolites predominate in blood, tissues and are mainly excreted via bile. *J. Funct. Foods.* **2017**, *35*, 332–340. [CrossRef]
74. Hayasaka, N.; Shimizu, N.; Komoda, T.; Mohri, S.; Tsushida, T.; Eitsuka, T.; Miyazawa, T.; Nakagawa, K. Absorption and Metabolism of Luteolin in Rats and Humans in Relation to in Vitro Anti-inflammatory Effects. *J. Agric. Food Chem.* **2018**, *66*, 11320–11329. [CrossRef] [PubMed]
75. Derynck, R.; Muthusamy, B.P.; Saeteurn, K.Y. Signaling pathway cooperation in TGF-β-induced epithelial-mesenchymal transition. *Curr. Opin. Cell Biol.* **2014**, *31*, 56–66. [CrossRef]
76. Gonzalez, D.M.; Medici, D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci. Signal* **2014**, *7*, re8. [CrossRef]
77. Singh, M.; Yelle, N.; Venugopal, C.; Singh, S.K. EMT: Mechanisms and therapeutic implications. *Pharmacol. Ther.* **2018**, *182*, 80–94. [CrossRef]
78. Scheel, C.; Onder, T.; Karnoub, A.; Weinberg, R.A. Adaptation versus selection: The origins of metastatic behavior. *Cancer Res.* **2007**, *67*, 11476–11480. [CrossRef]
79. Deshmukh, A.P.; Vasaikar, S.V.; Tomczak, K.; Tripathi, S.; den Hollander, P.; Arslan, E.; Chakraborty, P.; Soundararajan, R.; Jolly, M.K.; Rai, K.; et al. Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2102050118. [CrossRef]
80. Li, Z.; Wang, Y.; Zhu, Y.; Yuan, C.; Wang, D.; Zhang, W.; Qi, B.; Qiu, J.; Song, X.; Ye, J.; et al. The Hippo transducer TAZ promotes epithelial to mesenchymal transition and cancer stem cell maintenance in oral cancer. *Mol. Oncol.* **2015**, *9*, 1091–1105. [CrossRef]
81. Lamouille, S.; Xu, J.; Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 178–196. [CrossRef]
82. Hussain, Y.; Cui, J.H.; Khan, H.; Aschner, M.; Batiha, G.E.; Jeandet, P. Luteolin and cancer metastasis suppression: Focus on the role of epithelial to mesenchymal transition. *Med. Oncol.* **2021**, *38*, 66. [CrossRef] [PubMed]
83. Kim, V.N. Small RNAs: Classification, biogenesis, and function. *Mol. Cells* **2005**, *19*, 1–15.
84. Pan, G.; Liu, Y.; Shang, L.; Zhou, F.; Yang, S. EMT-associated microRNAs and their roles in cancer stemness and drug resistance. *Cancer Commun. (Lond.)* **2021**, *41*, 199–217. [CrossRef] [PubMed]
85. Lee, Y.S.; Dutta, A. MicroRNAs in cancer. *Annu. Rev. Pathol.* **2009**, *4*, 199–227. [CrossRef] [PubMed]
86. Fan, J.J.; Hsu, W.H.; Lee, K.H.; Chen, K.C.; Lin, C.W.; Lee, Y.A.; Ko, T.P.; Lee, L.T.; Lee, M.T.; Chang, M.S.; et al. Dietary Flavonoids Luteolin and Quercetin Inhibit Migration and Invasion of Squamous Carcinoma through Reduction of Src/Stat3/S100A7 Signaling. *Antioxidants* **2019**, *8*, 557. [CrossRef] [PubMed]
87. Chen, K.C.; Hsu, W.H.; Ho, J.Y.; Lin, C.W.; Chu, C.Y.; Kandaswami, C.C.; Lee, M.T.; Cheng, C.H. Flavonoids Luteolin and Quercetin Inhibit RPS19 and contributes to metastasis of cancer cells through c-Myc reduction. *J. Food Drug Anal.* **2018**, *26*, 1180–1191. [CrossRef] [PubMed]
88. Liekens, S.; De Clercq, E.; Neyts, J. Angiogenesis: Regulators and clinical applications. *Biochem. Pharmacol.* **2001**, *61*, 253–270. [CrossRef] [PubMed]
89. Carmeliet, P.J.R. Molecular mechanisms and clinical applications of angiogenesis. *Nature* **2011**, *473*, 298–307. [CrossRef]
90. Niu, G.; Chen, X. Vascular endothelial growth factor as an anti-angiogenic target for cancer therapy. *Curr. Drug Targets* **2010**, *11*, 1000–1017. [CrossRef]
91. Wang, Z.; Dabrosin, C.; Yin, X.; Fuster, M.M.; Arreola, A.; Rathmell, W.K.; Generali, D.; Nagaraju, G.P.; El-Rayes, B.; Ribatti, D.; et al. Broad targeting of angiogenesis for cancer prevention and therapy. *Semin. Cancer Biol.* **2015**, *35*, S224–S243. [CrossRef]
92. Mukherjee, A.; Madamsetty, V.S.; Paul, M.K.; Mukherjee, S. Recent Advancements of Nanomedicine towards Antiangiogenic Therapy in Cancer. *Int. J. Mol. Sci.* **2020**, *21*, 455. [CrossRef] [PubMed]

93. Martínez-Poveda, B.; Torres-Vargas, J.A.; Ocaña, M.D.C.; García-Caballero, M.; Medina, M.Á.; Quesada, A.R. The Mediterranean Diet, a Rich Source of Angiopreventive Compounds in Cancer. *Nutrients* **2019**, *11*, 2036. [CrossRef] [PubMed]
94. Tosetti, F.; Ferrari, N.; De Flora, S.; Albini, A. Angioprevention': Angiogenesis is a common and key target for cancer chemopreventive agents. *FASEB J.* **2002**, *16*, 2–14. [CrossRef] [PubMed]
95. Bagli, E.; Stefaniotou, M.; Morbidelli, L.; Ziche, M.; Psillas, K.; Murphy, C.; Fotsis, T. Luteolin inhibits vascular endothelial growth factor-induced angiogenesis; inhibition of endothelial cell survival and proliferation by targeting phosphatidylinositol 3'-kinase activity. *Cancer Res.* **2004**, *64*, 7936–7946. [CrossRef]
96. Fang, B.; Chen, X.; Wu, M.; Kong, H.; Chu, G.; Zhou, Z.; Zhang, C.; Chen, B. Luteolin inhibits angiogenesis of the M2-like TAMs via the downregulation of hypoxia inducible factor-1α and the STAT3 signalling pathway under hypoxia. *Mol. Med. Rep.* **2018**, *18*, 2914–2922. [CrossRef]
97. Basset, P.; Okada, A.; Chenard, M.P.; Kannan, R.; Stoll, I.; Anglard, P.; Bellocq, J.P.; Rio, M.C. Matrix metalloproteinases as stromal effectors of human carcinoma progression: Therapeutic implications. *Matrix Biol.* **1997**, *15*, 535–541. [CrossRef]
98. John, A.; Tuszynski, G. The role of matrix metalloproteinases in tumor angiogenesis and tumor metastasis. *Pathol. Oncol. Res.* **2001**, *7*, 14–23. [CrossRef]
99. Deryugina, E.I.; Quigley, J.P. Matrix metalloproteinases and tumor metastasis. *Cancer Metastasis Rev.* **2006**, *25*, 9–34. [CrossRef]
100. Niland, S.; Riscanevo, A.X.; Eble, J.A. Matrix Metalloproteinases Shape the Tumor Microenvironment in Cancer Progression. *Int. J. Mol. Sci.* **2021**, *23*, 146. [CrossRef]
101. Fuchs, Y.; Steller, H. Programmed cell death in animal development and disease. *Cell* **2011**, *147*, 742–758. [CrossRef]
102. Pistritto, G.; Trisciuoglio, D.; Ceci, C.; Garufi, A.; D'Orazi, G. Apoptosis as anticancer mechanism: Function and dysfunction of its modulators and targeted therapeutic strategies. *Aging* **2016**, *8*, 603–619. [CrossRef] [PubMed]
103. Elmore, S. Apoptosis: A review of programmed cell death. *Toxicol. Pathol.* **2007**, *35*, 495–516. [CrossRef] [PubMed]
104. Schulze, A.H.A. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* **2012**, *491*, 364–373. [CrossRef] [PubMed]
105. Kang, K.A.; Piao, M.J.; Ryu, Y.S.; Hyun, Y.J.; Park, J.E.; Shilnikova, K.; Zhen, A.X.; Kang, H.K.; Koh, Y.S.; Jeong, Y.J.; et al. Luteolin induces apoptotic cell death via antioxidant activity in human colon cancer cells. *Int. J. Oncol.* **2017**, *51*, 1169–1178. [CrossRef] [PubMed]
106. Raina, R.; Pramodh, S.; Rais, N.; Haque, S.; Shafarin, J.; Bajbouj, K.; Hamad, M.; Hussain, A. Luteolin inhibits proliferation, triggers apoptosis and modulates Akt/mTOR and MAP kinase pathways in HeLa cells. *Oncol. Lett.* **2021**, *21*, 192. [CrossRef] [PubMed]
107. Hill, M.T.N. miRNA interplay: Mechanisms and consequences in cancer. *Dis. Model. Mech.* **2021**, *14*, dmm047662. [CrossRef]
108. Osaki, M.; Oshimura, M.I.H. PI3K-Akt pathway: Its functions and alterations in human cancer. *Apoptosis* **2004**, *9*, 667–676. [CrossRef]
109. Fresno Vara, J.A.; Casado, E.; de Castro, J.; Cejas, P.; Belda-Iniesta, C.G.-B.M. PI3K/Akt signalling pathway and cancer. *Cancer Treat. Rev.* **2004**, *30*, 193–204. [CrossRef]
110. Luo, M.F.L.-W. Redundant kinase activation and resistance of EGFR-tyrosine kinase inhibitors. *Am. J. Cancer Res.* **2014**, *4*, 608–628.
111. Wee, P.W.Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers* **2017**, *9*, 52. [CrossRef]
112. Quatrale, A.E.; Porcelli, L.; Silvestris, N.; Colucci, G.; Angelo, A.A.A. EGFR tyrosine kinases inhibitors in cancer treatment: In vitro and in vivo evidence. *Front. Biosci. Landmark Ed.* **2011**, *16*, 1962–1972. [CrossRef] [PubMed]
113. Taliou, A.; Zintzaras, E.; Lykouras, L.; Francis, K. An open-label pilot study of a formulation containing the anti-inflammatory flavonoid luteolin and its effectson behavior in children with autism spectrum disorders. *Clin. Ther.* **2013**, *35*, 592–602. [CrossRef] [PubMed]
114. Di Stadio, A.; D'Ascanio, L.; Vaira, L.A.; Cantone, E.; De Luca, P.; Cingolani, C.; Motta, G.; De Riu, G.; Vitelli, F.; Spriano, G.; et al. Ultramicronized Palmitoylethanolamide and Luteolin. Supplement Combined with Olfactory Training to Treat Post-COVID-19 Olfactory Impairment: A Multi-Center Double-Blinded Randomized Placebo- Controlled Clinical Trial. *Curr. Neuropharmacol.* **2022**, *20*, 2001–2012. [CrossRef] [PubMed]
115. De Luca, P.; Camaioni, A.; Marra, P.; Salzano, G.; Carriere, G.; Ricciardi, L.; Pucci, R.; Montemurro, N.; Brenner, M.J.; Di Stadio, A. Effect of Ultra-Micronized Palmitoylethanolamide and Luteolin on Olfaction and Memory in Patients with Long COVID: Results of a Longitudinal Study. *Cells* **2022**, *11*, 2552. [CrossRef] [PubMed]
116. Versace, V.; Ortelli, P.; Dezi, S.; Ferrazzoli, D.; Alibardi, A.; Bonini, I.; Engl, M.; Maestri, R.; Assogna, M.; Ajello, V.; et al. Co-ultramicronized palmitoylethanolamide/luteolin normalizes GABAB-ergic activity and cortical plasticity in long COVID-19 syndrome. *Clin. Neurophysiol.* **2023**, *145*, 81–88. [CrossRef]
117. Terzo, S.; Amato, A.; Magán-Fernández, A.; Castellino, G.; Calvi, P.; Chianetta, R.; Giglio, R.V.; Patti, A.M.; Nikolic, D.; Firenze, A.; et al. Nutraceutical Containing Chlorogenic Acid and Luteolin Improves CardiometabolicParameters in Subjects with Pre-Obesity: A 6-Month Randomized, Double-Blind, Placebo-Controlled Study. *Nutrients* **2023**, *15*, 462. [CrossRef]
118. Tawornchat, P.; Pattarakankul, T.; Palaga, T.; Varol Intasanta, S.W. Polymerized Luteolin Nanoparticles: Synthesis, Structure Elucidation, and Anti-Inflammatory Activity. *ACS Omega* **2021**, *6*, 2846–2855. [CrossRef]
119. Shinde, P.; Agraval, H.; Singh, A.; Umesh, C.S.; Yadav, U.K. Synthesis of luteolin loaded zein nanoparticles for targeted cancer therapy improving bioavailability and efficacy. *J. Drug Deliv. Sci. Technol.* **2019**, *52*, 369–378. [CrossRef]

120. Wu, G.; Li, J.; Yue, J.; Zhang, S.Y.K. Liposome encapsulated luteolin showed enhanced antitumor efficacy to colorectal carcinoma. *Mol. Med. Rep.* **2018**, *17*, 2456–2464. [CrossRef]
121. Sinha, A.P.K.S. Enhanced Induction of Apoptosis in HaCaT Cells by Luteolin Encapsulated in PEGylated Liposomes-Role of Caspase-3/Caspase-14. *Appl. Biochem. Biotechnol.* **2019**, *188*, 147–164. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

Omics Analysis Unveils the Pathway Involved in the Anthocyanin Biosynthesis in Tomato Seedling and Fruits

Rui He, Kaizhe Liu, Shuchang Zhang, Jun Ju, Youzhi Hu, Yamin Li, Xiaojuan Liu and Houcheng Liu *

College of Horticulture, South China Agricultural University, Guangzhou 510642, China; ruihe@stu.scau.edu.cn (R.H.); 1836945107@stu.scau.edu.cn (K.L.); shuchangzhang@stu.scau.edu.cn (S.Z.); jujun-mail@stu.scau.edu.cn (J.J.); youzhihu@stu.scau.edu.cn (Y.H.); yaminli@stu.scau.edu.cn (Y.L.); liuxjjy628@stu.scau.edu.cn (X.L.)
* Correspondence: liuhch@scau.edu.cn; Tel.: +86-020-85-280

Abstract: The purple tomato variety 'Indigo Rose' (InR) is favored due to its bright appearance, abundant anthocyanins and outstanding antioxidant capacity. *SlHY5* is associated with anthocyanin biosynthesis in 'Indigo Rose' plants. However, residual anthocyanins still present in *Slhy5* seedlings and fruit peel indicated there was an anthocyanin induction pathway that is independent of HY5 in plants. The molecular mechanism of anthocyanins formation in 'Indigo Rose' and *Slhy5* mutants is unclear. In this study, we performed omics analysis to clarify the regulatory network underlying anthocyanin biosynthesis in seedling and fruit peel of 'Indigo Rose' and *Slhy5* mutant. Results showed that the total amount of anthocyanins in both seedling and fruit of InR was significantly higher than those in the *Slhy5* mutant, and most genes associated with anthocyanin biosynthesis exhibited higher expression levels in InR, suggesting that *SlHY5* play pivotal roles in flavonoid biosynthesis both in tomato seedlings and fruit. Yeast two-hybrid (Y2H) results revealed that *SlBBX24* physically interacts with *SlAN2-like* and *SlAN2*, while *SlWRKY44* could interact with *SlAN11* protein. Unexpectedly, both *SlPIF1* and *SlPIF3* were found to interact with *SlBBX24*, *SlAN1* and *SlJAF13* by yeast two-hybrid assay. Suppression of *SlBBX24* by virus-induced gene silencing (VIGS) retarded the purple coloration of the fruit peel, indicating an important role of *SlBBX24* in the regulation of anthocyanin accumulation. These results deepen the understanding of purple color formation in tomato seedlings and fruits in an *HY5*-dependent or independent manner via excavating the genes involved in anthocyanin biosynthesis based on omics analysis.

Keywords: anthocyanin biosynthesis; purple tomato; HY5; MBW; PIFs

Citation: He, R.; Liu, K.; Zhang, S.; Ju, J.; Hu, Y.; Li, Y.; Liu, X.; Liu, H. Omics Analysis Unveils the Pathway Involved in the Anthocyanin Biosynthesis in Tomato Seedling and Fruits. *Int. J. Mol. Sci.* **2023**, *24*, 8690. https://doi.org/10.3390/ijms24108690

Academic Editor: Wajid Zaman

Received: 9 April 2023
Revised: 8 May 2023
Accepted: 9 May 2023
Published: 12 May 2023

1. Introduction

Anthocyanins comprise a class of primary hydrosoluble pigments belonging to flavonoids, which are widely distributed in plants and confer various colorations in fruit, flower, seed and leaf. Anthocyanins are essential antioxidants that not only play their crucial roles in protecting plants from various biotic and abiotic stressors (i.e., cold, drought, UV irradiation, pathogen) but also contribute to decreasing the risk of certain types of cardiovascular and neurodegenerative diseases and cancer in the human body [1–3]. Anthocyanin biosynthetic pathways have been intensively studied in many species, and various structural genes and transcription factors have been well characterized in a strongly conserved pathway [4–6]. The regulation of anthocyanin biosynthesis is controlled by early biosynthetic genes (i.e., phenylalanine ammonia-lyase (*PAL*), 4-coumaryl:CoA ligase (*4CL*), chalcone synthase (*CHS*), chalcone isomerase (*CHI*) and flavanone 3-hydroxylase (*F3H*)) and late biosynthetic genes (i.e., flavonoid 3′5′-hydroxylase (*F3′5′H*), dihydroflavonol 4-reductase (*DFR*), anthocyanidin synthase (*ANS*), glutathione-S-transferase (*GST*) and flavonol-3-glucosyltransferase (*3GT*)) [7]. Most genes involved in anthocyanins biosynthesis could be activated or repressed by specific transcription factors as well as controlled by a ternary MYB–bHLH–WD repeat (MBW) transcriptional complex, which consists

of basic helix-loop-helix (bHLH) R2R3-MYB transcription factors and WD40-repeat proteins [8]. In addition, other regulatory factors, such as HY5, ERFs, PIFs, BBXs, and WRKYs, also participated in the regulation of anthocyanin biosynthesis [9–12]. Previous studies have proved that *AtBBX21*, *AtBBX22* and *AtBBX23* induced the accumulation of anthocyanins in Arabidopsis [13–16], while *AtBBX24*, *AtBBX25* and *AtBBX32* inhibit anthocyanin accumulation [17–19]. *SlBBX20* directly binds the promoter of the anthocyanin biosynthesis gene *SlDFR* to enhance anthocyanin biosynthesis in tomato fruits [20]. *MdBBX22* induced *mdm-miR858* expression via bounding to its promoter, thus governing anthocyanin accumulation in apples [21]. Additionally, members of the B-box (BBX) protein family (i.e., *BBX18/20/21/23/24/33*) directly conjunct with HY5 cooperatively regulate anthocyanin synthesis in Arabidopsis [22]. Furthermore, results showed that a WRKY gene negatively regulates the complex MYB-bHLH-WD40 petunia (*Petunia hybrida*) and Arabidopsis thaliana [23]. *PpWRKY44* could significantly promote light-induced anthocyanin accumulation in red pear fruit via binding to the promoters of *PpMYB10* [24].

The tomato (*Solanum lycopersicum*) is one of the most consumed vegetable products around the world. In most of the cultivated tomatoes, anthocyanins are generally undetectable in fruit. Cultivation attempts have been made to improve the anthocyanin content in tomato fruit. 'Indigo Rose' (InR), a purple tomato variety that contains the Aft locus and recessive atv locus, exhibits a high-level accumulation of anthocyanin on the fruit peel in a light-dependent manner [25]. Thus, 'Indigo Rose' has been frequently taken to underly the molecular mechanism of anthocyanin synthesis in purple tomato fruit [26–28]. Previous studies revealed R2R3-MYB transcription factor *SlAN2-like* as an active and critical regulator of anthocyanin biosynthesis, while *SlMYBATV* was identified as the regulatory repressor via competing for the binding of *SlAN2-like* to *SlAN1* [27].

HY5 (Elongated Hypocotyl 5), as a vital regulator of light-dependent development in higher plants, exhibits a dominant function in hypocotyl elongation and lateral root development as well as pigment accumulation [29]. To date, most of the genes and transcription factors involved in anthocyanin regulation were highly associated with HY5. The HY5 protein directly binds to either G-box or ACE-box in the promoters of anthocyanin biosynthetic genes such as *CHS* and *DFR,* then activates their expression, positively regulating anthocyanin accumulation [30]. *CaHY5* can bind to the promoter of *CaF3H*, *CaF3′5′H*, *CaDFR*, *CaANS* and *CaGST*, which are well related to anthocyanin biosynthesis or transport, and thereby promote anthocyanin accumulation in pepper hypocotyl [31]. However, residual anthocyanins have still been present in *hy5* mutants of Arabidopsis, which indicates that there is an anthocyanin induction pathway that is independent of HY5 in plants [30,32]. This result has been well proved in *Slhy5* mutants of tomatoes via analyzing the transcriptome of multiple tissues, which has found eight candidate transcription factors were likely involved in anthocyanin production in tomatoes in an HY5-independent manner [11]. In the present study, we found that anthocyanins accumulated on the surface of hypocotyls in InR tomato seedlings, but not *Slhy5* seedlings, at cotyledon emergence. Unexpectedly, residual anthocyanins also accumulated both in the cotyledon and hypocotyls of *Slhy5* seedling, which displayed obvious spatiotemporal specificity. Meanwhile, the InR fruit peel accumulated large amounts of anthocyanins, particularly in the light-exposing part, while *Slhy5* contained a lower anthocyanin content on the peel of the fruit shoulder and no anthocyanins accumulated in the peel of the shading part. Therefore, whether other transcription factors substitute or compensate for HY5 to regulate the anthocyanin accumulation? Advances in transcriptomics and metabolomics play pivotal roles in uncovering complex biological mechanisms of diversified pathways in plants [33,34].

The objectives of this study were to reveal the anthocyanin variation in seedlings and fruit of tomato ('Indigo Rose' (InR) and *Slhy5* mutants) at different developmental stages and excavate the candidate *HY5*-dependent or independent transcription factors involved in anthocyanin biosynthesis in the seedling and fruit via omics analysis.

2. Results

2.1. Morphological Characterization of InR and Slhy5 Seedling and Fruit

On the stage in which hypocotyl emergence and cotyledons still close, the surface of hypocotyls of InR seedlings exhibited purple color, whereas *Slhy5* seedlings displayed white color (hardly accumulate anthocyanins) and longer hypocotyl (Figure 1a and Figure S1). Considerable anthocyanins accumulation was observed in cotyledons and hypocotyls of InR seedlings once exposed to light (Figure 1a and Figure S1). However, *Slhy5* seedlings developed an opposite phenotype and displayed obvious spatiotemporal specificity in anthocyanin production. Briefly, in *Slhy5* seedlings, the anthocyanin accumulated first in the upper part of the hypocotyls and then gradually developed in the lower part in a light-dependent manner (Figure 1a and Figure S1). The anthocyanin content both in cotyledons and hypocotyls of *Slhy5* seedlings was lower than those of InR seedlings, respectively (Figure 1c and Figure S1).

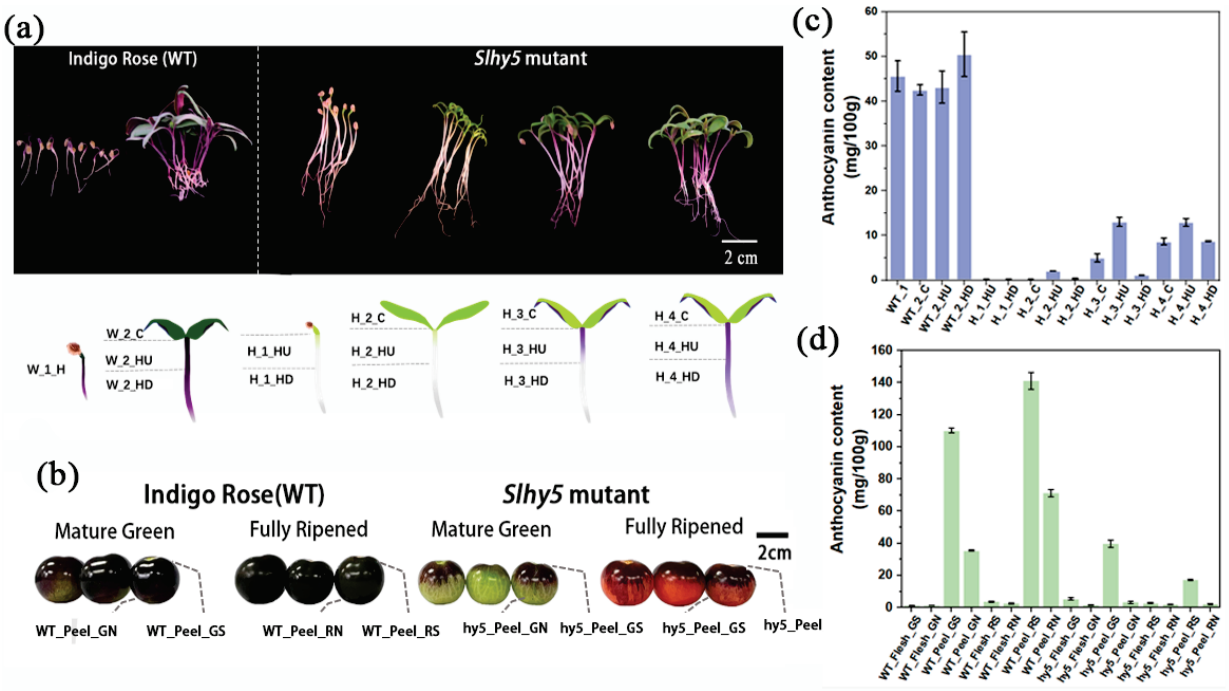


Figure 1. Morphological characterization of InR and *Slhy5* seedling and fruit. (**a**) Photograph showing the seedling phenotypes of different stages of InR and *Slhy5* seedlings for 4–7 days. (**b**) The phenotype of fruit from InR and *Slhy5* mutants at the mature green-mature stage and fully mature stage. (**c**) The anthocyanin content of InR and *Slhy5* seedling in different tissues. (**d**) The anthocyanin content of InR and *Slhy5* fruit s at the green-mature stage and fully ripened stage both in peel and flesh. Error bars indicate SD ($n > 3$).

Anthocyanin content just accumulated in peels at the green-mature stage and fully mature stage in InR and *Slhy5* fruits, and few anthocyanins were found in the fruit flesh (Figure 1b,d and Figure S2). Additionally, higher anthocyanins accumulated in the light-exposing peel part of InR fruits than in the shading part (Figure 1b,d). Unexpectedly, residual anthocyanins have still been present in the peel of the *Slhy5* fruit shoulder, though the anthocyanins contents were lower than the InR fruit peel. However, anthocyanin accumulation was hardly detected in the shading peel part of the *Slhy5* fruit (Figure 1b,d). These observations suggested that *Slhy5* had a predominant role in tomato pigmentation in a light-dependent manner, and there might be some regulators controlling anthocyanin biosynthesis in an HY5-independent manner.

2.2. Changes of Metabolites and Genes Expression in the Cotyledon of InR Seedlings and Slhy5 Seedlings

The color of cotyledon and hypocotyl was transformed continuously from green to purple during seedling development in InR seedlings and *Slhy5* seedlings. To explore the changes of metabolites and gene expression during InR and *Slhy5* seedling development, metabolic and transcriptome analysis of cotyledon, the upper and lower part of the hypocotyl were carried out at different developmental stages, respectively (Figure 1a). PCA was performed on detected metabolites to demonstrate the similarity in metabolic profiles among the samples (Figure S3). In the two-dimensional PCA plot, three biological replicates of each sample tended to group, indicating the high reproducibility of the generated metabolome data (Figure S3). Many metabolites and genes in the seedling varied considerably in terms of different tissues of different varieties. Therefore, variation of metabolites and gene expression in three parts (cotyledon, upper and lower part of the hypocotyl) of InR seedlings and *Slhy5* seedlings was investigated, respectively.

Four metabolite groups were observed in cotyledon based on the level of annotation of metabolite similarity. Group 1, which metabolites levels of InR seedlings cotyledon (W_2_C) were obviously higher than those in *Slhy5* seedlings (H_2_C, H_3_C and H_4_C), including tulipanin, rutin, butin etc., belonged to flavonoids, flavones and flavonols, anthocyanins according to KEGG analyses (Figure S4). So, anthocyanins and flavonoids might be responsible for the distinction of purple coloration of cotyledon between InR and *Slhy5* seedlings.

The transcriptome data validated the authenticity and accuracy of the metabolic analysis. A weighted gene co-expression network analysis (WGCNA) was performed on the genes of cotyledon. A slight relationship with purple coloration was displayed in the green module (Figure 2b), and the genes related to this module were annotated according to KEGG pathway enrichment analysis, which involved the flavonoid biosynthesis pathways (Figure 2c). Anthocyanins metabolism-related transcriptional factors (i.e., *SlAN1*, *SlAN2*, *SlAN2-like*) and the structural genes (i.e., *SlPAL*, *Sl4CL*, *SlCHS*, *SlCHI*, *SlDFR*, *SlANS*) displayed higher expression levels in the cotyledon of InR seedlings than *Slhy5* seedlings (Figure 2c,d). Meanwhile, the expression of these genes increased with the development of *Slhy5* seedlings cotyledon, which is consistent with the increasing trend of flavonoid and anthocyanins contents (Figure 2c,d).

2.3. Changes of Metabolites and Genes Expression in the Upper Part of the Hypocotyl of InR Seedlings and Slhy5 Seedlings

In the upper part of hypocotyl, the differential metabolites in group 1, which was positively correlated and had similar consistent patterns with the purple coloration in the hypocotyl, also contained a variety of flavonoids such as petunidin 3-O-glucoside, tulipanin, isotrifoliin (Figure 3a). The KEGG enrichment analysis displayed that the terms 'Flavone and flavonol biosynthesis', 'Photosynthesis', 'Anthocyanin biosynthesis', 'Vitamin B6 metabolism', and 'Starch and sucrose metabolism' were significantly enriched in the Group 1 (Figure S5). WGCNA identified genes in the black module with significant co-expression with the biosynthesis of the metabolite in the flavonoid pathway, which was responsible for the purple coloration observed in the upper part of the hypocotyl (Figure 3b).

KEGG enrichment analysis was performed on the genes in the black module and showed that these DEGs were enriched mainly in flavonoid biosynthesis pathways (Figure 3c). Consistent with the data from the transcriptomic analysis, both the anthocyanin positive regulatory genes, such as *SlAN2*, *SlAN2-like*, *SlAN1I* and *SlAN11* (except for the negative regulatory genes *SlMYB7*, *SlMYB3* and *SlMYB32*) and the anthocyanin biosynthetic genes, including *SlPAL*, *Sl4CL*, *SlCHS*, *SlCHI*, *SlF3H*, *SlF3'H*, *SlDFR*, *SlANS* and *Sl3GT* (Figure 3d,e), exhibited much higher expression in the upper part of InR seedlings hypocotyl than those of *Slhy5* seedlings. Compared with H_1_HU and H_2_HU, the genes involved in anthocyanins under H_4_HU and H_3_HU displayed higher expression levels (Figure 3d,e).

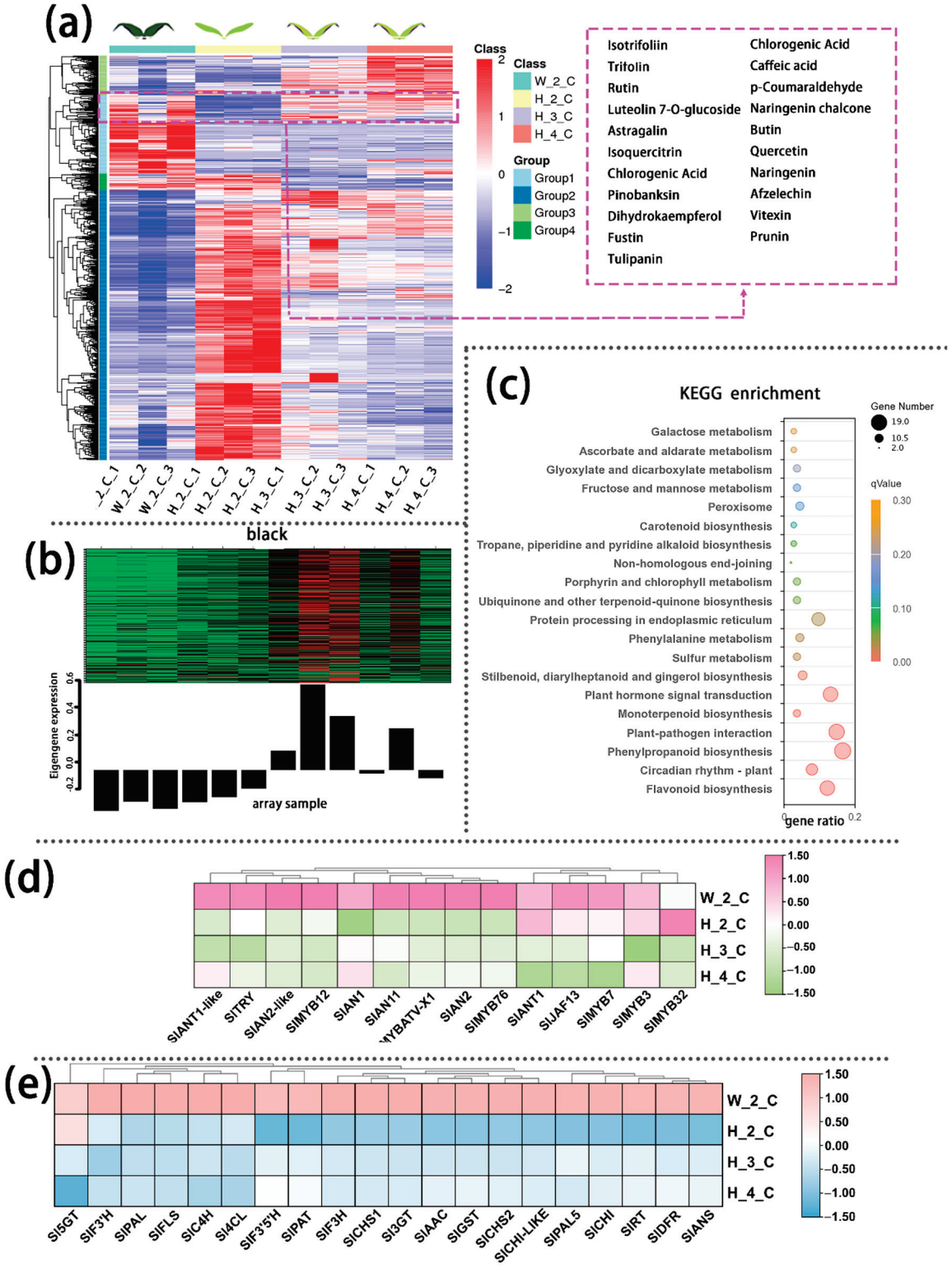


Figure 2. Differentially accumulated metabolites (DAMs) accumulation pattern consisted of the color variations during seedling development of cotyledon (**a**). The genes in the green module consisted of color variations during the seedling development of cotyledon (**b**). Top 20 enriched KEGG pathway enrichment of the genes in green module (**c**). The FKPM values of the transcriptional factors (**d**) and the structural genes (**e**) related to flavonoid and anthocyanin biosynthesis.

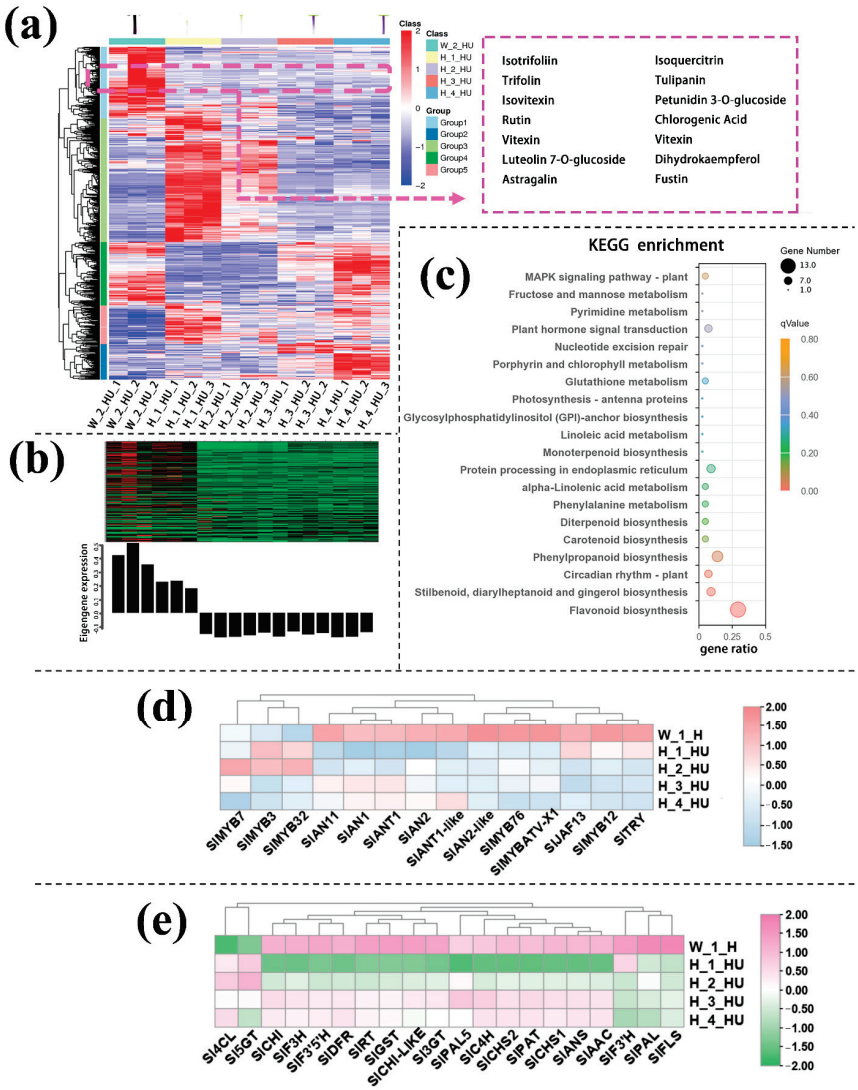


Figure 3. Differentially accumulated metabolites (DAMs) accumulation patterns consisted of color variations during seedling development of the upper part of the hypocotyl (**a**). The genes in the green module consisted of color variations during seedling development of the upper part of the hypocotyl (**b**). Top 20 enriched KEGG pathway enrichment of the genes in black module (**c**). The FKPM values of the transcriptional factors (**d**) and the structural genes (**e**) related to flavonoid and anthocyanin biosynthesis.

2.4. Changes of Metabolites and Genes Expression in the Lower Part of the Hypocotyl of InR Seedlings and Slhy5 Seedlings

Similarly, annotated metabolites in the lower part of the hypocotyl could be divided into several large groups based on similar variation tendencies, respectively. Among these metabolites, group 1 metabolites were present in higher levels in InR seedlings than in *Slhy5* seedlings, which were consistent with the color variations during seedling development. These metabolites include petunidin 3-O-glucoside, tulipanin and isotrifoliin (Figure 4a). KEGG analysis showed that these metabolites in different parts of the seedling were mainly enriched in flavone and flavonol biosynthesis and anthocyanin biosynthesis (Figure S6). The genes in mode marked by pale turquoise color according to WGCNA analysis of transcriptomic data from the lower part of the hypocotyl were consistent with the increasing trend of anthocyanin contents (Figure 4b). The KEGG enrichment analysis showed that the term 'flavonoid biosynthesis' pathway was significantly more pronounced (Figure 4c). Meanwhile, the expression levels of anthocyanin biosynthesis genes in the lower

part of InR hypocotyl were higher than *Slhy5* (Figure 4d,e). In addition, the expression of most anthocyanin biosynthetic genes in the lower part of the *Slhy5* seedling hypocotyl under different stages exhibited the following trend: H_4_HD ≈ H_3_HD > H_2_HD > H_1_HD.

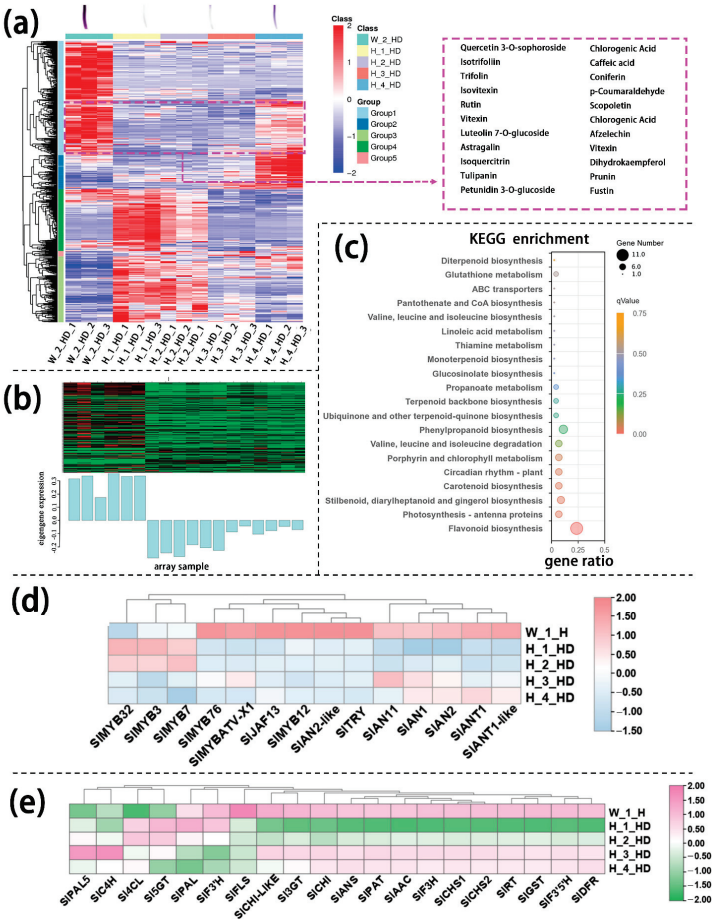


Figure 4. Differentially accumulated metabolites (DAMs) accumulation patterns consisted of color variations during seedling development of the lower part of the hypocotyl (**a**). The genes in the green module consisted of color variations during seedling development of the lower part of the hypocotyl (**b**). Top 20 enriched KEGG pathway enrichment of the genes in the pale turquoise module (**c**). The FKPM values of the transcriptional factors (**d**) and the structural genes (**e**) related to flavonoid and anthocyanin biosynthesis.

2.5. Changes of Metabolites and Genes Expression in Different Parts of Slhy5 Seedling

Based on WGCNA analysis, we identified a module (marked in cyan) whose gene expression pattern was associated with the phenotype of anthocyanin synthesis in *Slhy5* seedlings at the third and fourth development stages (Figure 5a,b). Twenty-one co-expressed genes in the cyan module were significantly correlated with pigment accumulation in *Slhy5* seedlings. These indicated that bHLH (*SlAN1*) was a hub gene involved in the positive regulation of flavonoid metabolism in *Slhy5* seedlings (Figure 5c), possibly by affecting structural node genes, such as *SlCHI*, *SlCHS*, *SlAN3*, *SlDFR* and *SlRT*, etc.

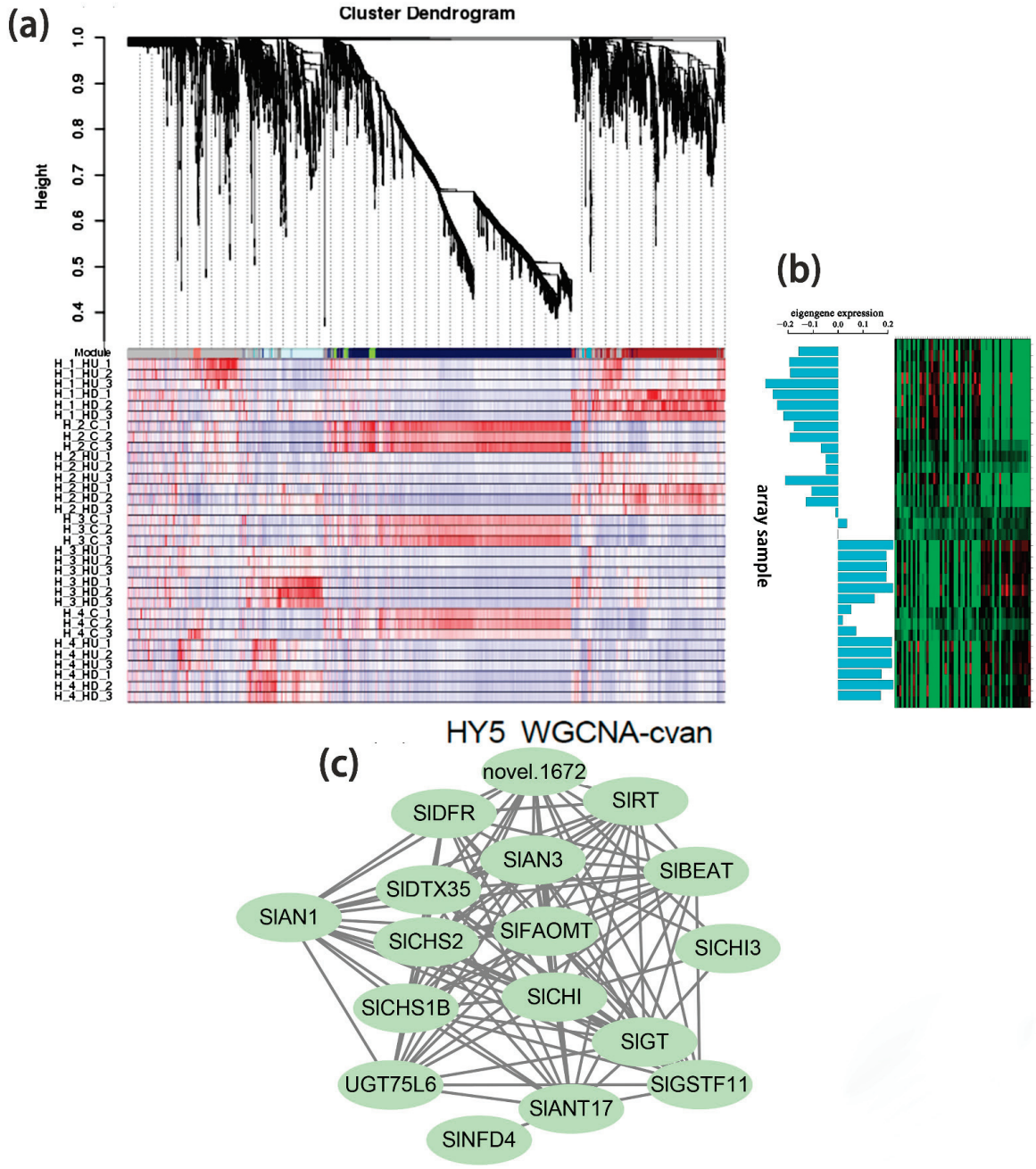


Figure 5. Hierarchical cluster tree showing seven modules obtained by WGCNA in *Slhy5* seedlings (**a**). Differential expression of genes in the accumulation pattern consisted of color variations during seedling development of *Slhy5* seedlings (**b**). Interaction network of DEG in the cyan model in *Slhy5* seedling (**c**).

2.6. Screening of Differentially Expressed Genes of Tomato Fruit

Similar to seedlings, residual anthocyanin production was also observed in the peel of the *Slhy5* fruit shoulder, whereas anthocyanin was almost undetectable in the shading part of the *Slhy5* fruit (Figure 1b,d). To underly gene expression changes over the fruit peel of InR and *Slhy5*, RNA-Seq analysis was conducted. The number of differentially expressed genes had a very high variance among InR and *Slhy5* fruit peel. Regarding WT-S-vs-slhy5-S (purple-colored peel of InR fruit compared to purple-colored peel of *Slhy5* fruit), a total of 2995 DEGs, including 1360 up- and 1635 down-regulated genes were detected (Figure S7). These genes were enriched in KEGG pathways related to photosynthesis, carbon fixation in photosynthetic organisms, phenylpropanoid biosynthesis, phenylalanine metabolism,

flavonoid biosynthesis and flavone and flavonol biosynthesis (Figure 6a), which the gene FKPM values in InR fruit peel was higher than those of *Slhy5* (Figure 6c,d). These results well demonstrated the importance of HY5 in the color formation (both green and purple) of tomato fruits.

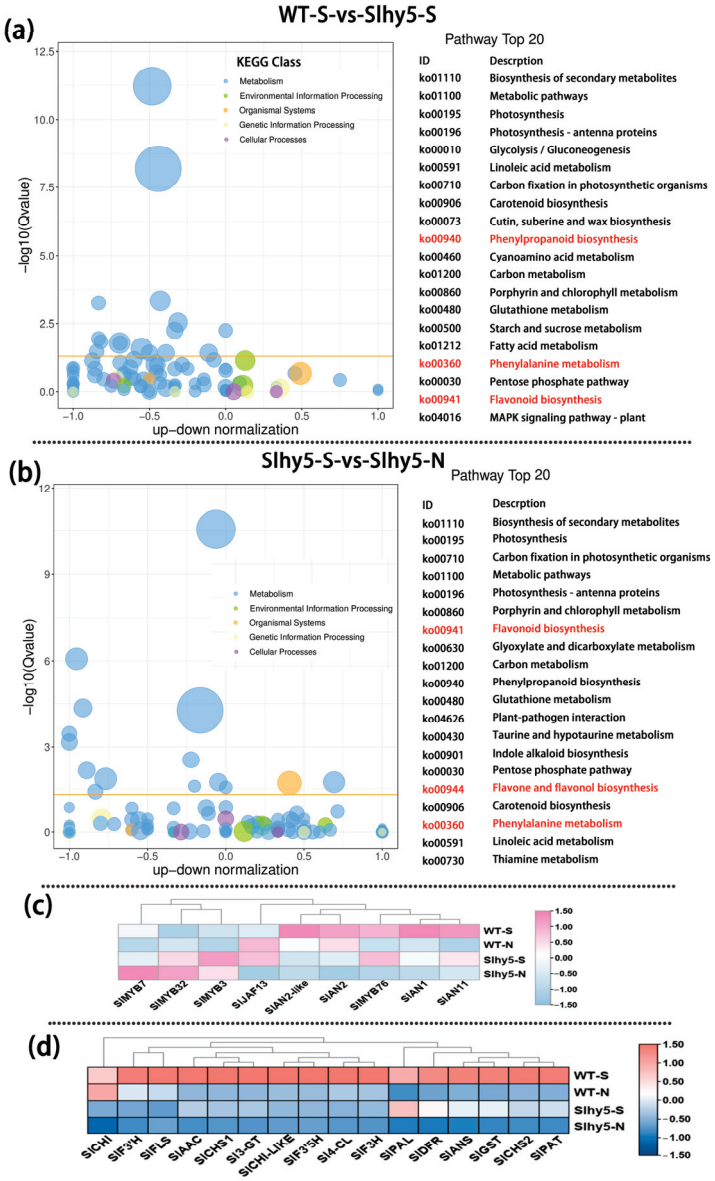


Figure 6. KEGG enrichment of differential genes in WT-S vs. Slhy5-S (**a**) and Slhy5-S vs. Slhy5-N (**b**). The FKPM values of the transcriptional factors (**c**) and the structural genes (**d**) related to flavonoid and anthocyanin biosynthesis.

Meanwhile, to further investigate the DEGs related to *Slhy5* fruit peel coloration, we compared the FPKM values of Slhy5-S-vs-Slhy5-N (purple-colored peel compared to white-colored peel in *Slhy5* fruit). A total of 2393 DEGs were identified from the groups of Slhy5-S-vs-Slhy5-N (Figure S7). According to KEGG enrichment analysis, the 20 top-ranked pathways contributed by these DEGs were photosynthesis, carbon fixation in photosynthetic organisms, porphyrin and chlorophyll metabolism, flavonoid biosynthesis, phenylpropanoid biosynthesis, flavone and flavonol biosynthesis, carotenoid biosynthesis and phenylalanine metabolism. A set of genes involved in flavonoid metabolism in the light-exposing part of the *Slhy5* fruit peel displayed a higher expression level than the shading part (Figure 6b). Therefore, we predicted that the flavonoid biosynthesis pathway

was leading to the purple peel coloration in *Slhy5* fruit peel, that was proved that other genes or TFs were involved in anthocyanin biosynthesis in an HY5-independent manner.

To further verify expression patterns of the DEGs in seedlings and fruit, the genes *SlAN1*, *SlPAL*, *SlCHS*, *SlCHI*, *SlF3H*, *SlDFR*, *SlANS*, *Sl3-GT*, *SlAAC* and *SlGST* were used for qRT-PCR verification (Figure S8). The qRT-PCR results were consistent with the transcriptomic analysis results.

2.7. The Genes Involved in the MYB-bHLH-WD40 (MBW) Complex That Activates Anthocyanidins in Tomato Fruit

To explore the regulation of flavonoid metabolism, the possible interaction of flavonoid metabolism-related genes was tested utilizing yeast two-hybrid (Y2H) assays. *SlPIF1*, *SlPIF3*, *SlAN2-like*, *SlAN2*, *SlAN11*, *SlAN1*, *SlHY5*, *SlJAF13*, *SlBBX24* and *SlWRKY44* were selected for the candidate genes based on the RNA Sequencing results, which might contribute to anthocyanin biosynthesis. We found that SlBBX24 physically interacts with SlAN2-Like and SlAN2 but not with SlAN11 in yeast, while SlWRKY44 could interact with SlAN11 protein only and showed no affinity for other genes. Additionally, these results indicated that SlBBX24 could interact with SlPIF1 and SlPIF3 but not SlPIF4. Unexpectedly, SlAN1 and SlJAF13 displayed the same physical interaction, which could interact with both SlPIF1 and SlPIF3 (Figure 7).

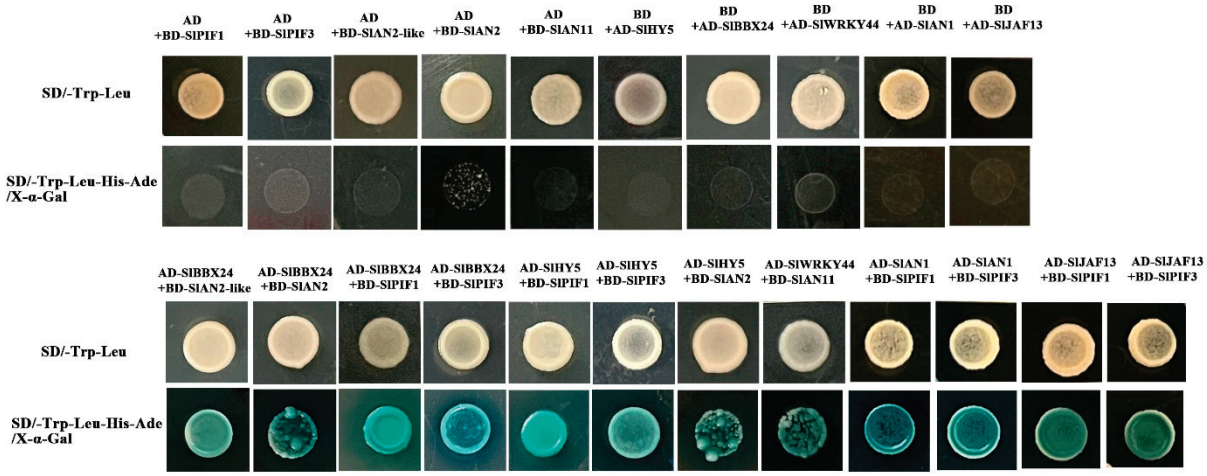


Figure 7. Yeast two-hybrid assay of the protein interactions between candidate genes involved in the components of the MBW complex that activates anthocyanidins in tomato.

2.8. SlBBX24 Physically Interacts with Regulators of Light Signaling and Anthocyanin Biosynthesis

To further explore whether SlBBX24 physically interacts with regulators of light signaling and anthocyanin biosynthesis, we conducted bimolecular luciferase complementation imaging (LCI) assays in Nicotiana benthamiana leaves. Luminescence was observed in leaves that co-expressed SlBBX20-nLUC and cLUC-SlHY5, cLUC-SlPIF1, cLUC-SlPIF3, cLUC-SlAN2-like, cLUC-SlAN2, cLUC-SlAN11. LUC signals were not detectable in the three controls. These results suggest that SlBBX24 associates with the transcription factors involved in the light-signaling pathway (SlHY5, SlPIF1, SlPIF3) and anthocyanin biosynthesis (SlAN2-like, SlAN2, SlAN11) in living plant cells (Figure 8).

2.9. SlBBX24 May Be Involved in Anthocyanin Accumulation in Tomato Fruit Peels

To investigate whether the biosynthesis of anthocyanin is regulated by candidate genes, a target gene of *SlBBX24* was silenced by the VIGS approach. Compared to the empty vector control, the expression of *SlBBX24* was reduced at 7 days after infiltration, and anthocyanin was less vibrant than that of control fruits (Figure 9). qRT-PCR analyses were performed to test the expression changes of the genes involved in anthocyanin biosynthesis. As shown in Figure 8, silencing of *SlBBX24* inhibited the expression levels of

anthocyanin structural genes, including EBGs (*SlC4L and SlCHS1*) and LBGs (*SlDFR* and *SlANS*) as well as the positive regulators (*SlAN1*, *SlAN2*, *SlAN2-like*, *SlAN11*, *SlHY5*), and increased the expression levels of negative regulators (*SlMYBATV*, *SlTRY* and *SlMYB76*) (Figure 9). Therefore, it suggests that *SlBBX24* might be one of the important regulators in the regulatory chain of anthocyanin biosynthesis in tomato fruit.

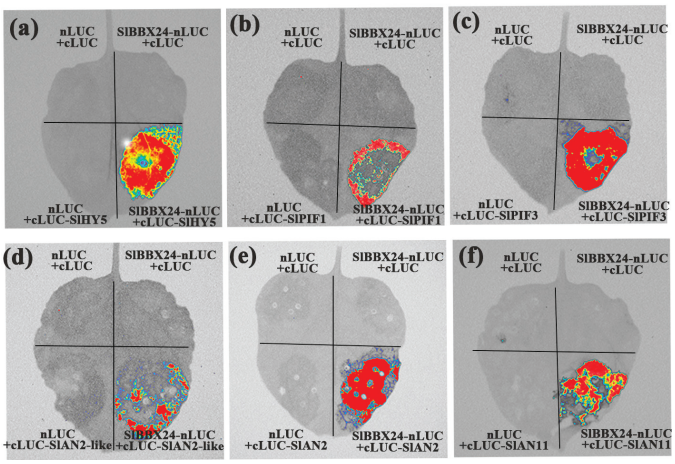


Figure 8. Interactions between *SlBBX20* and *SlHY5* (**a**), *SlPIF1* (**b**), *SlPIF3* (**c**), *SlAN2-like* (**d**), *SlAN2* (**e**), *SlAN11* (**f**) in firefly luciferase complementation imaging (LCI) assays in planta.

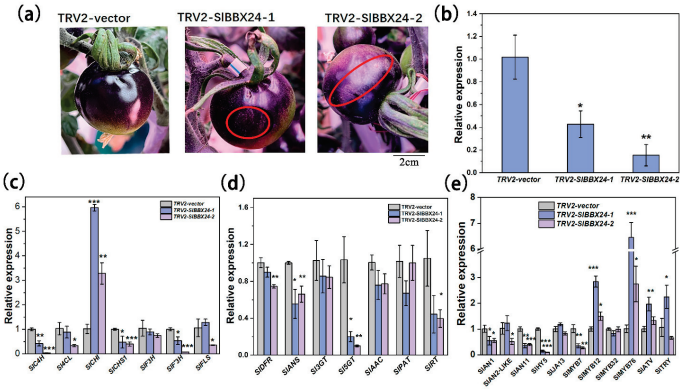


Figure 9. Analysis of VIGS of the *SlBBX24* gene. (**a**) Suppression of *SlBBX24* by virus-induced gene silencing (VIGS) retarded the purple coloration of the tomato fruit peel. (**b**) Relative expression of *SlBBX24* in fruit peels of TRV2-inoculated and TRV2-SlBBBX24-inoculated for 7 days. (**c**–**e**) Relative expression of anthocyanin structural genes of fruit peels of TRV2-inoculated and TRV2-*SlBBX24*-inoculated for 7 days. Statistically significant differences between the purple zone and the green zone were determined by Student's *t*-test. (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

3. Discussion

3.1. Flavonoids Might Be Attributed to Major Color Differences among InR and Slhy5 Mutant Seedlings and Fruit Peel

In recent years, integrated metabolomic and transcriptomic analyses as efficient tools to underly the molecular mechanisms of key metabolic pathways in plants [35–38]. A detailed description of secondary metabolic changes occurring in the whole germinated seeds as well as cotyledons, hypocotyls and roots from 3 to 9 days old tomato seedlings via LCMS profiling, which provided a new perspective to study metabolic networks controlling flavonoid biosynthesis in tomato [39]. The metabolite variants across 20 major tomato growth tissues and stages were explored by combining transcriptome and metabolome approaches, which verified novel transcription factors that control steroidal glycoalkaloids

and flavonoid pathways [40]. In the present study, an integrated analysis of the transcriptome and metabolome was conducted to reveal the differences between wild-type (InR) and *Slhy5* seedlings at different growth and development stages. A total of 987 metabolic components were accumulated specifically in InR and *Slhy5* seedlings, of which amino acids and organic acids, flavanones, flavones and isoflavonoids accounted for a large proportion (Table S2). Flavonoids, a product of the phenylpropanoid metabolism pathway, are extensively distributed in numerous plants and are composed of various subclasses, including flavanones, flavones, isoflavonoids, anthocyanins and flavonols. Anthocyanins, as a key flavonoid subgroup, are responsible for pigmentation in flowers, fruit, seed, and leaf [4]. Our metabolic profiling found metabolites in different parts of the seedling was mainly enriched in flavone and flavonol biosynthesis (Figure 2). Meanwhile, levels of flavonoids, including anthocyanins, in InR seedlings were obviously higher than those of *Slhy5* (Table S2). A KEGG analysis in WT-S-vs-Slhy5-S, Slhy5-S-vs-Slhy5-N were related to phenylpropanoid biosynthesis, phenylalanine metabolism, flavonoid biosynthesis and flavone and flavonol biosynthesis (Figure 5a). These genes were enriched in KEGG pathways related to photosynthesis, carbon fixation in photosynthetic organisms, phenylpropanoid biosynthesis, phenylalanine metabolism, flavonoid biosynthesis and flavone and flavonol biosynthesis (Figure 5a), and the gene FKPM values in InR fruit peel was higher than those of *Slhy5*. These results were well indicated that the main pigment components in InR and *Slhy5* seedlings and fruit peel were flavonoids.

3.2. SlHY5 Acts as a Master Regulator to Control Anthocyanin Biosynthetic in Seedlings and Fruit of Tomato

Anthocyanin biosynthetic genes are regulated directly by the MBW complex consisting of MYB, bHLH and WDR proteins. Aft and Atv are two important loci that are well-associated with anthocyanin biosynthesis in tomatoes. MYB TFs occupy the major determinant position in the control network of anthocyanin biosynthesis and have been well demonstrated [41,42]. Four R2R3 MYB TF genes (*SlAN2*, *SlANT1*, *SlANT1-like* and *SlAN2-like*) were previously identified to regulate anthocyanin biosynthesis in tomatoes [43–45]. Two bHLH TFs, *SlAN1* and *SlJAF13*, were recently reported to regulate anthocyanin production [46]. Otherwise, a tomato WDR protein, *SlAN11,* was also involved in anthocyanin synthesis [47].

HY5 was well characterized as a positive regulator of anthocyanin synthesis in a light-dependent manner. Knock-down *SlHY5* transcription significantly reduced the anthocyanin levels both in seedlings and fruit of tomatoes [11]. *HY5 is* involved in the regulation of anthocyanin biosynthesis by directly binding to MYB transcription factors, such as the production of flavonol glycosides *(MYB12/PFG1* and *MYB111/PGF3*), production of anthocyanin pigment1 (*MYB75/PAP1*) and MYB-like Domain (*MYBD*) [48]. *HY5* activates the expression of *PAP1* expression via direct binding to G-and ACE-boxes in the promoter region, positively inducing the accumulation of anthocyanin in Arabidopsis [22]. Consistently, *PyHY5* alone or interacted with *PyBBX18* activates the expression of *PyMYB10* and *PyWD40,* which subsequently regulate the anthocyanins accumulation in red pear [49]. In this study, expression levels of *SlAN2-like*, *SlAN2*, *SlAN1* and *SlAN11* were higher in InR than *Slhy5* both in seedlings and fruit, and the expression pattern of these genes was consistent with pigment accumulation (Figures 2–4 and 6). Two-hybrid (Y2H) assays determined *SlHY5* regulated anthocyanin biosynthesis through interaction with *SlAN2* (Figure 7), consistent with the result that *SlHY5* is a positive regulator of anthocyanin biosynthesis in vegetative tissues of tomato [50]. Therefore, *SlHY5* might be the master regulator to control anthocyanin accumulation in InR seedlings and fruit via mediating the transcriptional activity of an MBW complex and the enhanced expression of key genes, such as *SlCHI*, *SlCHS*, *SlF3H*, *SlDFR* and *SlANS*.

3.3. Possible Regulatory Mechanisms of Anthocyanin Biosynthesis in an HY5-Independent Manner in Tomato

Besides MYB and bHLH, other TF families also regulate anthocyanin biosynthesis. *SlBBX20* could directly bind to the promoter of *SlDFR* to activate its expression, thus pro-

moting anthocyanin accumulation in tomatoes [20]. In apple, *MdBBX37* interacted with *MdMYB1* and *MdMYB9*, two key positive regulators of anthocyanin biosynthesis, and inhibited the binding of those two proteins to their target genes and, therefore, negatively regulated anthocyanin biosynthesis [10]. *PpBBX18* and *PpBBX21* antagonistically regulate anthocyanin biosynthesis via competitive association with *PpHY5* in the peel of pear fruit. Also, discoveries have emphasized the importance of WRKY protein in the control of the flavonoid pathway and its relationship to the MBW complex [9]. *PpWRKY11* was able to bind to W-box cis-elements in the promoters of *PpMYB10,* then regulated anthocyanin synthesis in pear flesh [23]. In the present study, different from InR seedlings, some TFs were detected in *Slhy5* seedlings by RNA-Seq, such as WRKYs, BBXs and NACs, which might compensate for the function of HY5 and contribute to the expression of related genes involved in anthocyanin synthesis. Y2H assays revealed that *SlBBX24* could interact with *SlAN2-like* and *SlAN2*, which likely had a positive function in the regulation of anthocyanin biosynthesis (Figures 7 and 8). In addition, *SlWRKY44* also could interact with the *SlAN11* protein (Figure 7), which was consistent with the result that the WRKY factor physically interacted with the AN11 in yeast two-hybrid analysis [51]. Silencing of *SlBBX24* via virus-induced gene silencing (VIGS) led to the downregulation of the expression of structural genes and caused a decrease in anthocyanin accumulation (Figure 9). Additionally, PIFs play a role in the biosynthesis of plant pigments. *PIF3* could specifically bind to the G-box element of anthocyanin biosynthesis-related structural genes promoter to up-regulate anthocyanin accumulation in an HY5-dependent manner under far-red light [30]. Furthermore, *PIF4* negatively regulated anthocyanin accumulation by inhibiting *PAP1* transcription in Arabidopsis seedlings [52]. In this study, *SlPIF1* and *SlPIF3* could physically interact with *SlAN1* and *SlJAF13*, as well as *SlBBX24*. We speculated that *SlPIF1*, *SlPIF3*, *SlBBX24* and *SlWRKY44* might be involved in anthocyanin biosynthesis in a manner independent or dependent on *SlHY5*.

Taken together, given the PIFs and MBW were the key regulators of anthocyanin biosynthesis, we proposed a model to clearly illustrate this mechanism (Figure 10). In InR, *SlHY5* expression was induced by light and then activated the activity of the MBW complex to regulate the anthocyanin accumulation. While in *Slhy5* seedlings and fruit, PIFs or several other transcription factors might be involved in coordinating anthocyanin biosynthesis, such as BBXs and WRKYs. More thorough and rigorous molecular studies should be performed to explore the relationship of PIFs or other transcription factors which might be involved in anthocyanin biosynthesis.

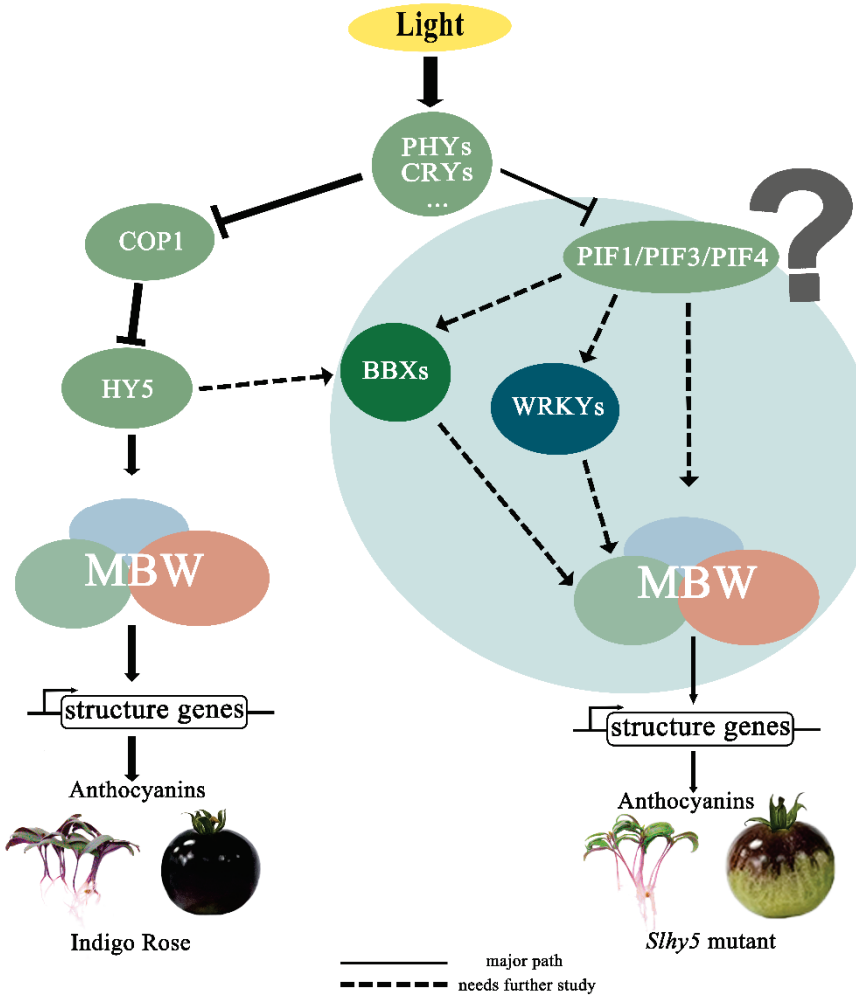


Figure 10. A putative model showing the anthocyanin induction pathways that might be dependent or independent of HY5 in tomatoes.

4. Materials and Methods

4.1. Plant Materials and Growth Conditions

The experiment was carried out in an artificial light plant factory at South China Agricultural University. The tomato seeds of wild-type (cv. 'Indigo Rose' (InR)) and *Slhy5* mutant were generously provided by Dr. Qiu of the College of Horticulture of South China Agriculture University [11]. Seeds were sanitized in 0.5% sodium hypochlorite for 15 min, then rinsed with distilled water. After being soaked in distilled water for 5 h at 25 °C, seeds were sowed in sponge cubes (2 cm × 2 cm × 2 cm) with distilled water in the plant growth chamber in the dark at 25 °C for 3 days. Then, the seedlings were grown under 300 μmol m^{-2} s^{-1} white LEDs (Chenghui Equipment Co., Ltd., Guangzhou, China; 150 cm × 30 cm), 10/14 h light/dark photoperiod, 24 ± 2 °C, and 65–75% relative humidity. The seedlings grown for the 4th, 5th, 6th and 7th days were sampled and divided into three parts: cotyledon (except for the samples on the 4th day, which cotyledons still closed), upper 1 cm of the hypocotyl, and bottom 1 cm of the hypocotyl (Figure 1a). Three biological replicates were collected for analyses, with each replicate composed of 60 seedlings. Fruits from the *Slhy5* mutants and 'Indigo Rose' were sampled at the green-mature stage and fully mature stage with three biological replicates (each replicate consisted of six fruits from different plants). The fruit peel and flesh were carefully split with a scalpel blade. The above-mentioned samples were immediately frozen in liquid nitrogen and stored at −80 °C for further analysis.

4.2. Anthocyanin Assay

The anthocyanin content was performed as described in a previous study with some modifications [53]. Samples (100 mg) were extracted with 1 mL buffers of pH 1.0 (50 mmol KCl and 150 mmol HCl) and pH 4.5 (400 mmol sodium acetate and 240 mmol HCl), respectively, and incubated overnight at 25 °C. The absorbance of the extract liquor was determined at 510 nm using a UV-spectrophotometer (UV-1600, Shimadzu, Kyoto, Japan).

4.3. RNA Sequencing and Data Analyses

Total RNA was extracted from different parts of the seedling (Figure 1a) and the peel of fruit (Figure 1b) using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany). Total amounts and integrity of RNA were assessed using the RNANano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Three independent biological replicates were performed. The RNA-seq sequencing and assembly of seedling and fruit peel were performed by NovoGene Science and Technology Corporation (Beijing, China) and Genedenovo Biotechnology Co., Ltd. (Guangzhou, China), respectively. A total of 3 μg RNA was prepared for sequencing libraries using the NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, Beverly, MA, USA) according to the manufacturer's instructions and sequences attributed to each sample by adding index codes. The library preparations were sequenced on an Illumina Hiseq 4000 platform to generate paired-end reads. The raw sequence reads were filtered by removing adaptor sequences and low-quality sequences, and raw sequences were changed into clean reads. Then the clean reads were then mapped to the tomato reference genome sequence (ITAG 4.0) (https://solgenomics.net/organism/Solanum_lycopersicum/genome/, accessed on 18 November 2021). Padj $\leq$ 0.05 and $|\log2(\text{foldchange})| \geq 1$ were set as the threshold for significantly differential expression. WGCNA analyses were constructed in the BioMarker cloud platform (http://www.biocloud.net, accessed on 10 March 2022).

4.4. Metabolite Extraction

Samples of different parts of seedling preparation, extract analysis, metabolite identification and quantification were performed by the NovoGene Database of NovoGene Co., Ltd. (Beijing, China). Tissues (100 mg) were individually ground with liquid nitrogen, and the homogenate was re-suspended with pre-chilled 80% methanol and 0.1% formic acid by vortexing. The supernatant was injected into an LC-MS/MS system. UHPLC-MS/MS analyses were performed using a Vanquish UHPLC system (ThermoFisher, Dreieich, Germany) coupled with an Orbitrap Q ExactiveTM HF mass spectrometer (ThermoFisher). Structural analysis of metabolites was determined using standard metabolic operating procedures. MRM was used to conduct metabolite quantification. All metabolites identified were subject to partial least squares discriminant analysis. Principle component analysis (PCA) and Orthogonal Partial Least Squares Discriminate Analysis (OPLS-DA) were carried out to identify potential biomarker variables. For potential biomarker selection, variable importance in projection (VIP) $\geq$ 1 and fold change (FC) $\geq$ 2 or $\leq$0.5 were set for metabolites with significant differences.

4.5. Yeast Two-Hybrid Assay

The amplified full-length CDSs of *SlPIF1*, *SlPIF3*, *SlAN2-like*, *SlAN2*, *SlAN11*, *SlAN1*, *SlHY5*, *SlJAF13*, *SlBBX24* and *SlWRKY44* were amplified and inserted into pGADT7 and pGBKT7 vectors, respectively. Primers used for amplified and plasmid construction are listed in Supplementary Table S1. Different combinations of bait and prey vectors were co-transformed into Y2Hgold and then cultured on SD/-Leu-Trp (SD-LT) medium supplemented at 28 °C for 2–3 days. Next, 2.5 μL of aliquots were patched on SD/Ade/His/Leu/Trp plates with 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-Gal) and incubated at 28 °C for 3 days.

4.6. Virus-Induced SlBBX24 Gene Silencing in Tomato

Specific coding regions *SlBBX24* fragment were selected for VIGS vector construction by the VIGS design tool (http://solgenomics.net/tools/vigs, accessed on 10 November 2022). A 300 bp fragment of the coding region of SlBBX24 was amplified using the forward primer (5′-gtgagtaaggttaccgaattc ATGAAGATACAGTGTGATGTG-3′) and the reverse primer (5′-cgtgagctcggtaccggatcc AGTGGCTAAGAAGCGTTGGTG-3′). The PCR product was ligated into the *pTRV2* vector to construct the *TRV2::SlBBX24* vector. The recombinant plasmid and *TRV1* vector were transferred into *Agrobacterium GV3101*. Resuspensions of *pTRV1* and *pTRV2* (as a negative control) or its derivative vectors were mixed at a 1:1 ratio and then infiltrated into a mature green-stage tomato. The injected fruits were grown in a growth chamber at 22 ± 2 °C, and 12 h of light/12 h of darkness, and relative humidity was controlled in the range of 70 ± 5%. After two weeks, fruits were harvested and stored at −80 °C for RNA and qRT-PCR analysis to assess the degree of silencing.

4.7. Luciferase Complementation Imaging Assays (LCI)

The full-length CDS of *SlBBX24* was amplified and cloned, and ligated into pCAMBIA1300-35S-cLUC to produce SlBBX24-nLUC. *SlHY5*, *SlPIF1*, *SlPIF3*, *SlAN2-like*, *SlAN2*, *SlAN11* were fused with *pCAMBIA1300-35S-cLUC* to generate the *cLUC-SlHY5*, *cLUC-SlPIF1*, *cLUC-SlPIF3*, *cLUC-SlAN2-like*, *cLUC-SlAN2* and *cLUC-SlAN11*. Agrobacterium strains GV3101 containing the above constructs were infiltrated into the leaves of 6-week-old N. *benthamiana* plants. The leaf images were captured after a 1 d incubation in darkness at 22 °C and an additional 1 d incubation under a 16-h light/8-h dark photoperiod.

4.8. RNA Extraction and Quantitative Reverse Transcription PCR

Total RNA was extracted from different parts of the seedling and the peel of fruit using RNAex Pro Reagent (Accurate Biotechnology Co., Ltd., Changsha, China), and its quality and quantity were evaluated using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The first cDNA strand was synthesized using Evo M-MLV RT for PCR Kit (Accurate Biotechnology Co., Ltd., Changsha, China). The relative expression levels were determined by performing a quantitative reverse transcription PCR (qRT-PCR) analysis using a LightCycler 480 system (Roche, Hercules, Switzerland) with an Evo M-MLV RT-PCR kit (Accurate Biotechnology Co., Ltd., Changsha, China). The PCR thermocycling protocol was as follows: 95 °C for 30 s, followed by 40 cycles of 95 °C for 15 s and annealing for 60 °C for 30 s. Three biological replicates and three qRT-PCR technical replicates were performed for each sample. The relative expression levels were normalized with the results of mean values of *SlUBI* using the $2^{-\Delta\Delta Ct}$ method [54]. The primer sequences used for the qRT-PCR analyses are listed in Table S1.

4.9. Statistical Analysis

All values are shown as the means of three replicates with standard error (SE). Analysis of variance was performed by Duncan's multiple range test using the SPSS 22.0 program (SPSS 22.0, SPSS Inc., Chicago, IL, USA). Different significance of means was tested by LSD at $p < 0.05$. Graphs were plotted using Origin 2021 (Origin Lab Corporation, Northampton, MA, USA). The heat map analysis was visualized by TBtools software (TBtools-II v1.120) [55].

5. Conclusions

HY5 has a pivotal role in regulating anthocyanin accumulation in tomatoes. *Slhy5* mutants were created via the CRISPR/Cas9 system from 'Indigo Rose'. *Slhy5* mutants displayed significantly lower anthocyanin accumulation than InR, both in seedlings and fruit. Interestingly, detectable levels of anthocyanins were present in hy5 mutant seedlings and fruit, and the pigment accumulation displayed obvious spatiotemporal specificity in the *Slhy5* seedling. These results indicated that other regulators existed to regulate anthocyanin biosynthesis in an HY5-independent manner. The total amount of flavonoids in InR was significantly higher

than in the *Slhy5* mutant, and most of the genes associated with anthocyanin biosynthesis displayed higher expression levels in InR. *SlBBX24* likely regulated anthocyanin biosynthesis by interacting with *SlAN2-like*, *SlAN2*, while *SlWRKY44* interacted with *SlAN11*. Moreover, *SlPIF1* and *SlPIF3* seemed to be involved in anthocyanin biosynthesis by interacting with *SlBBX24*. We identified *SlBBX24* as a target to silence to produce fewer anthocyanins in tomato fruit peel, indicating the important role of *SlBBX24* in the regulation of anthocyanin accumulation. These results deepened the understanding of purple color formation in tomato seedlings and fruits in an *HY5*-dependent or independent manner via excavating the genes involved in anthocyanin biosynthesis. This study will help with the functional analysis of candidate genes controlling the color components of tomatoes and provide a theoretical basis for the breeding of tomatoes with high anthocyanin accumulation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24108690/s1.

Author Contributions: R.H. and H.L. designed the project. R.H., K.L., J.J., Y.H. and S.Z. performed experiments. R.H., H.L., Y.L. and X.L. analyzed the data. R.H., K.L. and H.L. wrote the article. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the Key-Area Research and Development Program of Guangdong Province (2019B020214005 and 2019B020222003).

Institutional Review Board Statement: Animals. Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Tsuda, T. Dietary anthocyanin-rich plants: Biochemical basis and recent progress in health benefits studies. *Mol. Nutr. Food Res.* **2012**, *56*, 159–170. [CrossRef] [PubMed]
2. Koes, R.; Verweij, W.; Quattrocchio, F. Flavonoids: A colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* **2005**, *10*, 236–242. [CrossRef] [PubMed]
3. Rossi, A.; Serraino, I.; Dugo, P.; Paola, R.D.; Mondello, L.; Genovese, T.; Morabito, D.; Dugo, G.; Sautebin, L.; Caputi, A.P.; et al. Protective effects of anthocyanins from blackberry in a rat model of acute lung inflammation. *Free Radic. Res.* **2003**, *37*, 891–900. [CrossRef]
4. Gonzali, S.; Perata, P. Anthocyanins from purple tomatoes as novel antioxidants to promote human health. *Antioxidants* **2020**, *9*, 1017. [CrossRef] [PubMed]
5. Outchkourov, N.S.; Karlova, R.; Hölscher, M.; Schrama, X.; Blilou, I.; Beekwilder, J. Transcription factor-mediated control of anthocyanin biosynthesis in vegetative tissues 1. *Plant Physiol.* **2018**, *176*, 1862–1878. [CrossRef] [PubMed]
6. Butelli, E.; Bulling, K.; Hill, L.; Martin, C. Beyond purple tomatoes: Combined strategies targeting anthocyanins to generate crimson, magenta, and indigo fruit. *Horticulturae* **2021**, *7*, 327. [CrossRef]
7. Petroni, K.; Tonelli, C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* **2011**, *181*, 219–229. [CrossRef]
8. Boase, M.R.; Ngo, H.; Jameson, P.E.; Schwinn, K.E. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in Eudicots. *Plant Cell* **2014**, *26*, 962–980. [CrossRef]
9. Lloyd, A.; Brockman, A.; Aguirre, L.; Campbell, A.; Bean, A.; Cantero, A.; Gonzalez, A. Advances in the MYB-bHLH-WD Repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. *Plant Cell Physiol.* **2017**, *58*, 1431–1441. [CrossRef]
10. Bursch, K.; Niemann, E.T.; Nelson, D.C.; Johansson, H. Karrikins control seedling photomorphogenesis and anthocyanin biosynthesis through a HY5-BBX transcriptional module. *Plant J.* **2021**, *107*, 1346–1362. [CrossRef]
11. Qiu, Z.; Wang, H.; Li, D.; Yu, B.; Hui, Q.; Yan, S.; Huang, Z.; Cui, X.; Cao, B. Identification of candidate HY5-dependent and -independent regulators of anthocyanin biosynthesis in tomato. *Plant Cell Physiol.* **2019**, *60*, 643–656. [CrossRef] [PubMed]
12. Xiong, C.; Luo, D.; Lin, A.; Zhang, C.; Shan, L.; He, P.; Li, B.; Zhang, Q.; Hua, B.; Yuan, Z.; et al. A tomato B-box protein *SlBBX20* modulates carotenoid biosynthesis by directly activating PHYTOENE SYNTHASE 1, and is targeted for 26S proteasome-mediated degradation. *N. Phytol.* **2019**, *221*, 279–294. [CrossRef] [PubMed]
13. Datta, S.; Hettiarachchi, C.; Johansson, H.; Holm, M. Salt tolerance homolog 2, a B-box protein in Arabidopsis that activates transcription and positively regulates light-mediated development. *Plant Cell* **2007**, *19*, 3242–3255. [CrossRef]

14. Datta, S.; Johansson, H.; Hettiarachchi, C.; Irigoyen, M.L.; Desai, M.; Rubio, V.; Holm, M. LZF1/SALT TOLERANCE HOMOLOG3, an Arabidopsis B-box protein involved in light-dependent development and gene expression, undergoes COP1-mediated ubiquitination. *Plant Cell* **2008**, *20*, 2324–2338. [CrossRef] [PubMed]
15. Yadav, A.; Ravindran, N.; Singh, D.; Rahul, P.V.; Datta, S. Role of Arabidopsis BBX proteins in light signaling. *J. Plant Biochem. Biotechnol.* **2020**, *29*, 623–635. [CrossRef]
16. Zhang, X.; Huai, J.; Shang, F.; Xu, G.; Tang, W.; Jing, Y.; Lin, R. A PIF1/PIF3-HY5-BBX23 transcription factor cascade affects photomorphogenesis. *Plant Physiol.* **2017**, *174*, 2487–2500. [CrossRef]
17. Gangappa, S.N.; Crocco, C.D.; Johansson, H.; Datta, S.; Hettiarachchi, C.; Holm, M.; Botto, J.F. The Arabidopsis B-BOX protein BBX25 interacts with HY5, negatively regulating *BBX22* expression to suppress seedling photomorphogenesis. *Plant Cell* **2013**, *25*, 1243–1257. [CrossRef]
18. Job, N.; Yadukrishnan, P.; Bursch, K.; Datta, S.; Johansson, H. Two B-box proteins regulate photomorphogenesis by oppositely modulating HY5 through their diverse C-terminal domains. *Plant Physiol.* **2018**, *176*, 2963–2976. [CrossRef]
19. Holtan, H.E.; Bandong, S.; Marion, C.M.; Adam, L.; Tiwari, S.; Shen, Y.; Maloof, J.N.; Maszle, D.R.; Ohto, M.-A.; Preuss, S.; et al. BBX32, an arabidopsis b-box protein, functions in light signaling by suppressing HY5-regulated gene expression and interacting with STH2/BBX21. *Plant Physiol.* **2011**, *156*, 2109–2123. [CrossRef]
20. Luo, D.; Xiong, C.; Lin, A.; Zhang, C.; Sun, W.; Zhang, J.; Yang, C.; Lu, Y.; Li, H.; Ye, Z.; et al. *SlBBX20* interacts with the COP9 signalosome subunit *SlCSN5-2* to regulate anthocyanin biosynthesis by activating *SlDFR* expression in tomato. *Hortic. Res.* **2021**, *8*, 163. [CrossRef]
21. Zhang, B.; Yang, H.J.; Qu, D.; Zhu, Z.Z.; Yang, Y.Z.; Zhao, Z.Y. The MdBBX22–miR858–MdMYB9/11/12 module regulates proanthocyanidin biosynthesis in apple peel. *Plant Biotechnol. J.* **2022**, *20*, 1683–1700. [CrossRef] [PubMed]
22. Shin, D.H.; Choi, M.; Kim, K.; Bang, G.; Cho, M.; Choi, S.B.; Choi, G.; Park, Y.I. HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in Arabidopsis. *FEBS Lett.* **2013**, *587*, 1543–1547. [CrossRef] [PubMed]
23. Verweij, W.; Spelt, C.E.; Bliek, M.; de Vries, M.; Wit, N.; Faraco, M.; Koes, R.; Quattrocchio, F.M. Functionally similar WRKY proteins regulate vacuolar acidification in petunia and hair development in arabidopsis. *Plant Cell* **2016**, *28*, 786–803. [CrossRef] [PubMed]
24. Alabd, A.; Ahmad, M.; Zhang, X.; Gao, Y.; Peng, L.; Zhang, L.; Ni, J.; Bai, S.; Teng, Y. Light-responsive transcription factor PpWRKY44 induces anthocyanin accumulation by regulating PpMYB10 expression in pear. *Hortic. Res.* **2022**, *9*, 199. [CrossRef]
25. Mes, P.J.; Boches, P.; Myers, J.R.; Durst, R. Characterization of tomatoes expressing anthocyanin in the fruit. *J. Am. Soc. Hortic. Sci.* **2008**, *133*, 262–269. [CrossRef]
26. Vu, A.T.; Lee, J.M. Genetic variations underlying anthocyanin accumulation in tomato fruits. *Euphytica* **2019**, *215*, 196. [CrossRef]
27. Sun, C.; Deng, L.; Du, M.; Zhao, J.; Chen, Q.; Huang, T.; Jiang, H.; Li, C.B.; Li, C. A transcriptional network promotes anthocyanin biosynthesis in tomato flesh. *Mol. Plant* **2019**, *13*, 42–58. [CrossRef]
28. Ha, H.T.N.; Van Tai, N.; Thuy, N.M. Physicochemical characteristics and bioactive compounds of new black cherry tomato (*Solanum lycopersicum*) varieties grown in Vietnam. *Plants* **2021**, *10*, 2134. [CrossRef]
29. Gangappa, S.N.; Botto, J.F. The Multifaceted roles of HY5 in plant growth and development. *Mol. Plant* **2016**, *9*, 1353–1365. [CrossRef]
30. Shin, J.; Park, E.; Choi, G. PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. *Plant J.* **2007**, *49*, 981–994. [CrossRef]
31. Chen, R.; Yang, C.; Gao, H.; Shi, C.; Zhang, Z.; Lu, G.; Shen, X.; Tang, Y.; Li, F.; Lu, Y.; et al. Induced mutation in ELONGATED HYPOCOTYL5 abolishes anthocyanin accumulation in the hypocotyl of pepper. *Theor. Appl. Genet.* **2022**, *135*, 3455–3468. [CrossRef] [PubMed]
32. Lee, J.; He, K.; Stolc, V.; Lee, H.; Figueroa, P.; Gao, Y.; Tongprasit, W.; Zhao, H.; Lee, I.; Xing, W.D. Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* **2007**, *19*, 731–749. [CrossRef] [PubMed]
33. Mohammadi-Shemirani, P.; Sood, T.; Paré, G. From 'Omics to multi-omics technologies: The discovery of novel causal mediators. *Curr. Atheroscler. Rep.* **2023**, *25*, 55–65. [CrossRef] [PubMed]
34. Tiedge, K.; Li, X.; Merrill, A.T.; Davisson, D.; Chen, Y.; Yu, P.; Tantillo, D.J.; Last, R.L.; Zerbe, P. Comparative transcriptomics and metabolomics reveal specialized metabolite drought stress responses in switchgrass (*Panicum virgatum*). *N. Phytol.* **2022**, *236*, 1393–1408. [CrossRef]
35. Rothenberg, D.O.N.; Yang, H.; Chen, M.; Zhang, W.; Zhang, L. Metabolome and transcriptome sequencing analysis reveals anthocyanin metabolism in pink flowers of anthocyanin-rich tea (*Camellia sinensis*). *Molecules* **2019**, *24*, 1064. [CrossRef]
36. Datir, S.; Regan, S. Advances in physiological, transcriptomic, proteomic, metabolomic, and molecular genetic approaches for enhancing Mango fruit quality. *J. Agric. Food Chem.* **2022**, *71*, 20–34. [CrossRef]
37. Roumani, M.; Le Bot, J.; Boisbrun, M.; Magot, F.; Péré, A.; Robin, C.; Hilliou, F.; Larbat, R. Transcriptomics and metabolomics analyses reveal high induction of the phenolamide pathway in tomato plants attacked by the *Leafminer Tuta absoluta*. *Metabolites* **2022**, *12*, 484. [CrossRef]
38. Wang, Z.R.; Cui, Y.Y.; Vainstein, A.; Chen, S.W.; Ma, H.Q. Regulation of fig (*Ficus carica* L.) fruit color: Metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front. Plant. Sci.* **2017**, *8*, 1990. [CrossRef]

39. Roldan, M.V.G.; Engel, B.; de Vos, R.C.H.; Vereijken, P.; Astola, L.; Groenenboom, M.; van de Geest, H.; Bovy, A.; Molenaar, J.; van Eeuwijk, F.; et al. Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. *Metabolomics* **2014**, *10*, 958–974. [CrossRef]
40. Li, Y.; Chen, Y.; Zhou, L.; You, S.; Deng, H.; Chen, Y.; Alseekh, S.; Yuan, Y.; Fu, R.; Zhang, Z.; et al. MicroTom metabolic network: Rewiring tomato metabolic regulatory network throughout the growth cycle. *Mol. Plant* **2020**, *13*, 1203–1218. [CrossRef]
41. Colanero, S.; Tagliani, A.; Perata, P.; Gonzali, S. Alternative splicing in the anthocyanin fruit gene encoding an R2R3 MYB transcription factor affects anthocyanin biosynthesis in tomato fruits. *Plant Commun.* **2020**, *1*, 100006. [CrossRef] [PubMed]
42. Naing, A.H.; Kim, C.K. Roles of R2R3-MYB transcription factors in transcriptional regulation of anthocyanin biosynthesis in horticultural plants. *Plant Mol. Biol.* **2018**, *98*, 1–18. [CrossRef] [PubMed]
43. Schreiber, G.; Reuveni, M.; Evenor, D. Anthocyanin1 from *Solanum chilense* is more efficient in accumulating anthocyanin metabolites than its *Solanum lycopersicum* counterpart in association with the anthocyanin fruit phenotype of tomato. *Theor. Appl. Genet.* **2012**, *124*, 295–307. [CrossRef]
44. Kiferle, C.; Fantini, E.; Bassolino, L.; Povero, G.; Spelt, C.; Perata, P.; Gonzali, S. Tomato R2R3-MYB proteins SlANT1 and SlAN2: Same protein activity, different roles. *PLoS ONE* **2015**, *10*, e0136365. [CrossRef] [PubMed]
45. Meng, X.; Yang, D.; Li, X.; Zhao, S.; Sui, N.; Meng, Q. Physiological changes in fruit ripening caused by overexpression of tomato SlAN2, an R2R3-MYB factor. *Plant Physiol. Biochem.* **2015**, *89*, 24–30. [CrossRef] [PubMed]
46. Subban, P.; Prakash, S.; Mann, A.B.; Kutsher, Y.; Evenor, D.; Levin, I.; Reuveni, M. Functional analysis of MYB alleles from *Solanum chilense* and *Solanum lycopersicum* in controlling anthocyanin levels in heterologous tobacco plants. *Physiol. Plant.* **2021**, *172*, 1630–1640. [CrossRef]
47. Gao, Y.; Liu, J.; Chen, Y.; Tang, H.; Wang, Y.; He, Y.; Ou, Y.; Sun, X.; Wang, S.; Yao, Y. Tomato SlAN11 regulates flavonoid biosynthesis and seed dormancy by interaction with bHLH proteins but not with MYB proteins. *Hortic. Res.* **2018**, *5*, 27. [CrossRef]
48. Stracke, R.; Favory, J.J.; Gruber, H.; Bartelniewoehner, L.; Bartels, S.; Binkert, M.; Funk, M.; Weisshaar, B.; Ulm, R. The Arabidopsis bZIP transcription factor HY5 regulates expression of the *PFG1/MYB12* gene in response to light and ultraviolet-B radiation. *Plant Cell Environ.* **2010**, *33*, 88–103. [CrossRef]
49. Wang, Y.; Zhang, X.; Zhao, Y.; Yang, J.; He, Y.; Li, G.; Ma, W.; Huang, X.; Su, J. Transcription factor *PyHY5* binds to the promoters of PyWD40 and PyMYB10 and regulates its expression in red pear 'Yunhongli No. 1'. *Plant Physiol. Biochem.* **2020**, *154*, 665–674. [CrossRef]
50. Zhi, J.; Liu, X.; Li, D.; Huang, Y.; Yan, S.; Cao, B.; Qiu, Z. CRISPR/Cas9-mediated *SlAN2* mutants reveal various regulatory models of anthocyanin biosynthesis in tomato plant. *Plant Cell Rep.* **2020**, *39*, 799–809. [CrossRef] [PubMed]
51. Quattrocchio, F.; Verweij, W.; Kroon, A.; Spelt, C.; Mol, J.; Koes, R. PH4 of petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* **2006**, *18*, 1274–1291. [CrossRef] [PubMed]
52. Qin, J.; Zhao, C.; Wang, S.; Gao, N.; Wang, X.; Na, X.; Wang, X.; Bi, Y. PIF4-PAP1 interaction affects MYB-bHLH-WD40 complex formation and anthocyanin accumulation in Arabidopsis. *J. Plant Physiol.* **2022**, *268*, 153558. [CrossRef] [PubMed]
53. Rapisarda, P.; Fanella, F.; Maccarone, E. Reliability of analytical methods for determining anthocyanins in blood orange juices. *J. Agric. Food Chem.* **2000**, *48*, 2249–2252. [CrossRef] [PubMed]
54. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]
55. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

GmTCP and *GmNLP* Underlying Nodulation Character in Soybean Depending on Nitrogen

Yunchol Kim †, Jinhui Wang †, Chao Ma, Cholnam Jong, Myongil Jin, Jinmyong Cha, Jing Wang, Yang Peng, Hejia Ni, Haibo Li, Mingliang Yang, Qingshan Chen * and Dawei Xin *

College of Agriculture, Northeast Agricultural University, Harbin 150036, China; 13134518306@163.com (Y.K.); jinhuiwang113@126.com (J.W.); mcneau@163.com (C.M.); 13030046075@163.com (C.J.); j15504506210@163.com (M.J.); chejinming1983@163.com (J.C.); wangjing20211029@163.com (J.W.); b210301014@neau.edu.cn (Y.P.); nhjwinner@163.com (H.N.); lhb2172637105@163.com (H.L.); yml5419@126.com (M.Y.)

* Correspondence: qshchen@neau.edu.cn (Q.C.); dwxin@neau.edu.cn (D.X.)

† These authors contributed equally to this work.

Abstract: Soybean is a cereal crop with high protein and oil content which serves as the main source of plant-based protein and oil for human consumption. The symbiotic relationship between legumes and rhizobia contributes significantly to soybean yield and quality, but the underlying molecular mechanisms remain poorly understood, hindering efforts to improve soybean productivity. In this study, we conducted a transcriptome analysis and identified 22 differentially expressed genes (DEGs) from nodule-related quantitative trait loci (QTL) located in chromosomes 12 and 19. Subsequently, we performed functional characterisation and haplotype analysis to identify key candidate genes among the 22 DEGs that are responsive to nitrate. Our findings identified *GmTCP* (TEOSINTE-BRANCHED1/CYCLOIDEA/PCF) and *GmNLP* (NIN-LIKE PROTEIN) as the key candidate genes that regulate the soybean nodule phenotype in response to nitrogen concentration. We conducted homologous gene mutant analysis in *Arabidopsis thaliana*, which revealed that the homologous genes of *GmTCP* and *GmNLP* play a vital role in regulating root development in response to nitrogen concentration. We further performed overexpression and gene knockout of *GmTCP* and *GmNLP* through hairy root transformation in soybeans and analysed the effects of *GmTCP* and *GmNLP* on nodulation under different nitrogen concentrations using transgenic lines. Overexpressing *GmTCP* and *GmNLP* resulted in significant differences in soybean hairy root nodulation phenotypes, such as nodule number (NN) and nodule dry weight (NDW), under varying nitrate conditions. Our results demonstrate that *GmTCP* and *GmNLP* are involved in regulating soybean nodulation in response to nitrogen concentration, providing new insights into the mechanism of soybean symbiosis establishment underlying different nitrogen concentrations.

Keywords: TCP transcription factor; nodulation; nitrate concentration

Citation: Kim, Y.; Wang, J.; Ma, C.; Jong, C.; Jin, M.; Cha, J.; Wang, J.; Peng, Y.; Ni, H.; Li, H.; et al. *GmTCP* and *GmNLP* Underlying Nodulation Character in Soybean Depending on Nitrogen. *Int. J. Mol. Sci.* **2023**, *24*, 7750. https://doi.org/10.3390/ijms24097750

Academic Editor: Wajid Zaman

Received: 14 March 2023
Revised: 19 April 2023
Accepted: 21 April 2023
Published: 24 April 2023

1. Introduction

Soybean is considered one of the most important crops due to the high oil and protein content in its seeds [1]. As the demand for soybeans continues to increase in many countries around the world, scientific cultivation and management of soybean farming have become essential issues. Legumes have a unique feature that distinguishes them from other plants: symbiosis with *Rhizobium*. In other words, legumes have the ability to fix nitrogen for root nodule formation through symbiosis with *Rhizobium* [2]. Biological nitrogen fixation is a process in which nitrogen is converted into a usable form with the help of living organisms. Soybean has the capacity to form symbiosis with *Rhizobium*, which allows it to acquire nitrogen from the air [3,4].

Identifying the genes responsible for establishing symbiotic interactions between soybean and *rhizobium* is essential for understanding the molecular mechanism underlying

their interaction. In recent years, the development of molecular marker analysis techniques and the construction of molecular genetic maps, quantitative trait-related QTL analysis and gene cloning of many crops have become a hot spot in molecular breeding [5,6]. The in-depth analysis of quantitative genetic traits is closely related to the development of modern molecular biology techniques [7]. The application and development of molecular markers, population mapping and statistical analysis methods have greatly improved the efficiency of candidate gene discovery [8,9]. In one study, a total of 22 QTLs for the two nodule traits (nodule number and nodule dry weight) were mapped to 12 different chromosomes in the soybean genome, and 17 candidate genes were selected by the previous RNA-seq analysis after infection with *Sinorhizobium fredii* HH103 (a fast-growing rhizobial strain) [10]. *S. fredii* strain HH103 was isolated from a soil sample of Honghu County (Hubei, China) and was first described in 1985, and through its complete genome sequencing and annotation provided a valuable basis for the study of *S. fredii* HH103 and soybean [11].

The TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factors are a family of plant-specific proteins that play critical roles in plant growth, development and stress response [12,13]. TCP transcription factors are characterised by a conserved region, a non-canonical basic helix-loop-helix (bHLH) DNA-binding domain called the TCP domain [14,15]. TCP proteins can be divided into two classes. Class I lacks four conservative amino acids in the structure, whereas class II has four conservative amino acids inserted. Class II can be further subdivided into CIN and CYC based on sequence differences [16,17]. In recent years, the regulatory roles of various TCP transcription factors have been extensively studied [18,19]. *Arabidopsis TCP14* and *TCP15* can regulate plant height [20], while *TCP11* and *TCP16* can regulate the vascular bundle and pollen development [21,22]. In addition, *TCP3* can regulate leaf development [23], and *TCP17* can regulate hypocotyl elongation by regulating light and auxin signalling [24]. *Arabidopsis TCP20* is involved in the regulation of shoot, flower and embryo development. In particular, *TCP20* has a tubular function in the systemic signalling pathway of nitrate uptake by the *Arabidopsis* root system [25,26]. *TCP20* mutants have normal primary and lateral root growth in uniform nitrate media, but lateral root growth is inhibited in non-uniform media [17]. In addition, even when ammonium ions are uniformly present in the culture medium, the preferential lateral root growth of the mutant is still clearly inhibited. In the *TCP20* mutant complemented by the transformation of the *TCP20* cDNA clone, the preferential lateral root growth was restored to WT levels [17]. These data demonstrate that *TCP20* regulates nitrate responses and root development. A total of 54 GmTCPs have been identified in soybean, and the GmTCPs have been divided into two homology classes: Class I and Class II. Class II was further subdivided into two subclasses: CIN and CYC/TB1 [27,28]. *GmTCP1*, *GmTCP2*, *GmTCP6*, *GmTCP8*, *GmTCP12*, *GmTCP17* and *GmTCP43* were the best orthologues of *Arabidopsis AtTCP14* and *AtTCP15*, suggesting their potential functions in plant height regulation [27]. One researcher found that overexpression of the flowering repressor gene *E1* could affect leaf development in soybean by directly repressing a large number of leaf development-related CIN-type TCP genes, thus regulating leaf development and flowering time [29]. *GmTCP30* and *GmTCP33* respond to drought stress, *GmTCP42* to heat stress and *GmTCP47* to salt and heat stress [27]. In a study, *Agrobacterium*-mediated transformation was used to introduce the CRISPR/Cas9 expression vector into the soybean cultivar 'Williams 82' and generate targeted mutants of the *GmTCP19L* gene, which has been implicated in soybean responses to *Phytophthora sojae* [30]. Although the TCP gene has recently been found to play an important role in abiotic stress and hormone signalling responses in many plant species, it is still unexplored how TCP family genes affect growth and responses to abiotic stress, and whether TCP genes involved in nitrate response affect nodulation or not.

NIN-LIKE PROTEIN (NLP) consists of three domains including the N-terminal conserved domain, an RWP-RK DNA-binding domain and a PB1 domain. The N-terminal conserved domain is required for the nitrate response, while the RWP-RK domain is essential for protein–protein interaction [31]. The PB1 domains adopt a ubiquitin-like β-grasp

fold containing two α-helices and a mixed five-stranded β-sheet, and are classified into groups containing an acidic OPCA motif (type I), the invariant lysine residue on the first β-strand (type II), or both (type I/II). The PB1 domain is also a conserved domain required for protein–protein interaction [32]. Although the PB1 domain of NLP transcription factors appears to mediate protein–protein interactions associated with nitrate-inducible gene expression in higher plants, its precise role in nitrate-inducible gene expression has not been characterised [33]. In the model legume *Lotus japonicus*, LjNLP1—previously identified as a NIN-like protein—directly regulates the expression of *LjNRT2.1* and *LjNLP4*, which are involved in regulating the expression of nodulation-related genes [34,35]. In *Medicago truncatula*, *MtNLP1* also inhibits nodulation by directly inducing a negative regulator of nodulation, *MtCLE35* [34]. *NLP6* and *NLP7* are activators of nitrate assimilation genes [26]. *AtNLP6* and *AtNLP7* have been identified as key transcription factors for the primary nitrate response (PNR) in *Arabidopsis* [33]. Under N starvation, TCP20-NLP6&7 heterodimers accumulate in the nucleus and then down-regulate the expression of the G2/M cell cycle marker gene *CYCB1;1* and up-regulate nitrate assimilation [26]. The NLP protein is a homologue of NIN (NODULE INCEPTION) and is found in a variety of plants. *GmNINa* interacted with *NNC1* (Nodule Number Control 1) and predominantly inhibited the transcriptional activation of *GmRIC1* (Rhizobia-Induced CLE1), *GmRIC2* and *miR172c* by *GmNINa*, thereby controlling nodulation [36]. The NIN-like protein (NLP) transcription factors *NLP2* and Nodule Inception (NIN) directly activate the expression of leghemoglobins through a promoter motif that resembles a 'double' version of the nitrate responsive elements (NREs) targeted by other NLPs, and has conserved orientation and position across legumes [37]. Few studies have been carried out on NLP in soybean. TCP and NLP proteins belong to an evolutionarily conserved family that appears to play ancient roles. TCP and NLP proteins may be an important factor in the nitrate response. In soybean, TCP and NLP are involved in the nitrate response pathway and affect symbiotic nitrogen fixation.

There is a significant interaction between soil nitrogen content and symbiosis establishment. In this study, genes' responses to *rhizobium* infection were identified by mining the different expression genes induced by *rhizobium*. Two candidate genes, *GmTCP* and *GmNLP*, were identified. Complementation analysis of *GmTCP* and *GmNLP* in *Arabidopsis* response to nitrogen content was performed. Moreover, the roles of *GmTCP* and *GmNLP* regulating the nodule phenotype depending on different nitrogen contents were elucidated. Our results will contribute to the functional analyses of *GmTCP* and *GmNLP* genes and genetic modification in the soybean symbiotic character.

2. Results

2.1. Identification of Candidate Genes Associated with Nodulation Traits

In the previous RNA-seq analysis, we determined two chromosome regions that might respond to rhizobium induction and identified 22 DEGs in those two regions (Figure 1a). Among the DEGs, *Glyma.19G095300*, *Glyma.19G095400* and *Glyma.19G094800* exhibited a tendency for their expression levels to gradually increase until 6 h of infection by *Sinorhizobium fredii* HH103. Conversely, the expression levels of *Glyma.12G185700*, *Glyma.12G184900*, *Glyma.19G095900*, *Glyma.12G185200* and *Glyma.12G186000* decreased after infection with *S. fredii* HH103. Furthermore, we also detected the single nucleotide polymorphism support of *GmTCP* and *GmNLP* in the chromosome segments' substituted lines [10]. A high degree of single nucleotide polymorphism (SNP) between the parents and lines of chromosome segments' substituted lines was identified [10] (Figure 1b). These results led us to the hypotheses that *GmTCP* and *GmNLP* could respond to rhizobium.

2.2. TCP and NLP Homolog Genes' Responses to the Nitrogen Concentration in Arabidopsis

Firstly, in order to study the function of candidate genes in *Arabidopsis thaliana*, sequence blast was used to identify the homolog protein with *GmTCP/GmNLP* in *Arabidopsis thaliana.* The homology between these genes and *GmTCP/GmNLP* was found to be 61.0%

and 86.5%, respectively. Since the TCP and NLP family are related to nitrogen response, it is important to understand their function. We detected the gene response to different nitrogen levels in *Arabidopsis thaliana*. Using the knockout mutant, we could also detect the complementation effect of *GmTCP/GmNLP* on different nitrogen contents. Furthermore, we investigated how the primary root growth of the *tcp20* mutant responded to nitrate concentration. Specifically, the change in the primary root length phenotype of the *tcp20* mutant was examined across nitrate concentrations ranging from 0 to 5 mM. Under nitrate starvation, the primary root length of the *TCP20* mutant was significantly different from that of the wild type. However, at high nitrate concentrations (5 mM KNO_3), the difference in primary root length between the mutant and the wild types was not apparent. At lower nitrate concentrations, the primary root length of the *TCP20* mutant was significantly shorter, but this difference gradually decreased at higher concentrations (Figure 2a,c).

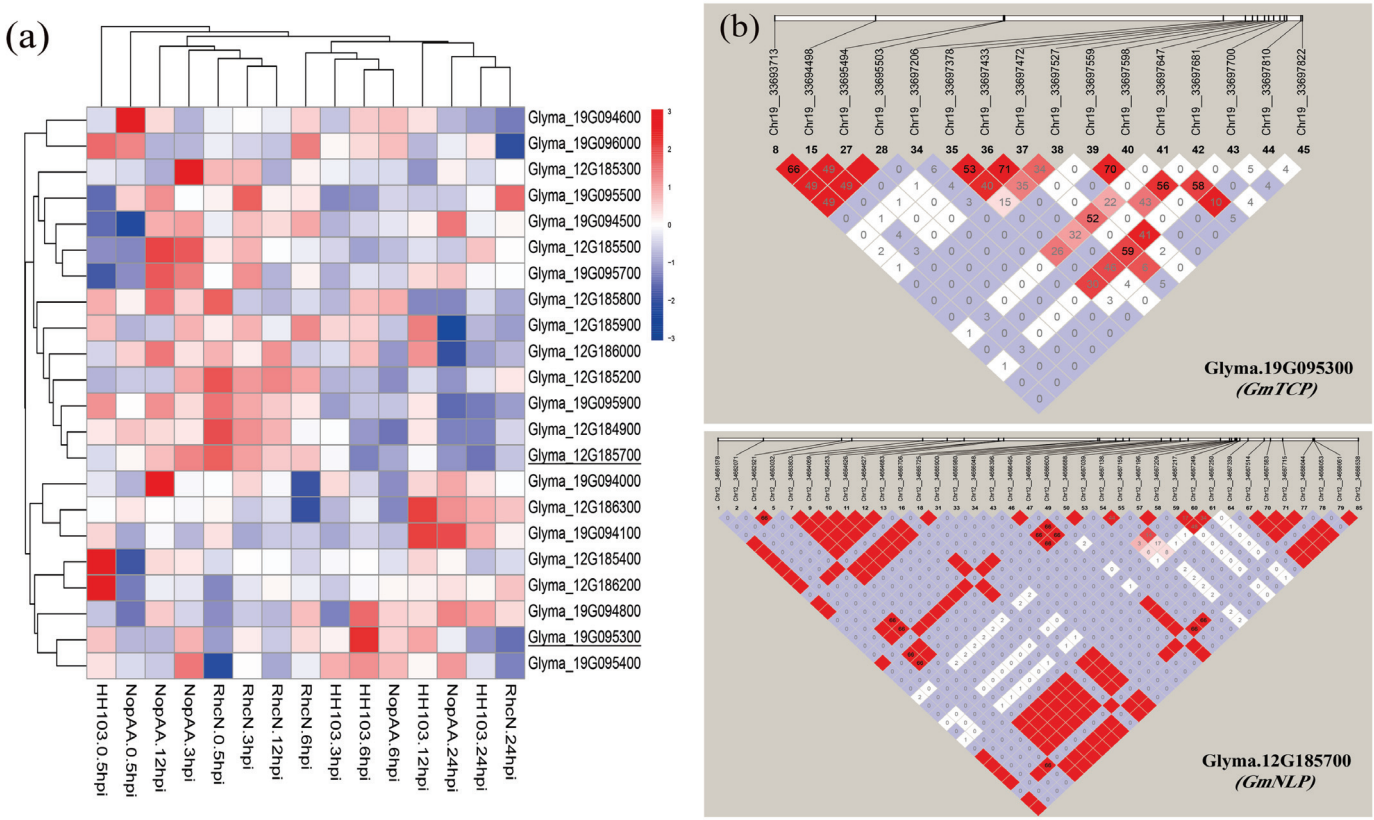


Figure 1. The expression profiles of DEGs and haplotype analysis. (**a**) Heat map of DEGs located in QTL locus. (**b**) Haplotype analysis of two candidate genes in the chromosome segments' substituted lines.

It has recently been reported that the NIN-like protein LjNLP1 directly regulates the expression of *LjNRT2.1* and *LjNLP4*, as well as nodulation-related genes. It has also been shown that certain NLP genes encode the factors that activate the nitrate assimilation genes. To further investigate the response of *NLP1* to nitrate concentrations, primary root length was examined in the nlp1 mutant. The results showed a significantly reduced primary root growth under N starvation and low nitrate concentrations; however, this difference was not significant under high nitrate conditions (Figure 2b,d). Multiple comparison analyses showed that there was no correlation between *AtTCP20* and *AtNLP1*. Two-way ANOVA analysis showed that the interaction effect between N treatment and genotypes was statistically significant in regulating growth phenotypes, such as primary root length ($p < 0.05$).

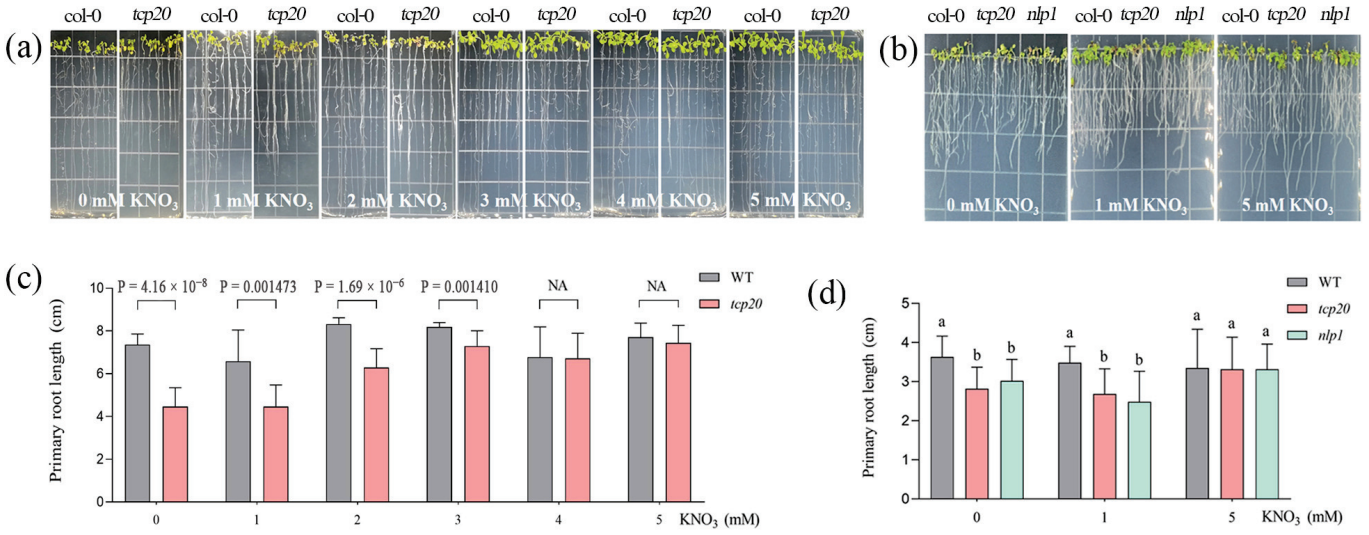


Figure 2. Root phenotype and primary root length of WT, single mutants under 0~5 mM nitrate treatment. (**a**) Root phenotype of WT, *tcp20* mutants under 0~5 mM nitrate treatment. (**b**) Root phenotype of WT, *nlp1* mutants under 0, 1, and 5 mM nitrate treatment. (**c**) Primary root length of WT, *tcp20* mutants under 0~5 mM nitrate treatment. (**d**) Primary root length of WT, *nlp1* mutants under 0, 1, and 5 mM nitrate treatment. For root phenotype and primary root length, *Arabidopsis* seedlings were grown on 5 mM KNO_3 plates for 6 days and then were transferred to new 0~5 mM KNO_3 plates for three days. Error bars show SEM (n = 10). One-way ANOVA was performed and followed by t-test to calculate the p-value (using WT as control). Different letters represent statistic difference between WT and mutants under same nitrate condition ($p < 0.05$).

Finally, the mutation of these genes resulted in a limitation of the primary root length under N starvation and low nitrate concentrations, suggesting that they may encode key factors in the development of nitrate-responsive root formation.

2.3. Response of GmTCP and GmNLP to Different Nitrate Concentrations

Previous experiments on *AtTCP20* and *AtNLP1* confirmed that these genes are involved in the nitrate response signalling pathway. We further investigated the functions of *GmTCP* and *GmNLP* in the soybean nitrate response pathway using *GmTCP/GmNLP* transgenic hairy roots. Six plants were selected from each overexpressed line and used for RNA extraction and the subsequent evaluation of *GmTCP/GmNLP* expression by qRT-PCR. The expression of *GmTCP* and *GmNLP* was significantly higher than that of the empty vector (Figure 3a,d). Three lines with the highest *GmTCP/GmNLP* expression levels were then used for further phenotype analysis. In the knockout transgenic lines, a mutation effect of two to seven base deletions in a target region was observed, and the nodulation phenotype of plants with this mutation was observed at different nitrogen concentrations. The overexpression of *GmTCP* had a negative effect on nodule number and nodule dry weight (Figure 3b), although the nodule number was reduced more at high nitrogen concentrations than at low nitrogen concentrations (0 mM and 1 mM KNO_3) (Figure 3b). When *GmTCP* was knocked out, the nodule number and nodule dry weight had no significant difference but at the high nitrogen concentration (5mM KNO_3) (Figure 3b,c). These results support *GmTCP* being able to regulate the nodule phenotype depending on the high nitrogen concentration.

The overexpression of *GmNLP* also showed a slight difference with increasing nitrate concentrations, but it was significantly reduced compared to the wild type. The *GmNLP* mutant lines showed some reduction in nodule formation with increasing nitrate concentrations (Figure 3e,f). Under different nitrogen concentrations, the difference in nodule phenotype was not significant, however, compared with the wild type at high nitrogen concentration being able to inhibit the nodule number, and the knockout mutant of

GmNLP could recover the nodule phenotype as wild type at a lower nitrogen concentration (Figure 3e,f). Taken together, these results support *GmTCP* and *GmNLP* being involved in the formation and regulation of nodulation at a high nitrate concentration but not at a low nitrogen concentration.

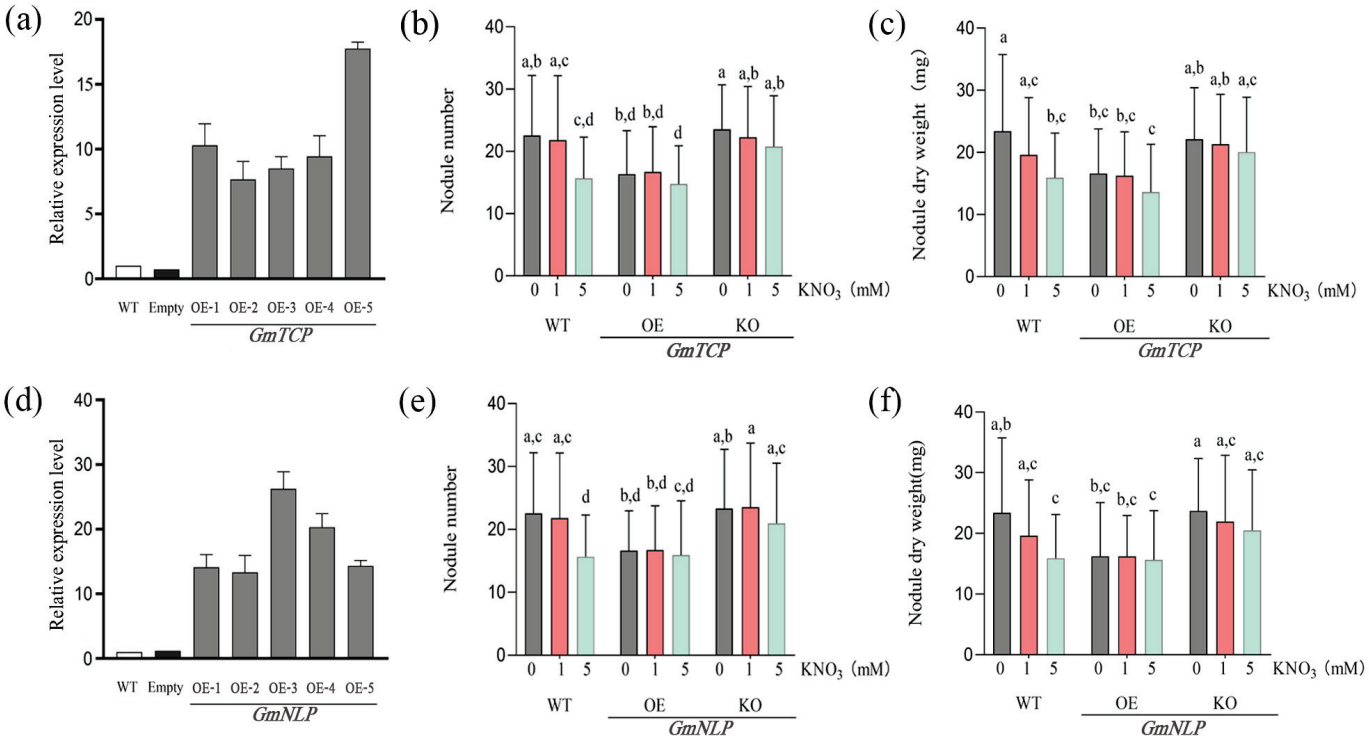


Figure 3. WT, overexpression, and mutant lines' response to different nitrate concentrations in soybean. (**a**,**d**) The identification of *GmTCP* and *GmNLP* overexpression. (**b**,**c**) Overexpression, knockout of *GmTCP* response to different nitrate concentrations. (**e**,**f**) Overexpression, knockout of *GmNLP* response to different nitrate concentrations. The nodule phenotype (nodule number, nodule dry weight) of WT, overexpression and mutant lines was observed 28 days after inoculation with *S. fredii* HH103. Error bars show SEM (n = 20). Different letters indicate statistically significant differences ($p < 0.05$, one-way ANOVA followed by multiple comparisons).

2.4. Effects of GmTCP and GmNLP on the Nodulation

We demonstrated that *GmTCP* and *GmNLP* were involved in the formation and regulation of nitrate-induced nodulation. To investigate the effects of *GmTCP* and *GmNLP* on nodule phenotype, overexpression and knockout lines were constructed via the soybean hair root system. The effects of *GmTCP* and *GmNLP* on the nodule number and nodule dry weight were detected at 28 post inoculation with *S. fredii* HH103. The five plants were selected from the overexpressed line and used for RNA extraction and subsequent evaluation of the *GmTCP*/*GmNLP* expression through qRT-PCR. The expression was significantly higher in *GmTCP* and *GmNLP* than in the empty vector (Figure 4b,e). These lines were used for further phenotype analysis. In knockout transgenic lines, a mutant effect of two to seven base deletions was observed and the nodulation phenotype of plants with this mutation was observed at different concentrations. It was observed that when compared with the wild type, the overexpression of *GmTCP* resulted in a decrease in the nodule number and the nodule dry weight compared to the wild type. This finding was consistent with the above experimental data that showed the inhibition of the nodulation in the overexpression lines. In contrast, the mutant had a nodule phenotype that was almost indistinguishable from or slightly higher than that of the wild type. This indicates that *GmTCP* function loss resulted in the removal of the inhibition of nodulation (Figure 4a,c,d). Consistent results were obtained for *GmNLP*; the overexpression resulted in a more reduced

nodule phenotype than that of the wild type, while the mutant had a higher nodule phenotype than that of the wild type (Figure 4a,f,g). These results support the upregulation of *GmTCP* and *GmNLP* being able to inhibit nodule formation, and the result of the nodule number and nodule dry weight, suggesting that *GmTCP* and *GmNLP* are involved in the regulation of nodulation.

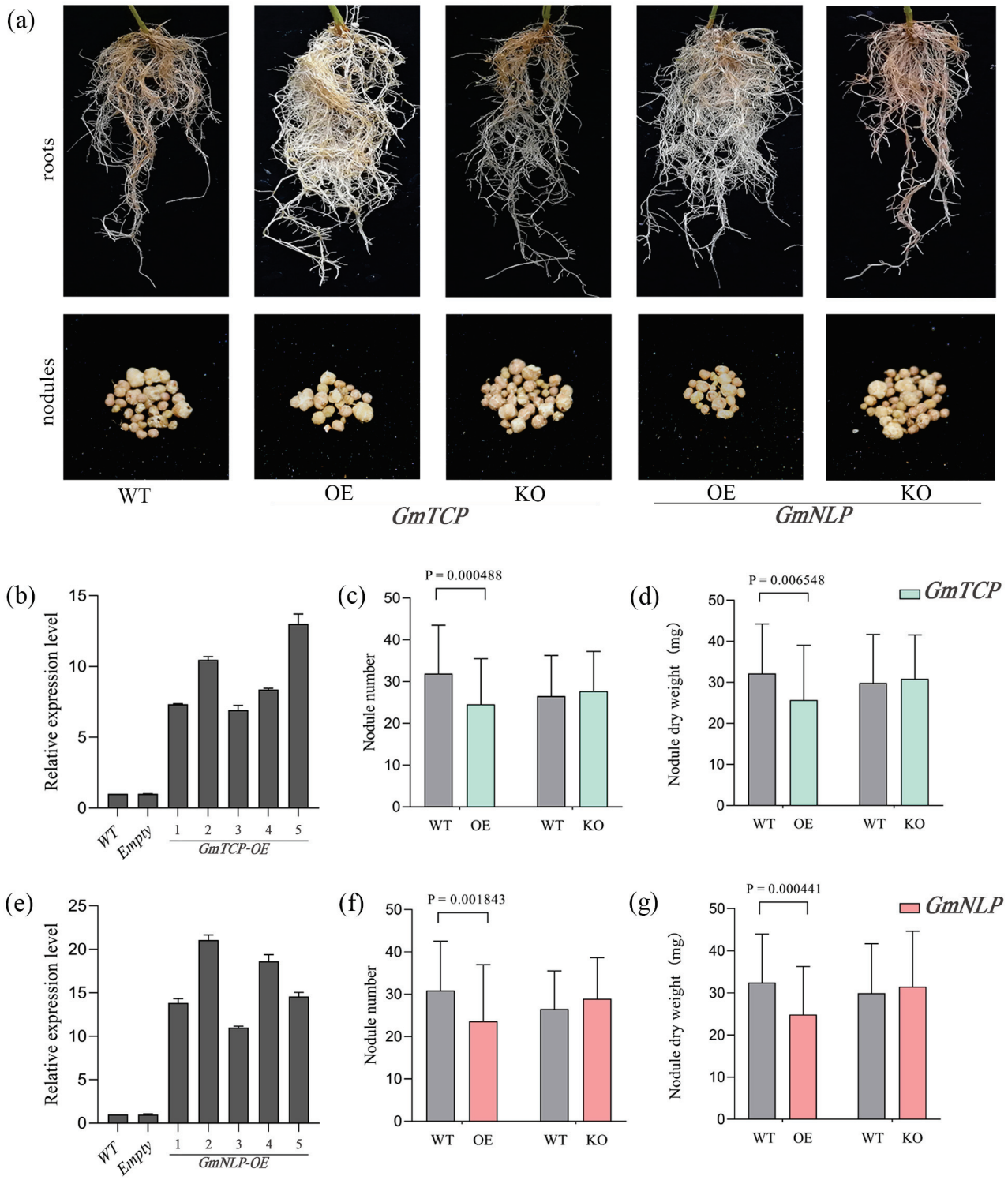


Figure 4. The influence of WT, overexpression and knockout of *GmTCP* and *GmNLP* in nodule. The growth of transgenic hairy roots 30 d after inoculation with *S. fredii* HH103. KO showed that the target gene was silenced in hairy roots, OE showed that the target gene was overexpressed in hairy

roots, WT, and soybean hairy roots with the empty vector pSOY1-hairyred. (**a**) Photographs of root development and growth performance of nodules. (**b**) The identification of overexpression of *GmTCP*. (**c**) The nodule number of overexpression, knockout of *GmTCP*. (**d**) The nodule dry weight of overexpression, knockout of *GmTCP*. (**e**) The identification of overexpression of *GmNLP*, (**f**) the nodule number of overexpression, knockout of *GmNLP*. (**g**) The nodule dry weight of overexpression, knockout of *GmNLP*. The nodule phenotype (nodule number, nodule dry weight) of WT, overexpression and mutant were observed 28 days after inoculation with *S. fredii* HH103. Error bars show SEM (n = 20). Different letters indicate statistically significant differences ($p < 0.05$, one-way ANOVA followed by multiple comparisons).

2.5. Expression Pattern of GmTCP during Root Growth and Nodulation

To determine the spatial expression pattern of *GmTCP*, we performed a promoter-GUS reporter analysis. First, the promoter fragment (3.0 Kb) of *GmTCP* was constructed on the pCAMBIA3301 vector. The *GmTCPpro:GUS* construct showed that *GmTCP* was expressed throughout root development and nodulation. At the stage of nodulation, the *GmTCPpro:GUS* construct showed strong GUS expression throughout the nodule. In addition to its expression in the nodulation cell line, *GmTCP* was also expressed in the roots, especially in the root tips (Figure 5). Furthermore, the higher the nitrate concentration, the higher the GUS expression. These results, that *GmTCP* was expressed in roots and nodules, support that *GmTCP* was involved in the formation and regulation of nodulation under varying nitrate concentration. The promoter activity analysis showed as a primary result that the high nitrogen concentration might have more activity to active the promoter function than lower nitrogen concentration at 5 d post inoculation with *rhizobium*.

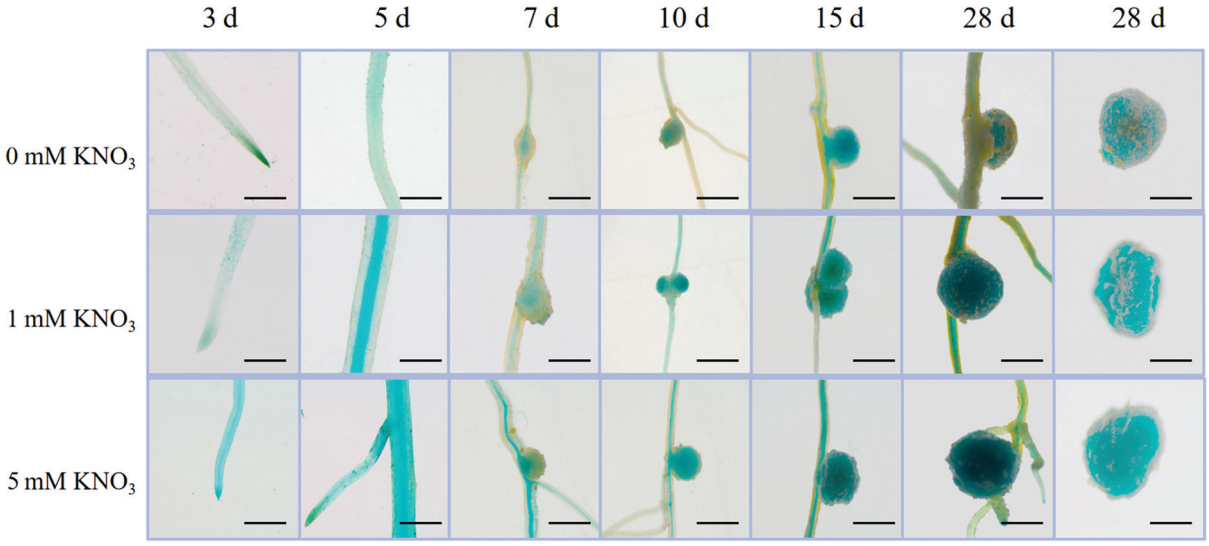


Figure 5. After inoculation with rhizobia, at different nitrate concentration expressions and localisation patterns of *GmTCPpro::GUS* during nodule and lateral root formation.

3. Discussion

In this study, we identified the candidate genes under soybean nodule character response to nitrogen concentration by integrating QTL mapping, RNA-seq data and single nucleotide polymorphism analysis, and we characterised the function of two candidate genes.

AtTCP20, the *Arabidopsis* homologous gene of *GmTCP*, has been known to regulate genes related to bud, embryo, and flower development. In particular, *AtTCP20* can bind to 100 or more nitrate-induced genes, being a crucial factor in the nitrate signalling pathway [17,25,26]. In this study, there was found a significant difference in primary root length

between the *AtTCP20* mutant and wild type under nitrogen-free conditions, while under high concentrations of nitrate (5 mM KNO_3), no significant difference in primary root length was observed between the mutant and the wild type. Additionally, at low nitrate concentrations, the primary root length of *AtTCP20* mutants and wild types was significantly different, but the difference decreased as the concentration of nitrate increased. These results demonstrate that *AtTCP20* responds specifically to nitrate starvation or imbalance and then regulates root growth. Previous research on single mutants of *AtTCP20* showed normal primary and lateral root growth on a homogenous nitrate medium, but lateral root growth was restricted on a heterogeneous nitrate medium. *AtTCP20* mutants still inhibited lateral root growth on a uniform ammonium salt medium, suggesting that *AtTCP20* may respond to other nitrogen sources such as ammonium salt [17]. *TCP20* binds to the GCCCR motif in the promoter of the mitotic cyclin gene *CYCB1;1*, regulating its expression. Under nitrogen-free conditions, the mRNA level of *CYCB1;1* was significantly increased in the whole root of *TCP20* mutants, indicating that nitrogen starvation induced the expression of the *TCP20* gene. The mutation of *TCP20* resulted in the improper regulation of *CYCB1;1*, thereby causing a premature exit from the cell cycle and eventually inhibiting root growth [26]. In summary, *AtTCP20* is an important factor that responds to nitrogen starvation or imbalance. We also studied the response of *AtNLP1* mutant plants to nitrate concentration in *Arabidopsis*. Under nitrogen-free and low nitrate conditions, the primary root length of *NLP1* mutants was significantly limited, whereas under high concentrations of nitrate the difference was not significant. *NLP6/7* regulates the expression of *CYCB1;1* in the same way as *TCP20*, ultimately leading to limited primary root development [26]. To date, there have been few reports on the roles of *NLP*, particularly on the molecular mechanism of *NLP* in plant growth and abiotic stress.

The *GmTCP* gene may also respond to nitrogen status. In order to verify the responses of *GmTCP* overexpression lines to different nitrate concentrations, this study constructed *GmTCP* overexpression and knockout transgenic lines via hairy root transformation and examined their nodulation responses after treatment with different concentrations of nitrate and the inoculation of *S. fredii* HH103. The results showed that in hairy roots transformed with empty vector, NN and NDW were significantly reduced under high nitrate concentration. Under low nitrate concentration, compared with nitrogen deficiency, NN and NDW decreased, but the effect was not significant. In *GmTCP* overexpressing hairy roots, NN and NDW decreased with the increase in nitrate concentration. However, compared with the wild type, the nodulation of *GmTCP* overexpressing hairy roots was significantly limited, and NN and NDW were significantly reduced. This experiment shows that *GmTCP* gene overexpression can inhibit nodulation at different nitrate concentrations; in *GmTCP* knockout hairy roots, NN and NDW decreased slightly with the increase in nitrate concentration, but the nodule phenotype of *GmTCP* knockout hairy roots did not differ much from that of hairy roots transformed with the empty vector under nitrogen-starved conditions, and there was no restriction on the nodule phenotype. Overall, the *GmTCP* gene is involved in the response to different nitrate concentrations. There are many studies on how plants absorb and transport nitrate, and the representative proteins in this process are nitrate transport proteins NRT1 and NRT2. Under N starvation, nitrate can induce the expression of nitrate transporter protein GmNRT2 in soybean roots, thereby increasing the absorption and transport of nitrate by roots [34]. Different nitrogen concentrations in the soil will affect the dry weight and number of nodules, and the activity of nitrogenase will decrease when the nitrogen source is abundant. The activity of nitrogenase is highest under nitrogen-free conditions. This indicates that the nitrogen-fixing activity of *rhizobia* is related to the concentration of nitrate. In hairy roots transformed with empty vector, the *GmTCP* gene may respond to high concentrations of nitrate, thereby regulating the expression of *NRT1* and *NRT2* genes and restricting the nodule phenotype. This hypothesis can explain the results obtained from our experiments on *GmTCP* gene overexpression and mutant strains.

This study also constructed *GmNLP* gene overexpression and knockout transgenic roots. The gene's response to different nitrate concentrations was tested by treating the

roots with various concentrations of nitrate and inoculating them with *S. fredii* HH103. The results showed that in *GmNLP*-overexpressing hairy roots, NN and NDW decreased as the nitrate concentration increased. However, compared with the wild type, overexpression significantly limited the nodulation and reduced both NN and NDW. In *GmNLP*-knockout hairy roots, the nodule phenotype slightly decreased with increasing nitrate concentration, but the nodule phenotype of *GmNLP*-knockout hairy roots did not differ significantly from that of empty vector-transformed hairy roots under nitrogen starvation conditions. In summary, the *GmNLP* gene is involved in responding to different nitrate concentrations. In *L. japonicus*, *LjNLP4* and *LjNLP1* play critical roles in the response to nitrate. High nitrate concentrations can induce *LjNLP1*, which, in turn, can induce *LjNRT2.1* expression [35]. The expression of *MtNRT2.1* is impaired due to mutations in the homologous gene *MtNLP1*. Such *NLP1-NRT2.1* regulatory modules may be conserved in many plants [38]. *LjNLP1* may induce the expression of *LjNRT2.1* when nitrate is present in the culture medium. In turn, the expression of *LjNRT2.1* may promote nitrate uptake and transport, ultimately triggering the accumulation of *LjNLP4* in the nucleus. Then, *LjNLP4* may induce or suppress the expression of symbiosis genes to regulate nodule formation [35,39]. This may be a way to respond to nitrate and regulate symbiotic relationships. The tissue-specificity experiment results of *GmTCP* promoter expression showed that the *GmTCP* gene was strongly expressed throughout nodule formation, supporting the hypothesis that the *GmTCP* gene is involved in symbiotic nodule formation. Further experiments are needed to confirm this hypothesis.

In summary, *GmTCP* and *GmNLP* encode important regulatory factors that respond to nitrogen concentrations for soybean nodulation. Further analysis of these genes would be of great significance for better understanding the symbiotic mechanism between soybean and rhizobia, as well as for improving soybean's nitrogen fixation ability.

4. Materials and Methods

4.1. Candidate Gene Identification Based on RNA-Seq Data

DEGs (with logFC > 1 and FDR < 0.05) were extracted from the previous RNA-seq data [10] for soybean roots infected with the rhizobium of one wild type and two mutants. Specifically, DEGs were identified in chromosomes 12 and 19, and the chromosome substitution segments were further compared to screen the key root-nodulation-associated candidate genes [7].

RNA-seq was performed on the root samples obtained from the soybean cultivar Suinong 14 (SN14) inoculated with different *rhizobia* at various time points [40]. A total of 180 samples were taken from the roots of SN14 at various points (0.5, 3, 6, 12, and 24 h) after infection with rhizobia injected with HH103, ΩNopAA, or ΩRhcN. For each injection, 3 biological replicates of the plants were used. The TruSeq Stranded mRNA Library Prep Kit from Illumina was used to create PolyA$^+$ libraries, and the mRNA sequencing was carried out on the Illumina HiSeq 4000 PE 150 platform [7]. The transcriptome raw sequencing data have been submitted to the NCBI (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA854816, accessed on 15 July 2022) database as individual BioProjects: PRJNA854816.

FastQc and MultiQc were employed to detect the raw Fastq sequence files for the initial quality check. The low-quality reads were trimmed with Trim galore at the criterion of [-q 20] [-length 140] [-e 0.1] [-stringency 3]. Then, FastQc and MultiQc were used again for quality testing to screen qualified reads ($19.7 \leq$ M SEQS ≤ 32.7 after treatment). Indexes were built by hisat2 with the *Glycine max* genome annotation file Gmax_275_Wm82.a2.v1.gene_exons.gff3 (http://www.phytozome.net/, accessed on 15 July 2022) [41].

4.2. Strains, Plant Materials and Growth Conditions

The *E. coli* DH5α cells used were grown at 37 °C in Luria–Bertani (LB) medium [42]. *Sinorhizobium fredii* HH103 were grown at 28 °C in tryptone yeast (TY) medium [43].

All *Arabidopsis thaliana* plants were of Columbia ecotype. The *tcp20* (SALK_041906), *nlp1*(SALK_125809) mutants were obtained from ABRC. The *tcp20* and *nlp1* mutants were produced by selfing the F1 plants and selecting F2 homozygous progeny by PCR analysis.

For primary root length, *Arabidopsis* seedlings were grown on 5mM KNO_3 plates for 6–7 days and then were transferred to different nitrate concentration plates or N-free plates for three days. The medium used in this experiment was N-free 1/2MS medium.

4.3. Plasmid Construct

The CDS fragment of *GmTCP* and *GmNLP* was constructed into the vector Fu28-GFP, and then LR recombination reactions were used to construct the overexpression vector pSOY1-*GmTCP*-GFP and pSOY1-*GmNLP*-GFP (Supplemental Table S1, Figures S1−S10). The online *Crispr-P* tool was used to design the sgRNA primer for these genes, and the designed sqRNA was attached to the pGES201 vectors. The designed sgRNA sequence and the experimental results of the pGES201 vector construction process are shown in Supplemental Table S1, Figures S6–S8.

This fragment, including a 3.0-kb sequence directly upstream of the initiation codon of *GmTCP*, was cloned into pCAMBIA3301-GFP. For promoter-GUS analysis using hairy roots on soybean, a 3.0 kb fragment of the *GmTCP* promoter region was amplified by PCR from WT genomic DNA and cloned upstream of the GUS gene in the pCAMBIA3301-GUS vector (Supplemental Table S1, Figures S9 and S10).

4.4. Agrobacterium Rhizogenes-Mediated Transformation of Soybean to Study Root Biology

The sterilised seeds were placed into wet vermiculite at a depth of 1–2 cm which was then transferred into a growth chamber or greenhouse at 28 °C. Meanwhile, *Agrobacterium* strain(s) harbouring the desired construct(s) (from glycerol stock) were inoculated onto the surface of LB plates containing the appropriate antibiotics and incubated at 28 °C for 2 days, then a single colony was transferred onto a fresh plate. The fresh bacterial culture from the plate was suspended in 1 mL of liquid LB medium containing 15% (*v*/*v*) glycerol, and then each 200 μL of the suspension culture liquid was spread onto the surface of four LB plates containing the appropriate antibiotics and incubated at 28 °C overnight. Bacteria from the plates were collected and then inoculated into the healthy plantlets with unfolded green cotyledons by stabbing at the cotyledonary node and/or at the hypocotyl proximal to the cotyledon. The plants were then placed under 12 h light/12 h dark at 28 °C/25 °C in a growth chamber (humid chamber only) for 2–3 weeks until the hairy roots grew to approximately 5–10 cm in length and were long enough to support growth of the plant. During this period, the plants were supplied with sterile B&D solution containing 0~2 mM KNO_3 as a nitrogen source. When the hairy roots were approximately 5–10 cm in length, the primary root was removed by cutting the hypocotyl ~1 cm under the wounding site where the hairy roots were formed. The six plants were selected from each overexpressed line and used for RNA extraction and subsequent evaluation of *GmTCP*/*GmNLP* expression through qRT-PCR. The highest *GmTCP*/*GmNLP* expression levels were then used for further phenotype analysis. These lines moved into new pots after 5 days' infection with *S. fredii* HH103. After 28 days, the two phenotypes of the nodules in all soybean materials were investigated [44].

4.5. Biological Nitrogen Fixation Traits

At 28 days of *S. fredii* HH103 inoculation, the two phenotypes of the nodules in all soybean materials were investigated. The NN and NDW were evaluated. To determine the NDW, the nodules were placed in an oven at 65 °C for 36 hr.

Root samples were harvested at 28 days post-inoculation with *S. fredii* HH103, and wild-type plants were used as the control. In overexpression lines, the total RNA was isolated from the roots using the TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), and then each RNA sample was converted into cDNA using the HiScript®II Q RT SuperMix (Vazyme Biotech Co., Nanjing, China). The qRT-PCR was performed with the TB Green

Fast qPCR Mix (TaKaRa, Dalian, China) on the Roche LightCycler 480 II System (Roche, Basel, Switzerland). The qRT-PCR program was as follows: denaturation at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s, 58 °C for 20 s and 72 °C for 30 s. All RNA extractions were performed in three biological replicates, and each cDNA sample was analysed three times. Specific primers were designed to the Williams 82 gene sequences retrieved from the Phytozome website [45]. *GmActin* was used as the reference gene to calibrate the transcript abundance values among different samples. Ct values were calculated by the Roche LightCycler 480 II software. In knockout transgenic lines, genome DNA was extracted from the root, the corresponding gene region was amplified with a specific primer, and then sequencing was conducted to see if there was a mutation in the region, and the NN and NDM were measured for the mutated roots.

4.6. GUS Staining

Root tissues were gently pre-fixed in 90% acetone (pre-cooled at −20 °C) under vacuum infiltration for 5 min, and then placed on ice for 20 min. Next, they were rinsed with cold water for 5 min and vacuum infiltrated for 5 min on ice with a pre-cooled staining solution. After that, they were incubated at 37 °C for 12 h. Samples were changed through 30 min steps of 20% ethanol, 30% ethanol, 50% ethanol, and FAA. The FAA was then removed, and 70% ethanol was added and the solution was left for two h, during which ethanol was replaced with fresh twice. Finally, the samples were stored at 4 °C before observation with a dissecting microscope [46].

5. Conclusions

In this study, we identified two candidate genes, *GmTCP* and *GmNLP*, that regulate soybean nodule phenotype in response to nitrogen. Homologous genes of these two candidates have been shown to be important factors in regulating root development in response to nitrogen concentration in *Arabidopsis thaliana*. Overexpression and gene knockout in *GmTCP* and *GmNLP* in soybean resulted in different responses to different nitrate concentrations, leading to different nodule phenotypes. These results highlight the involvement of *GmTCP* and *GmNLP* in the regulation of soybean nodulation under nitrogen concentration. The role of these genes in the symbiotic mechanism between soybean and rhizobia is of great importance for improving nitrogen fixation efficiency, and therefore further analysis of their functions may provide valuable insights into the mechanism of nodule formation in response to nitrate.

Supplementary Materials: The supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24097750/s1.

Author Contributions: Conceptualisation, D.X. and Q.C.; methodology, Y.K., J.W. (Jinhui Wang), H.N. and C.M.; software, J.W. (Jing Wang); validation, C.J., M.J. and J.C.; formal analysis, Y.K.; investigation, M.J., J.C. and H.L.; resources, J.W. (Jinhui Wang); data curation, M.Y.; writing—original draft preparation, Y.K. and J.W. (Jinhui Wang); writing—review and editing, D.X.; visualisation, Y.K. and Y.P.; supervision, D.X. and Q.C.; project administration, M.Y.; funding acquisition, D.X., Q.C. and J.W. (Jinhui Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (Grant numbers: 32070274, 31771882, 32072014 and U20A2027).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Dawei Xin for the technical help.

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

References

1. Nicolás, M.; Noelia, C.; Gabriel, G.; Carolina, A.; Diana, L.; Damián, M.; Carlos, U. Soybean yield, protein content and oil quality in response to interaction of arbuscular mycorrhizal fungi and native microbial populations from mono- and rotation-cropped soils. *Appl. Soil Ecol.* **2020**, *152*, 103575.
2. Jinxia, Y.; Can, W.; Fengrong, Z.; Li, C.; Zhengxi, S.; Yupeng, C.; Yuanqing, L.; Junwen, L.; Yongliang, W.; Yanyan, C.; et al. A nitrogen fixing symbiosis-specific pathway required for legume flowering. *Sci. Adv.* **2023**, *9*, eade1150.
3. Schwember, A.R.; Schulze, J.; Del Pozo, A.; Cabeza, R.A. Regulation of symbiotic nitrogen fixation in legume root nodules. *Plants* **2019**, *8*, 333. [CrossRef] [PubMed]
4. Massawe, P.; Mtei, K.; Munishi, L.; Ndakidemi, P. Improving soil fertility and crops yield through maize-legumes (*Common bean and Dolichos lablab*) intercropping systems. *J. Agric. Sci.* **2016**, *8*, 148–163. [CrossRef]
5. Jinggui, F.; Xudong, Z.; Chen, W.; Lingfei, S. Applications of DNA technologies in agriculture. *Curr. Genom.* **2016**, *17*, 379–386.
6. Ayaz, A.; Saqib, S.; Huang, H.; Zaman, W.; Lü, S.; Zhao, H. Genome-wide comparative analysis of long-chain acyl-CoA synthetases (LACSs) gene family: A focus on identification, evolution and expression profiling related to lipid synthesis. *Plant Physiol. Biochem.* **2021**, *161*, 1–11. [CrossRef]
7. Shi, Y.; Zhang, Z.; Wen, Y.; Yu, G.; Zou, J.; Huang, S.; Wang, J.; Zhu, J.; Wang, J.; Chen, L.; et al. RNA sequencing-associated study identifies *GmDRR1* as positively regulating the establishment of symbiosis in soybean. *Mol. Plant-Microbe Interact.* **2020**, *33*, 798–807. [CrossRef]
8. Ayaz, A.; Huang, H.; Zheng, M.; Zaman, W.; Li, D.; Saqib, S.; Zhao, H.; Lü, S. Molecular cloning and functional analysis of *GmLACS2-3* reveals its involvement in cutin and suberin biosynthesis along with abiotic stress tolerance. *Int. Journal. Mol. Sci.* **2021**, *22*, 9175. [CrossRef]
9. Liu, Y.; Sun, M.; Chen, Q.; Xin, D.; Sun, X. Mapping quantitative trait loci related to nodule number in soybean (*Glycine max*(L.) Merr.) in response to the Sinorhizobium (Ensifer) fredii HH103 NopT typeIII effector. *J. Plant. Interact.* **2021**, *16*, 126–135. [CrossRef]
10. Hejia, N.; Yang, P.; Jinhui, W.; Jing, W.; Yantong, Y.; Tingting, F.; Zikun, Z.; Jialin, Z.; Xipeng, P.; Zhuoling, C.; et al. Mapping of Quantitative Trait Loci Underlying Nodule Traits in Soybean (*Glycine max* (L.) Merr.) and Identification of Genes. Whose Expression Is. Affected by the Sinorhizobium fredii HH103 Effector Proteins NopL and NopT. *Agronomy* **2022**, *12*, 946.
11. Margaret, I.; Becker, A.; Blom, J.; Bonilla, I.; Goesmann, A.; Göttfert, M.; Lloret, J.; Mittard-Runte, V.; Rückert, C.; Ruiz-Sainz, Z.; et al. Symbiotic properties and first analyses of the genomic sequence of the fast growing model strain Sinorhizobium fredii HH103 nodulating soybean. *J. Biotechnol.* **2011**, *155*, 11–19. [CrossRef]
12. Manassero, N.G.U.; Viola, I.L.; Welchen, E.; Gonzalez, D.H. TCP transcription factors: Architectures of plant form. *Biomol. Concepts* **2013**, *4*, 111–127. [CrossRef] [PubMed]
13. Cubas, P.; Lauter, N.; Doebley, J.; Coen, E. The TCP domain: A motif found in proteins regulating plant growth and development. *Plant J.* **1999**, *18*, 215–222. [CrossRef]
14. Horn, S.; Pabón-mora, N.; Theuß, V.S.; Busch, A.; Zachgo, S. Analysis of the CYC/TB1 class of TCP transcription factors in basal angiosperms and magnoliids. *Plant J.* **2015**, *81*, 559–571. [CrossRef] [PubMed]
15. Danisman, S.; van Dijk, A.D.; Bimbo, A.; van der Wal, F.; Hennig, L.; de Folter, S.; Angenent, C.; Immink, R.G. Analysis of functional redundancies within the *Arabidopsis* TCP transcription factor family. *J. Exp. Bot.* **2013**, *64*, 5673–5685. [CrossRef]
16. Shutian, L. The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. *Plant Signal. Behav.* **2015**, *10*, e1044192.
17. Guan, P.; Wang, R.; Nacry, P.; Breton, G.; Kayd, S.A.; Pruneda-Paz, J.L.; Davani, A.; Crawford, N.M. Nitrate foraging by *Arabidopsis* roots is mediated by the transcription factor TCP20 through the systemic signaling pathway. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15267–15272. [CrossRef]
18. Lei, L.; Wenrui, Z.; Yimin, A.; Binghao, D.; Dan, W.; Changhong, G. Genome-wide analysis of the TCP transcription factor genes in five legume genomes and their response to salt and drought stresses. *Funct. Integr. Genom.* **2020**, *20*, 537–550.
19. Liu, Y.; Guan, X.; Liu, S.; Yang, M.; Ren, J.; Guo, M.; Huang, Z.; Zhang, Y. Genome-Wide Identification and Analysis of TCP Transcription Factors Involved in the Formation of Leafy Head in Chinese Cabbage. *Int. J. Mol. Sci.* **2018**, *19*, 847. [CrossRef]
20. Lopez, J.A.; Sun, Y.; Blair, P.B.; Mukhtar, M.S. TCP three-way handshake: Linking developmental processes with plant immunity. *Trends Plant Sci.* **2015**, *20*, 238–245. [CrossRef]
21. Takeda, T.; Amano, K.; Ohto, M.A.; Nakamura, K.; Sato, S.; Kato, T.; Tabata, S.; Ueguchi, C. RNA Interference of the *Arabidopsis* Putative Transcription Factor *TCP16* Gene Results in Abortion of Early Pollen Development. *Plant Mol. Biol.* **2006**, *61*, 165–177. [CrossRef] [PubMed]
22. Huang, T.; Irish, V.F. Temporal Control of Plant Organ Growth by TCP Transcription Factors. *Curr. Biol.* **2015**, *25*, 1765–1770. [CrossRef]
23. Palatnik, J.F.; Allen, E.; Wu, X.; Schommer, C.; Schwab, R.; Carrington, J.C.; Weigel, D. Control of leaf morphogenesis by micro RNAs. *Nature* **2003**, *425*, 257–263. [CrossRef] [PubMed]
24. Danisman, S. TCP Transcription Factors at the Interface between Environmental Challenges and the Plant's Growth Responses. *Front. Plant Sci.* **2016**, *7*, 1930–1942. [CrossRef]

25. Herve, C.; Dabos, P.; Bardet, C.; Jauneau, A.; Auriac, M.C.; Ramboer, A.; Lacout, F.; Tremousaygue, D. In Vivo Interference with AtTCP20 Function Induces Severe Plant Growth Alterations and Deregulates the Expression of Many Genes Important for Development. *Am. Soc. Plant. Biol.* **2009**, *3*, 1462–1477. [CrossRef]
26. Guan, P.; Ripoll, J.J.; Wang, R.; Vuong, L.; Bailey-Steinitz, L.J.; Ye, D.; Crawford, N.M. Interacting TCP and NLP transcription factors control plant responses to nitrate availability. *Plant Biol.* **2017**, *114*, 2419–2424. [CrossRef]
27. Feng, Z.J.; Xu, S.C.; Liu, N.; Zhang, G.W.; Hu, Q.Z.; Gong, Y.M. Soybean TCP transcription factors: Evolution, classification, protein interaction and stress and hormone responsiveness. *Plant Physiol. Biochem.* **2018**, *127*, 129–142. [CrossRef] [PubMed]
28. Davière, J.M.; Wild, M.; Regnault, T.; Baumberger, N.; Eisler, H.; Genschik, P.; Achard, P. Class I TCP-DELLA Interactions in Inflorescence Shoot Apex Determine Plant Height. *Curr. Biol.* **2014**, *24*, 1923–1928.
29. Li, Y.; Hou, Z.; Li, W.; Li, H.; Lu, S.; Gan, Z.; Du, H.; Li, H.; Zhang, Y.; Kong, F.; et al. The legume-specific transcription factor E1 controls leaf morphology in soybean. *BMC Plant Biol.* **2021**, *21*, 1–12. [CrossRef]
30. Fan, S.; Zhang, Z.; Song, Y.; Zhang, J.; Wang, P. CRISPR/Cas9-mediated targeted mutagenesis of GmTCP19L increasing susceptibility to *Phytophthora sojae* in soybean. *PLoS ONE* **2022**, *17*, e0267502. [CrossRef] [PubMed]
31. Chardin, C.; Girin, T.; Roudier, F.; Meyer, C.; Krapp, A. The plant RWP-RK transcription factors: Key regulators of nitrogen responses and of gametophyte development. *J. Exp. Bot.* **2014**, *65*, 5577–5587. [CrossRef]
32. Sumimoto, H.; Kamakura, S.; Ito, T. Structure and function of the PB1 domain, a protein interaction module conserved in animals, fungi, amoebas, and plants. *Sci. STKE* **2007**, *401*, re6. [CrossRef] [PubMed]
33. Konishi, M.; Yanagisawa, S. The role of protein-protein interactions mediated by the PB1 domain of NLP transcription factors in nitrate-inducible gene expression. *BMC Plant Biol.* **2019**, *19*, 90. [CrossRef]
34. Orsel, M.; Krapp, A.; Daniel-Vedele, F. Analysis of the NRT2 nitrate transporter family in *Arabidopsis*. Structure and gene expression. *Plant Physiol.* **2002**, *129*, 886–896. [CrossRef] [PubMed]
35. Nishida, H.; Nosaki, S.; Suzuki, T.; Ito, M.; Miyakawa, T.; Nomoto, M.; Tada, Y.; Miura, K.; Tanokura, M.; Kawaguchi, M.; et al. Different DNA-binding specificities of NLP and NIN transcription factors underlie nitrate-induced control of root nodulation. *Plant Cell* **2021**, *33*, 2340–2359. [CrossRef] [PubMed]
36. Wang, L.; Sun, Z.; Su, C.; Wang, Y.; Yan, Q.; Chen, J.; Thomas, O.; Li, X. A GmNINa-miR172c-NNC1 regulatory network coordinates the nodulation and autoregulation of nodulation pathways in soybean. *Mol. Plant* **2019**, *12*, 1211–1226. [CrossRef] [PubMed]
37. Jiang, S.; Jardinaud, M.F.; Gao, J.; Pecrix, Y.; Wen, J.; Mysore, K.; Murray, J.D. NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. *Science* **2021**, *374*, 625–628. [CrossRef] [PubMed]
38. Lin, J.S.; Li, X.; Luo, Z.; Mysore, K.S.; Wen, J.; Xie, F. NIN interacts with NLPs to mediate nitrate inhibition of nodulation in *Medicago Truncatula*. *Nat. Plants* **2018**, *4*, 942–952. [CrossRef]
39. Fumika, M.; Momoyo, I.; Shohei, N.; Hanna, N.; Masahiro, W.; Takamasa, S.; Kenji, M.; Masayoshi, K.; Takuya, S. Nitrate transport via NRT2.1 mediates NIN-LIKE PROTEIN-dependent suppression of root nodulation in Lotus japonicus. *Plant Cell* **2022**, *34*, 1844–1862.
40. Baker, M. Similarities between legume–Rhizobium communication and steroid-mediated intercellular communication in vertebrates. *Can. J. Microbiol.* **1992**, *38*, 541–547. [CrossRef] [PubMed]
41. Purcell, L.C.; Salmeron, M.; Ashlock, L. Soybean growth and development. *Ark. Soybean Prod. Handb.* **2014**, *197*, 1–8.
42. Green, R.; Rogers, E.J. Transformation of chemically competent *E. coli. Methods Enzym.* **2013**, *529*, 329–336.
43. Nagesh, K.T.; Ambuj, S.; Karttik, C.B.; Rao, P.V.L. Optimization of culture medium for production of recombinant dengue protein in *Escherichia coli*. *Ind. Biotechnol.* **2009**, *5*, 179–183.
44. Attila, K.; Dongxue, L.; Arief, I.; Nguyen, C.D.; Sureeporn, N.; Mark, K.; Peter, M.G. *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat. Protoc.* **2007**, *2*, 948–952.
45. Zhu, J.; Wang, J.; Li, Q.; Wang, J.; Liu, Y.; Li, J.; Chen, L.; Shi, Y.; Li, S.; Zhang, Y.; et al. QTL analysis of nodule traits and the identification of loci interacting with the type III secretion system in soybean. *Mol. Genet. Genom.* **2019**, *294*, 1049–1058. [CrossRef] [PubMed]
46. Lauren, K.D.; Emily, O.; Braybrook, S.A. Fake news blues: A GUS staining protocol to reduce false-negative data. *Am. Soc. Plant Biol.* **2022**, *6*, e367.

International Journal of
Molecular Sciences

MDPI

Article

First Draft Genome Assembly of Root-Lesion Nematode *Pratylenchus scribneri* Generated Using Long-Read Sequencing

Deepika Arora [1,†], Alvaro G. Hernandez [2], Kimberly K. O. Walden [2], Christopher J. Fields [2] and Guiping Yan [1,*]

[1] Department of Plant Pathology, North Dakota State University, Fargo, ND 58105, USA; deepika.arora@ndus.edu or darora@genemarkersllc.com
[2] Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA; aghernan@illinois.edu (A.G.H.); kosterbu@illinois.edu (K.K.O.W.); cjfields@illinois.edu (C.J.F.)
* Correspondence: guiping.yan@ndsu.edu
† Current address: Genemarkers, Kalamazoo, MI 49007, USA.

Abstract: Root-lesion nematodes (genus *Pratylenchus*) belong to a diverse group of plant-parasitic nematodes (PPN) with a worldwide distribution. Despite being an economically important PPN group of more than 100 species, genome information related to *Pratylenchus* genus is scarcely available. Here, we report the draft genome assembly of *Pratylenchus scribneri* generated on the PacBio Sequel IIe System using the ultra-low DNA input HiFi sequencing workflow. The final assembly created using 500 nematodes consisted of 276 decontaminated contigs, with an average contig N50 of 1.72 Mb and an assembled draft genome size of 227.24 Mb consisting of 51,146 predicted protein sequences. The benchmarking universal single-copy ortholog (BUSCO) analysis with 3131 nematode BUSCO groups indicated that 65.4% of the BUSCOs were complete, whereas 24.0%, 41.4%, and 1.8% were single-copy, duplicated, and fragmented, respectively, and 32.8% were missing. The outputs from GenomeScope2 and Smudgeplots converged towards a diploid genome for *P. scribneri*. The data provided here will facilitate future studies on host plant-nematode interactions and crop protection at the molecular level.

Keywords: root-lesion nematode; genome; ultra-low DNA input; contigs; sequencing; BUSCO; diploid

Citation: Arora, D.; Hernandez, A.G.; Walden, K.K.O.; Fields, C.J.; Yan, G. First Draft Genome Assembly of Root-Lesion Nematode *Pratylenchus scribneri* Generated Using Long-Read Sequencing. *Int. J. Mol. Sci.* **2023**, *24*, 7311. https://doi.org/10.3390/ijms24087311

Academic Editor: Manickam Sugumaran

Received: 17 March 2023
Revised: 6 April 2023
Accepted: 13 April 2023
Published: 15 April 2023

1. Introduction

Root-lesion nematodes (RLN) are recognized as the top three plant-parasitic nematode groups of the world and lie behind root-knot (*Meloidogyne*) and cyst nematodes (*Heterodera*) in terms of economic importance [1]. Unlike sedentary endoparasites (*Meloidogyne* and *Heterodera*), nematodes in the genus *Pratylenchus* do not establish permanent feeding sites (giant cells or syncytia) and move freely between soil and roots throughout their lifecycle. This feeding behavior results in the development of characteristic reddish-brown necrotic lesions on root tissue, which construct pathways for secondary invaders, such as bacteria, fungi, etc., resulting in severe yield losses [1,2]. *Pratylenchus* spp. have a cosmopolitan distribution and are known to parasitize more than 400 host plant species. They have a wide host range and impact crops of major economic importance, such as potato, wheat, corn, and soybean [3]. The use of chemical nematicides is a common management strategy for controlling RLNs, but their toxic effects on non-target organisms and environmental health call for alternative control strategies based on novel gene targets [4].

The root-lesion nematode, *Pratylenchus scribneri* Steiner 1943, is found commonly associated with the major crops grown in the northern Great Plains region of North America [5–7]. Considering the importance of *P. scribneri*, several molecular detection methods for identifying and quantifying this nematode in soil and roots have been developed to enhance the management decisions and practices [8–10]. More recently, it has been reported as an important nematode pest infecting corn, soybean, wheat, and tomato in China [11–14].

In Nigeria, yield losses in corn have been reported to be between 26 to 37% based on the population densities of *P. scribneri* [15].

Despite the ubiquity and significant importance of *Pratylenchus* species in agriculture, the availability of genome information on these species is limited as compared to the sedentary PPN. So far, the genome sequence of only one *Pratylenchus* species, *P. coffeae*, is available, and the published genome size of 19.7 Mb is the smallest among the metazoans [16]. This raises questions regarding its entirety, especially due to the lack of a complete physical map. Additionally, the genomes *of Pratylenchus* spp. are expected to have complex ploidy similar to root-knot nematodes (*Meloidogyne* spp.) [17], but this is still an unexplored area. Ploidy determination could provide important insights into the developmental and evolutionary biology of *Pratylenchus* spp. Moreover, the genome data for different *Pratylenchus* spp. are needed to help better understand the parasitism mechanism of the nematodes in this genus. Jointly, this information could help reveal novel gene targets for employing specific control strategies for *P. scribneri* and other related species. In light of this, the main objective of the current study is to generate a contiguous and relatively complete annotated genome of one of the important root-lesion nematodes, *P. scribneri*, and make it publicly available. We also explored two different genomic DNA isolation methods to prepare sequencing quality DNA for *P. scribneri*, and the advantages of one over the other have been presented.

2. Results

2.1. Library Preparation and Assembly

The raw PacBio HiFi sequencing yields and assembly metrics for the libraries prepared using two different DNA isolation methods are presented in detail in Table 1. The library resulting from DNA extracted using 500 handpicked nematodes (Method 2) had greater sequencing yield (19.3 Gb) as compared to Method 1 (14.9 Gb), which involved DNA isolation directly from nematode suspension. Fewer contigs were assembled in Method 2. The contig N50 increased from 222 kb in the first method to 1.70 Mb in the second method. The average GC% was 37.6% and 32.8% for sequencing libraries prepared using Method 1 and Method 2, respectively.

Table 1. Sequencing yield and assembly metrics for *Pratylenchus scribneri* raw hifiasm libraries constructed using two different DNA isolation methods.

Sequencing Yield	Method 1 [a]	Method 2 [b]
Total Number of Bases (bp)	14,985,187,925	19,351,261,593
Total Number of Reads	1,665,292	2,749,471
Maximum Read Length (bp)	42,009	43,956
Minimum Read Length (bp)	295	68
Mean Read Length (bp)	8999	7038
Assembly Metrics		
Number of Contigs	9447	492
Total number of bp	714,324,379	361,758,083
Shortest (bp)	5119	5897
Longest (bp)	6,863,581	6,221,503
Average length (bp)	75,613	735,280
Median (bp)	30,718	334,760
Average GC%	37.60%	32.80%
Non-ACGT bases	0	0
Contig N50 (bp)	222,427	1,701,422

[a] 8000 to 10,000 *P. scribneri* directly from nematode suspension harvested from carrot cultures used for DNA extraction using the DNAeasy Plant Mini Kit. [b] 500 *P. scribneri* juveniles and adults handpicked and DNA extracted using the MagAttract HMW DNA kit.

2.2. Assembly Decontamination and Completeness Analysis

The Blobplot in Figure 1 represents the preliminary contig taxonomic assignments, contig read coverage, contig lengths, and contig GC% obtained for the two DNA extraction

methods. Figure 1a,b show the Blobplot analyses from the "core" nematode genome contigs as red points, and that the library constructed using Method 2 showed >30× *P. scribneri* read coverage than the library constructed using Method 1. In Method 2, out of the 39 contigs classified into phyla other than Nematoda, only one contig was determined to be a bacterial contaminant sequence which was removed prior to further genome analysis. All others had very weak hits to a variety of phyla tested or no hits at all. The assembly obtained from Method 2 was used for further analyses and subject to haplotig purging. The final purged decontaminated assembly contained 276 contigs summing to ~227 Mb. The BUSCO scores for raw, purged, and decontaminated assembly are provided in Table 2. The final decontaminated assembly had a completeness score of 65.4%, with 24.0% being single copy, 41.4% duplicated, 1.8% fragmented, and 32.8% missing BUSCOs out of the total 3131 orthologs used.

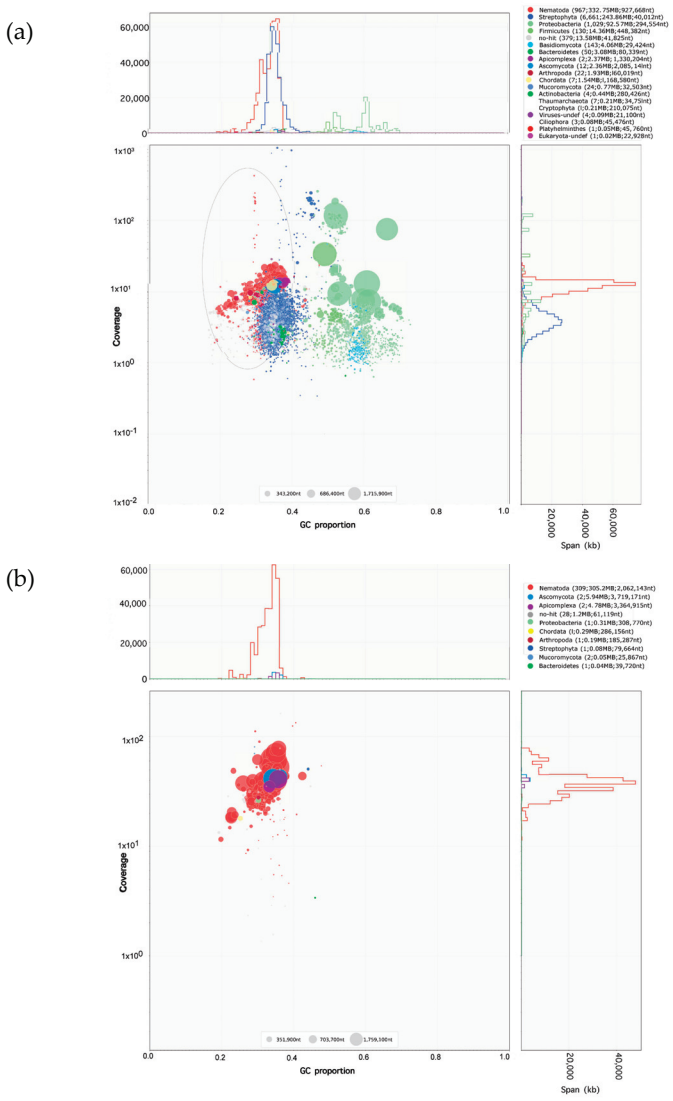


Figure 1. Blobplot contaminant analysis of *Pratylenchus scribneri* assembly obtained using two DNA extraction methods, before removing contaminants. The contig length is represented by circles proportionally scaled and colored by taxonomic annotation based on BLAST similarity search results. Nematode genome contigs are represented by red circles for: (**a**) Method 1 (library prepared from 8000 to 10,000 mixed-stage *P. scribneri* in nematode suspension obtained from carrot culture); (**b**) Method 2 (library prepared from 500 hand-picked *P. scribneri* juveniles and adults). Contigs are placed based on the GC proportion (*X*-axis) and the coverage of reads (*Y*-axis). The legend box on the top right of each Blobplot provides a color-coded description of the different taxonomic groups identified.

Table 2. Genome summary statistics and benchmarking universal single-copy ortholog (BUSCO) score estimation for raw, purged, and decontaminated *Pratylenchus scribneri* genome assemblies obtained using DNA extraction Method 2.

BUSCO Category	Raw Assembly	Purged Assembly [a]	Decontaminated Assembly [b]
Complete (C)	2084 [c] (66.5%) [d]	2048 (65.4%)	2047 (65.4%)
Complete and single copy (S)	530 (16.9%)	753 (24.0%)	752 (24.0%)
Complete and duplicated (D)	1554 (49.6%)	1295 (41.4%)	1296 (41.4%)
Fragmented (F)	36 (1.1%)	57 (1.8%)	56 (1.8%)
Missing (M)	1011 (32.3%)	1026 (32.8%)	1025 (32.7%)
Total BUSCO groups searched	3131	3131	3131

[a] Assembly with haplotypic and duplicate content removed from the primary set of contigs with purge_dups v.1.2.5 using default parameters. [b] Purged assembly with a single 40 kb contig determined to be a bacterial contaminant sequence removed. [c] Number of single-copy orthologues represented in each category out of the total BUSCO groups searched in Nematoda lineage (Nematoda_odb10). [d] Percentage of single-copy orthologues versus total BUSCO groups searched in Nematoda lineage (3131) for each category.

2.3. Genome Ploidy Estimation and Structural Annotation

The *P. scribneri* genome ploidy estimations were made using the final assembly constructed using Method 2. After manually setting the estimated monoploid peak from GenomeScope and Smudgeplot, the models and output converged on ploidy = 2. The presence of two major peaks in GenomeScope output and coverage and distribution of k-mer pairs into two bright smudges in Smudgeplot pointed towards a diploid genome for *P. scribneri* (Figure 2).

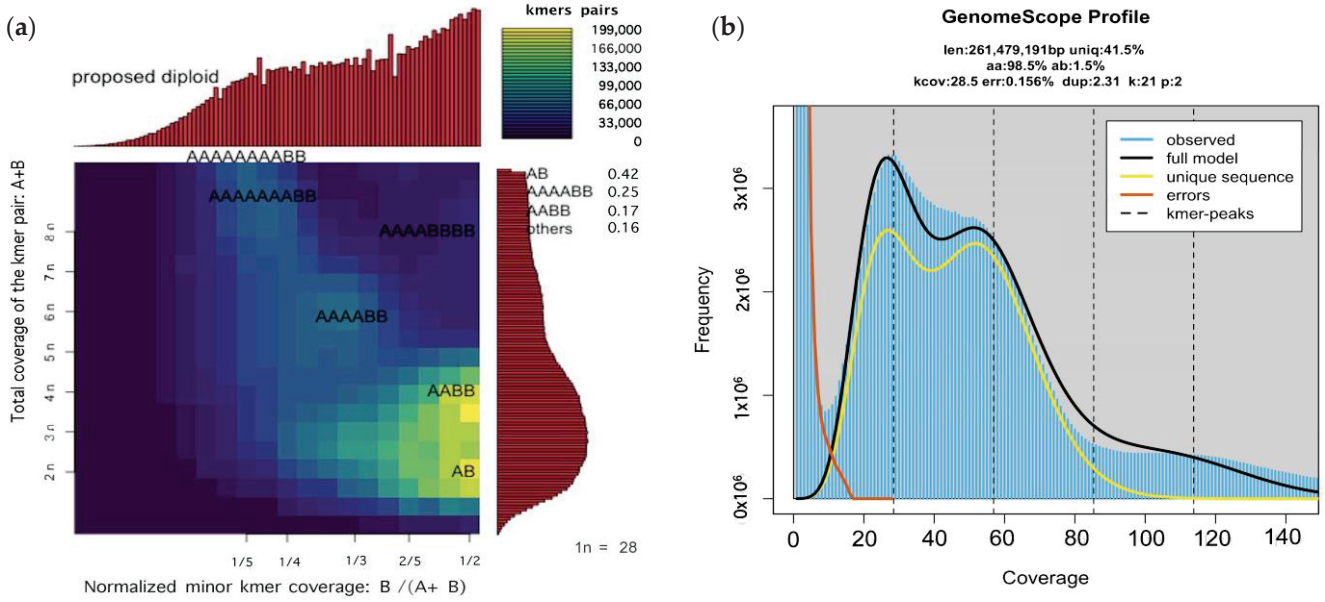


Figure 2. Genome ploidy estimation of *Pratylenchus scribneri* genome assembly: (**a**) Smudgeplots showing the coverage and distribution of k-mer pairs that fit to diploid genome model; (**b**) GenomeScope k-mer (k = 21) profile plot of the *P. scribneri* showing the fit of the GenomeScope model (black) to the observed k-mer frequencies (blue) with two major peaks.

For predicting proteins, the BUSCO analysis was run in "long mode" on the final decontaminated assembly. The predicted proteins (single copy (23.5%)/multi-copy (42.7%)/fragmented (1.4%)) were then merged together into 3966 protein sequences as additional

protein input for BRAKER annotation. The structural annotation ran using BRAKER and TSERBA with RNA-seq and protein evidence resulted in a total of 51,146 predicted protein sequences. The BUSCO scores for the TSEBRA-combined BRAKER2 predicted protein had 72% complete BUSCOs (single copy: 12.0%, duplicate: 60.0%, fragmented: 1.2%) and 26.8% missing BUSCOs.

3. Discussion

Recently, long-read PacBio HiFi sequencing has emerged as a powerful tool to generate high-quality de novo genome assemblies. However, the technique has not been explored to its full potential for microorganisms such as nematodes, due to the relatively high DNA input requiring a larger number of nematodes. Here, we present, for the first time, the de novo draft genome assembly of the root-lesion nematode *Pratylenchus scribneri* constructed using 5 ng of input DNA from 500 nematodes as an example. In this study, the libraries prepared from two different DNA extraction methods were sequenced using the PacBio ultra-low DNA input protocol which uses as low as 5 ng of DNA as input and therefore, significantly fewer nematodes. The library prepared using DNA extracted from 500 handpicked nematodes (Method 2) was much cleaner (fewer contaminants) as compared to the library prepared from DNA extracted directly from nematode suspension (Method 1). This could be attributed to the presence of carrot tissue and other organisms in the library prepared using the nematode suspension obtained directly from carrot cultures. As a result, the number of assembled contigs specific to different taxonomical groups were reduced from 9.5 k in Method 1 to 492 in Method 2. It is difficult to get the sequencing library from nematodes grown on carrots to be completely free of any contaminants. For optimal results, the handpicking of nematodes required for Method 2 should be carried out in a laminar flow hood under sterile conditions, but Method 2 is also tedious due to the handpicking of 500 nematodes, which could be a limitation for preparing the nematode sample for DNA extraction.

There are only a few reports available on the de novo genome assembly construction of small organisms based on the ultra-low DNA input protocol. Kingan et al. [18] generated a high-quality de novo genome assembly using a single mosquito with the ultra-low DNA protocol. DNA isolation for this study was performed using the MagAttract HMW kit as used in Method 2 for the current study. However, no reports on de novo genome assembly of a plant-parasitic nematode using the ultra-low DNA input protocol are available thus far. The use of MagAttract HMW kit for genomic DNA isolation for whole genome sequencing has mostly been observed for bacterial pathogens and insects [18–20]. In the present study, the library prepared using 5 ng of input DNA from 500 active *P. scribneri* individuals (Method 2) was sequenced with a single PacBio SMRT cell 8 M chip on the Sequel IIe System. Nearly 2.7 million reads producing 19.3 Gb of raw data with mean read length of 7 kb were obtained and the genome size was predicted to be around 227 Mb. Similar raw sequencing yield results were obtained for the sand fly (*Phlebotomus papatasi*) sequenced on the Sequel IIe System using just 5 ng of input DNA following PacBio's ultralow DNA input protocol (https://www.pacb.com/wp-content/uploads/Korlach-PAG-2020-High-Quality-PacBio-Insect-Genome-from-5-ng-of-Input-DNA.pdf, accessed on 30 November 2022).

The BUSCO completeness scores for the final decontaminated assembly tested against 3131 orthologs were fairly average and the presence of more than 65% complete genes in our study are comparable with previous studies [21,22]. However, the gene duplication content of our assembly was higher, and could be related to the likely repetitive gene content of *P. scribneri* genome and usage of orthologs from different reference datasets in other studies on PPNs [19,22]. This in part may also explain the presence of fewer complete and duplicated BUSCOs in genome assemblies of other migratory endoparasitic nematodes such as *Radopholus similis* and *Bursaphelenchus xylophilus* [22,23]. Additionally, the genome assembly of *P. scribneri* generated in this study had a GC content of 32.8%, which is consistent with other PPNs sequenced on a PacBio RSII platform [24,25]. However, the completeness and quality of the set of 51,146 predicted proteins in *P. scribneri* assessed

using BUSCO showed 72% completeness. The BUSCO scores for predicted proteins were improved over the genome assembly; however, duplicated complete genes were still high (60%). Similar observations related to higher duplicated genes in the protein set of a root-knot nematode *Meloidogyne enterolobii* were noted by Koutsovoulos et al. [26]. The 227.2 Mb genome assembly size of *P. scribneri* is much bigger than 19.6 Mb of *P. coffeae*, the only *Pratylenchus* species sequenced and reported so far [16]. The above observation is supported by the underlying genome organization of *P. scribneri* containing a fair amount of gene duplication. This study suggests a diploid genome for *P. scribneri* which is consistent with ploidy estimations of another migratory endoparasitic nematode, *B. xylophilus* [27]. Further studies on comparative and functional genomics analysis are needed to validate these results and take a deeper look into the common genes and their functions between the two species.

To the best of our knowledge, this is the first report of de novo draft assembly of the *P. scribneri* genome generated using the PacBio ultra-low DNA input protocol. The new workflow described here will facilitate the sequencing and assembly of new and existing species of plant-parasitic nematodes that are important for crop production. In vitro culturing of nematodes is a common technique to increase pure nematode population. The DNA preparation and sequencing method described here is recommended for sequencing genome of the nematodes that are difficult to be mass-produced. Overall, the strategy of hand-picking nematodes under sterile conditions performed better in our study and resulted in an almost contaminant free library and can be considered for other endoparasitic and ectoparasitic plant-parasitic nematodes. The long-read-based sequencing and genome assembly of *P. scribneri* would enable the identification of parasitism genes/effectors involved in the host and nematode interaction mechanism. Additionally, the comparative genome analysis of *Pratylenchus* species with that of sedentary endoparasites could facilitate studies associated with evolutionary and lifestyle mechanisms of plant-parasitic nematodes based on the effectors present in different groups.

4. Materials and Methods

4.1. Nematode Collection, DNA Isolation, and Evaluation

Pratylenchus scribneri isolate was originally collected from a potato field (in rotation with corn) located in Sargent County, ND and, thereafter, maintained aseptically on carrot discs in our lab as suggested by Lawn and Noel [28]. For preparing nematode suspension, the inoculated carrot discs in the Petri plates were cut into thin slices using a sterile razor blade. The carrot pieces were left suspended in water for 3 h for the nematodes to move out of the carrot tissues. The nematodes along with the carrot tissues were then passed through a 60-mesh sieve placed on a 635-mesh sieve. The nematodes retained on the 635-mesh sieve were later collected into a 50 mL tube using a wash bottle containing distilled water with 50 mg/L carbenicillin and 50 mg/L kanamycin. Starting with nematode suspension, two different methods were used to isolate *P. scribneri* genomic DNA. In the first method, a 1 mL nematode suspension containing 8000 to 10,000 mixed-stage individuals of *P. scribneri* was ground to a fine powder in liquid nitrogen. The crushed nematodes were then subject to DNA extraction using the DNAeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNA quality was assessed using a ND-1000 Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA, USA) and the samples with 260/280 above 1.8 were selected for further analyses (Method 1). For the second method, 500 *P. scribneri* adults and juveniles from the original nematode suspension were handpicked and transferred under a microscope on a sterile glass slide containing 100 µL double-distilled water with antibiotics (50 mg/L carbenicillin and 50 mg/L kanamycin). Using a 200 µL pipette, the nematodes were then transferred to a 1.5 mL tube and centrifuged at 4000 rpm at 4 °C for 10 min. The nematode pellet was snap frozen in liquid nitrogen and stored in a −80 °C freezer (Method 2). For subsequent DNA extraction and sequencing, the frozen sample was sent off to the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign on dry ice through overnight shipping. The nematode pellet was then pulverized under

liquid nitrogen. Two hundred microliters of CTAB (OPS Diagnostics, Lebanon, NJ, USA) were added, mixed with the powder, and the sample was incubated at 60 °C for 60 min and cooled to room temperature. The MagAttract HMW DNA (Qiagen, Hilden, Germany) kit was used for the rest of the procedure, with some modifications. Briefly, proteinase K and RNAase were added to the mixture and incubated overnight at room temperature. The mixture was cleaned up with 1× volume of chloroform/isoamyl alcohol, gently inverted 10 times to mix, and incubated at room temperature for 5 min, followed by centrifugation for 2 min at 10,000× *g*. The supernatant was transferred to a new 1.5 mL tube, and magnetics beads from the MagAttract kit were added, washed twice, and the DNA was eluted from the beads in AE buffer at 37 °C for 30 min. The DNA was quantitated with Qubit using the High-Sensitivity dsDNA kit (ThermoFisher Scientific, Waltham, MA, USA) and the integrity was evaluated in a Femto Pulse system (Agilent, Santa Clara, CA, USA).

4.2. Library Preparation and Sequencing

Five nanograms of purified genomic DNA were sheared with Megaruptor 3 (Diagenode, Denville, NJ, USA) to an average size of 10 kb. The sheared gDNA was amplified using the PacBio Ultra-Low DNA Input kit following the manufacture's recommendations (Pacific Biosciences, Menlo Park, CA, USA). The amplified gDNA was converted into a PacBio library with the SMRTbell Express Template Prep kit version 2.0 (Pacific Biosciences, Menlo Park, CA, USA). Briefly, the sheared genomic DNA was first added to an enzymatic reaction to remove single-stranded overhangs followed by treatment with repair enzymes to repair the DNA ends. The ends of the repaired, double-stranded fragments were then tailed with an A-overhang. After that, the T-overhang SMRTbell adapters were ligated and the SMRTbell library was purified with two rounds of AMPure PB bead clean up (Pacific Biosciences, Menlo Park, CA, USA). The library was quantitated with Qubit and then run on a Femto Pulse to confirm the presence of DNA fragments of the expected size. The library was sequenced on one SMRT cell 8 M on a PacBio Sequel IIe system with 30 h movie time. The circular consensus analysis was performed in real time in the instrument with SMRT Link V10.1 software (Pacific Biosciences, Menlo Park, CA, USA) using default parameters, which include 3 minimum passes and a minimum accuracy of 99%.

4.3. Genome Assembly, Decontamination, and Completeness Analysis

Raw PacBio HiFi reads were filtered with Seqkit v2.0.0 [29] to remove reads < 5 kb. Filtered reads were assembled with Hifiasm-v0.16.1 [30], turning off the parameter for automatic internal haplotig purging, and the GFA data output was converted to FASTA format with Gfatools v0.4 (https://github.com/lh3/gfatools, accessed on 25 April 2022). Primary contig assembly contamination by non-nematode sequences was assessed using Blobtools v1.1.1 [31]. Read depth was calculated by aligning the filtered PacBio HiFi reads to the primary contig assembly with Minimap v2.21 [32] and processing the bam file with SAMtools v1.12 [33]. Contigs were taxonomically assigned to phyla using Diamond v2.0.9 blastx [34] with the long reads option and e-value limit of 1×10^{-25} against the UniProt Reference Clusters UniRef100 dataset [35]. Thirty-nine contigs classified into phyla other than Nematoda were then further analyzed manually by BLASTN/BLASTX against the NCBI non-redundant (nr) datasets. The remaining primary contigs were then purged of haplotypic and duplicate content using the purge_dups v.1.2.5 tool [36]. The ortholog completeness was evaluated with benchmarking universal single-copy ortholog (BUSCO) v4.1.4 [37] using the nematoda_odb10 lineage dataset.

4.4. Genome Ploidy Estimations

To estimate the ploidy level of the genome, KMC v3.1.1 [38] was used to count k-mers (k = 21) in the filtered PacBio HiFi reads (>5 kb). The k-mer histogram was analyzed by GenomeScope 2 [39] with ploidy tested from $p = 2$ to $p = 6$. Haplotypic and duplicate genome content were also analyzed using heterozygous kmer pairs with the tool Smudgeplot [34].

4.5. Genome Structural Annotation

RepeatMasker v4.1.2 (http://www.repeatmasker.org, accessed on 5 July 2022) was used to soft-mask repeats from the final purged and decontaminated genome assembly file. Two separate runs of BRAKER2 [40] used RNA-seq and protein evidence with default parameters. The raw *P. scribneri* RNA-seq reads (read length = 100 bp) used for the above analysis were generated in our previous study through paired-end sequencing of cDNA libraries prepared from good quality total RNA extracts of several thousand pre-parasitic and parasitic nematodes on an Illumina NovaSeq 6000 sequencer (unpublished work). The reads were trimmed with Fastp v0.20.0 [41] and then aligned to the masked genome assembly using STAR v.2.7.10a [42]. Protein evidence was provided using NCBI's Nematoda RefSeq dataset containing 18,060 protein sequences. In addition, BUSCO v5.3.2 running in "long mode" [43] was used to predict protein sequences using the 3131 "lineage nematoda_odb10" orthologs from the final purged/decontaminated/unmasked genome assembly file. The outputs from the two independent BRAKER runs were then combined with the tool TSEBRA v.1.0.3 [38] to obtain GTF and amino acid formats using Cufflinks v2.2.1 [44].

Author Contributions: Conceptualization, D.A. and G.Y.; methodology, D.A., A.G.H., K.K.O.W. and G.Y.; data analysis, C.J.F. and K.K.O.W.; writing—original draft preparation, D.A.; writing—review and editing, G.Y., A.G.H., C.J.F. and K.K.O.W.; supervision, G.Y.; project administration, G.Y.; funding acquisition, G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the North Dakota State Board of Agricultural Research and Education under funding number FARG090473 and the North Dakota Corn Council under funding number FAR0035599.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the raw PacBio sequence data supporting the results of this article as well as the genome assembly have been deposited at the NCBI with the following accession numbers: BioProject: PRJNA932437; BioSample: SAMN33192627; Genome Assembly: JAQSEB000000000; SRA raw data: SRP421680/SRS16720223/SRX19322091/SRR23381582. The data will be made publicly available with the published version of this manuscript.

Acknowledgments: We thank the North Dakota Corn Council and the North Dakota State Board of Agricultural Research and Education for funding this project. The authors would also like to thank Ekta Ojha for her assistance in providing nematode carrot cultures for Method 1.

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. Jones, J.T.; Haegeman, A.; Danchin, E.G.J.; Gaur, H.S.; Helder, J.; Jones, M.G.K.; Kikuchi, T.; Manzanilla-López, R.; Palomares-Rius, J.E.; Wesemael, W.M.L.; et al. Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* **2013**, *14*, 946–961. [CrossRef] [PubMed]
2. Smiley, R.W. Multiplication of *Pratylenchus neglectus* and *P. thornei* on plants other than cereals. In Proceedings of the Fifth International Cereal Nematode Initiative Workshop, Ankara, Turkey, 12–15 September 2015.
3. Castillo, P.; Vovlas, N. *Pratylenchus* (Nematoda: Pratylenchidae): Diagnosis, biology, pathogenicity and management. *Nematol. Monogr. Perspect.* **2007**, *6*, 1–543. [CrossRef]
4. Grynberg, P.; Coiti Togawa, R.; Dias de Freitas, L.; Antonino, J.D.; Rancurel, C.; Mota do Carmo Costa, M.; Grossi-de-Sa, M.F.; Miller, R.N.; Brasileiro, A.C.M.; Messenberg Guimaraes, P.; et al. Comparative genomics reveals novel target genes towards specific control of plant-parasitic nematodes. *Genes* **2020**, *11*, 1347. [CrossRef]
5. Akhter, N. Molecular Characterization of Root-Lesion Nematode Species from Corn Fields in North Dakota and Evaluation of Resistance in Corn Hybrids. Master's Thesis, North Dakota State University, Fargo, North Dakota, October 2019.
6. Ozbayrak, M. DNA Barcoding of *Pratylenchus* from Agroecosystems in the Northern Great Plains of North America. Master's Thesis, University of Nebraska-Lincoln, Lincoln, Nebraska, May 2019.

7. Yan, G.P.; Plaisance, A.; Huang, D.; Liu, Z.; Chapara, V.; Handoo, Z.A. First report of the root-lesion nematode *Pratylenchus scribneri* infecting potato in North Dakota. *Plant Dis.* **2016**, *100*, 1023. [CrossRef]
8. Arora, D.; Yan, G.P.; Baidoo, R. Developing a real-time PCR assay for direct detection and quantification of *Pratylenchus scribneri* in field soil. *Nematology* **2020**, *22*, 733–744. [CrossRef]
9. Arora, D.; Yan, G.P. Early detection and temporal dynamics of *Pratylenchus scribneri* infection in potato roots determined using quantitative PCR and root staining. *Phytopathology* **2022**, *112*, 1776–1782. [CrossRef] [PubMed]
10. Huang, D.; Yan, G.P. Specific detection of the root-lesion nematode *Pratylenchus scribneri* using conventional and real-time PCR. *Plant Dis.* **2017**, *101*, 359–365. [CrossRef]
11. Li, Y.; Lu, Q.S.; Wang, S.; Liu, Y.K.; Wang, K.; Yuan, H.X.; Li, H.L. Discovery of a root-lesion nematode, *Pratylenchus scribneri*, infecting corn in inner Mongolia, China. *Plant Dis.* **2019**, *103*, 1792. [CrossRef]
12. Li, Y.; Lu, Q.S.; Wang, S.; Guo, F.; Xia, Y.H.; Wang, K.; Li, H.L. Occurrence of *Pratylenchus scribneri* on soybean in Henan Province, China. *Plant Dis.* **2019**, *103*, 774–775. [CrossRef]
13. Xia, Y.; Li, J.; Hao, P.; Wang, K.; Lei, B.; Li, H.L.; Li, Y.U. Discovery of root-lesion nematode (*Pratylenchus scribneri*) on corn in Hainan Province of China. *Plant Dis.* **2022**, *106*, 1999. [CrossRef]
14. Xia, Y.H.; Li, J.; Sun, M.R.; Lei, B.; Li, H.L.; Li, Y.; Wang, K. Identification and pathogenicity of *Pratylenchus scribneri* on tomato in Sichuan Province of People's Republic of China. *J. Helminthol.* **2022**, *96*, E5. [CrossRef]
15. Olowe, T. Relationship between inoculum density levels of *Pratylenchus scribneri*, and growth and yield of maize, *Zea mays*. *Int. J. Nematol.* **2011**, *21*, 73–78.
16. Burke, M.; Scholl, E.H.; Bird, D.M.; Schaff, J.E.; Colman, S.D.; Crowell, R.; Diener, S.; Gordon, O.; Graham, S.; Wang, X.; et al. The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. *Nematology* **2015**, *17*, 621–637. [CrossRef]
17. Fosu-Nyarko, J.; Jones, M.G.K. Advances in understanding the molecular mechanisms of root lesion nematode host interactions. *Annu. Rev. Phytopathol.* **2016**, *54*, 253–278. [CrossRef] [PubMed]
18. Kingan, S.B.; Heaton, H.; Cudini, J.; Lambert, C.C.; Baybayan, P.; Galvin, B.D.; Durbin, R.; Korlach, J.; Lawniczak, M.K. A high-quality de novo genome assembly from a single mosquito using PacBio sequencing. *Genes* **2019**, *10*, 62. [CrossRef]
19. Rufai, S.B.; McIntosh, F.; Poojary, I.; Chothe, S.; Sebastian, A.; Albert, I.; Praul, C.; Venkatesan, M.; Mahata, G.; Maity, H.; et al. Complete Genome Sequence of *Mycobacterium orygis* Strain 51145. *Microbiol. Resour. Announc.* **2021**, *10*, e01279-20. [CrossRef]
20. Crowley, L.M. Darwin Tree of Life Consortium. The genome sequence of the seven-spotted ladybird, *Coccinella septempunctata* Linnaeus, 1758. *Wellcome Open Res.* **2021**, *6*, 319. [CrossRef]
21. Bali, S.; Hu, S.; Vining, K.; Brown, C.; Mojtahedi, H.; Zhang, L.; Gleason, C.; Sathuvalli, V. Nematode Genome Announcement: Draft genome of *Meloidogyne chitwoodi*, an economically important pest of potato in the Pacific Northwest. *Mol. Plant-Microbe Interact.* **2021**, *34*, 981–986. [CrossRef]
22. Wram, C.L.; Hesse, C.N.; Wasala, S.K.; Howe, D.K.; Peetz, A.B.; Denver, D.R.; Humphreys-Pereira, D.; Zasada, I.A. Genome announcement: The draft genomes of two *Radopholus similis* populations from Costa Rica. *J. Nematol.* **2019**, *51*. [CrossRef]
23. Ding, X.; Guo, Y.; Ye, J.; Wu, X.; Lin, S.; Chen, F.; Zhu, L.; Huang, L.; Song, X.; Zhang, Y.; et al. Population differentiation and epidemic tracking of *Bursaphelenchus xylophilus* in China based on chromosome-level assembly and whole\-\genome sequencing data. *Pest Manag. Sci.* **2022**, *78*, 1213–1226. [CrossRef]
24. Sato, K.; Kadota, Y.; Gan, P.; Bino, T.; Uehara, T.; Yamaguchi, K.; Ichihashi, Y.; Maki, N.; Iwahori, H.; Suzuki, T.; et al. High-quality genome sequence of the root-knot nematode *Meloidogyne arenaria* genotype A2-O. *Genome Announc.* **2018**, *6*, e00519-18. [CrossRef] [PubMed]
25. Susič, N.; Koutsovoulos, G.D.; Riccio, C.; Danchin, E.G.; Blaxter, M.L.; Lunt, D.H.; Strajnar, P.; Širca, S.; Urek, G.; Stare, B.G.; et al. Genome sequence of the root-knot nematode *Meloidogyne luci*. *J. Nematol.* **2020**, *52*, 1–5. [CrossRef] [PubMed]
26. Koutsovoulos, G.D.; Poullet, M.; Elashry, A.; Kozlowski, D.K.; Sallet, E.; Da Rocha, M.; Perfus-Barbeoch, L.; Martin-Jimenez, C.; Frey, J.E.; Ahrens, C.H.; et al. Genome assembly and annotation of *Meloidogyne enterolobii*, an emerging parthenogenetic root-knot nematode. *Sci. Data* **2020**, *7*, 324. [CrossRef]
27. Shinya, R.; Sun, S.; Dayi, M.; Tsai, I.J.; Miyama, A.; Chen, A.F.; Hasegawa, K.; Antoshechkin, I.; Kikuchi, T.; Sternberg, P.W.; et al. Possible stochastic sex determination in *Bursaphelenchus* nematodes. *Nat. Commun.* **2022**, *13*, 2574. [CrossRef] [PubMed]
28. Lawn, D.A.; Noel, G.R. Gnotobiotic culture of *Pratylenchus scribneri* on carrot discs. *Nematropica* **1986**, *16*, 45–51.
29. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef]
30. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [CrossRef]
31. Laetsch, D.R.; Blaxter, M.L. BlobTools: Interrogation of genome assemblies. *F1000Research* **2017**, *6*, 1287. [CrossRef]
32. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **2016**, *32*, 2103–2110. [CrossRef]
33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
34. Buchfink, B.; Reuter, K.; Drost, H.G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [CrossRef] [PubMed]

35. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2008**, *10*, 421. [CrossRef] [PubMed]
36. Guan, D.; McCarthy, S.A.; Wood, J.; Howe, K.; Wang, Y.; Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **2020**, *36*, 2896–2898. [CrossRef] [PubMed]
37. Manni, M.; Berkeley, M.R.; Seppey, M.; Simão, F.A.; Zdobnov, E.M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [CrossRef]
38. Kokot, M.; Dlugosz, M.; Deorowicz, S. KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* **2017**, *33*, 2759–2761. [CrossRef]
39. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [CrossRef]
40. Bruna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with genemark-ep+ and augustus supported by a protein database. *NAR Genomic. Bioinform.* **2021**, *3*, 1. [CrossRef]
41. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, 884–890. [CrossRef]
42. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
43. Gabriel, L.; Hoff, K.J.; Brůna, T.; Borodovsky, M.; Stanke, M. TSEBRA: Transcript selector for BRAKER. *BMC Bioinform.* **2021**, *22*, 1–12. [CrossRef]
44. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotech.* **2010**, *28*, 511–515. [CrossRef] [PubMed]

International Journal of
Molecular Sciences

MDPI

Article

Systematical Characterization of the *AT-Hook* Gene Family in *Juglans regia* L. and the Functional Analysis of the *JrAHL2* in Flower Induction and Hypocotyl Elongation

Peng Jia, Jiale Liu, Rui Yan, Kaiyu Yang, Qinglong Dong, Haoan Luan, Xuemei Zhang, Han Li, Suping Guo and Guohui Qi *

College of Forestry, Hebei Agricultural University, Baoding 071000, China; yangkaiyu2023@163.com (K.Y.)
* Correspondence: bdqgh@hebau.edu.cn

Abstract: AT-hook motif nuclear localization (AHL) proteins play essential roles in various plant biological processes. Yet, a comprehensive understanding of *AHL* transcription factors in walnut (*Juglans regia* L.) is missing. In this study, 37 *AHL* gene family members were first identified in the walnut genome. Based on the evolutionary analysis, *JrAHL* genes were grouped into two clades, and their expansion may occur due to segmental duplication. The stress-responsive nature and driving of developmental activities of *JrAHL* genes were revealed by *cis*-acting elements and transcriptomic data, respectively. Tissue-specific expression analysis showed that *JrAHLs* had a profound transcription in flower and shoot tip, *JrAHL2* in particular. Subcellular localization showed that JrAHL2 is anchored to the nucleus. Overexpression of *JrAHL2* in Arabidopsis adversely affected hypocotyl elongation and delayed flowering. Our study, for the first time, presented a detailed analysis of *JrAHL* genes in walnut and provided theoretical knowledge for future genetic breeding programs.

Keywords: AT-hook gene family; *Juglans regia*; hypocotyl; flowering; Phylogenetics

Citation: Jia, P.; Liu, J.; Yan, R.; Yang, K.; Dong, Q.; Luan, H.; Zhang, X.; Li, H.; Guo, S.; Qi, G. Systematical Characterization of the *AT-Hook* Gene Family in *Juglans regia* L. and the Functional Analysis of the *JrAHL2* in Flower Induction and Hypocotyl Elongation. *Int. J. Mol. Sci.* **2023**, *24*, 7244. https://doi.org/10.3390/ijms24087244

Academic Editor: Wajid Zaman

Received: 9 March 2023
Revised: 31 March 2023
Accepted: 3 April 2023
Published: 14 April 2023

1. Introduction

The *AT-hook motif nuclear localized* (*AHL*) gene family encodes the conserved transcription factor that widely distributed in plants species [1]. The AHL proteins are generally distinguished by two characteristic domains: the AT-hook motif and the Plant and Prokaryote Conserved (PPC) domain. The AT-hook motif was first described as a high mobility group (HMG) non-histone chromosomal protein HMG-I/Y [2]. The AT-hook motif contains conserved core sequence 'Arg-Gly-Arg' that can bind to the minor groove of the AT-rich B form of DNA [3]. AT-hook motifs can be divided into two categories according to the conserved amino acid sequence: Type-I AT-hook motif contains 'Gly-Ser-Lys-Asn-Lys' consensus sequences at the end of the 'Arg-Gly-Arg' core, while the Type-II contains 'Arg-Lys-Tyr.' The PPC domain, also known as the Domain of Unknown Function #296 (DUF296), is also found in, archaea, prokaryotes, and higher plants. The PPC domain was involved in the physical interaction with other AHL (or themselves) or nuclear proteins [1,4]. According to sequence differences, the PPC can be divided into two categories: Type-A and Type-B. Type-B contains the conserved core 'Phe-Thr-Pro-His' in the upstream region, while the core sequence within the PPC domain in Type-A is more variable ('Leu-Arg-Ser-His' or 'Leu-Arg-Ala-His', etc.) [5]. The *AHL* genes in land plants have been grouped into Clades based on the phylogenetic relationship: Clades-A and Clades-B. The Clade-A *AHLs* are intron-free and encode proteins containing only one Type-I AT-hook motif and Type-A PPC domain. The Clade-B *AHLs* contain introns, and Clade-B AHL proteins can be further divided into two types: one type of AHLs contains a Type-II AT-hook motif and a Type-B PPC domain, and the other type of AHLs contains two AT-hooks (Type-I and Type-II) and a Type-B PPC domain [6].

The AHL gene family plays vital roles in plant growth, development, and stimulus response. In Arabidopsis (Arabidopsis thaliana), the biological functions of several AHLs have been extensively studied. A number of AHL genes redundantly inhibit the elongation of hypocotyl, such as AHL22 [7], SOB3/AHL29 [6,8], and ESC/AHL27 [9]. *AHL18*, which is involved in the regulation of the length of the proliferation domain and the division in the root apical meristem, and inhibiting primary root growth and lateral root development [10]. In *M. truncatula*, the *MtAHL1* and *MtAHL2* were involved in the formation of nitrogen-fixing nodules [11].

The *AHL* gene is related to regulating flowering time and reproductive development differently. The TEK/AHL16 directly binds to the floral suppressor gene FLC's repressive regulatory region, leading to its transcriptional silencing. Knockout of the *TEK/AHL16* activates a variety of transposons, and up-regulate the expression of floral suppressor gene *FLC* through epigenetic regulation (histone acetylation and DNA methylation), thereby delaying flowering [12]. TEK/AHL16 also contributes to nexine formation and pollen wall development by directly binding to the nuclear matrix attachment region (MAR) and the promoter of *AGP6*, an arabinogalactan protein coding gene [13,14]. In addition, overexpressing *AHL* delays flowering by inhibiting *FT* expression [7,15]. AHL22 binds to the AT-rich sequence in the *FT* locus and then alters the structural design of the chromatin region by regulating both histone acetylation and methylation [15]. *AHL18*, which is in the same phylogenetic clade as *AHL22*, conferred a similar late-flowering phenotype when overexpression in Arabidopsis to that of *AHL22* [7]. AHLs were also reported to be involved in hormone-related pathways. SOB3/AHL29 inhibited PIF4 protein accumulation, thereby reducing the binding of PIF4 to target loci involved in hormone transport, homeostasis, or response such as auxin, brassinosteroid, and ethylene [16]. SOB3/AHL29 also targeted downstream genes associated with auxin-related genes, such as *YUCCA8* and *SAUR19* [8]. AGF1/AHL25 regulated gibberellins homeostasis in Arabidopsis by targeting the GA 3-oxidase gene AtGA3ox1 [17]. In addition, AHLs participated in the biotic and abiotic stress responses. AHL13, AHL15, AHL19, AHL20, and AHL27 were reported to implicate in PAMP-triggered immunity [18–20]. Overexpression of AHL20 and its closely related family members (AHL15, AHL19, and AHL27) in protoplasts inhibited the innate immune response [19]. Especially in Arabidopsis, AHL13 regulated key factors in the jasmonic acid (JA) pathway and affected immunity toward Botrytis cinerea and Pseudomonas syringae pathogens [20]. AHL10 played a critical in response to low-water potential stress, and in the regulation of auxin- and jasmonic acid -related gene expression [21]. AtAHL4 was involved in the regulation of lipid catabolism during seedling establishment [22] and of root xylem development [4]. Interactions between AHL proteins are also crucial for their functions. In Arabidopsis roots, the AHL3 physically interacted with AHL4 and facilitated intercellular trafficking [23]. In maize (*Zea mays*), the AT-hook protein BAF1 is capable of forming homodimers and heterodimers with other members of the AT-hook family, controlling the formation of maize ears [24].

In recent years, the AHL gene family has been identified and analyzed in several species, such as Arabidopsis, soybean (*Glycine max*) [25], cotton (*Gossypium raimondii*) [26], *Brassica napus* [27], grape (*Vitis vinifera*) [28], carrot (*Daucus carota subsp. sativus*) [29], and so on. Walnut (*Juglans regia* L.) is an economically vital nut trees in China and other temperate parts of the world, but the long period before flowering and bearing fruiting after seed germination limits the progress of walnut breeding work. In addition, functional characterization of AHL genes in walnut is still missing. In this study, the AHL gene family was characterized in walnut based on the high-quality genome data [30]. The chromosomal locations, phylogenetic relationships, gene expansions and structures, conserved domains, putative protein interaction, and cis-elements in the promoter region were analyzed. Furthermore, the *JrAHL2*, a Clade-A member highly expressed in flower, was isolated, its protein subcellular localization was analyzed, and its functions involved in floral transition and hypocotyl elongation were dissected. The current study will provide a theoretical basis for further investigation of the functions of the *JrAHL* gene family, and an improved

understanding of the molecular mechanisms underlying its role in regulating growth and development in walnut.

2. Results

2.1. AHL Genes in Walnut

The HMMscan was executed against the protein databases using the PPC/DUF296 (PF03479) and AT-hook as queries. After manual checking, the 37 unique *AHL* genes were obtained from walnut (*Juglans regia* L.). The *JrAHL* genes were named according to their chromosomal locations (*JrAHL1*–*JrAHL37*) (Table 1). The gene lengths of *JrAHL* genes ranged from 667 bp to 7791 bp, which encode polypeptides from 221 to 406 amino acids with predicted molecular weights ranging from 20.04 kDa to 42.63 kDa. The theoretical pI ranged from 5.08 to 10.61. Approximately 92% of JrAHL proteins (except for JrAHL10/11/33) were predicted to be located in the nucleus (Supplementary Table S2), consistent with the previous report [1].

Table 1. Information of the *AHL* genes in *J. regia* L.

Name	Gene ID	Location (Start)	Location (End)	Strand	Length	CDS (bp)	Peptide (aa)	pI	MW (kDa)
JrAHL1	JreChr01G12984	39,611,176	39,611,994	+	819	819	272	6.08	28.02
JrAHL2	JreChr01G12992	47,390,686	47,391,612	−	927	927	308	6.59	32.84
JrAHL3	JreChr01G13333	47,470,900	47,477,455	+	6556	1005	334	9.59	34.79
JrAHL4	JreChr01G13761	50,056,338	50,057,327	−	990	990	329	5.08	34.44
JrAHL5	JreChr02G10898	7,160,919	7,167,725	−	6807	1035	344	9.91	35.36
JrAHL6	JreChr02G11467	21,601,290	21,602,144	+	855	855	284	6.45	29.38
JrAHL7	JreChr02G11474	26,318,938	26,319,843	−	906	906	301	6.80	31.86
JrAHL8	JreChr02G11687	26,371,944	26,376,879	+	4936	1068	355	8.83	36.83
JrAHL9	JreChr02G11978	27,859,904	27,860,884	−	981	981	326	5.60	34.15
JrAHL10	JreChr03G13350	4,426,812	4,430,966	−	4155	795	264	10.46	27.04
JrAHL11	JreChr04G10741	8,368,413	8,372,787	+	4375	1014	337	9.21	35.15
JrAHL12	JreChr04G12210	20,642,675	20,643,481	+	807	807	268	9.24	28.25
JrAHL13	JreChr06G10970	6,003,983	6,011,773	+	7791	930	309	9.61	32.44
JrAHL14	JreChr07G10387	23,875,413	23,876,234	−	822	822	273	9.74	28.01
JrAHL15	JreChr07G11230	16,643,389	16,644,297	−	909	909	302	5.77	31.30
JrAHL16	JreChr07G11239	27,147,611	27,148,277	−	667	666	221	10.09	23.04
JrAHL17	JreChr07G11616	27,252,429	27,256,966	+	4538	1107	368	10.09	36.49
JrAHL18	JreChr08G11927	30,383,998	30,384,978	+	981	981	326	7.07	34.60
JrAHL19	JreChr08G12220	5,871,308	5,872,264	−	957	957	318	6.36	33.92
JrAHL20	JreChr08G12225	8,400,985	8,408,164	−	7180	1182	393	10.33	39.80
JrAHL21	JreChr09G12239	8,457,851	8,459,672	+	1822	696	231	9.87	23.85
JrAHL22	JreChr10G11635	644,863	648,220	+	3358	1089	362	8.49	37.48
JrAHL23	JreChr11G10743	358,990	361,864	+	2875	1068	355	9.12	36.87
JrAHL24	JreChr11G11355	23,940,903	23,941,778	−	876	876	291	5.96	30.42
JrAHL25	JreChr11G11477	30,455,540	30,459,915	+	4376	1017	338	10.61	35.04
JrAHL26	JreChr11G11481	31,577,477	31,578,376	−	900	900	299	6.16	32.00
JrAHL27	JreChr12G10636	31,622,448	31,628,198	+	5751	939	312	9.63	31.89
JrAHL28	JreChr12G10638	18,010,455	18,017,606	+	7152	1134	377	9.37	38.78
JrAHL29	JreChr12G10640	18,021,521	18,027,959	−	6439	1140	379	9.59	38.92
JrAHL30	JreChr12G10998	18,039,195	18,040,085	−	891	891	296	6.10	31.07
JrAHL31	JreChr12G11092	21,387,711	21,392,430	+	4720	1020	339	10.00	35.22
JrAHL32	JreChr12G11095	22,118,819	22,119,700	−	882	882	293	6.30	31.42
JrAHL33	JreChr13G10464	22,173,675	22,179,704	+	6030	1221	406	9.59	42.63
JrAHL34	JreChr13G10846	1,739,954	1,740,790	−	837	837	278	5.80	28.67
JrAHL35	JreChr13G10963	23,001,456	23,006,865	−	5410	963	320	9.73	33.66
JrAHL36	JreChr14G10308	24,429,650	24,430,551	−	902	795	264	6.37	26.99
JrAHL37	JreChr01G12984	1,416,877	1,417,719	−	843	843	280	6.04	29.02

2.2. Phylogenetic Analysis of JrAHL Genes

To infer the evolutionary relationship among the *JrAHL* proteins, phylogenetic analysis was performed on the full-length protein sequences. The result showed that the JrAHL proteins in walnut could be divided into two clades, as in Arabidopsis (Figure 1)-Clade-A and Clade-B. JrAHL were further classified into Type-I, Type-II, and Type-III (Figure 2 left panel), based on the presence and combination of the two characteristic functional units-the PPC domain and the AT-hook motif (Figure 2 right panel). Clade-A contained 20 JrAHLs, and it was also classified into Type-I, harboring the Clade-A PPC domain and the Type-I AT-hook motifs in their putative protein sequence (Figure 2 right panel). The higher abundance of Type I members (54%) in walnut is also consistent with observations in other land plants [25,27,31]. There were 17 JrAHL members in Clade-B, which all harbored the Clade-B PPC domain. Clade-B was further divided into two types: 11 Type-II members and 6 Type-III members. Type-II members contained Type-I and Type-II AT-hook motifs, whereas the Type-III JrAHLs only contained the Type-II AT-hook motif (Figure 2 right panel).

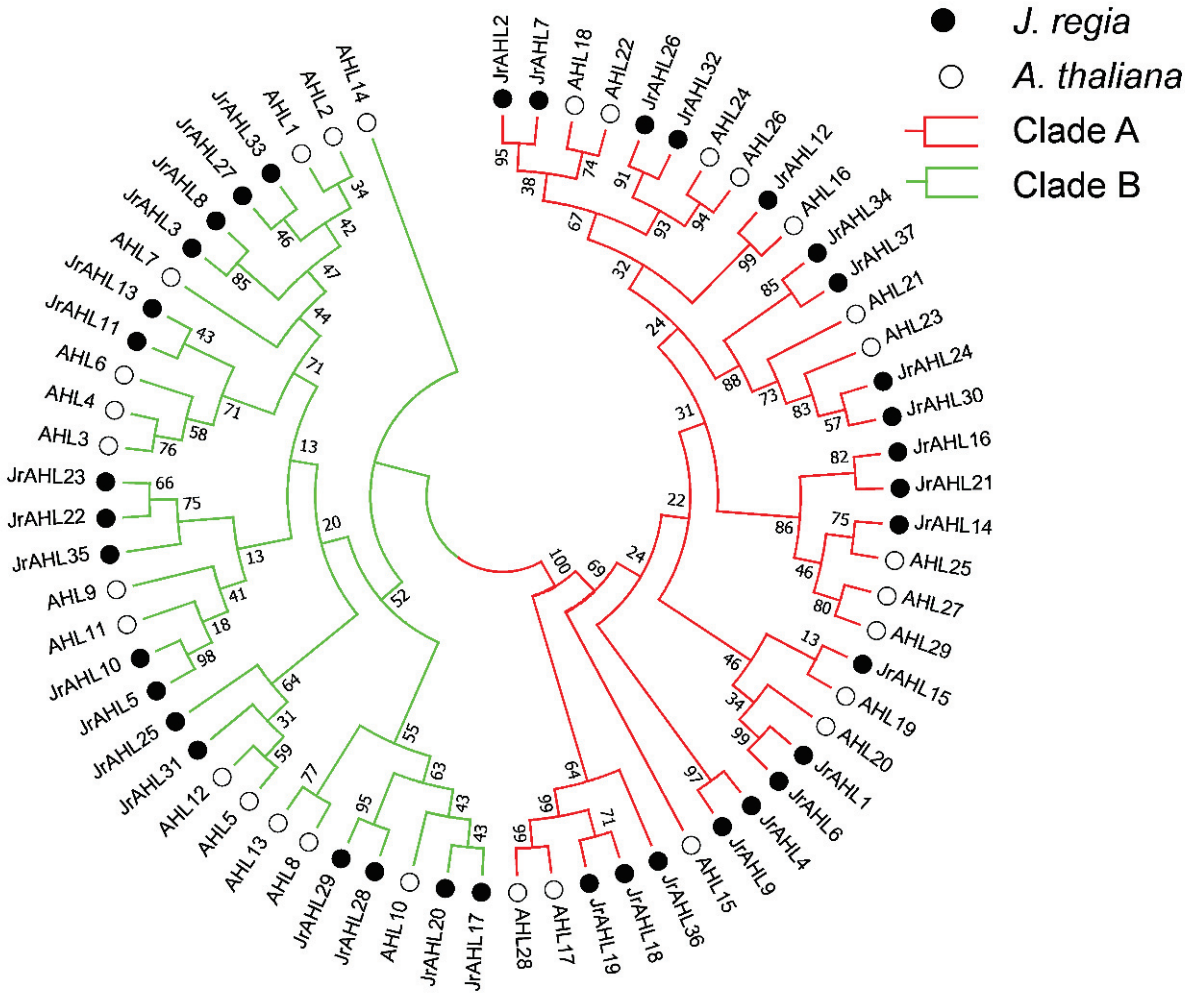


Figure 1. Neighbor-joining tree representing phylogenetic relationships among *AHL* genes from walnut and Arabidopsis. The evolutionary history was inferred using the Neighbor-Joining matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution. Evolutionary analyses were conducted in MEGA7.

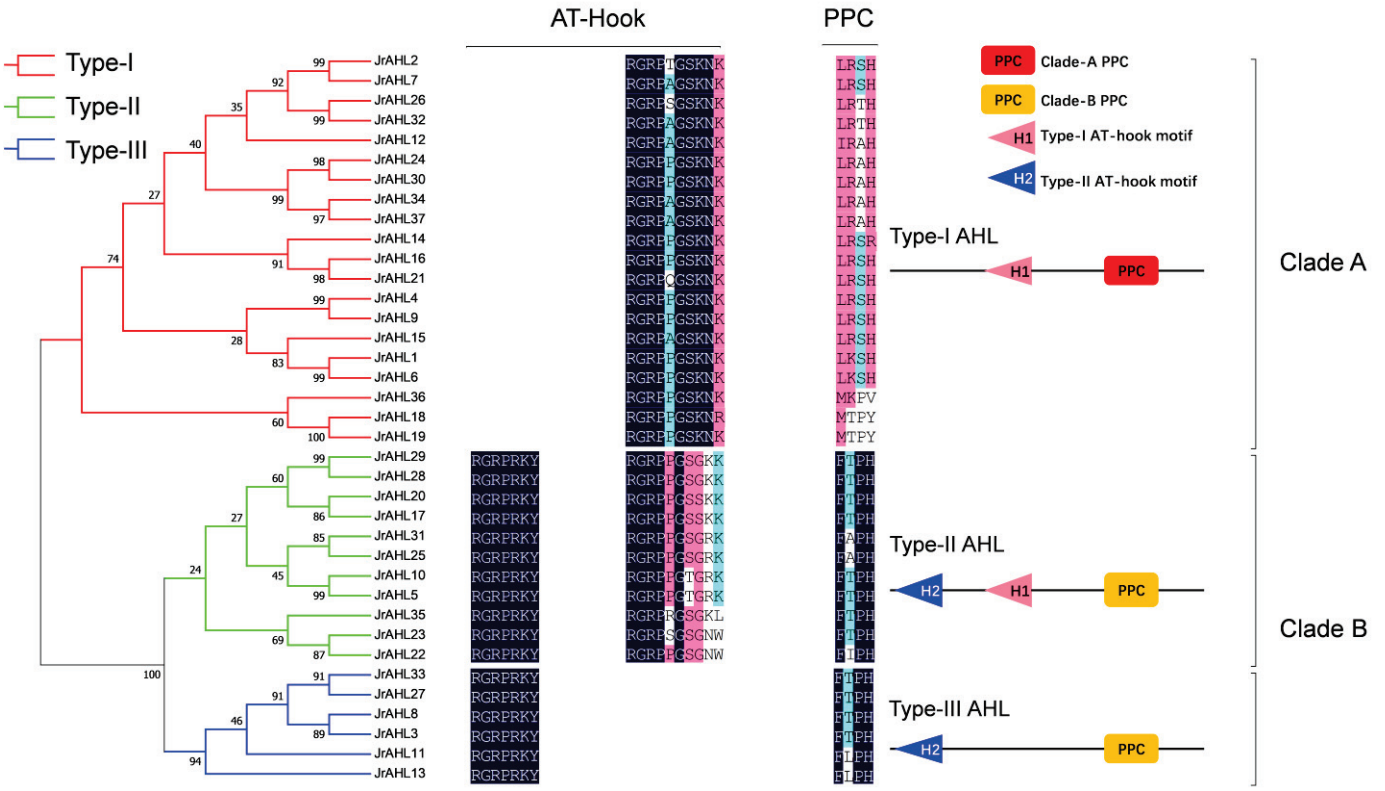


Figure 2. Phylogenetic (**left panel**) and sequence alignment (**right panel**) analysis of the walnut AHL proteins. The obtained phylogenetic tree by MEGA7 is shown on the left. Multiple sequence alignment analysis was performed using DNAMAN software, and the conserved domain is displayed on the right.

2.3. Gene Structure and Duplication of JrAHLs

The distribution of introns and exons was investigated to explore the diversity of gene structure. The structures of *JrAHL* genes could be divided into two types, one contained no or only one intron, and the other contained multiple introns (Figure 3 left panel). All Type-II and Type-III *JrAHLs* contained four to five introns, while the *JrAHLs* in Type-I were intronless except for *JrAHL21* and *JrAHL26*. Furthermore, the number and arrangement of introns are relatively conserved within the same clade, implicating the evolutionary similarity between these members. A total of seven conserved motifs in which two motifs (motif 1 and motif 2) have highly conserved 'R-G-R-P' amino acid sequences (Figure 3 right panel), which are the typical characteristic sequences of the AT-hook motif family.

Gene duplication leads to the expansion of AHL family through segmental and tandem duplications [1]. We speculated the exact mechanism has driven the expansion of the *AHL* gene family in walnut. We analyzed gene duplication events in the *JrAHL* gene family. As shown in Figure 4, a total of 37 *JrAHL* genes were unevenly mapped onto 16 chromosomes of walnut. Chromosomes 00 and 14 did not contain duplicated genes, whereas Chromosome 12 had the highest number of duplications. More than twenty pairs of *JrAHL* genes, such as *JrAHL1/6*, *JrAHL4/9*, *JrAHL24/34*, and *JrAHL31/25*, were located in duplicated genomic regions. Duplicated pairs showed similar motifs (Figure 2), patterns of gene structure (Figure 3), and clustered tightly in the evolutionary tree (Figure 1), suggesting the *JrAHL* genes have duplicated segmentally during the evolution.

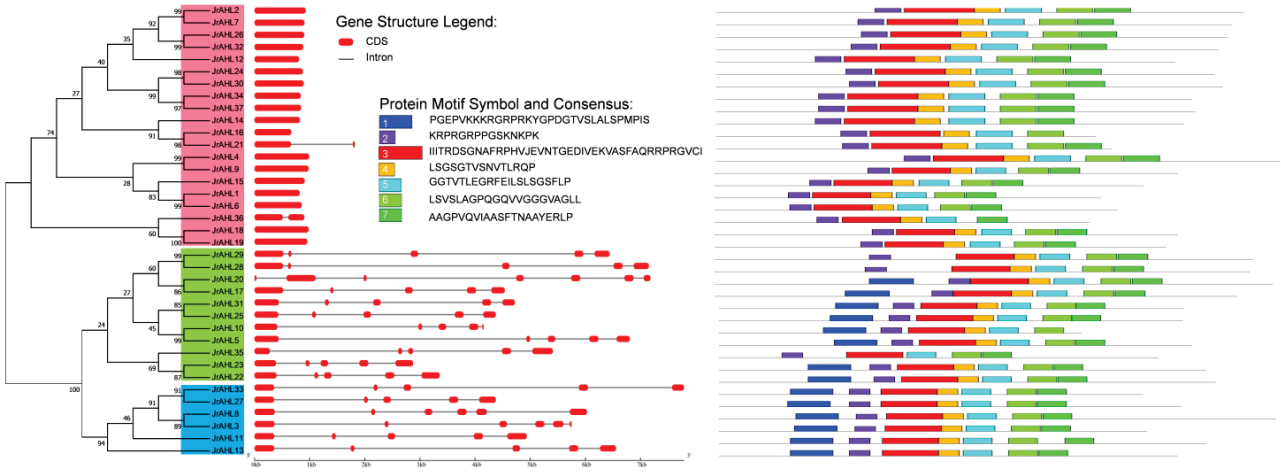


Figure 3. Gene structure analysis and conservative motif prediction of the AT-hook motif gene family in walnut. The gene structure map was obtained with the GSDS online tool (**left**). The *x*-axis shows the inferred length of the different genes (5′ to 3′) and their respective CDS (red). Domain structure and motif were elucidated with the MEME platform (**right**).

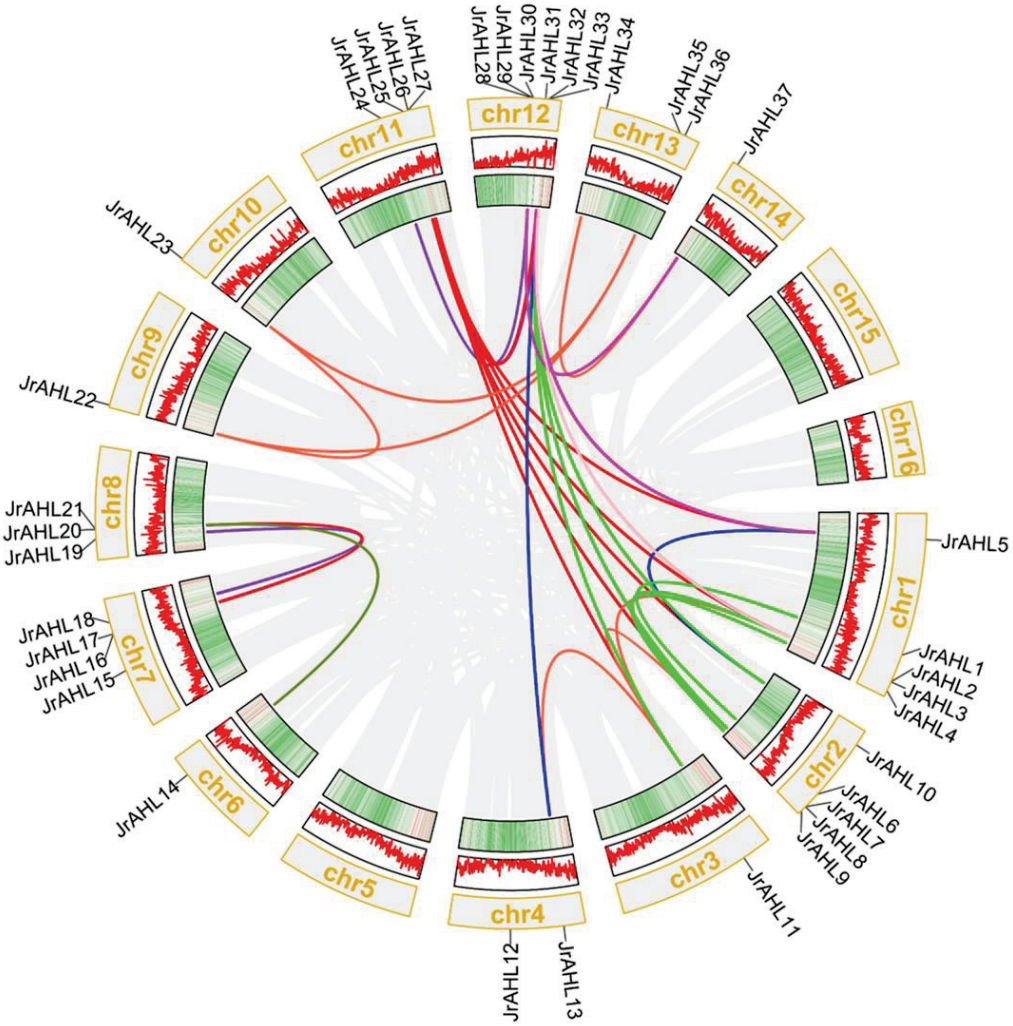


Figure 4. Chromosomal distribution and collinearity of walnut *AHL* genes. The gray lines in the background indicate the collinear blocks in the genomes of *J. regia*, while the colorful lines emphasize the segmental duplication of the walnut *AHLs*.

2.4. Cis-Acting Element Analysis of JrAHLs Promoter

The promoter is critical for transcriptional regulation because of its responsiveness to specific hormone levels and environmental cues and plays a vital role in the biological regulation of gene expression under stress. We investigated the possible regulatory mechanisms by searching the promoter region of 1500 bp upstream of the *JrAHL* gene family. Four categories of *cis*-elements were identified related to hormone response, stress response, tissue-specific expression, and transcription factor binding (Figure 5). Hormone response elements, such as gibberellin response element (GARE-motif), salicylic acid response element (TCA-element), abscisic acid-responsive element (ABRE), auxin-responsive (AUXRR-core), and MeJA-responsiveness (CGTCA-motif) were abundant in *JrAHL* promoter regions. Stress response elements, including defense responsive element (TC-rich repeat), low-temperature response (LTR), and drought response element (MBS), were discovered in most of the promoter regions of *JrAHL* genes. As for *cis*-acting regulatory elements related to tissue-specific expression, endosperm- and meristem-expression elements were found in 4 and 14 *JrAHL* promoter regions, respectively (Figure 5 and Figure S2). Four *JrAHL* genes contained the AT-rich element, which provided the binding site of ATBP-1, a class of AT-rich DNA binding proteins. These results imply that *JrAHL*s are involved in many biological processes.

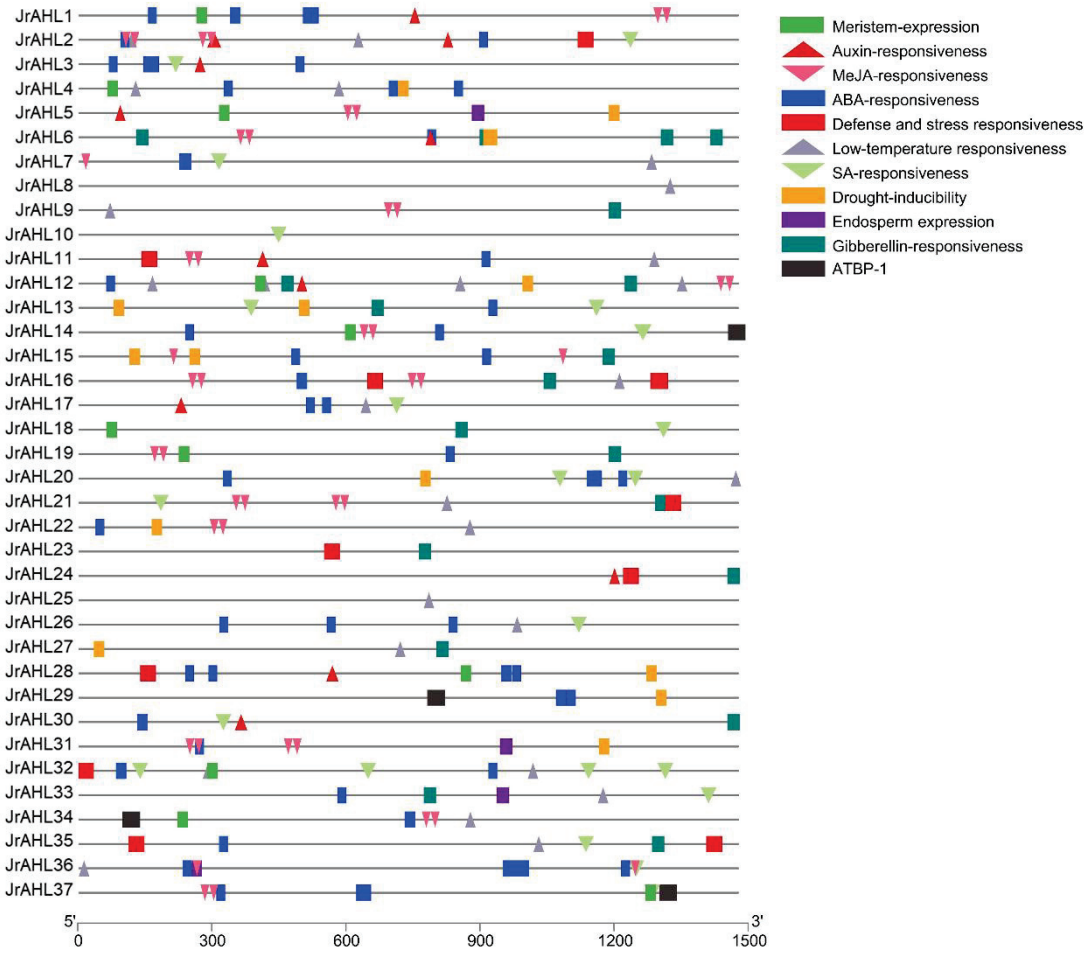


Figure 5. Predicted *cis*-elements in the *JrAHL* promoters. The 1.5-kb sequence upstream from the start codon of *JrAHL* genes was analyzed using the PlantCARE database. Different colored graphics represent different *cis*-acting elements.

2.5. Protein-Protein Interaction Analysis JrAHLs

Members of the *AHL* family are partially involved in various plant biological processes via protein-protein interaction [23]. We constructed the protein interaction networks between the JrAHL proteins. Consistent with the previous report [16,24], the JrAHLs are shown to interact with each other (Figure 6). Additionally, JrAHL3/JrAHL8/JrAHL11/JrAHL13/JrAHL27/JrAHL33 were predicted in the networks, which were clustered with the physically-interacting proteins AHL3/AHL4 from Arabidopsis [23]. In *Poncirus trifoliata*, both PtrAHL14 and PtrAHL17 could interact with other proteins to target the same downstream genes [32]. However, the interaction between JrAHL proteins must be further verified in vitro and in vivo.

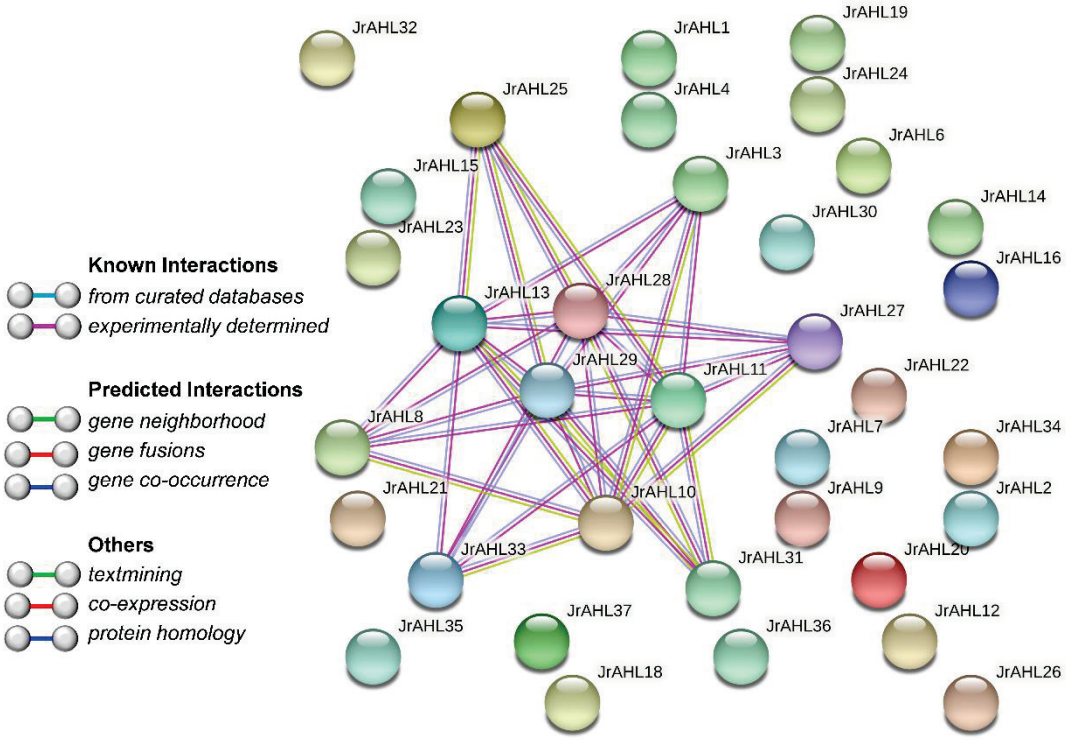


Figure 6. Predicted protein-protein interaction network of JrAHL protein. The network nodes represent proteins. The line color indicates the type of interaction evidence.

2.6. Gene Expression Patterns of JrAHLs

The *JrAHLs* expression profiles of F26 (anthracnose-resistant) and F423 (anthracnose-susceptible) fruit bracts at five-time points after infection were analyzed based on the transcriptomic data. In the anthracnose-resistant F26, more than half of the *JrAHLs* gene members were up-regulated within 24 hours after inoculation. However, the up-regulation of these *JrAHL* genes was delayed in the susceptible F423 (Figure 7). This result suggested that *JrAHLs* are involved in the immune regulation of pathogens. Expression profiles of *JrAHLs* at different developmental stages of endopleura, one of the walnut's most commercially essential organs, were analyzed. More than half of *JrAHL* genes showed high transcription levels in the early endopleura development stage (Figure 8). The tissue-specific expression profile of some *JrAHL* genes from different clades was determined by real-time quantitative PCR (Figure 9). *JrAHL1*, *JrAHL2*, *JrAHL7*, *JrAHL19*, and *JrAHL28* were highly expressed in the shoot tip. *JrAHL1*, *JrAHL2*, and *JrAHL7* also showed high expression levels in pistillate flowers. Two genes were preferentially expressed in the vegetative organs, such as *JrAHL10* in root and *JrAHL33* in leaf. The *JrAHL11* is highly expressed in fruit. In addition, *JrAHL14* and *JrAHL31* showed no tissue-specific expression.

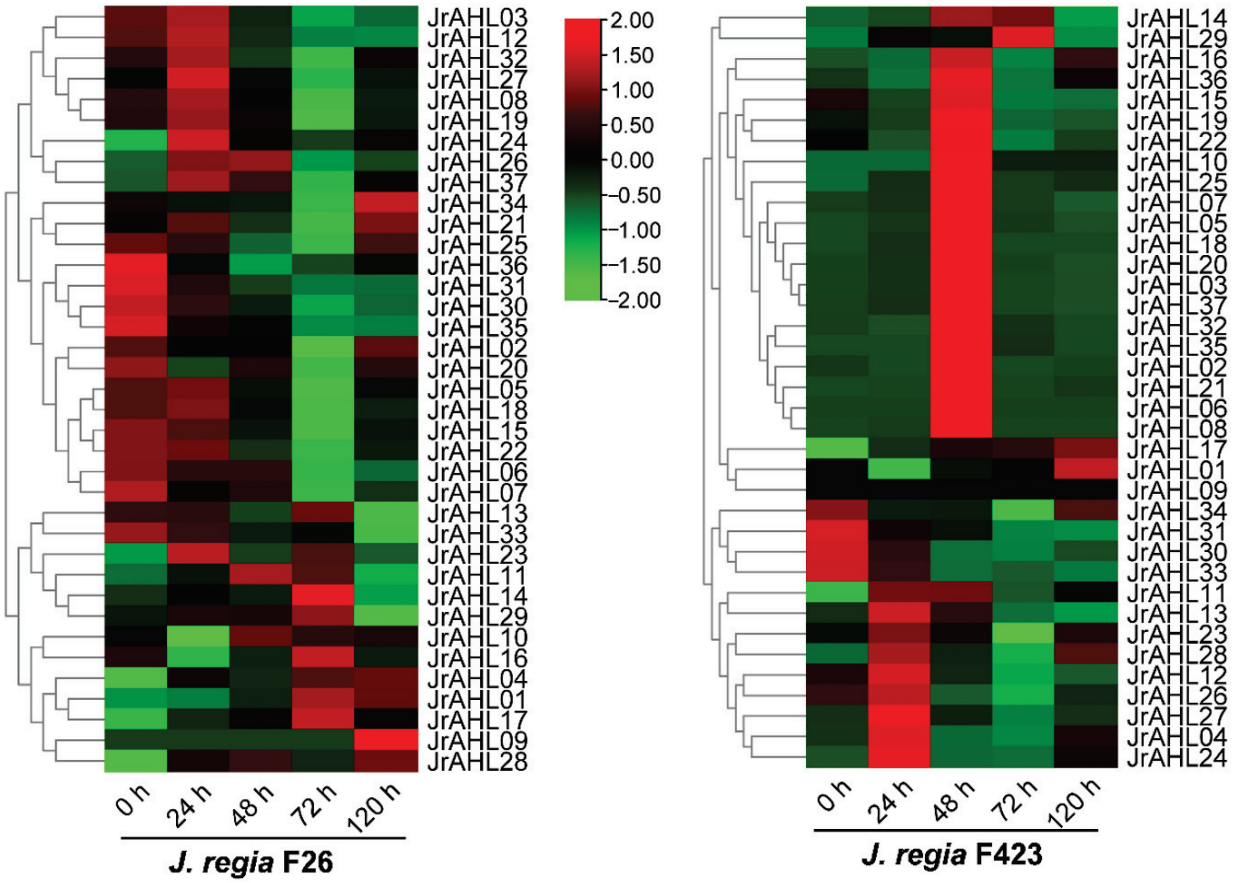


Figure 7. *JrAHL* gene expression profiles in F26 and the F423 fruits in response to anthracnose infection. The heat map was generated based on the transcriptomic data of the anthracnose-resistant F26 fruit bracts and anthracnose-susceptible F423 fruit bracts at five time points after infection with *C. gloeosporioides*.

2.7. Gene Cloning and Subcellular Localization of JrAHL2

JrAHL2 was cloned from walnut and found to contain a 927-bp CDS (Figure 10A) that encodes a protein of 307 amino acids. Multiple sequence alignment analysis results showed that the JrAHL2 shared 59.75% sequence identity with AtAHL22. The amino acid sequence within the AT-hook motif and the large PPC/DUF296 domain are highly conserved (Figure 10B). JrAHL2 protein is predicted to be localized in the nucleus (Supplementary Table S2). JrAHL2 was fused to a green fluorescent protein (EGFP) (Figure 10C) and transiently expressed in tobacco leaf cells. The GFP signal was targeted to the nucleus (Figure 10D), suggesting that JrAHL2 functions as a transcription factor.

2.8. Functional Analysis of JrAHL2 in Transgenic Arabidopsis

To dissect the function of *JrAHL2*, we generated *JrAHL2*-overexpressing (OE) transgenic Arabidopsis plants. The homozygous *JrAHL2*-OE plants flowered significantly later than the WT controls (Figure 11A,B). Additionally, the flowering genes *AtFT* is expressed at a lower level in the *JrAHL2*-OE Arabidopsis plants (Figure 11C). Overexpression of *JrAHL2* also leads to a dominant-negative long hypocotyl phenotype under both long day and dark conditions (Figure 12). Thus, *JrAHL2* inhibits floral transition and hypocotyl elongation.

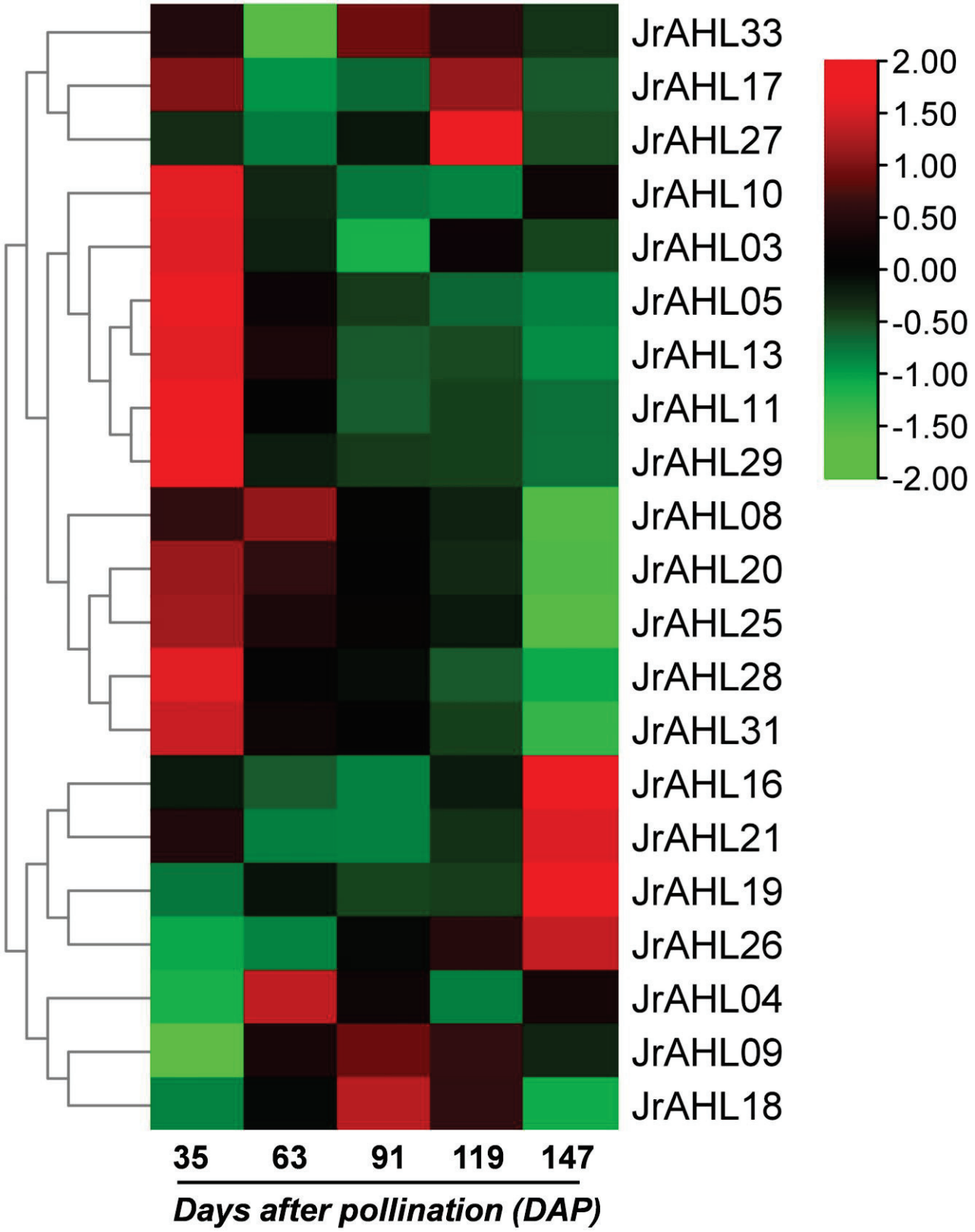


Figure 8. *JrAHL* gene expression profiles at different developmental stages of endopleura in walnut. The heat map was generated based on the transcriptomic data of the endopleura at different (35, 63, 91, 119, and 147) days after pollination (DAP).

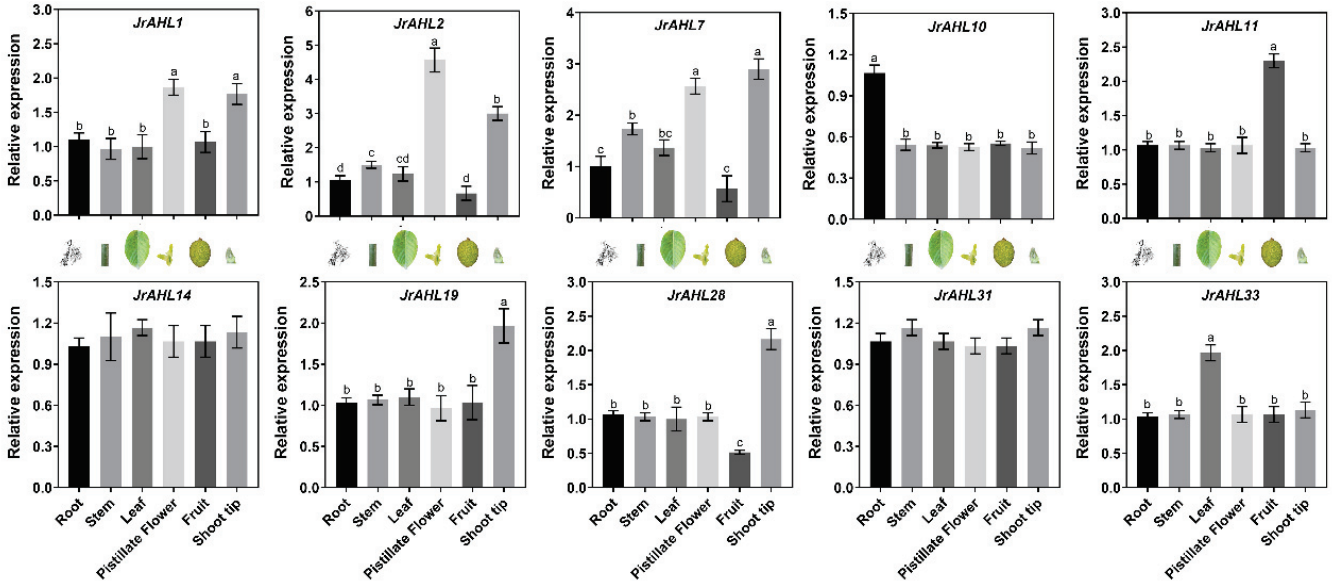


Figure 9. *JrAHL* gene expression profiles in different tissues. Each value represents the mean ± standard error of three replicates. Different letters (a, b, c, d) indicate significant differences, as determined by Student's *t*-test at the 0.05 level.

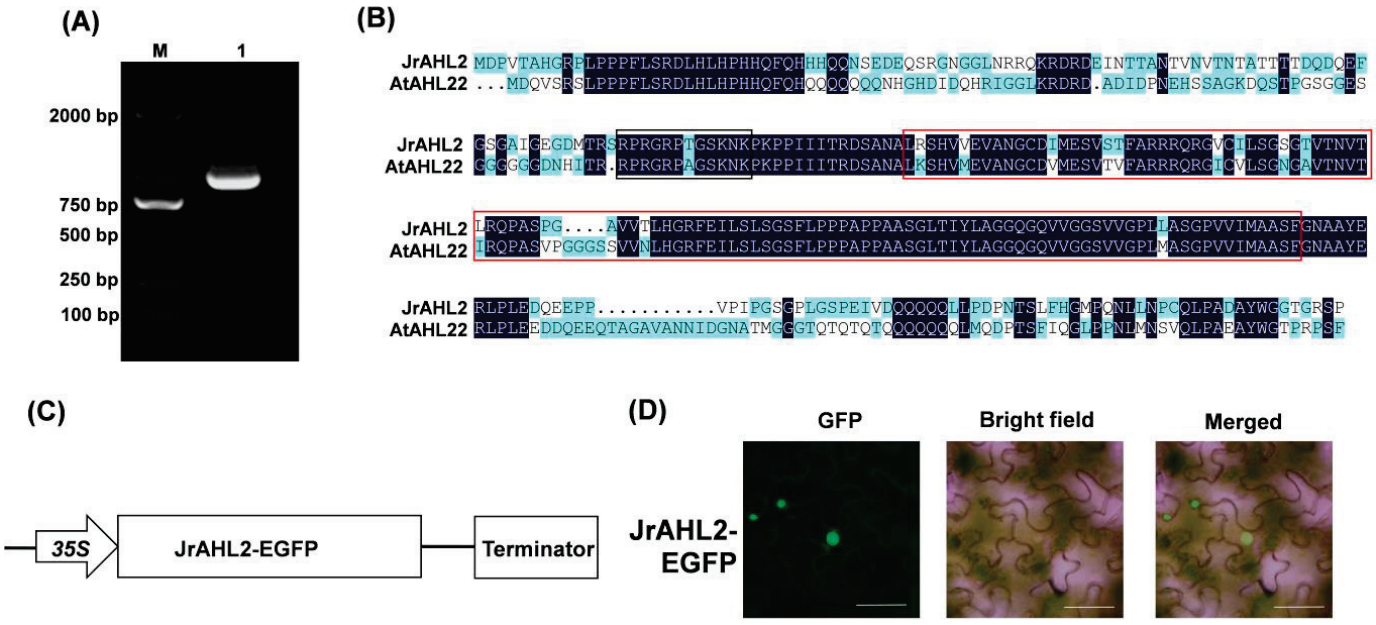


Figure 10. *JrAHL2* gene cloning and subcellular localization. (**A**) Gene cloning of *JrAHL2* coding region. (**B**) Putative protein sequence analysis of the cloned *JrAHL2* gene. (**C**) Schematic diagram of the *JrAHL2* overexpression vector for subcellular localization. (**D**) Subcellular localization of JrAHL2 in tobacco leaves.

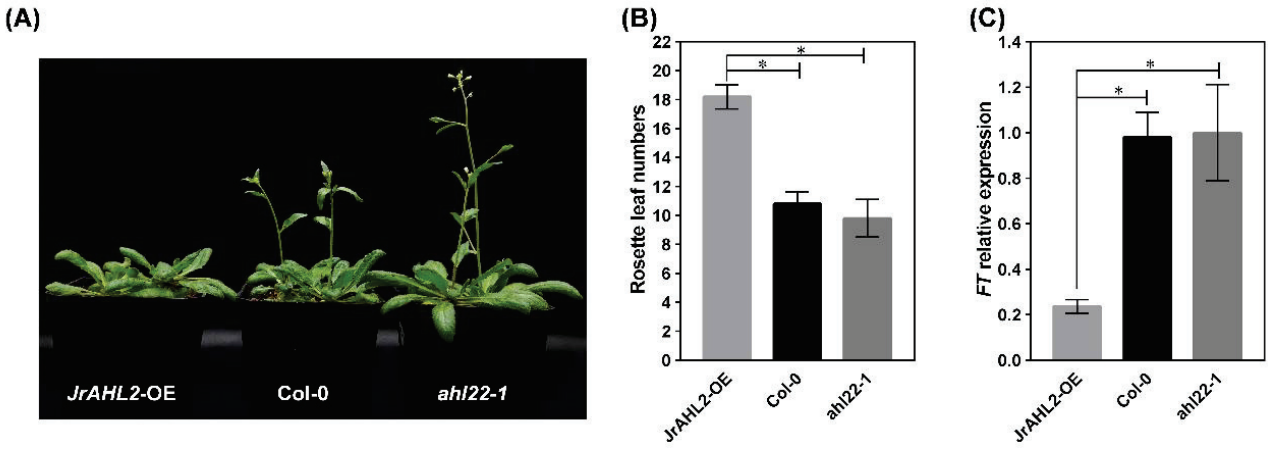


Figure 11. *JrAHL2* affects flowering in Arabidopsis. (**A**) Flowering phenotype of *JrAHL2*-overexpressing (*JrAHL2*-OE), wide type (Col-0), and *AtAHL* mutant (*ahl22-1*) Arabidopsis lines, respectively. (**B**) The number of rosette leaves at flowering. (**C**) The expression level of flowering-related gene *FT*. The columns represent the mean values of three replicates ± standard deviations. The asterisk (*) indicate significant differences, as determined by Student's *t*-test at the 0.05 level.

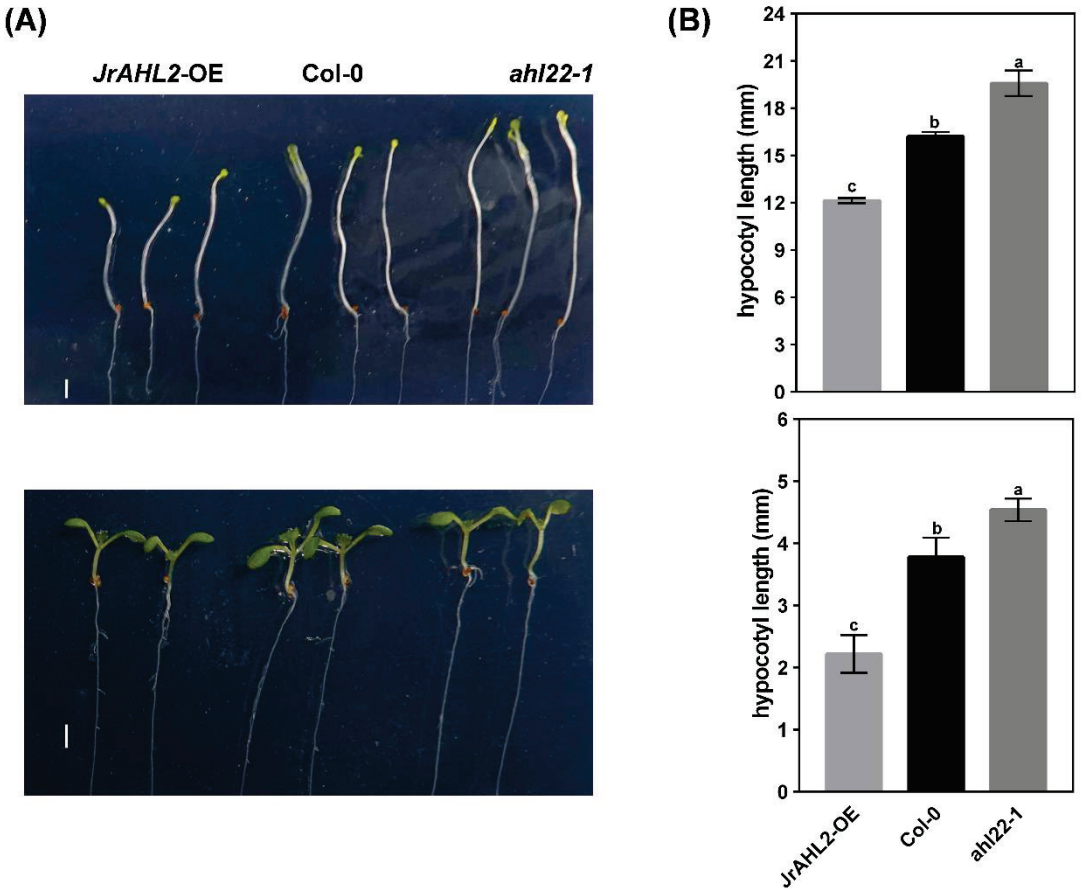


Figure 12. *JrAHL2* affects hypocotyl elongation. Samples were photographed (**A**), (**upper panel**) for darkness, and (**lower panel**) for normal photoperiod and the length of the hypocotyl was analyzed (**B**). Bar = 2 mm. The columns represent the mean values of three replicates ± standard deviations. Different letters (a, b, c) indicate significant differences, as determined by Student's *t*-test at the 0.05 level.

3. Discussion

3.1. Identification and Analysis of JrAHL Gene Family Members

The walnut (*Juglans regia* L.) is one of the most economically essential nut trees, and the research on walnut focused on flowering and molecular breeding in recent years. However, less attention was paid to walnut than other plants because of the problems in the walnut industry development, such as the long juvenile phase, difficulty in flowering, and hypocotyl elongation through the husk. The high-quality walnut genome has been published [30], providing a reference for systematic research on the genetic composition and potential function. To date, some gene families from walnut have been identified, such as the *Basic leucine zipper* (*bZIP*) [33] and *General Regulatory Factor* (*GRF*) [34] genes, which may be involved in the regulation of flowering. However, the role of the *AT-Hook Motif Containing Nuclear Localized* (*AHL*) gene family has not been explored in higher plants beyond herbaceous plants. In this work, 37 *JrAHL* members were identified in the walnut genome (Table 1). More member was identified in walnut compared to Arabidopsis (29 members) [6], probably because the walnut genome (540 Mb) is larger than that of Arabidopsis (125 Mb). The *JrAHL* members were divided into two Clades, based on the phylogenetic tree (Figure 1) and PPC/DUF domain (Figure 2) as reported in other species [1]. AHL proteins interact with each other through their PPC domain [35]. Thus, the variability increase in PPC domains may extend the range of biological functions. The Clade-B is further divided into two types based on the AT-hook motif, Type-II and Type-III. At the same time, the Clade-A is referred to as Type-I (Figure 2). Because the AT-hook motif is responsible for recognizing and binding a specific nucleotide sequence as well as changing the chromatin structure [36], it is, of course, possible that JrAHLs from the different branch can target different downstream genes. The sequences of conserved motifs (Figure 2), the gene structure (Figure 3), and the arrangement of distinct domains (Figure 3) and are highly consistent within each evolutionary branch. High consistency may be caused by segmental duplications (Figure 4) or the whole-genome duplication process [27], which are the major driving forces of *AHL* family evolution. Similarly, domain-swapping analyses showed that the N-terminus of AHL4 is required for its DNA-binding activity [4]. All *JrAHL* genes in Clade-B contain multiple exons, whereas all the Clade-A members are intronless (Figure 3). Introns are ubiquitous in eukaryotic genomes, which affect gene function by affecting gene expression regulation [37]. For example, the number of introns in soybeans certainly affects the expression of *AHL* genes [25]. In addition to regulating the rate of gene evolution [38], introns also provide specific functional elements for selective splicing and exon shuffling, and some functions of non-coding DNA [39], and in so doing allow the occurrence of diversity in JrAHLs protein accumulation, structures (Supplementary Figure S1), and physically interactions (Figure 6). Because the eukaryote genomes gained introns during or after prokaryote-eukaryote divergence [13], an early divergence occurred between the Type-I *JrAHLs* and the other two types (II and III) of *JrAHL* genes, and the *JrAHLs* of Clade-B in the walnut genome have diverged from Clade-A. Gene duplication, one of the most important features of plant genome structure, contributed to the evolution of novel functions [40] (Panchy et al., 2016). The ratio of non-synonymous to synonymous mutations in genes (Ka/Ks) of duplicated gene pairs was < 1, suggesting that purifying selection was the main source [40] of evolution for the *JrAHL* gene family. It is thus plausible that introns acquisition may cause the structural and functional changes of *JrAHLs* during duplication (Figures 3 and 4).

In addition to introns, the promoter, another important non-coding region, plays a critical role in regulating gene transcription. The *cis*-acting elements of some duplicated gene pairs are not the same, such as *JrAHL1/6* and *JrAHL4/9*. These differences would allow the regulation of different aspects of plant development and the responses to hormones and stresses. Studies of promoters, especially *cis*-acting elements (*CREs*) are crucial for improving our fundamental understanding of gene regulation and function towards different environmental cues and stress. The *CREs* that regulate plant development and confer resistance stresses are found in the *JrAHL* promoters (Figures 5 and S2). Stress-related hormone response elements, including

ABA, MeJA, and SA, are widely distributed on multiple *JrAHL* promoters. Fifteen *JrAHLs* contained GARE-motif, especially *JrAHL6,* contains up to four of these elements, suggesting a cross-talk between *JrAHLs* and GA hormones, such as negative feedback regulation as previously reported [17]. Aux-responsive *CREs* are predicted in ten *JrAHL* promoters. Specifically, *JrAHL28* is homologous to *AHL10* in Arabidopsis, which regulates genes related to the auxin signaling pathway, suggesting a cross-talk between *JrAHLs* and auxin [8,21]. AHL also negatively regulates jasmonic acid (JA) levels and affects the gene expression in the JA pathway [20]. Notably, two AT-Hook proteins, PtrAHL14/17 in *Poncirus trifoliata,* were reported to modulate cold tolerance [32]. Adversity stress-related *CREs,* including low temperature, drought, and defense, suggest that *JrAHLs* are relevant to controlling these processes. Four *JrAHL* genes contained the AT-rich element, which provided the binding site of AT-rich DNA binding protein (ATBP-1) [41]. It is plausible that there is reciprocal transcriptional regulation among the *JrAHLs* family members, in addition to the predicted protein-protein interactions (Figure 6). In addition, tissue-specific elements such as endosperm- and meristem-expression (Figures 5 and S2) suggested that *JrAHLs* regulate embryogenesis and meristem maturation like previously reported [35,42]. Specifically, *JrAHL17* contains both endosperm- and meristem-expression *CREs* and is homologous to *AtAHL10* from Arabidopsis, which affects the expression of the transcription factor *STM* (*Shootmeristemless*) required for meristem maintenance [21]. qRT-PCR revealed that the *JrAHL* genes containing meristem-expression tissue-specific elements in the promoter were highly expressed in the shoot tip (Figures 5 and 9). Transcriptome data facilitate a comprehensive understanding of expression levels of multiple genes. Multiple *JrAHL* genes showed differential expression by analyzing the transcriptome data of two walnut materials after inoculation with *Colletotrichum gloeosporioides*. In response to *C. gloeosporioides* infection, more than half of the *JrAHL* genes were rapidly activated in F26 at the initial stage of inoculation. At the same time, the activation of *JrAHL* in F234 was relatively delayed, suggesting that *JrAHL* is involved in the immune regulation of pathogens and that early expression may confer high resistance to anthracnose in F26 fruits (Figure 7). In fact, there is already established evidence that *AHL* is involved in plant immunity and positively regulates plant resistance to pathogen infection [20,43]. More than half of the *JrAHL* genes are highly expressed at the beginning of endocarp development (Figure 8), including *JrAHL11*, which showed a profound expression in walnut fruit (Figure 9). *JrAHL11* was clustered into the same clade with the Arabidopsis *AHL4*, and the latter is involved in the regulation of lipid mobilization and fatty acid β-oxidation [22], suggesting that the *JrAHL11* plays an important role in fruit development. The three detected Type I *JrAHL* genes, *JrAHL1*, *JrAHL2*, and *JrAHL7*, were highly expressed in pistillate flowers and shoot tip (Figure 9). *JrAHL1* was homologous to the *AHL20*, *JrAHL2* and *JrAHL7* were homologous to the *AHL22* (Figure 1) in Arabidopsis, respectively. Both *AHL20* and *AHL22* were involved in controlling flowering time in Arabidopsis [7,15,44], suggesting that these three genes are involved in floral regulation.

3.2. JrAHL2 Delayed Flowering and Inhibited Hypocotyl Elongation

Hypocotyl elongation is an essential event in the photomorphogenesis of walnut seeds, and flowering is an important trait for economic forest species. The functions and mechanisms of *AHL* genes from Arabidopsis have been extensively studied, whereas the function dissection information about the specific *JrAHL* gene in walnut is still unavailable. The *JrAHL2* were highly expressed in the pistillate flower and was homologous to the Arabidopsis *AHL22*, which stimulates us to investigate the function of *JrAHL2* in controlling flowering and hypocotyl elongation. The transgenic Arabidopsis with constitutive expression of the *JrAHL2* gene was prepared. Overexpression of *JrAHL2* caused a significant inhibition of hypocotyl elongation, whereas the Arabidopsis *ahl22* (homologous to *JrAHL2*) mutant is longer in the hypocotyl length than wide type (Figure 10) under both long-day conditions and darkness. It reported that AHL22, SOB3/AHL29, ESC/AHL27 [6–8] are negative modulators of hypocotyl growth in Arabidopsis [9]. This result indicates that *JrAHL2* plays a negative role in controlling hypocotyl elongation and functions in both

photomorphogensis and skotomorphogenesis. In addition to repressing the hypocotyl growth, *JrAHL2* also inhibits the transcription of *FT* and delays flowering (Figure 9). Nevertheless, no perturbations to flower phenotype and *FT* expression were recorded in the *ahl22* mutant (Figure 9), possibly because of extensive functional redundancy between *AtAHL22* and other *AHL* genes [15]. These results suggested that the *JrAHL2* has a much stronger effect on hypocotyl elongation than its effect on flowering regulation (Xiao et al., 2009). Additionally, despite the report that *AHL* regulates the expression of target genes, including *FT*, by directly binding to a specific nuclear MAR containing A/T-rich sequences within their promoter [15,32,44], the ubiquity of A/T-rich sequences in the promoter region makes it difficult to discriminate the elements specifically bound by *JrAHL2*. Thus, a systems approach such as ChIP-seq/qPCR could be used further to dissect the network of *JrAHL*.

4. Materials and Methods

4.1. Identification of the AHL Gene Family

For the discovery of the AHL family members in walnut, the high-quality common walnut (*Juglans regia* L.) reference genome data was downloaded [30]. The Hidden Markov Model scan (HMMscan) was used on the complete walnut protein database using PPC/DUF296 (PF03479) as well as AT-hook (PF02178) Pfam IDs obtained from the Pfam database (https://pfam.xfam.org/ (accessed on 7 January 2022)) [45] as queries. All non-redundant putative protein sequences were manually checked with the NCBI Conserved Domains Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi (accessed on 12 January 2022)). The unique sequences containing the AT-hook motif and the PPC domain were selected as putative JrAHLs. The online ExPASy program (http://www.expasy.org/tools/ (accessed on 12 February 2022)) was used to predict protein physicochemical characteristics. The online system Cell-PLoc (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/ (accessed on 15 February 2022)) [46] was used to predict the subcellular localization of these JrAHL proteins.

4.2. Phylogenetic Analysis of AHL Proteins

A Neighbor-Joining tree was constructed to infer the phylogenetic relationship between the AHL proteins from Arabidopsis and walnut. Sequence alignment was carried out using ClustalW in MEGA7 [47] with default parameters. The 1000 Bootstrap replicates were applied to present the evolutionary history. Multiple alignments were also performed with DNAMAN software (LynnonBiosoft, San Ramon, CA, USA. version 8.0).

4.3. Genes and Protein Structure Analysis

The GSDS 2.0 (Gene Structure Display Server, http://gsds.cbi.pku.edu.cn) was utilized to draw the structure of JrAHL genes. MEME (http://meme-suite.org/ (accessed on 15 April 2022)) was employed for functional motif analysis. The three-dimensional structure of JrAHL proteins was predicted by the PHYRE server v2.0 (http://www.sbg.bio.ic.ac.uk/phyre2 (accessed on 22 April 2022)).

4.4. Collinearity Analysis

The OrthoMCL algorithm [48] was used to identify paralogous genes within the walnut genome, and then the MCScan [49] algorithm was applied to detect syntenic blocks containing walnut *AHL*. The Circos [50] was used to visualize the syntenic relationships between the walnut genomes. The Ka/Ks ratio was calculated to estimate the selection pressure among duplicated *JrAHL* genes as previously described [51]. All plugins are used in TB tool software (TBtools-II v1.108) [52].

4.5. Promoter Sequence Analysis

The 1.5-kb upstream region of the start codon of each JrAHL gene was retrieved. The online PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/ (accessed on 16 July 2022)) was used to identify the potential cis-acting elements.

4.6. Interactive Protein Network Analysis

The STRING database predicted the interaction network of AHL proteins (https://string-db.org/ (accessed on 25 July 2022)). All JrAHL proteins were used as the query to search for the walnut data in the STRING database. Both the functional and physical protein associations were included in the network. The minimum required interaction score was selected as medium confidence (0.400).

4.7. Expression Profiles of JrAHL Genes

Gene expression data from the anthracnose-resistant F26 and the anthracnose-susceptible F423 fruits in response to anthracnose (*Colletotrichum gloeosporioides*) infection were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/ (accessed on 20 July 2022)) at accession GSE147083 [53]. Transcriptomic data of the walnut endopleura during its development was also downloaded from GEO database at accession GSE185230 [54]. The log2 transformed FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) values were used for preparing the heatmaps. Root, stem, leaf, flowers, fruits, and shoot tip were taken, frozen immediately in liquid nitrogen, and stored at −80 °C until use.

4.8. Quantitative RT-PCR (qRT-PCR)

Total RNA was extracted from different tissues using the RNA extraction kit (OMEGA, Doraville, GA, USA). The RNA concentration was determined by spectrophotometer, and RNA integrity was verified by electrophoresis. First-strand cDNA was synthesized using the EasyScript® One-Step cDNA Synthesis SuperMix (TransGen Biotech, China) following the manufacturer's instructions. Primer pairs for qRT-PCR were designed using Primer Premier 6.0 (Premier Biosoft, San Francisco, CA, USA) (Supplementary Table S1). qRT-PCR was performed on a Bio-Rad CFX96 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA). The *18S* rRNA gene was used as a standard internal control. The relative expression levels were calculated by the comparative $2^{-\Delta\Delta Ct}$ method [55].

4.9. Gene Cloning, Vector Construction, and Subcellular Localization

The coding sequence (CDS) of the *JrAHL2* was cloned from the walnut pistillate flower's cDNA and then inserted into the modified pCAMBIA2300-EGFP vector. Subcellular localization was carried out according to our previously published article [56]. The leaves of 5-week-old tobacco plants were infiltrated with the *Agrobacterium* strain GV3103 cells containing the recombinant vector. After 3 d, the green fluorescent protein signals were detected with an Olympus fluorescence microscope (Japan). All primers used for gene cloning and vector construction were listed in Supplementary Table S1.

4.10. Plant Transformation and Phenotype Analysis

Arabidopsis was transformed with the *Agrobacterium* strain GV3103 mentioned above, according to the *A. tumefaciens* mediated floral dip method [57]. The transformants were identified based on kanamycin resistance, and the T_3 generation seedlings were used for functional analyses. The flowering phenotype of different Arabidopsis lines, including *JrAHL2*-overexpressing (*JrAHL2-OE*), wild type (Col-0), and *AtAHL22* mutant (*ahl22-1*, Salk_018866), were identified according to the previous description [58]. To analyze the hypocotyl growth phenotype, Arabidopsis seeds were surface-sterilized and sown on Murashige & Skoog (MS) medium [56]. The plate was held at 4 °C for 3 days, exposed to white light for 4 h to stimulate germination, and then cultured for 6 days at 22 °C with a 16-h light/8-h dark photoperiod (light intensity of 150 μmol m^{-2} s^{-1}) (long day) or darkness.

5. Conclusions

Thirty-seven *AHL* genes were identified in the walnut (*Juglans regia* L.) genome and found to be unequally distributed on 16 chromosomes. The phylogenetic tree divided *JrAHLs* into two clades. Segmental duplications were the major driving forces of *AHL* gene family expansion and evolution. Promoter and expression analysis indicated that

JrAHLs were involved in multiple biological processes. Functional analysis showed that the nuclear-localized *JrAHL2* plays a negative role in floral transition and hypocotyl elongation. The first systematic exploration of the *AHL* gene family in walnut will provide helpful information for future work in walnut breeding.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24087244/s1.

Author Contributions: Conceptualization: P.J. and G.Q.; Data curation, P.J.; Formal analysis, K.Y., Q.D., H.L. (Haoan Luan) and S.G.; Funding acquisition, G.Q.; Investigation, H.L. (Han Li); Methodology, J.L.; Project administration, G.Q.; Resources, H.L. (Haoan Luan); Software, R.Y.; Supervision, G.Q.; Validation, X.Z.; Writing—review & editing, P.J. and G.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Key Research and Development Program grant number 2022YFD1600402. And The APC was funded by The National Key Research and Development Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are available from the corresponding author on request (bdqgh@hebau.edu.cn).

Acknowledgments: The authors would like to thank Rahat Sharif for the technical help.

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

References

1. Zhao, J.; Favero, D.S.; Qiu, J.; Roalson, E.H.; Neff, M.M. Insights into the evolution and diversification of the AT-hook Motif Nuclear Localized gene family in land plants. *BMC Plant Biol.* **2014**, *14*, 266. [CrossRef] [PubMed]
2. Zhao, K.; Kas, E.; Gonzalez, E.; Laemmli, U.K. SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. *EMBO J.* **1993**, *12*, 3237–3247. [CrossRef] [PubMed]
3. Huth, J.R.; Bewley, C.A.; Nissen, M.S.; Evans, J.N.; Reeves, R.; Gronenborn, A.M.; Clore, G.M. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **1997**, *4*, 657–665. [CrossRef] [PubMed]
4. Seo, M.; Lee, J.Y. Dissection of Functional Modules of AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN 4 in the Development of the Root Xylem. *Front. Plant Sci.* **2021**, *12*, 632078. [CrossRef]
5. Zhang, W.M.; Cheng, X.Z.; Fang, D.; Cao, J. AT-HOOK MOTIF NUCLEAR LOCALIZED (AHL) proteins of ancient origin radiate new functions. *Int. J. Biol. Macromol* **2022**, *214*, 290–300. [CrossRef]
6. Zhao, J.; Favero, D.S.; Peng, H.; Neff, M.M. *Arabidopsis thaliana* AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4688–E4697. [CrossRef]
7. Xiao, C.; Chen, F.; Yu, X.; Lin, C.; Fu, Y.F. Over-expression of an *AT-hook* gene, *AHL22*, delays flowering and inhibits the elongation of the hypocotyl in Arabidopsis thaliana. *Plant Mol. Biol.* **2009**, *71*, 39–50. [CrossRef]
8. Favero, D.S.; Jacques, C.N.; Iwase, A.; Le, K.N.; Zhao, J.; Sugimoto, K.; Neff, M.M. SUPPRESSOR OF PHYTOCHROME B4-#3 Represses Genes Associated with Auxin Signaling to Modulate Hypocotyl Growth. *Plant Physiol.* **2016**, *171*, 2701–2716. [CrossRef]
9. Street, I.H.; Shah, P.K.; Smith, A.M.; Avery, N.; Neff, M.M. The AT-hook-containing proteins SOB3/AHL29 and ESC/AHL27 are negative modulators of hypocotyl growth in Arabidopsis. *Plant J.* **2008**, *54*, 1–14. [CrossRef]
10. Sirl, M.; Snajdrova, T.; Gutierrez-Alanis, D.; Dubrovsky, J.G.; Vielle-Calzada, J.P.; Kulich, I.; Soukup, A. At-Hook Motif Nuclear Localised Protein 18 as a Novel Modulator of Root System Architecture. *Int. J. Mol. Sci.* **2020**, *21*, 1886. [CrossRef]
11. Zhang, S.; Wang, T.; Lima, R.M.; Pettko-Szandtner, A.; Kereszt, A.; Downie, J.A.; Kondorosi, E. Widely conserved AHL transcription factors are essential for *NCR* gene expression and nodule development in *Medicago*. *Nat. Plants* **2023**, *9*, 280–288. [CrossRef]
12. Xu, Y.; Gan, E.S.; Ito, T. The AT-hook/PPC domain protein TEK negatively regulates floral repressors including *MAF4* and *MAF5*. *Plant Signal. Behav.* **2013**, *8*, e25006. [CrossRef]
13. Jo, B.S.; Choi, S.S. Introns: The Functional Benefits of Introns in Genomes. *Genomics. Inform.* **2015**, *13*, 112–118. [CrossRef]
14. Jia, Q.S.; Zhu, J.; Xu, X.F.; Lou, Y.; Zhang, Z.L.; Zhang, Z.P.; Yang, Z.N. Arabidopsis AT-hook protein TEK positively regulates the expression of arabinogalactan proteins for Nexine formation. *Mol. Plant* **2015**, *8*, 251–260. [CrossRef]
15. Yun, J.; Kim, Y.S.; Jung, J.H.; Seo, P.J.; Park, C.M. The AT-hook motif-containing protein AHL22 regulates flowering initiation by modifying *FLOWERING LOCUS T* chromatin in Arabidopsis. *J. Biol. Chem.* **2012**, *287*, 15307–15316. [CrossRef]

16. Favero, D.S.; Kawamura, A.; Shibata, M.; Takebayashi, A.; Jung, J.H.; Suzuki, T.; Jaeger, K.E.; Ishida, T.; Iwase, A.; Wigge, P.A.; et al. AT-Hook Transcription Factors Restrict Petiole Growth by Antagonizing PIFs. *Curr. Biol.* **2020**, *30*, 1454–1466.e1456. [CrossRef]
17. Matsushita, A.; Furumoto, T.; Ishida, S.; Takahashi, Y. AGF1, an AT-hook protein, is necessary for the negative feedback of *AtGA3ox1* encoding GA 3-oxidase. *Plant Physiol.* **2007**, *143*, 1152–1162. [CrossRef]
18. Kim, S.Y.; Kim, Y.C.; Seong, E.S.; Lee, Y.H.; Park, J.M.; Choi, D. The chili pepper CaATL1: An AT-hook motif-containing transcription factor implicated in defence responses against pathogens. *Mol. Plant Pathol.* **2007**, *8*, 761–771. [CrossRef]
19. Lu, H.; Zou, Y.; Feng, N. Overexpression of AHL20 negatively regulates defenses in Arabidopsis. *J. Integr. Plant Biol.* **2010**, *52*, 801–808. [CrossRef]
20. Rayapuram, N.; Jarad, M.; Alhoraibi, H.M.; Bigeard, J.; Abulfaraj, A.A.; Volz, R.; Mariappan, K.G.; Almeida-Trapp, M.; Schloffel, M.; Lastrucci, E.; et al. Chromatin phosphoproteomics unravels a function for AT-hook motif nuclear localized protein AHL13 in PAMP-triggered immunity. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2004670118. [CrossRef]
21. Wong, M.M.; Bhaskara, G.B.; Wen, T.N.; Lin, W.D.; Nguyen, T.T.; Chong, G.L.; Verslues, P.E. Phosphoproteomics of *Arabidopsis* Highly ABA-Induced1 identifies AT-Hook-Like10 phosphorylation required for stress growth regulation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2354–2363. [CrossRef] [PubMed]
22. Cai, G.; Kim, S.C.; Li, J.; Zhou, Y.; Wang, X. Transcriptional Regulation of Lipid Catabolism during Seedling Establishment. *Mol. Plant* **2020**, *13*, 984–1000. [CrossRef] [PubMed]
23. Zhou, J.; Wang, X.; Lee, J.Y.; Lee, J.Y. Cell-to-cell movement of two interacting AT-hook factors in Arabidopsis root vascular tissue patterning. *Plant Cell* **2013**, *25*, 187–201. [CrossRef] [PubMed]
24. Gallavotti, A.; Malcomber, S.; Gaines, C.; Stanfield, S.; Whipple, C.; Kellogg, E.; Schmidt, R.J. BARREN STALK FASTIGIATE1 is an AT-hook protein required for the formation of maize ears. *Plant Cell* **2011**, *23*, 1756–1771. [CrossRef] [PubMed]
25. Wang, M.; Chen, B.; Zhou, W.; Xie, L.; Wang, L.; Zhang, Y.; Zhang, Q. Genome-wide identification and expression analysis of the *AT-hook Motif Nuclear Localized* gene family in soybean. *BMC Genomics* **2021**, *22*, 361. [CrossRef]
26. Zhao, L.; Lu, Y.; Chen, W.; Yao, J.; Li, Y.; Li, Q.; Pan, J.; Fang, S.; Sun, J.; Zhang, Y. Genome-wide identification and analyses of the *AHL* gene family in cotton (*Gossypium*). *BMC Genom.* **2020**, *21*, 69. [CrossRef]
27. Zhang, W.M.; Fang, D.; Cheng, X.Z.; Cao, J.; Tan, X.L. Insights Into the Molecular Evolution of AT-Hook Motif Nuclear Localization Genes in *Brassica napus*. *Front. Plant Sci.* **2021**, *12*, 714305. [CrossRef]
28. Li, X.; He, H.; Wang, H.; Wu, X.; Wang, H.; Mao, J. Identification and expression analysis of the AHL gene family in grape (*Vitis vinifera*). *Plant Gene* **2021**, *26*, 100285. [CrossRef]
29. Machaj, G.; Grzebelus, D. Characteristics of the *AT-Hook Motif Containing Nuclear Localized* (*AHL*) Genes in Carrot Provides Insight into Their Role in Plant Growth and Storage Root Development. *Genes* **2021**, *12*, 764. [CrossRef]
30. Zhang, J.; Zhang, W.; Ji, F.; Qiu, J.; Song, X.; Bu, D.; Pan, G.; Ma, Q.; Chen, J.; Huang, R.; et al. A high-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. *Plant Biotechnol. J.* **2020**, *18*, 1848–1850. [CrossRef]
31. Bishop, E.H.; Kumar, R.; Luo, F.; Saski, C.; Sekhon, R.S. Genome-wide identification, expression profiling, and network analysis of AT-hook gene family in maize. *Genomics* **2020**, *112*, 1233–1244. [CrossRef]
32. Dahro, B.; Wang, Y.; Khan, M.; Zhang, Y.; Fang, T.; Ming, R.; Li, C.; Liu, J.H. Two AT-Hook proteins regulate A/NINV7 expression to modulate sucrose catabolism for cold tolerance in *Poncirus trifoliata*. *New Phytol.* **2022**, *235*, 2331–2349. [CrossRef]
33. Zhang, Z.; Quan, S.; Niu, J.; Guo, C.; Kang, C.; Liu, J.; Yuan, X. Genome-Wide Identification, Classification, Expression and Duplication Analysis of bZIP Family Genes in *Juglans regia* L. *Int. J. Mol. Sci.* **2022**, *23*, 5961. [CrossRef]
34. Zhang, Z.; Quan, S.; Niu, J.; Guo, C.; Kang, C.; Liu, J.; Yuan, X. Comprehensive Identification and Analyses of the *GRF* Gene Family in the Whole-Genome of Four *Juglandaceae* Species. *Int. J. Mol. Sci.* **2022**, *23*, 12663. [CrossRef]
35. Karami, O.; Rahimi, A.; Khan, M.; Bemer, M.; Hazarika, R.R.; Mak, P.; Compier, M.; van Noort, V.; Offringa, R. A suppressor of axillary meristem maturation promotes longevity in flowering plants. *Nat. Plants* **2020**, *6*, 368–376. [CrossRef]
36. Aravind, L.; Landsman, D. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* **1998**, *26*, 4413–4421. [CrossRef]
37. Castillo-Davis, C.I.; Mekhedov, S.L.; Hartl, D.L.; Koonin, E.V.; Kondrashov, F.A. Selection for short introns in highly expressed genes. *Nat. Genet.* **2002**, *31*, 415–418. [CrossRef]
38. Roy, S.W.; Gilbert, W. The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat. Rev. Genet.* **2006**, *7*, 211–221. [CrossRef]
39. North, K.; Benbarche, S.; Liu, B.; Pangallo, J.; Chen, S.; Stahl, M.; Bewersdorf, J.P.; Stanley, R.F.; Erickson, C.; Cho, H.; et al. Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat. Biotechnol.* **2022**, *40*, 1103–1113. [CrossRef]
40. Panchy, N.; Lehti-Shiu, M.; Shiu, S.H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef]
41. Tjaden, G.; Coruzzi, G.M. A novel AT-rich DNA binding protein that combines an HMG I-like DNA binding domain with a putative transcription domain. *Plant Cell* **1994**, *6*, 107–118. [CrossRef] [PubMed]
42. Karami, O.; Rahimi, A.; Mak, P.; Horstman, A.; Boutilier, K.; Compier, M.; van der Zaal, B.; Offringa, R. An Arabidopsis AT-hook motif nuclear protein mediates somatic embryogenesis and coinciding genome duplication. *Nat. Commun.* **2021**, *12*, 2508. [CrossRef] [PubMed]

43. Liu, H.; Dong, S.; Li, M.; Gu, F.; Yang, G.; Guo, T.; Chen, Z.; Wang, J. The Class III peroxidase gene *OsPrx30*, transcriptionally modulated by the AT-hook protein OsATH1, mediates rice bacterial blight-induced ROS accumulation. *J. Integr Plant Biol.* **2021**, *63*, 393–408. [CrossRef] [PubMed]
44. Tayengwa, R.; Sharma Koirala, P.; Pierce, C.F.; Werner, B.E.; Neff, M.M. Overexpression of *AtAHL20* causes delayed flowering in Arabidopsis via repression of *FT* expression. *BMC Plant Biol.* **2020**, *20*, 559. [CrossRef]
45. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic. Acids Res.* **2021**, *49*, D412–D419. [CrossRef]
46. Chou, K.C.; Shen, H.B. Cell-PLoc: A package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.* **2008**, *3*, 153–162. [CrossRef]
47. 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]
48. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef]
49. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]
50. 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]
51. Zhang, Z. KaKs_Calculator 3.0: Calculating Selective Pressure on Coding and Non-coding Sequences. *Genom. Proteom. Bioinform.* **2022**, *20*, 536–540. [CrossRef]
52. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]
53. Fang, H.; Liu, X.; Dong, Y.; Feng, S.; Zhou, R.; Wang, C.; Ma, X.; Liu, J.; Yang, K.Q. Transcriptome and proteome analysis of walnut (*Juglans regia* L.) fruit in response to infection by *Colletotrichum gloeosporioides*. *BMC Plant Biol.* **2021**, *21*, 249. [CrossRef]
54. Ji, F.; Ma, Q.; Zhang, W.; Liu, J.; Feng, Y.; Zhao, P.; Song, X.; Chen, J.; Zhang, J.; Wei, X.; et al. A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. *Genome. Biol.* **2021**, *22*, 300. [CrossRef]
55. 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]
56. Jia, P.; Xing, L.; Zhang, C.; Zhang, D.; Ma, J.; Zhao, C.; Han, M.; Ren, X.; An, N. MdKNOX19, a class II knotted-like transcription factor of apple, plays roles in ABA signalling/sensitivity by targeting ABI5 during organ development. *Plant Sci.* **2021**, *302*, 110701. [CrossRef]
57. Zhang, X.; Henriques, R.; Lin, S.S.; Niu, Q.W.; Chua, N.H. *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **2006**, *1*, 641–646. [CrossRef]
58. Jia, P.; Xing, L.B.; Zhang, C.G.; Chen, H.; Li, Y.M.; Zhang, D.; Ma, J.J.; Zhao, C.P.; Han, M.Y.; Ren, X.L.; et al. MdKNOX15, a class I knotted-like transcription factor of apple, controls flowering and plant height by regulating GA levels through promoting the *MdGA2ox7* transcription. *Environ. Exp. Bot.* **2021**, *185*, 104411. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

Identification and Characterization of *PRE* Genes in Moso Bamboo (*Phyllostachys edulis*)

Sujin Zheng [1,2], Kihye Shin [1,3], Wenxiong Lin [1], Wenfei Wang [1,*] and Xuelian Yang [1,*]

[1] College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2] College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[3] Department of Microbiology and Immunology, Jeju National University College of Medicine, Jeju 63243, Republic of Korea
* Correspondence: wenfeiwang@fafu.edu.cn (W.W.); xlyang@fafu.edu.cn (X.Y.)

Abstract: Basic helix–loop–helix (bHLH)/HLH transcription factors are involved in various aspects of the growth and development of plants. Here, we identified four HLH genes, *PePRE1-4*, in moso bamboo plants that are homologous to *Arabidopsis PRE* genes. In bamboo seedlings, *PePRE1/3* were found to be highly expressed in the internode and lamina joint by using quantitative RT-PCR analysis. In the elongating internode of bamboo shoots, *PePRE* genes are expressed at higher levels in the basal segment than in the mature top segment. Overexpression of *PePRE*s (*PePRE*s-*OX*) in *Arabidopsis* showed longer petioles and hypocotyls, as well as earlier flowering. *PePRE1* overexpression restored the phenotype due to the deficiency of *AtPRE* genes caused by artificial micro-RNA. *PePRE1-OX* plants showed hypersensitivity to propiconazole treatment compared with the wild type. In addition, PePRE1/3 but not PePRE2/4 proteins accumulated as punctate structures in the cytosol, which was disrupted by the vesicle recycling inhibitor brefeldin A (BFA). *PePRE* genes have a positive function in the internode elongation of moso bamboo shoots, and overexpression of PePREs genes promotes flowering and growth in *Arabidopsis*. Our findings provided new insights about the fast-growing mechanism of bamboo shoots and the application of *PRE* genes from bamboo.

Keywords: HLH; PRE; fast-growth; shoot elongation; moso bamboo; transcription factor; flowering

Citation: Zheng, S.; Shin, K.; Lin, W.; Wang, W.; Yang, X. Identification and Characterization of *PRE* Genes in Moso Bamboo (*Phyllostachys edulis*). *Int. J. Mol. Sci.* **2023**, *24*, 6886. https://doi.org/10.3390/ijms24086886

Academic Editor: Wajid Zaman

Received: 24 February 2023
Revised: 24 March 2023
Accepted: 4 April 2023
Published: 7 April 2023

1. Introduction

Moso bamboo (*Phyllostachys edulis*) is one of the fastest growing non-timber tree plants all over the world, with considerable ecological, economic, and cultural value [1,2]. In spring, it can grow up to 1 m in less than 24 h and reach a final height of 20 m in 45–60 days [3]. The rapid expansion of bamboo stems is driven by the cell division and elongation of internodes, which are regulated by a combination of endogenous phytohormones and environmental factors such as auxin, gibberellin acid (GA), brassinosteroids (BRs), and light [4]. Recent studies demonstrate that the BR, auxin, GA, and phytochrome pathways converge through direct interactions among their transcription factors/regulators, then pass to a tripartite module of helix-loop-helix (HLH) and basic helix–loop–helix (bHLH) factors, which is named the HHbH module [5]. The HLH/bHLH cascade regulates cell elongation downstream of various hormonal and environmental signaling pathways [6,7].

Paclobutrazol-resistant (PRE) proteins are a typical bHLH transcription factors, homologs to the human Id-1 (inhibitor of DNA binding 1) protein which lacks the basic domain required for DNA binding. PRE proteins dimerize with other bHLH factors to inhibit their DNA-binding activity [8,9]. IBH1 (ILI1 binding bHLH protein) interacts with HBI1 (homolog of *BEE2* interacting with IBH1), a positive regulator of cell elongation, and inhibits its transcriptional activity, thereby promoting the hypocotyl elongation by inhibiting the DNA-binding activity of HBI1 [10,11]. PRE1 and ILI1 promote cell elongation both in *Arabidopsis* and rice by interacting with IBH1 and forming a pair of antagonistic

HLH/bHLH transcriptional factors that function downstream of BZR1 (brassinazole resistant 1) to mediate BRs regulation of cell elongation [12]. PIF4 (phytochrome-interacting factor 4) had a major role in the multiple signal integration for plant growth regulation [13]. PIF4 and BZR1 are direct targets of PRE1, PRE5, and PRE6/KIDARI [14]. *PRE3/TOM7* is involved in regulating a plant's growth response to light signals [9]. PAR1 (phytochrome rapidly regulated 1)–PRE1 and PAR1–PIF4 heterodimers form a complex HLH/bHLH network that controls cell elongation and plant development in response to light [15]. PRE6 regulates photomorphogenesis by inhibiting the activity of HFR1 (the long hypocotyl in far-red 1) [16,17]. Another study showed the CIB1 (cryptochrome-interacting bHLH 1)–PAR1 and PIF4–PAR1/HFR1 systems antagonistically regulate cell elongation in response to light and high temperature [18]. Auxin works independently of and in conjunction with the PIF and GA pathways to regulate expression of growth-associated genes in cell elongation [19,20].

On the other hand, GA, BR, and auxin enhance cell elongation by inhibiting many inhibitory bHLH factors by inducing the expression of *PRE* family genes [6,16]. Overexpression of *PRE1* suppressed GA-deficient phenotypes of the *ga2* mutant and impacted several aspects of GA-dependent response, indicating that PRE1 plays a regulatory function in GA-dependent development in *Arabidopsis* [21]. Overexpression of *PRE3* suppresses the dwarf phenotype of the *bri1-301* mutant, indicating *PRE*s are involved and functional redundancy in BR signaling and response [9]. PRE1 acts downstream of ARF10 (auxin response factor 10) in regulating hypocotyl elongation, whereas PRE6 is a transcriptional repressor that is directly regulated by ARF5 and ARF8 in *Arabidopsis* [22].

PRE homologs are also involved in the regulation of plant development in other species. In rice, the *PRE* homologous gene *OsILI1* (*increased leaf inclination 1*) stimulates cell elongation in the lamina joint and *OsILI1* overexpression results in a large leaf angle phenotype [10]. FaPRE (*Fragaria* × *ananassa* paclobutrazol resistant) promotes the expression of genes involved in the ripening process while suppressing the expression of growth-promoting genes in the receptacle of octoploid strawberries [23]. *GhPRE1* (*Gossypium hirsutum* paclobutrazol resistant 1) overexpression results in longer fibers with higher quality characteristics in cotton [24]. In a recent study, moso bamboo *PRE* homolog genes were downregulated by a BR biosynthesis inhibitor (propiconazole (PPZ)) treatment, according to the transcriptome profile [25].

Recently, transcriptome sequencing and epigenetic-modification profiling in fast-growing moso bamboo shoots identified a large number of putative fast-growing genes [3]. Among these genes, the transcription factors act as regulatory switches for gene expression, which will deepen and expand understanding the fast-growth mechanism. However, the role of moso bamboo *PRE* genes, putative growth-promoting transcription factors, remains unclear. In this study, we identified four *PRE* genes from the moso bamboo genome. Specific expression patterns of *PePRE* genes were observed in the seedling and elongating bamboo shoots, as well as in the different parts of the internodes. Overexpressing *PePRE*1 in *Arabidopsis* resulted in phenotypes including a long petiole, slightly pale green leaves, and early flowering. PRE1 and PRE3 proteins accumulated as punctate structures in the cytosol. Our study suggests that *PePRE*s may play a positive role in the fast-growth processes of moso bamboo. Our results will benefit the future identification of more growth-promoting genes from moso bamboo genome and will provide basis for further functional characterizations of *PRE* family genes.

2. Results

2.1. Identification of PePREs in Moso Bamboo

Four homologous *PRE* genes were identified from the moso bamboo genome by BLAST-P searching in the moso bamboo protein database (http://www.bamboogdb.org/, accessed on 8 May 2017) [26] with AtPREs amino acids sequences. These were named as *PePRE1* (PH01000065G2010), *PePRE2* (PH01000068G1560), *PePRE3* (PH01000519G0840), and *PePRE4* (PH01000960G0260). When compared with rice and *Arabidopsis* PREs, all

bamboo PRE proteins are clustered with rice PREs but not AtPREs (Figure 1b). Protein sequence alignment of the PREs showed that all bamboo PREs proteins have highly conserved helix–loop–helix (HLH) domains but not basic domains that are critical for DNA binding (Figure 1a). The expression patterns of *PePRE* genes were investigated in different tissues of moso bamboo seedlings by quantitative reverse transcription PCR (qRT-PCR) (Figure 1c,d). *PePRE2* and *PePRE3* are found highly expressed in roots, whereas *PePRE1* and *PePRE4* were hardly detected (Figure 1d). Interestingly, only the *PePRE1* gene was detected in leaves, suggesting its unique function in leaf development (Figure 1d). Consistent with a previous report about the critical function of *PRE* genes on the lamina joint bending [12], four bamboo *PRE* genes were found to be expressed in the lamina joint (Figure 1d). As all *PePREs* were expressed in the internode and sheath, *PePRE1* and *PePRE3* showed predominant expressions in the sheath and internode, respectively (Figure 1d). Together, these results suggested that *PePRE* genes showed tissue-specific patterns in moso bamboo seedlings.

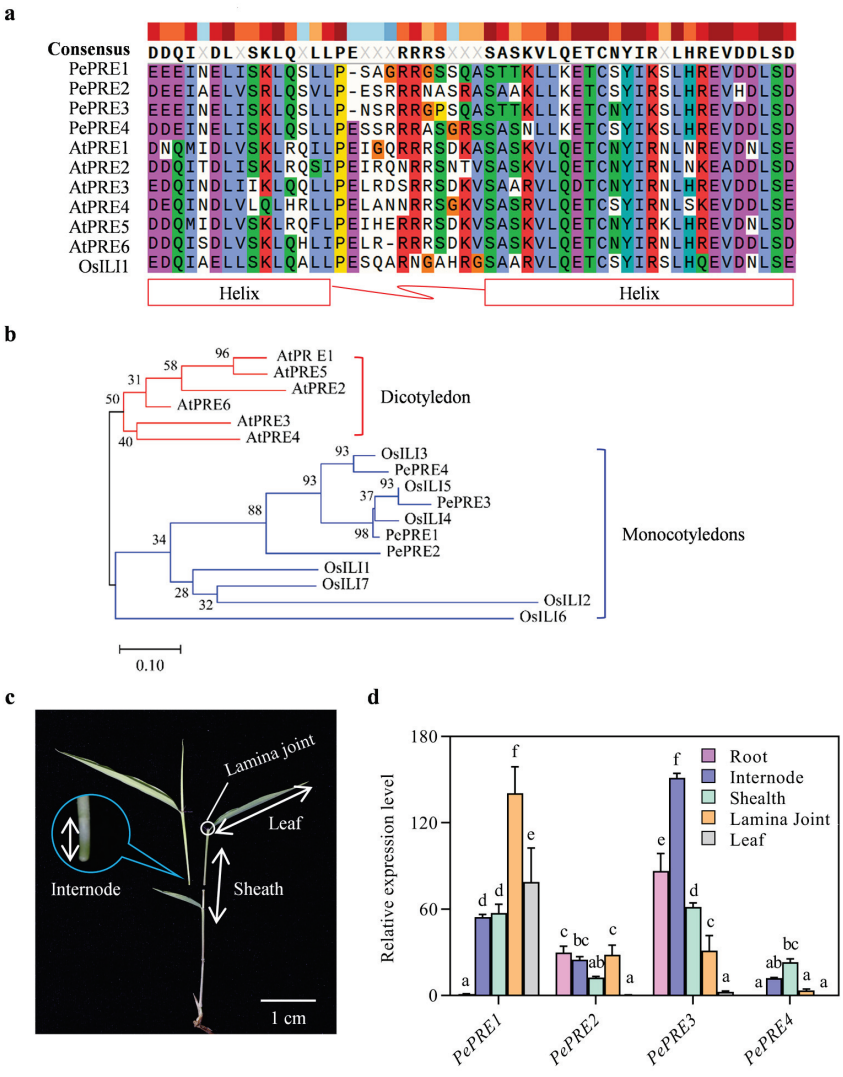


Figure 1. *PePREs* in moso bamboo. (**a**) Sequences alignment of PRE proteins. Identical or similar amino acid residues are indicated by colorful shading and the helix and loop domains are highlighted. (**b**) Phylogenetic analysis of PRE proteins from rice and *Arabidopsis*. Full amino acid sequences were used. Bootstrap values of 1000 replications were shown. Pe, *Phyllostachys edulis*; *Os*, *Oryza sativa*; *At*, *Arabidopsis thaliana*. (**c**) The tissue of the aerial part of moso bamboo seedling. Scale bar represents 1 cm. (**d**) Expression level of *PePREs* in different tissues of 3-week-old seedlings by qRT-PCR analysis. Different letters above the data columns indicate significant differences ($p < 0.05$; Duncan's test).

2.2. Expression of PePRE Genes in Elongating Bamboo Shoots

Bamboo shoot growth is attributed to cell proliferation in the intercalary meristem and subsequent cell elongation in the elongation zone of the internodes. The expression patterns of bamboo *PRE* genes were investigated in the elongating bamboo shoot (Figure 2a). *PePRE1* and *PePRE3* transcript levels were highest in the elongating internode (EIN) and scale leaves, whereas *PePRE2* and *PePRE4* transcript levels were highest in the scale leaves (Figure 2b). All *PePRE*s showed higher expression levels in the internodes than in the nodes (Figure 2b). To further explore the role of *PePRE* genes in internode growth, the elongating internode of bamboo shoots was divided into top, middle, and basal regions (Figure 2a bottom right). Generally, expression of *PePRE* genes accumulated in the middle and basal regions, whereas *PePRE*s were hardly detected in the upper parts. *PePRE1* and *PePRE3* were predominantly expressed in the middle and basal regions (Figure 2c). There was high *PePRE2* expression in scale leaves, followed by the unelongated internodes. Interestingly, *PePRE4* showed the lowest expression levels in the internodes and had relatively higher expression in the upper parts. Collectively, these data suggest that *PePRE*s have a prevalent accumulation in the elongating tissue of bamboo shoots.

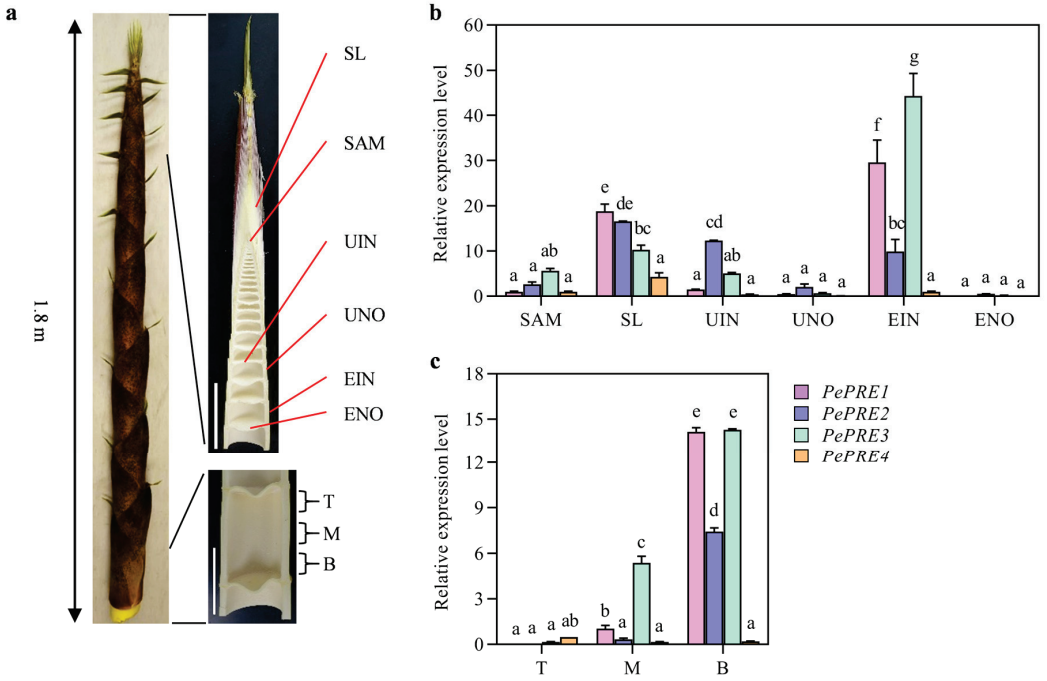


Figure 2. Expression pattern of *PePRE*s in the elongating bamboo shoot. (**a**) Structure of an elongating moso bamboo shoot. Samples are collected from a 1.8 m bamboo shoot. SAM, shoot apical meristem; SL, scale leaf; UIN, unelongated internode; UNO, unelongated node; EIN, elongating internode; ENO, elongating node. Scale bar represents 10 cm. (**b**) Expression levels of *PePRE* genes in the elongating bamboo shoot. (**c**) Expression levels of *PePRE* genes in different parts of an elongating internode. T, top part; M, middle part; B, bottom part. Different letters above the data columns indicate significant differences ($p < 0.05$; Duncan's test).

2.3. Overexpressing PePREs in Arabidopsis

To further understand the biological function of *PePRE*s, we generated overexpression transgenic lines with the 35S promoter driving the *PePRE* cDNAs fused with green fluorescent protein (GFP) in a Col-0 background. A total of 59, 42, 64, and 14 transgenic lines for PePRE1, 2, 3, and 4 were generated, respectively, with 31 (52.5%), 34 (81%), 22 (34.4%), and 3 (21.4%) lines for each gene, respectively, showing an early flowering phenotype. Days to bolting for *PePRE1-OX* #19 and *PePRE2-OX* #21 were found to be reduced by 13.3% and 14.0%, respectively, whereas the total rosette leaf numbers of *PePRE-OX* plants were also reduced with exception of *PePRE4-OX*. Under long-day conditions, *PePRE1-OX* transgenic

plants exhibited longer petioles and had paler green leaves than control plants (Figure 3a), which were similar to the *AtPRE1* overexpression lines reported in a previous study [27]. This suggested that *PePREs* have a conserved function similar to *AtPRE1* on flowering control. A total of 45% and 82.5% of *PePRE1* and *PePRE3* overexpression transgenic lines, respectively, showed a stem-bending phenotype, which was mainly due to a broken stem with a longitudinal crack (Figure A2). We further crossed *PePRE1-OX* transgenic plants with *pre-amiR* transgenic lines, in which four *AtPREs* (*AtPRE1/2/5/6*) were knocked-down using artificial microRNA [14]. Due to the lack of *AtPREs*, *pre-amiR* exhibited extreme dwarfism, delayed flowering, and had a reduced fertility phenotype [21]. All F1 plants showed a normal phenotype similar to the wild type (Figure 3d). Overexpression of *PePRE1* rescued the *AtPRE* (*1/2/5/6*)-deficient phenotypes of the *pre-amiR* mutant, including late flowering and curled leaves. To exclude *AtPRE1* expression interference, the expression level of *AtPRE1* was detected in all plants. Only in the *AtPRE1-OX* line did the *AtPRE1* expression level increase; there was no difference in the Col-0 and *PePRE1-OX* lines (Figure 3e). These studies suggest that *PePRE* genes play a conserved role similar to *AtPRE* genes on flowering promotion and cell elongation.

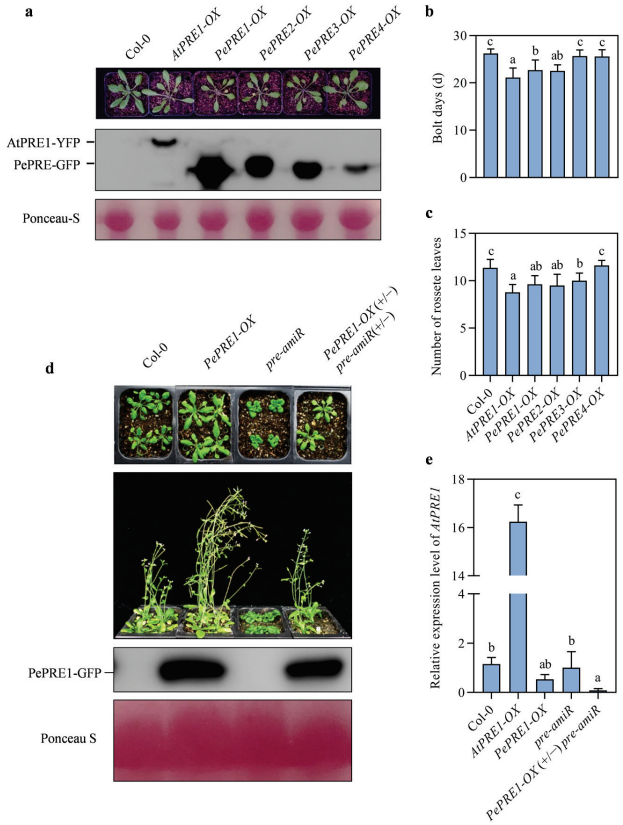


Figure 3. Overexpression of *PePREs* promoted plant flowering and growth in *Arabidopsis.* (**a**) Overexpression of *PePREs* in *Arabidopsis* increased petiole elongation. Panels from top to bottom are pictures of plants grown in soil for 4 weeks (left to right: Col-0, *AtPRE1-OX* (35S::*AtPRE1-YFP*), and *PePREs-OX* (35S::*PePREs-GFP*)), Western bolt of PRE proteins using GFP antibody, and protein loading control by Ponceau S. Days of bolting (**b**) and rosette leaf numbers (**c**) when plants started bolting. A total of 10 plants were used for calculations. (**d**) Overexpression of *PePRE1* suppressed the dwarf phenotype of *pre-amiR*. Panels from top to bottom are morphology of 4-week-old plants of Col-0, *PePRE1-OX*, *pre-amiR*, and *PePRE1-OX* (+/−) *pre-amiR* (+/−), 5-weeks-old plants of Col-0, *PePRE1-OX*, *pre-amiR*, and *PePRE1-OX* (+/−) *pre-amiR* (+/−), expression level of PRE1 using GFP antibody, and the protein loading control by Ponceau S. (**e**) Quantitative analysis of *AtPRE1* in Col-0, *PePRE1-OX*, *pre-amiR*, and *PePRE1-OX* (+/−) *pre-amiR* (+/−). Different letters above the data columns indicate significant differences compared between Col-0 and transgenic lines ($p < 0.05$; Duncan's test).

2.4. Subcellular Localization of PePREs in Arabidopsis

To determine the intracellular localization of PePRE proteins, the fluorescence signals of 35S::*PePRE-GFP* were examined in transgenic *Arabidopsis* plants. PePRE-GFP signals were mainly located in the nucleus, cytosol, and plasma membranes, whereas 4′, 6-diamidino-2-phenylindole dihydrochloride (DAPI) staining was used as the indicator of nuclear area. Interestingly, PePRE1-GFP and PePRE3-GFP proteins specifically displayed small punctate structures in the cytosol (Figure 4a), which are different from the AtPRE1 protein. When treated with BFA, a vesicle trafficking inhibitor [28], aggregation of PePRE1 fusion proteins in the punctate structures was reduced. Moreover, these signals were recovered by removing BFA (Figure 4b). Taken together, PePREs are located in the nucleus, cytosol, and membrane, whereas PePRE1 and PePRE3 are distributed in the cytosol in punctate structures.

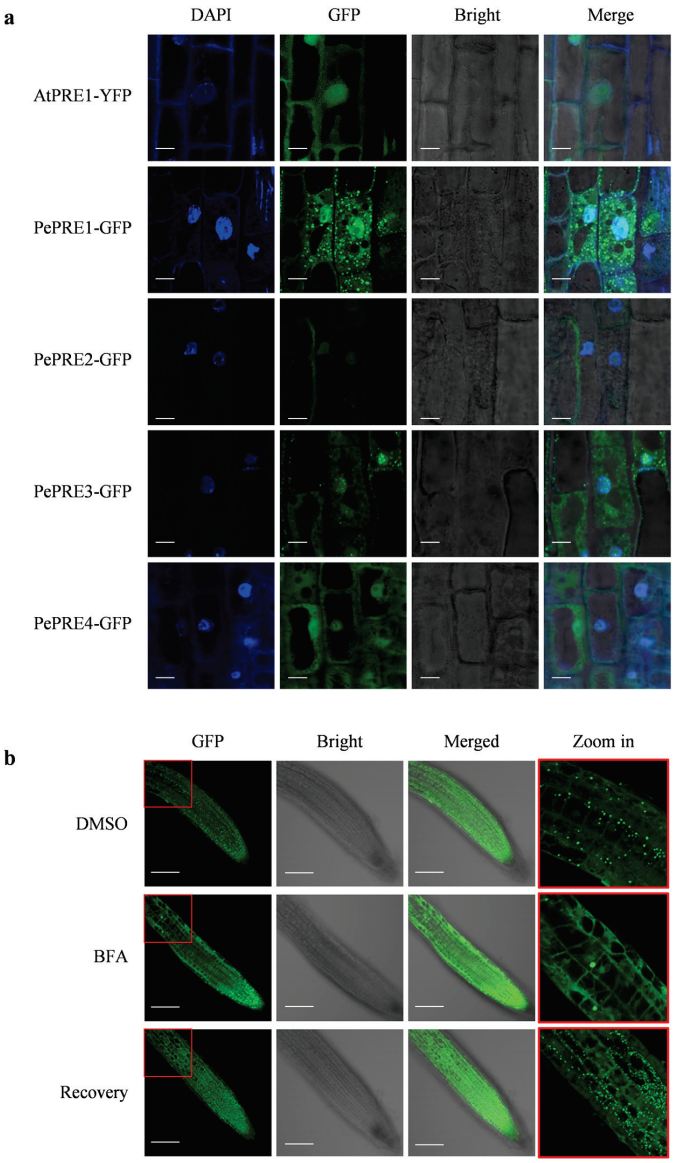


Figure 4. Subcellular localization of PRE proteins in *Arabidopsis*. (**a**) The root tip of the *PePREs*′ transgenic lines was observed by confocal microscope. Seedlings were grown on half MS medium with constant light for 7 days. Cells were stained with DAPI. Scale bar indicates 10 μm. (**b**) Subcellular localization of PePRE1-GFP by treating with or without BFA. The root tips of *PePRE1*-OX #19 were observed by confocal microscope. Seedlings were grown on half MS medium with constant light for 4 days vertically. Scale bar indicates 100 μm.

2.5. BR Regulates the Expression Levels of PePREs

The expression levels of *PePRE1* and *PePRE2* dramatically decreased in the aerial part, whereas BR increased the expression of *PePRE1* and *PePRE2* in bamboo shoot [25]. We further examined the expression profile of *PRE* homologous genes in various tissues of bamboo seedlings with PPZ and eBL (Figure 5). Expression of *PRE1*, *3*, and *4* was decreased by PPZ but recovered by subsequent BR treatment in most tissues. However, mRNA levels of *PePRE2* increased in the internode and sheath and decreased in the lamina joint and root. In addition, *PePRE2* did not further respond to eBL treatment. Overexpression of the *PePRE1* gene resulted in longer hypocotyl in comparison with the wild type (Figure 6a,b; Figure A1). Compared with the wild type, *PePRE1* overexpressing plants have longer hypocotyls and roots (Figure 6a,b), similar to *AtPRE1* overexpressing plants. Additionally, compared with *AtPRE1-OX*, the shoot and root parts of *PePRE1-OX* plants showed more sensitivity and less sensitivity to PPZ, respectively (Figures 6c,d and A3).

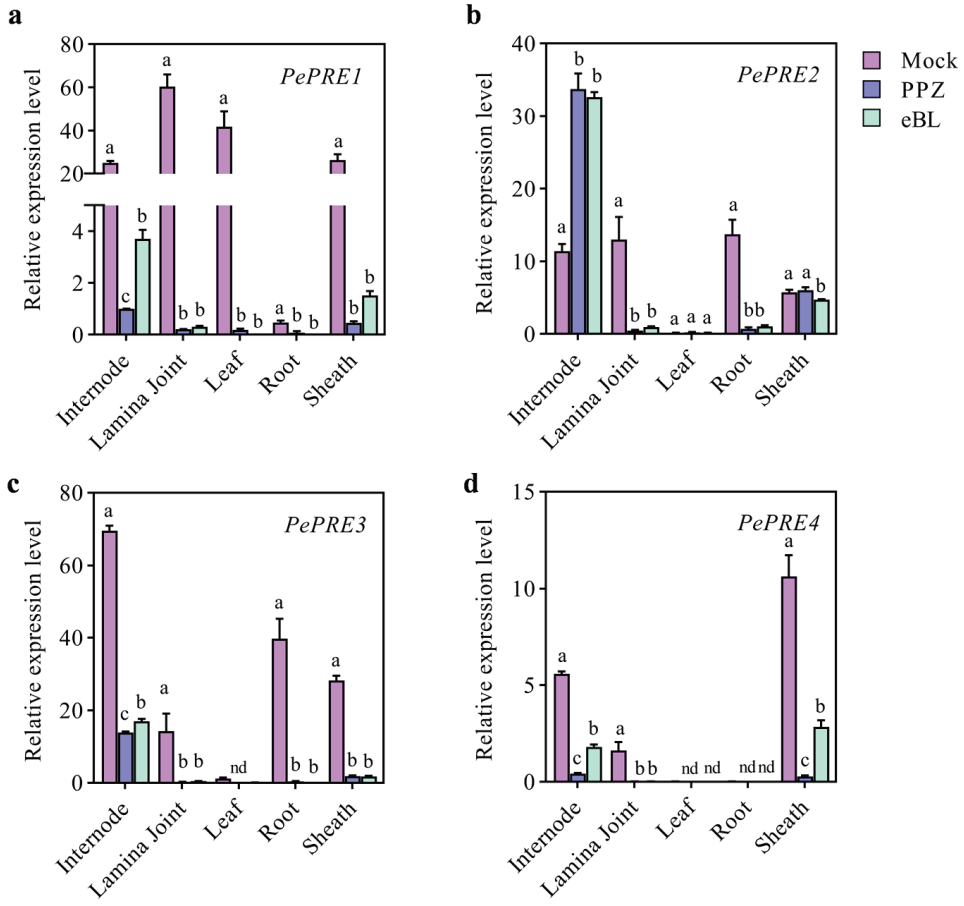


Figure 5. The expression levels of *PePRE*s after eBL or PPZ treatment in bamboo seedlings. Quantitative RT-PCR analysis of *PePRE*s (**a**–**d**) in different bamboo seedling organs after PPZ or eBL treatment. The bamboo seedlings were grown on half MS medium for 18 days. For PPZ treatment, plants were transferred to half MS medium containing 50 μM PPZ for another 4 days. For eBL treatment, the seedlings were transferred to half MS medium containing 10 μM PPZ and 1 μM eBL solution for 4 h. Different lowercase letters above the data columns indicate significant differences ($p < 0.05$; Duncan's test); nd, no data available.

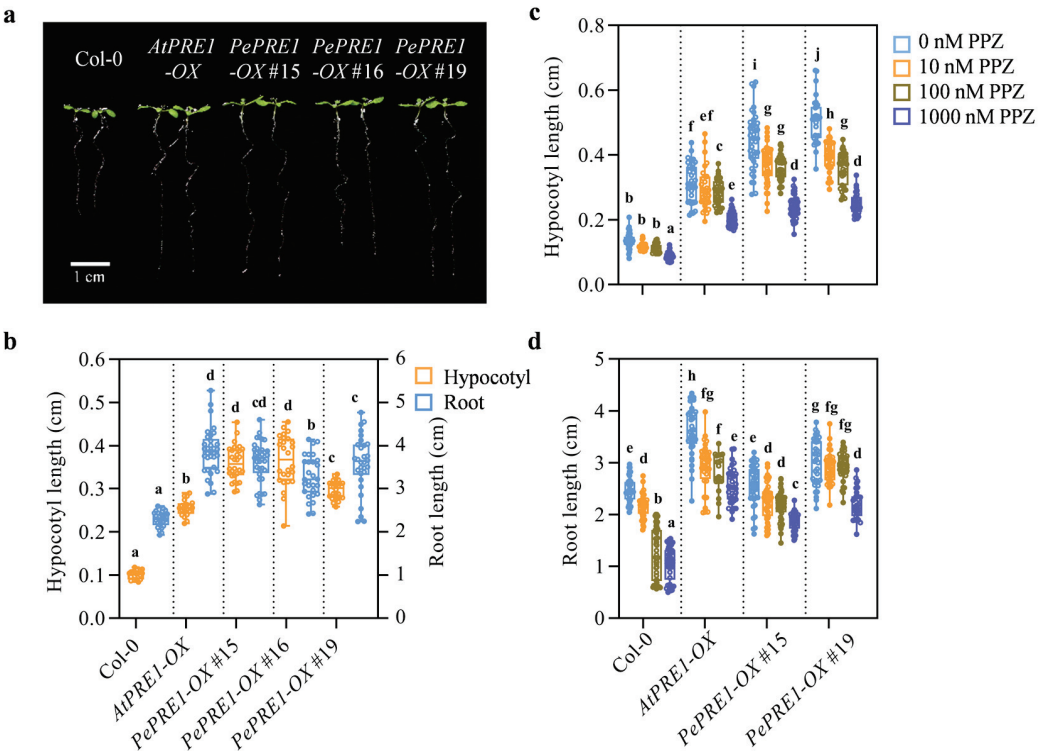


Figure 6. *PePRE1-OX* transgenic *Arabidopsis* exhibited longer hypocotyls and roots that were hypersensitive to PPZ treatment. (**a**) Representative seedlings of Col-0, *AtPRE1-OX*, and *PePRE1-OX*. The seeds were sown on half MS medium and grown for 7 days vertically. Scale bar presents 1 cm. (**b**) The hypocotyl and root lengths of the control and *PePRE1-OX* plants. Hypocotyl length (**c**) and root length (**d**) of different lines of *PePRE1-OX* compared with Col-0 with a gradient concentration (0 nM, 10 nM, 100 nM, and 1000 nM) of PPZ treatment. Values are the means calculated from at least 20 seedlings. Error bars represent mix/max values. Different letters above the data column indicate significant differences ($p < 0.05$; Duncan's test).

3. Discussion

Although the rapid growth of woody bamboo plants has been widely studied, little is known about the molecular mechanism underlying the elongation of moso bamboo. In this study, we investigated the unique patterns of *PePRE* genes, the conserved growth-promoting function in *Arabidopsis*, and the potential important role for *PePRE*s in the elongation of moso bamboo.

3.1. PePREs Have Tissue-Specific Expression Patterns in Moso Bamboo

PePRE1 and *PePRE4* are highly expressed in the shoot, whereas *PePRE2* and *PePRE4* are specifically expressed in the roots of seedlings. Consistent with these tissue-specific expressions, a rice homologous gene of *PePRE*, *PGL1* (positive regulator of grain length 1), was expressed in the floral organs, young panicle, and predominantly in the root but not the leaf [29]. The expression level of *BU1* (brassinosteroid upregulated 1) is high in the lamina joint in vegetative organs and the panicle at the heading stage [30]. *OsILI1* is ubiquitously expressed in rice, whereas the highest expression of *OsILI1* was observed in the lamina joint [12]. *OsBUL1* (*O. sativa* brassinosteroid upregulated 1-like1), an *AtPRE* homolog in rice, is preferentially expressed in the lamina joint where it controls cell elongation and positively affects leaf angles [31,32]. *FaPRE1* was proposed as a ripening-associated gene and showed a rapid increase in expression in the receptacle during fruit enlargement [33]. In our study, all four *PRE* genes were highly expressed in the elongating tissues but not in the mature tissues such as the nodes of the bamboo shoot (Figures 1 and 2), which is consistent with their function in promoting cell elongation. During the rapid growth of monocots,

elongation generally occurs from top to bottom in each individual internode [34]. From our results, *PePRE1* is highly expressed in the basal part of the elongating internodes, with a gradient distribution from the basal to the top parts (Figure 7), which may contribute to the fast growth of bamboo shoot. A previous report showed that *GRF*s (growth-regulating factors) and *ARF* genes were highly expressed in the basal region of the elongating internode, in which *ARF6* and two *ARF8*s targeted by *miR167* and 11 *GRFs* targeted by *miR396* were in the top region [35]. The *DsEXLA2* (*Dendrocalamus sinicus* expansin-like A2) gene is highly expressed in the elongating internode and accelerates the plant growth rate of *Arabidopsis* [36]. *PeGT43*s (*P. edulis* Glycosyltransferase 43) and lignin biosynthesis are significantly upregulated within the shoot [3]. Considering the roles for *PRE* in *Arabidopsis*, *PePRE* genes may work downstream of *ARF* or *GRF* genes to regulate target genes in the elongating bamboo.

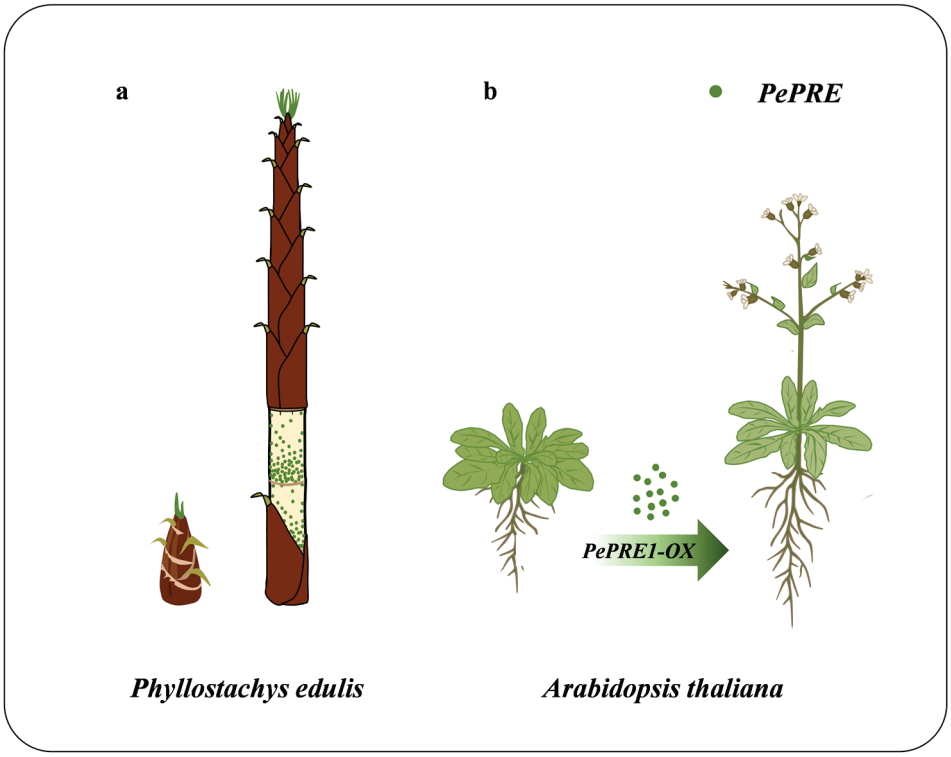


Figure 7. Diagram of *PePRE1* gene expression and function. (**a**) Gradient expression levels of *PePRE1* in the elongating internodes of moso bamboo shoots. (**b**) Overexpression of *PePRE1* genes leads to early flowering, longer hypocotyl, and growth promotion in *Arabidopsis*. Green dots indicate *PePRE* genes.

3.2. Phytohormones Regulates PRE Function in Bamboo Elongation

The stem elongation was contributed to by the division and expansion of individual internodes [37]. Transcriptome analysis revealed that multiple signaling pathways, including GA, auxin, and ABA, may play a role in regulating internode elongation [38]. GA plays an antagonistic regulatory role in regulating internode stem elongation in rice [39]. *OsbHLH073* encodes an atypical bHLH protein and regulates plant height, internode elongation, and panicle extension by regulating GA biosynthesis genes [40]. The content of GAs in dwarf bamboo varieties is also lower than that in normal bamboos, implying that GA plays a major role in the height of Shidu bamboo [41]. Exogenous application of GA resulted in a significant increase in internode length in bamboo seedlings [42]. Those results hint that GA may play a dual role in internode elongation between shoots and seedlings of bamboo. PREs were reported to act as hostile antagonists of the bHLH family of transcription factors, which positively regulate cell elongation via multiple signaling pathways [6,7].

Bamboo PRE proteins showed different subcellular localizations, and only PRE1 and PRE3 showed punctate structures that were disturbed by BFA treatment. BFA treatment may lead to rapid protein aggregation within the endoplasmic reticulum and collapse of the Golgi apparatus [28]. The specific localizations of PePRE1 and PePRE3 may be related to the tuning of protein stability or function. PePREs are localized in both the nucleus and cytoplasm, whereas the punctate structures in cytoplasm were only observed in PePRE1 and PePRE3 in transgenic lines (Figure 4). Considering synchrony in phenotype, expression pattern, and subcellular localization, we proposed that the difference in subcellular localization may contribute to its functional variation. Certain *PePRE* overexpressing plants were easily dislodged caused by a broken stem (Figure A2). PRE1 was reported to promote cell elongation by preventing IBH1 from inhibiting HBI1, which directly activates genes encoding cell-wall-loosening enzymes (e.g., EXP, etc.) [10,43]. The accumulation of PePRE may change the stem segment structure, especially the composition of the cell wall, leading to the rupture and hollowness of the stem (Figure A2).

3.3. Various Functions of PRE in Regulating Plant Growth through Multiple Signaling Pathways

PRE proteins are important parts of the signaling pathways involved in physiological development and reproduction [21,27,44]. Due to the limitation of the transformation of moso bamboo, the function of *PePRE*s was checked in *Arabidopsis* in this work. The overexpression of *PePREs* was associated with early flowering and promoted growth, such as longer petioles, hypocotyls, and roots (Figure 7). *PREs* are involved in regulating the growth of floral organs in *Arabidopsis* [27,45]. *FaPRE1* antagonistically modulates the transcription of genes related to both receptacle growth and ripening [23,33]. *GhPRE1* has contributed to spinnable fiber formation in cotton; overexpressing *GhPRE1* leads to longer fibers with improved quality parameters, indicating that this bHLH gene is useful for improving cotton fiber quality [24]. *SlPRE2* was reported to regulate fruit development via the gibberellin pathway and tomato fruit pigment accumulation in tomatoes [46]. Those results suggest that *PRE* affects multiple aspects of the development of plants. PRE6 is a positive regulator of shade avoidance and interacts with a number of negative growth regulators (PAR1, etc.) [17]. Transcriptional regulators (ARFs and BZR1) and post-transcriptional regulators (HFR1, etc.) were key modules of the signaling network controlling shade avoidance [47,48]. Interestingly, *PePRE* genes are regulated by brassinosteroid levels (Figure 5) in the elongating bamboo, and the transgenic lines also shown an altered response to PPZ treatment (Figure 6). Shade avoidance syndrome (SAS) allows plants that are grown in densely populated environments to maximize their sunlight access [48]. As mentioned above in the relationship between those TFs and PRE, PePREs may play a positive role in rhizome elongation underground and shoot elongation aboveground. Functional analyses of PePREs will help to elucidate the mechanism of fast growth in plants as well. Light affects the dynamic growth and development of the *P. pygmaeus* rhizome–root system [49]. Taken together, *PePRE* genes show a conserved function in controlling flowering and promoting growth by responding to multiple signaling pathways. This is similar to the *PRE* genes from other species.

Our analysis identified *PePRE* genes with specific expression patterns in the seedling and shooting stage and demonstrated that *PePREs* are brassinosteroid-regulated genes. Overexpression of the *PePRE1* gene will promote flowering, hypocotyl elongation, and root growth. The current results indicate a key role for *PePRE* genes in bamboo growth and development. Our findings will provide the basis for further functional characterizations of *PRE* family genes and the molecular mechanism of bamboo fast growth and will benefit the application of growth-promoting gene resources from bamboo.

4. Materials and Methods

4.1. Plant Material and Growth Conditions

The moso bamboo shoots (approximately 1.8 m above ground height) were obtained from the Bamboo Garden of Fujian Agriculture and Forestry University, Fuzhou (coordinate

119°14′ E, 26°50′ N). Moso bamboo seeds were collected from Gong city, Guanxi province, China (118°48′ E, 24°51′ N). The seeds of moso bamboo were soaked in tap water for 24 h to induce seed germination and then sown on the soil. Seedlings were cultivated for 3 weeks in a greenhouse (long-day conditions, 22 °C).

Columbia (Col-0) wild-type seeds of *Arabidopsis* were germinated on half MS medium (pH = 5.8) with 1% sucrose, then transferred to a greenhouse (long-day conditions, 22 °C). *PePRE-OX* was crossed with *pre-amiR* to generate *PePRE-OX pre-amiR* plants.

4.2. Protein Sequence Alignment and Phylogenic Tree Construction

Alignment of the protein sequences was performed using Clustal Omega and analyzed in the GENEDOC program with the default settings. A phylogenic tree based on the sequence alignment was generated using MEGA-X by the neighbor-joining method [50]. Accession number: The protein sequences reported in this article can be found in the database (http://www.bamboogdb.org/, accessed on 8 May 2017) [26], the rice genome database (http://rice.plantbiology.msu.edu/, accessed on 8 May 2017), and TAIR (http://www.arabidopsis.org/, accessed on 8 May 2017) under the following accession numbers: *OsILI1* (Os04g54900.1), *OsILI2* (Os11g39000.1), *OsILI3* (Os03g07540.1), *OsILI4* (Os06g12210.1), *OsILI5* (Os02g51320.1), *OsILI6* (Os03g07510.1), *OsILI7* (Os10g26460.1), *PePRE1* (PH01000065G2010), *PePRE2* PH01000068G1560), *PePRE3* (PH01000519G0840), *PePRE4* (PH01000960G0260), *AtPRE1* (AT5G39860.1), *AtPRE2* (AT5G15160.1), *AtPRE3* (AT1G74500.1), *AtPRE4* (AT3G47710.1), *AtPRE5* (AT3G28857.1), and *AtPRE6* (AT1G26945.1).

4.3. Gene Expression Analysis

Total RNA was isolated from various tissues with the Plant Total RNA Kit (Sigma, ATRN50) and extensively treated with RNase-free DNase I (Sigma, St. Louis, MO, USA, DNASE70-1SET). cDNA was generated by RT-PCR using the PrimeScript^{TM} RT reagent Kit (Takara, Kusatsu, Japan, RR047A). qRT-PCR analysis was performed with SYBR Premix Ex Taq II (Takara, RR820) on a QuanStudio 6 Flex instrument (Applied Biosystems, Waltham, MA, USA). For each sample, qPCR was performed with three technical replicates on three biological replicates. The bamboo *PeTIP41* gene [51] was used as an internal control for the qRT-PCR. The relative expression levels were calculated as $E^{-\Delta Cq}$ and normalized to *PeTIP41*. The primer sequences used in this study are listed in Table A1.

4.4. Vector Construction and Transformation

The full-length *PePRE1* sequence was amplified and ligated into pAGM1311. A 35S-promotor-driven PePRE C-terminal fused with a GFP tag was constructed into a pAGM4673 backbone. Through *Agrobacterium tumefaciens s*train GV3101, these constructions were transformed into *Arabidopsis* (Col-0). T1 seeds were screened by an RFP selection marker and then further confirmed by qRT-PCR and Western blot.

4.5. Hypocotyl Measurements and Statistical Analysis

For hypocotyl and root length measurements, seedlings were grown for 7 days on vertically oriented plates. Seedlings were flattened and photographed before taking quantitative measurements using ImageJ software (http://rsb.info.nih.gov/, accessed on 8 April 2018) to analyze the scanned images of the seedlings. The differences among groups were assessed by one-way ANOVA and Duncan's multiple comparisons test using SPSS 23.0. GraphPad Prism 8.0 (http://www.graphpad.com/, accessed on 26 April 2018) was used to plot figures. At least 20 seedlings were measured, and experiments were repeated more than two times.

4.6. BFA Treatment

For BFA treatment, 4-day-old seedlings of *PePRE1-OX* were incubated in 50 μM BFA for 90 min before viewing the seedlings; control seedlings were incubated in a solution without BFA but with DMSO at the same concentration as the BFA-treated seedlings. After

BFA treatment, the seedlings were washed with half MS liquid media several times and the root tips were observed for the formation of BFA compartments using a confocal laser scanning microscope (Leica Microsystems, Wetzlar, Germany). For recovery, BFA was removed from seedlings and supplied with half MS media for another 40 min followed by immediate observation with the confocal microscope.

5. Conclusions

Our study identified an atypical bHLH transcription factor (*PRE* homologs) in moso bamboo via gene expression analysis, heterologous overexpression, etc. We verified that *PePREs* function as a positive regulator in the promotion of internode elongation during the fast-growth process. Overexpressing *PePRE* promoted *Arabidopsis* growth and petiole/hypocotyl elongation. Our findings shed light on bHLH-mediated fast growth to provide preliminary knowledge for fast-growing plants.

Author Contributions: Conceptualization, S.Z., K.S., W.W. and X.Y.; methodology, data curation, and validation, S.Z., K.S., W.W. and X.Y.; writing-original draft preparation, S.Z., W.W. and X.Y.; writing-review and editing, K.S., W.L. and X.Y.; funding acquisition, W.W.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Innovation Program of FAFU (CXZX2020046A) and the program of First-class Ecology Disciplines Construction of FAFU to W.W.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Mingyi Bai (Shandong University) for providing the *Arabidopsis* seeds of the *AtPRE1-OX* and *pre-amiR* plants.

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

Appendix A

Table A1. The primers used for quantitative PCR (RT-qPCR) and amplification of full-length coding sequences of *PePREs* in *Phyllostachys edulis*.

Primers	Sequences (5′-3′)	Note
PePRE1 qPCR F1	ATTATGTCCATGGATCTCTG	Gene expression level detection of *PePRE1*
PePRE1 qPCR R1	ACTTAGTCCTTAGGTATCGA	Gene expression level detection of *PePRE1*
PePRE2 qPCR F1	CCCGCAGGCAGACATCAT	Gene expression level detection of *PePRE2*
PePRE2 qPCR R1	GTGCTTCGAACTGAGTAAAT	Gene expression level detection of *PePRE2*
PePRE3 qPCR F1	CAAGACCTCACCCAGCTTAG	Gene expression level detection of *PePRE3*
PePRE3 qPCR R1	TGCCGCGCGACGACCTTC	Gene expression level detection of *PePRE3*
PePRE4 qPCR F1	TGAGTTTAAGCTCACACTACTC	Gene expression level detection of *PePRE4*
PePRE4 qPCR R1	AGAGGTCGCGATCTCCTAGC	Gene expression level detection of PePRE4
qPeTIP41-F1	AAAATCATTGTAGGCCATTGTCG	Internal control gene in *Ph. edulis*
qPeTIP41-R1	ACTAAATTAAGCCAGCGGGAGTG	Internal control gene in *Ph. edulis*
AtPRE1 qPCR F	GTTCTGATAAGGCATCAGCCTCG	Gene expression level detection of *AtPRE1*
AtPRE1 qPCR R	CATGAGTAGGCTTCTAATAACGG	Gene expression level detection of *AtPRE1*
UBQ10-F	ATCACCCTTGAAGTGGA	Internal control gene in *Arabidopsis*
UBQ10-F	ATCACCCTTGAAGTGGA	Internal control gene in *Arabidopsis*
PePRE1 MoClo F	tttgaagacaaAATGTCGAGCCGGAGGTCGCG	For amplifying CDS of *PePRE1*
PePRE1 MoClo R	tttgaagacaacgaaccGCGGAGGAGGCTGCGGATGA	For amplifying CDS of *PePRE1*
PePRE2 MoClo F1	tttgaagacaaAATGTCGAGGCGGCGGGGG	For amplifying CDS of *PePRE2*
PePRE2 MoClo R	tttgaagacaacgaaccGCTAGGAGGACCGGAGCTGA	For amplifying CDS of *PePRE2*
PePRE3 MoClo F	tttgaagacaaAATGTCGGGCCGAAGGTCGTC	For amplifying CDS of *PePRE3*
PePRE3 MoClo R	tttgaagacaacgaaccGGAGCGGAGGATGTTGCGGA	For amplifying CDS of *PePRE3*
PePRE4 MoClo F	tttgaagacaaAATGTCGGGGCGCAGAGCCGG	For amplifying CDS of *PePRE4*
PePRE4 MoClo R	tttgaagacaacgaaccGGATGAGCGGAGGAGACTGC	For amplifying CDS of *PePRE4*
PePRE4 MoClo F(-R)1	tttgaagacaaGTCaTCGGCGTCGAACCTGCTG	For amplifying CDS of *PePRE4*
PePRE4 MoClo (F-)R	tttgaagacaatGACTGCAACAAAAACATTA	For amplifying CDS of *PePRE4*
level 0 F	CGTTATCCCCTGATTCTGTGGATAAC	MoClo cloning general primer
level 0 R	GTCTCATGAGCGGATACATATTTGAATG	MoClo cloning general primer
level 1 F	GAACCCTGTGGTTGGCATGCACATAC	MoClo cloning general primer
level 1 R	CTGGTGGCAGGATATATTGTGGTG	MoClo cloning general primer
level 2 F	GTGGTGTAAACAAATTGACGC	MoClo cloning general primer
level 2 R	GGATAAACCTTTTCACGCCC	MoClo cloning general primer

Appendix B

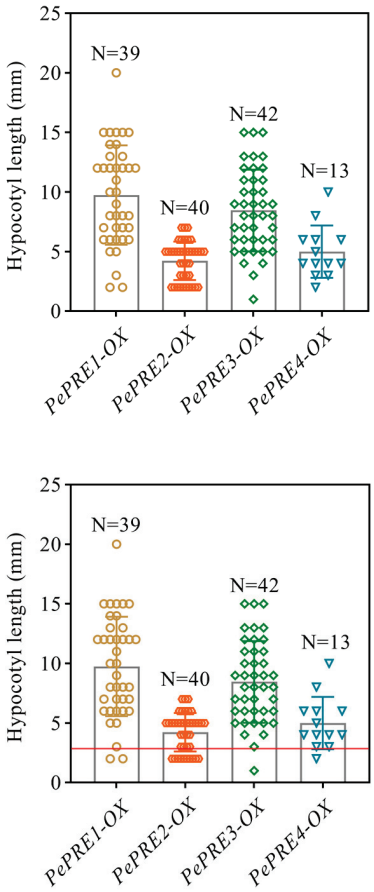


Figure A1. The phenotypes of T1 transgenic plants overexpressing *PePRE*s in Col-0 background. Overexpression PePRE-GFP caused longer hypocotyls in Col-0. Hypocotyl measurement of light-grown PePRE-GFP transgenic plants compared with Col-0 (the red line indicates the average hypocotyl length of Col-0).

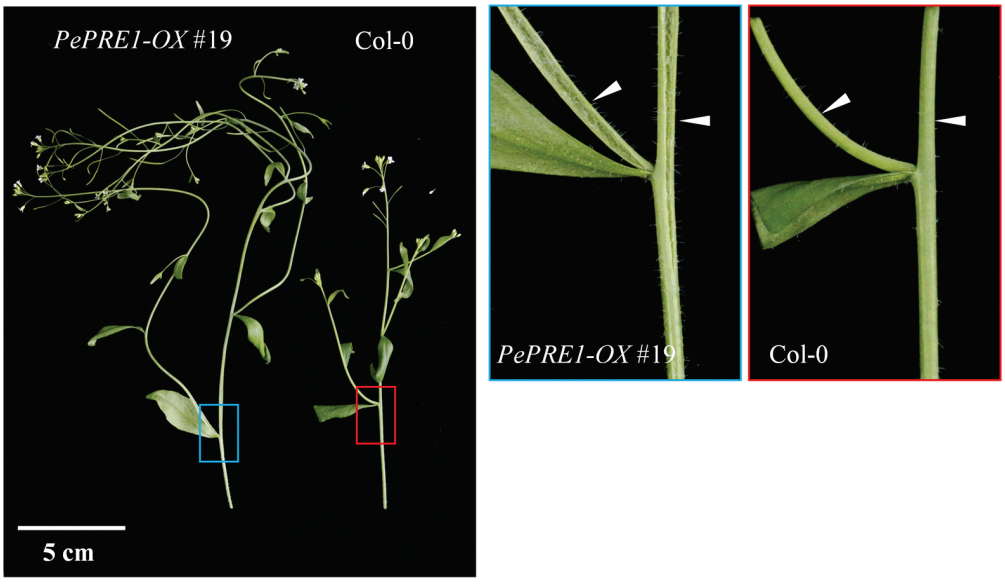


Figure A2. The morphology of T1 transgenic plants overexpressing PePREs (35S::*PePRE1-GFP* #19 phenotype) in *Arabidopsis*. The plants were grown in a growth room for 4 weeks.

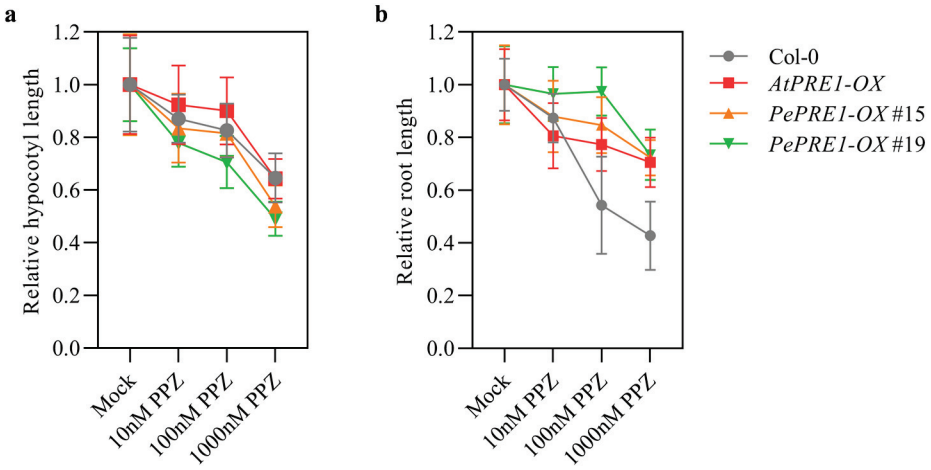


Figure A3. The *PePRE1-OX* transgenic seedlings were treated with PPZ. The relative hypocotyl length (**a**) and relative root length (**b**) of *PePRE1-OX* and control plants were treated with mock solution and a gradient concentration of PPZ. Seedlings were grown on medium containing different concentrations of PPZ for 7 days under constant light. The relative hypocotyl/root length was the average of at least 20 seedlings and normalized to the untreated seedlings. Error bars represent standard deviations.

References

1. Kelchner, S.A.; Bamboo Phylogeny, G. Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. *Mol. Phylogenet. Evol.* **2013**, *67*, 404–413. [CrossRef] [PubMed]
2. Wysocki, W.P.; Clark, L.G.; Attigala, L.; Ruiz-Sanchez, E.; Duvall, M.R. Evolution of the bamboos (Bambusoideae; Poaceae): A full plastome phylogenomic analysis. *BMC Evol. Biol.* **2015**, *15*, 50. [CrossRef] [PubMed]
3. Li, T.; Wang, H.; Zhang, Y.; Wang, H.; Zhang, Z.; Liu, X.; Zhang, Z.; Liu, K.; Yang, D.; Zhang, H.; et al. Comprehensive profiling of epigenetic modifications in fast-growing Moso bamboo shoots. *Plant Physiol.* **2023**, *191*, 1017–1035. [CrossRef] [PubMed]
4. Gamuyao, R.; Nagai, K.; Ayano, M.; Mori, Y.; Minami, A.; Kojima, M.; Suzuki, T.; Sakakibara, H.; Higashiyama, T.; Ashikari, M.; et al. Hormone distribution and transcriptome profiles in bamboo shoots provide insights on bamboo stem emergence and growth. *Plant Cell Physiol.* **2017**, *58*, 702–716. [CrossRef] [PubMed]
5. Chaiwanon, J.; Wang, W.; Zhu, J.Y.; Oh, E.; Wang, Z.Y. Information integration and communication in plant growth regulation. *Cell* **2016**, *164*, 1257–1268. [CrossRef]
6. Oh, E.; Zhu, J.Y.; Bai, M.Y.; Arenhart, R.A.; Sun, Y.; Wang, Z.Y. Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *Elife* **2014**, *3*, e03031. [CrossRef]
7. Zhiponova, M.K.; Morohashi, K.; Vanhoutte, I.; Machemer-Noonan, K.; Revalska, M.; Van Montagu, M.; Grotewold, E.; Russinova, E. Helix-loop-helix/basic helix-loop-helix transcription factor network represses cell elongation in Arabidopsis through an apparent incoherent feed-forward loop. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2824–2829. [CrossRef]
8. Ruzinova, M.B.; Benezra, R. Id proteins in development, cell cycle and cancer. *Trends Cell Biol.* **2003**, *13*, 410–418. [CrossRef]
9. Castelain, M.; Le Hir, R.; Bellini, C. The non-DNA-binding bHLH transcription factor PRE3/bHLH135/ATBS1/TMO7 is involved in the regulation of light signaling pathway in Arabidopsis. *Physiol. Plant* **2012**, *145*, 450–460. [CrossRef]
10. Bai, M.Y.; Fan, M.; Oh, E.; Wang, Z.Y. A triple Helix-Loop-Helix/basic Helix-Loop-Helix cascade controls cell elongation downstream of multiple hormonal and environmental signaling pathways in Arabidopsis. *Plant Cell* **2012**, *24*, 4917–4929. [CrossRef]
11. Malinovsky, F.G.; Batoux, M.; Schwessinger, B.; Youn, J.H.; Stransfeld, L.; Win, J.; Kim, S.K.; Zipfel, C. Antagonistic regulation of growth and immunity by the Arabidopsis basic helix-loop-helix transcription factor homolog of brassinosteroid enhanced expression2 interacting with increased leaf inclination1 binding bHLH1. *Plant Physiol.* **2014**, *164*, 1443–1455. [CrossRef]
12. Zhang, L.Y.; Bai, M.Y.; Wu, J.; Zhu, J.Y.; Wang, H.; Zhang, Z.; Wang, W.; Sun, Y.; Zhao, J.; Sun, X.; et al. Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and *Arabidopsis*. *Plant Cell* **2009**, *21*, 3767–3780. [CrossRef]
13. Choi, H.; Oh, E. PIF4 integrates multiple environmental and hormonal signals for plant growth regulation in *Arabidopsis*. *Mol. Cells* **2016**, *39*, 587–593. [CrossRef]
14. Oh, E.; Zhu, J.Y.; Wang, Z.Y. Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. *Nat. Cell Biol.* **2012**, *14*, 802–809. [CrossRef]
15. Hao, Y.; Oh, E.; Choi, G.; Liang, Z.; Wang, Z.Y. Interactions between HLH and bHLH factors modulate light-regulated plant development. *Mol. Plant* **2012**, *5*, 688–697. [CrossRef]

16. Hyun, Y.; Lee, I. *KIDARI*, encoding a non-DNA Binding bHLH protein, represses light signal transduction in *Arabidopsis thaliana*. *Plant Mo.l Biol.* **2006**, *61*, 283–296. [CrossRef]
17. Buti, S.; Pantazopoulou, C.K.; van Gelderen, K.; Hoogers, V.; Reinen, E.; Pierik, R. A Gas-and-Brake mechanism of bHLH proteins modulates shade avoidance. *Plant Physiol.* **2020**, *184*, 2137–2153. [CrossRef]
18. Ikeda, M.; Mitsuda, N.; Ishizuka, T.; Satoh, M.; Ohme-Takagi, M. The CIB1 transcription factor regulates light- and heat-inducible cell elongation via a two-step HLH/bHLH system. *J. Exp. Bot.* **2021**, *72*, 1795–1808. [CrossRef]
19. Chapman, E.J.; Greenham, K.; Castillejo, C.; Sartor, R.; Bialy, A.; Sun, T.P.; Estelle, M. Hypocotyl transcriptome reveals auxin regulation of growth-promoting genes through GA-dependent and -independent pathways. *PLoS ONE* **2012**, *7*, e36210. [CrossRef]
20. Reed, J.W.; Wu, M.F.; Reeves, P.H.; Hodgens, C.; Yadav, V.; Hayes, S.; Pierik, R. Three auxin response factors promote hypocotyl elongation. *Plant Physiol.* **2018**, *178*, 864–875. [CrossRef]
21. Lee, S.; Lee, S.; Yang, K.Y.; Kim, Y.M.; Park, S.Y.; Kim, S.Y.; Soh, M.S. Overexpression of *PRE1* and its homologous genes activates Gibberellin-dependent responses in *Arabidopsis thaliana*. *Plant Cell Physiol.* **2006**, *47*, 591–600. [CrossRef] [PubMed]
22. Dai, X.; Lu, Q.; Wang, J.; Wang, L.; Xiang, F.; Liu, Z. MiR160 and its target genes *ARF10*, *ARF16* and *ARF17* modulate hypocotyl elongation in a light, BRZ, or PAC-dependent manner in *Arabidopsis*: miR160 promotes hypocotyl elongation. *Plant Sci.* **2021**, *303*, 110686. [CrossRef] [PubMed]
23. Medina-Puche, L.; Martinez-Rivas, F.J.; Molina-Hidalgo, F.J.; Garcia-Gago, J.A.; Mercado, J.A.; Caballero, J.L.; Munoz-Blanco, J.; Blanco-Portales, R. Ectopic expression of the atypical HLH *FaPRE1* gene determines changes in cell size and morphology. *Plant Sci.* **2021**, *305*, 110830. [CrossRef] [PubMed]
24. Zhao, B.; Cao, J.F.; Hu, G.J.; Chen, Z.W.; Wang, L.Y.; Shangguan, X.X.; Wang, L.J.; Mao, Y.B.; Zhang, T.Z.; Wendel, J.F.; et al. Core cis-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *New Phytol.* **2018**, *218*, 1061–1075. [CrossRef]
25. Zhang, Z.; Yang, X.; Cheng, L.; Guo, Z.; Wang, H.; Wu, W.; Shin, K.; Zhu, J.; Zheng, X.; Bian, J.; et al. Physiological and transcriptomic analyses of brassinosteroid function in moso bamboo (*Phyllostachys edulis*) seedlings. *Planta* **2020**, *252*, 27. [CrossRef]
26. Zhao, H.; Peng, Z.; Fei, B.; Li, L.; Hu, T.; Gao, Z.; Jiang, Z. BambooGDB: A bamboo genome database with functional annotation and an analysis platform. *Database* **2014**, *2014*, bau006. [CrossRef]
27. Mara, C.D.; Huang, T.; Irish, V.F. The *Arabidopsis* floral homeotic proteins APETALA3 and PISTILLATA negatively regulate the *BANQUO* genes implicated in light signaling. *Plant Cell* **2010**, *22*, 690–702. [CrossRef]
28. Klausner, R.D.; Donaldson, J.G.; Lippincott-Schwartz, J. Brefeldin A: Insights into the control of membrane traffic and organelle structure. *J. Cell Biol.* **1992**, *116*, 1071–1080. [CrossRef]
29. Heang, D.; Sassa, H. Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. *PLoS ONE* **2012**, *7*, e31325. [CrossRef]
30. Tanaka, A.; Nakagawa, H.; Tomita, C.; Shimatani, Z.; Ohtake, M.; Nomura, T.; Jiang, C.J.; Dubouzet, J.G.; Kikuchi, S.; Sekimoto, H.; et al. BRASSINOSTEROID UPREGULATED1, encoding a Helix-Loop-Helix protein, is a novel gene involved in brassinosteroid signaling and controls bending of the lamina joint in rice. *Plant Physiol.* **2009**, *151*, 669–680. [CrossRef]
31. Jang, S.; Cho, J.Y.; Do, G.R.; Kang, Y.; Li, H.Y.; Song, J.; Kim, H.Y.; Kim, B.G.; Hsing, Y.I. Modulation of rice leaf angle and grain size by expressing *OsBCL1* and *OsBCL2* under the control of *OsBUL1* promoter. *Int. J. Mol. Sci.* **2021**, *22*, 7792. [CrossRef]
32. Jang, S.; An, G.; Li, H.Y. Rice leaf angle and grain size are affected by the OsBUL1 transcriptional activator complex. *Plant Physiol.* **2017**, *173*, 688–702. [CrossRef]
33. Medina-Puche, L.; Martinez-Rivas, F.J.; Molina-Hidalgo, F.J.; Mercado, J.A.; Moyano, E.; Rodriguez-Franco, A.; Caballero, J.L.; Munoz-Blanco, J.; Blanco-Portales, R. An atypical HLH transcriptional regulator plays a novel and important role in strawberry ripened receptacle. *BMC Plant Biol.* **2019**, *19*, 586. [CrossRef]
34. McKim, S.M. Moving on up—Controlling internode growth. *New Phytol.* **2020**, *226*, 672–678. [CrossRef]
35. Wang, K.L.; Zhang, Y.; Zhang, H.M.; Lin, X.C.; Xia, R.; Song, L.; Wu, A.M. MicroRNAs play important roles in regulating the rapid growth of the *Phyllostachys edulis* culm internode. *New Phytol.* **2021**, *231*, 2215–2230. [CrossRef]
36. Li, J.; Liu, Z.; Gao, C.; Miao, Y.; Cui, K. Overexpression of *DsEXLA2* gene from Dendrocalamus sinicus accelerates the plant growth rate of *Arabidopsis*. *Phytochemistry* **2022**, *199*, 113178. [CrossRef]
37. McKim, S.M. How plants grow up. *J. Integr. Plant Biol.* **2019**, *61*, 257–277. [CrossRef]
38. Yeh, S.H.; Lee, B.H.; Liao, S.C.; Tsai, M.H.; Tseng, Y.H.; Chang, H.C.; Yang, C.C.; Jan, H.C.; Chiu, Y.C.; Wang, A.Y. Identification of genes differentially expressed during the growth of *Bambusa oldhamii*. *Plant Physiol. Biochem.* **2013**, *63*, 217–226. [CrossRef]
39. Nagai, K.; Mori, Y.; Ishikawa, S.; Furuta, T.; Gamuyao, R.; Niimi, Y.; Hobo, T.; Fukuda, M.; Kojima, M.; Takebayashi, Y.; et al. Antagonistic regulation of the gibberellic acid response during stem growth in rice. *Nature* **2020**, *584*, 109–114. [CrossRef]
40. Lee, J.; Moon, S.; Jang, S.; Lee, S.; An, G.; Jung, K.-H.; Park, S.K. OsbHLH073 negatively regulates internode elongation and plant height by modulating ga homeostasis in rice. *Plants* **2020**, *9*, 547. [CrossRef]
41. Wang, Y.; Qiao, G.; Xu, J.; Jin, K.; Fan, M.; Ding, Y.; Wei, Q.; Zhuo, R. Anatomical characteristics and variation mechanisms on the thick-walled and dwarfed culm of shidu bamboo (*Phyllostachys nidularia* f. farcta). *Front. Plant Sci.* **2022**, *13*, 876658. [CrossRef] [PubMed]

42. Chen, M.; Guo, L.; Ramakrishnan, M.; Fei, Z.; Vinod, K.K.; Ding, Y.; Jiao, C.; Gao, Z.; Zha, R.; Wang, C.; et al. Rapid growth of Moso bamboo (*Phyllostachys edulis*): Cellular roadmaps, transcriptome dynamics, and environmental factors. *Plant Cell* **2022**, *34*, 3577–3610. [CrossRef] [PubMed]
43. Hofmann, N.R. A tripartite growth regulatory cascade of basic Helix-Loop-Helix transcription factors. *Plant Cell* **2012**, *24*, 4774. [CrossRef] [PubMed]
44. Ikeda, M.; Fujiwara, S.; Mitsuda, N.; Ohme-Takagi, M. A triantagonistic basic helix-loop-helix system regulates cell elongation in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4483–4497. [CrossRef] [PubMed]
45. Shin, K.; Lee, I.; Kim, E.; Park, S.K.; Soh, M.S.; Lee, S. PACLOBUTRAZOL-RESISTANCE gene family regulates floral organ growth with unequal genetic redundancy in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* **2019**, *20*, 869. [CrossRef]
46. Zhu, Z.; Liang, H.; Chen, G.; Li, F.; Wang, Y.; Liao, C.; Hu, Z. The bHLH transcription factor SlPRE2 regulates tomato fruit development and modulates plant response to gibberellin. *Plant Cell Rep.* **2019**, *38*, 1053–1064. [CrossRef]
47. Casal, J.J.; Fankhauser, C. Shade avoidance in the context of climate change. *Plant Physiol.* **2023**, *191*, 1475–1491. [CrossRef]
48. Galstyan, A.; Cifuentes-Esquivel, N.; Bou-Torrent, J.; Martinez-Garcia, J.F. The shade avoidance syndrome in *Arabidopsis*: A fundamental role for atypical basic helix-loop-helix proteins as transcriptional cofactors. *Plant J.* **2011**, *66*, 258–267. [CrossRef]
49. Huang, W.; Ding, Y.; Wang, S.; Song, C.; Wang, F. Growth and development responses of the Rhizome-Root system in *Pleioblastus pygmaeus* to light intensity. *Plants* **2022**, *11*, 2204. [CrossRef]
50. 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]
51. Fan, C.; Ma, J.; Guo, Q.; Li, X.; Wang, H.; Lu, M. Selection of reference genes for quantitative real-time PCR in bamboo (*Phyllostachys edulis*). *PLoS ONE* **2013**, *8*, e56573. [CrossRef]

International Journal of
Molecular Sciences

MDPI

Article

Morphological and Physio-Chemical Responses to PEG-Induced Water Stress in *Vanilla planifolia* and *V. pompona* Hybrids

José Martín Barreda-Castillo [1], Juan L. Monribot-Villanueva [2], Noé Velázquez-Rosas [1], Paul Bayman [3], José A. Guerrero-Analco [2,*] and Rebeca Alicia Menchaca-García [1,*]

[1] Centro de Investigaciones Tropicales, Universidad Veracruzana, Xalapa 91090, Veracruz, Mexico
[2] Red de Estudios Moleculares Avanzados, Instituto de Ecología A.C., Clúster Científico y Tecnológico BioMimic®, Xalapa 91073, Veracruz, Mexico
[3] Department of Biology, University of Puerto Rico, Río Piedras, San Juan 00925, Puerto Rico
* Correspondence: joseantonio.guerrero@inecol.mx (J.A.G.-A.); rmenchaca@uv.mx (R.A.M.-G.)

Abstract: *Vanilla planifolia* is an orchid of cultural and economic value. However, its cultivation in many tropical countries is threatened by water stress. In contrast, *V. pompona* is a species that is tolerant of prolonged periods of drought. Due to the need for plants' resistant to water stress, the use of hybrids of these two species is considered. Therefore, the objective of this study was to evaluate the morphological and physio-chemical responses of in vitro vanilla seedlings of the parental genotype *V. planifolia*, and the hybrids *V. planifolia* × *V. pompona* and *V. pompona* × *V. planifolia,* which were then exposed over five weeks to polyethylene glycol-induced water stress (−0.49 mPa). Stem and root length, relative growth rate, number of leaves and roots, stomatal conductance, specific leaf area, and leaf water content were determined. Metabolites potentially associated with the response to water stress were identified in leaves, through untargeted and targeted metabolomics. Both hybrids exhibited a smaller decrease in the morphophysiological responses compared to *V. planifolia* and exhibited an enrichment of metabolites such as carbohydrates, amino acids, purines, phenols, and organic acids. Hybrids of these two species are considered as a potential alternative to the traditional cultivation of vanilla to face drought in a global warming scenario.

Keywords: climate change; crops; genetic improvement

Citation: Barreda-Castillo, J.M.; Monribot-Villanueva, J.L.; Velázquez-Rosas, N.; Bayman, P.; Guerrero-Analco, J.A.; Menchaca-García, R.A. Morphological and Physio-Chemical Responses to PEG-Induced Water Stress in *Vanilla planifolia* and *V. pompona* Hybrids. *Int. J. Mol. Sci.* **2023**, *24*, 4690. https://doi.org/10.3390/ijms24054690

Academic Editor: Wajid Zaman

Received: 28 January 2023
Revised: 17 February 2023
Accepted: 22 February 2023
Published: 28 February 2023

1. Introduction

Vanilla (*Vanilla planifolia* Andrews) is an important crop worldwide [1]. Its fruit, whose commercial value recently reached USD 600 per kg, is used in food, perfumes, and pharmaceuticals [1,2]. Vanilla's fruit is appreciated because of its aromatic quality, which is provided by phenolic compounds, of which vanillin is the most important [2]. Despite its economic importance, in Mexico, *V. planifolia* is subject to special protection (NOM-059-SEMARNAT-2019) and it is also protected worldwide by the Convention on International Trade in Endangered Species of Wild Flora and Fauna, and the International Union for Conservation of Nature Red List [3]. *V. planifolia* is seriously threatened due to overexploitation and habitat loss [4]. In addition, genetic erosion has occurred due to the fact that plants for cultivation are always propagated asexually [4,5], making vanilla more susceptible to several biotic and abiotic factors, including water deficiency [5].

Vanilla, like many other crops, is threatened by water stress due to drought, which is expected to increase in the face of climate change [6]. Originally cultivated in Mexico, vanilla is mainly cultivated in tropical countries such as Madagascar and Indonesia [3,4]. Tropical countries will be seriously affected by climate change, with increased temperatures and longer drought periods [6,7]. In a future scenario of climate change, *V. planifolia* will

be affected in terms of its growth and development [2,4,8], because it lacks the efficient response mechanisms to deal with water stress [3–5]. Therefore, vanilla plants with improved tolerance to water stress are needed.

Plants that are resistant to water stress are capable of adapting their physiology, growth, and anatomy in order to cope with the lack of available water [9]. Roots production and elongation must be increased [10], and stomata are closed in order to reduce the water lost by transpiration [11,12]. Stomatal closure also reduces CO_2 uptake and consequently decreases photosynthetic activity, which reduces the specific leaf area (SLA) and therefore, the relative growth rate (RGR) [11,13,14]. For this reason, water-stress-tolerant plants usually are smaller in surface area than drought-sensitive plants under full hydration [14].

In addition to the anatomical and physiological responses, plants must modify their metabolism in order to tolerate water stress [11,14]. Due to the decrease in turgor, plants accumulate within their cells' metabolites, which act as osmolytes to keep membranes and proteins stable [15,16]. Due to the effects of water stress on photosynthesis, reactive oxygen species (ROS) accumulate within the cells causing cell damage [17]. For this reason, the accumulation of metabolites with antioxidant activity is also necessary [17,18]. Consequently, the accumulation of osmolytes and antioxidants, such as carbohydrates [16,19], amino acids [15,20,21], purines [15,20,22], organic acids [23], and phenols [24,25], among other specialized metabolites [26,27], has been reported in plants that are tolerant to water stress.

Through genetic improvement it is possible to obtain plants with enhanced morphological and physio-chemical responses to water stress [28,29]. The generation of hybrids might allow a combination of phenotypical characteristics of the parental species and also the appearance of new characters [28–30]. In the case of vanilla, hybrids between local Mexican species *V. planifolia* (PL) (favored for aroma and flavor) [1] and *V. pompona* (PO) (described as more resistant to water stress) [31] could display the necessary tolerance to water stress, while preserving the aromatic qualities required by the industry [4,28,29]. These hybrids have already been shown to have improved resistance to *Fusarium oxysporum* f. sp. *vanillae* [32], vanilla's most important pathogen, but their response to water stress has not been addressed yet.

For the study of water stress, previous research has used osmotically active substances such as polyethylene glycol (PEG) to reduce the water potential in the medium in vitro [33,34]. Few studies have simultaneously focused on the study of morphological and physio-chemical responses to understand the possible mechanisms of tolerance to induced water stress [15,16,35–37]. Thus, the objective of this work was to determine the morphological and physio-chemical responses to PEG-induced water stress of *V. planifolia* (PL) and *V. pompona* (PO) hybrids as a potential alternative to face the challenges of climate change.

2. Results

2.1. Morphological and Physiological Responses to Water Stress

The morphological and physiological parameters evaluated in the parental genotype PL, and in the hybrids *V. planifolia* × *V. pompona* (PL × PO) and *V. pompona* × *V. planifolia* (PO × PL) exposed to PEG-induced water stress (−0.49 mPa) during five weeks were stem and root length, relative growth rate (RGR), number of leaves and roots, stomatal conductance (SC), specific leaf area (SLA), and leaf water content (LWC). Stem length (SL) was reduced due to water stress in PL and PL × PO ($\chi^2_{(2,57)} = 8.78$, $p < 0.05$), whereas no reduction was observed in PO × PL (Figure 1A). RGR decreased due to water stress in PL and PL × PO ($F_{(2,57)} = 6.37$, $p < 0.01$). The most affected genotype was PL, while PL × PO RGR was not affected (Figure 1B). The root length (RL) was modified due to water stress in all genotypes evaluated; it decreased in PL and PL × PO and increased in PO × PL ($\chi^2_{(2,57)} = 19.76$, $p < 0.01$) (Figure 1C).

The number of new leaves (NL) was significantly reduced by water stress in PL and PL × PO ($\chi^2_{(2,57)} = 6.35$, $p < 0.01$) while in PO × PL, it remained unchanged (Figure 1D). In contrast, there was a higher number of new roots (NR) that were produced under water

stress in PL ($\chi^2_{(2,57)} = 0.015$, $p < 0.01$), which doubled the number of roots in comparison to its controls. However, in the hybrids, differences between the two conditions were not significant (Figure 1E). The leaves of all vanilla genotypes were smaller under water stress (Figure 2D–F) than the controls (Figure 2A–C).

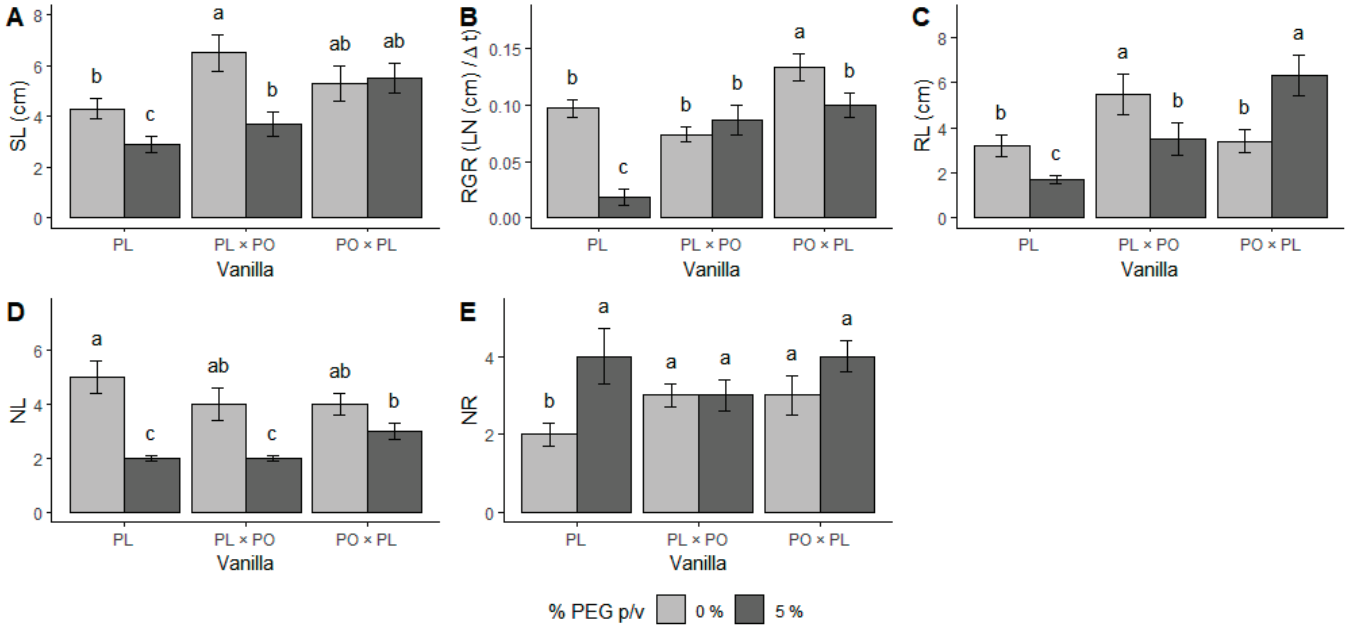


Figure 1. Water stress affects growth of *Vanilla planifolia* (PL) seedlings and *V. planifolia* × *V. pompona* (PL × PO) and *V. pompona* × *V. planifolia* (PO × PL) hybrids. (**A**) Stem length (SL); (**B**) relative growth rate (RGR); (**C**) root length (RL); (**D**) number of new leaves (NL); and (**E**) number of new roots (NR). Gray bars: 0% PEG (control group, −0.08 mPa). Black bars: 5% PEG and hydric stress (−0.49 mPa). Bars with different letters are significantly different ($p < 0.05$), according to GLM test.

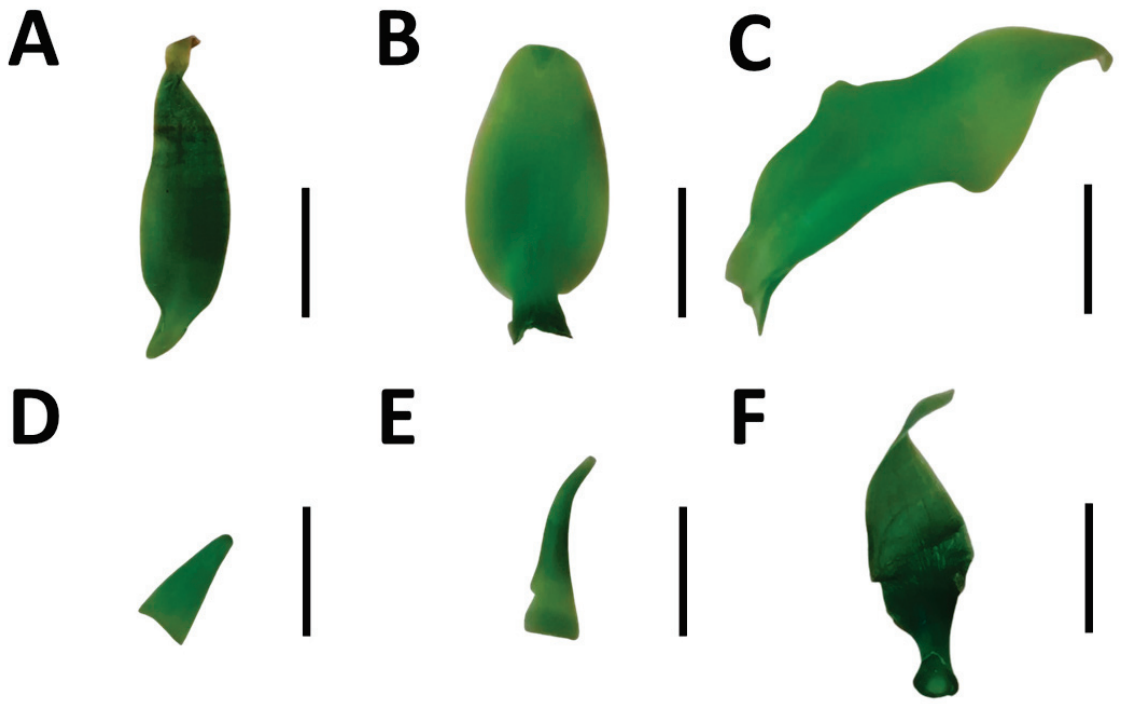


Figure 2. Leaves of vanilla seedlings without and with PEG-induced water stress. (**A**,**D**) *Vanilla planifolia* (PL); (**B**,**E**) *V. planifolia* × *V. pompona* (PL × PO); (**C**,**F**) *V. pompona* × *V. planifolia* (PO × PL); (**A**–**C**) control leaves (−0.08 mPa); and (**D**–**F**) leaves produced under water stress (−0.49 mPa). Scale bars = 10 mm.

The ratio of stem length/longest root length (SL/RL) was modified by water stress only in the PO × PL hybrid ($\chi^2_{(2,57)} = 16.77$, $p < 0.01$), suggesting a sign of the prioritization of roots compared to the leaves (Figure 3A). The ratio of the number of leaves/number of roots (NL/NR) was significantly reduced by water stress in all the vanillas studied ($\chi^2_{(2,57)} = 3.88$, $p < 0.05$) (Figure 3B). Stomatal conductivity (SC) decreased due to water stress in PL and in PL × PO ($F_{(2,57)} = 41.1$, $p < 0.01$). In contrast, the reduction in SC in PO × PL by water stress was not significant (Figure 3C). Similarly, the specific leaf area (SLA) was significantly reduced under water stress in PL and PL × PO, but not in PO × PL ($F_{(2,57)} = 12.28$, $p < 0.01$) (Figure 3D). All vanillas exhibited a significant decrease in leaf water content (LWC) in water stress ($F_{(2,57)} = 14.7$, $p < 0.01$) (Figure 3E).

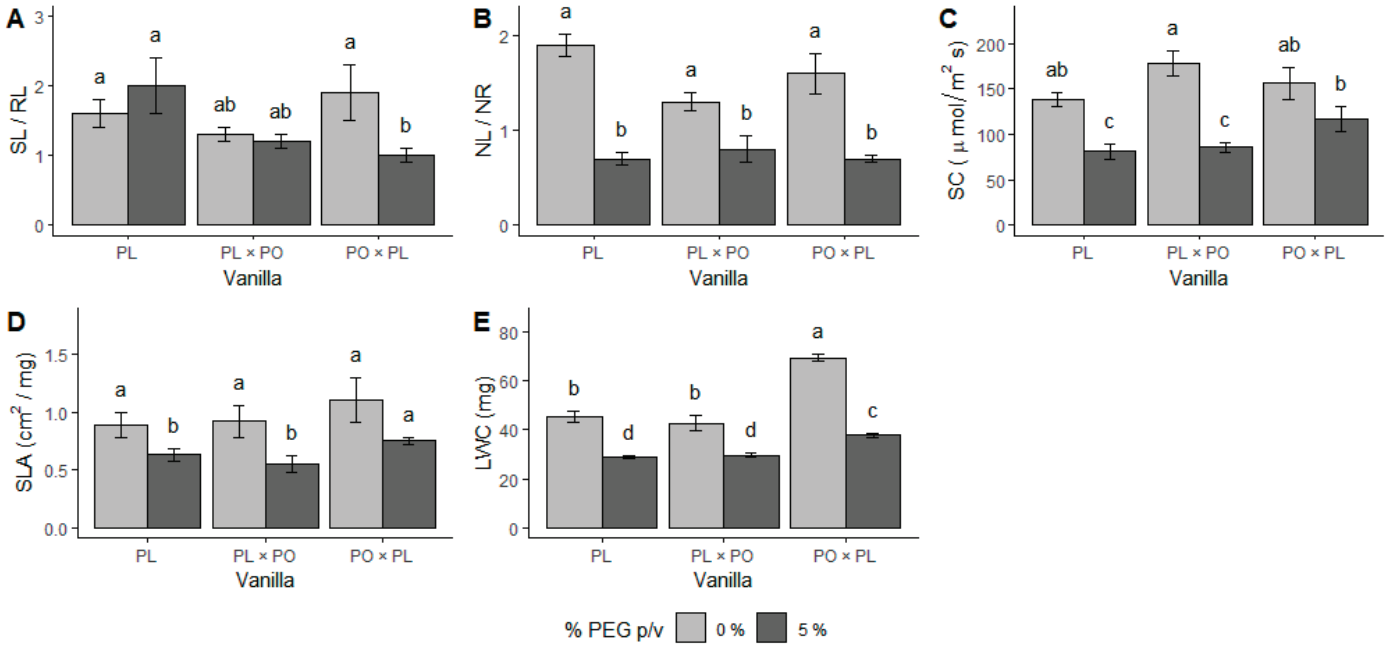


Figure 3. Water stress affects growth ratios and physiology of *Vanilla planifolia* (PL) seedlings and *V. planifolia* × *V. pompona* (PL × PO) and *V. pompona* × *V. planifolia* (PO × PL) hybrids. (**A**) Stem length/longest root length ratio (SL/RL); (**B**) number of leaves/number of roots ratio (NL/NR); (**C**) stomatal conductance (SC); (**D**) specific leaf area (SLA); and (**E**) leaf water content (LWC). Gray bars: 0% PEG (control group, −0.08 mPa). Black bars: 5% PEG, hydric stress (−0.49 mPa). Bars with different letters are significantly different ($p < 0.05$), according to GLM or ANOVA and post hoc Tukey test.

2.2. Physio-Chemical Response to Water Stress

Untargeted metabolomic analysis based on mass spectrometry detected 812 signals (retention time–mass/charge features; RT-m/z) in the positive mode of ionization (electrospray) and 382 signals in the negative mode. The effect of water stress on the vanilla plants was evidenced by a heatmap (Figure 4A). Two major clusters were identified with distinct patterns of altered metabolite abundances: all the vanillas under full hydration (control groups) were grouped in one cluster, and those exposed to water stress in another, except for PL at 20 days of water stress, which grouped in the clade with all the vanillas under full hydration. Metabolites belonging to the pathways of amino acids, carbohydrates, lipids, phenols, and organic acids, among others, were found (Figure 4B). The names of the identified metabolic pathways, as well as the names of the tentatively identified metabolites, are shown in Tables S1 and S2, respectively. Principal component analysis (PCA) was performed for drought and full hydration plants at the two exposure times. The first component (PC1) explained 58.4% of total variation, whereas the second component (PC2) explained 12.6% variation. The scores between PC1 and PC2 revealed two distinct groups associated with drought and hydration conditions, suggesting a physio-chemical change due to water stress (Figure 4C). Similar to the heatmap (Figure 4A), PL at 20 days of water stress exposure also grouped with all the vanillas under full hydration. The metabolite interaction network analysis (Figure 4D) allowed us to identify the amino acids L-tryptophan, L-phenylalanine, L-arginine, L-glutamic acid, and ornithine, as well the purine adenosine, as main nodes among all the metabolites were detected and tentatively identified.

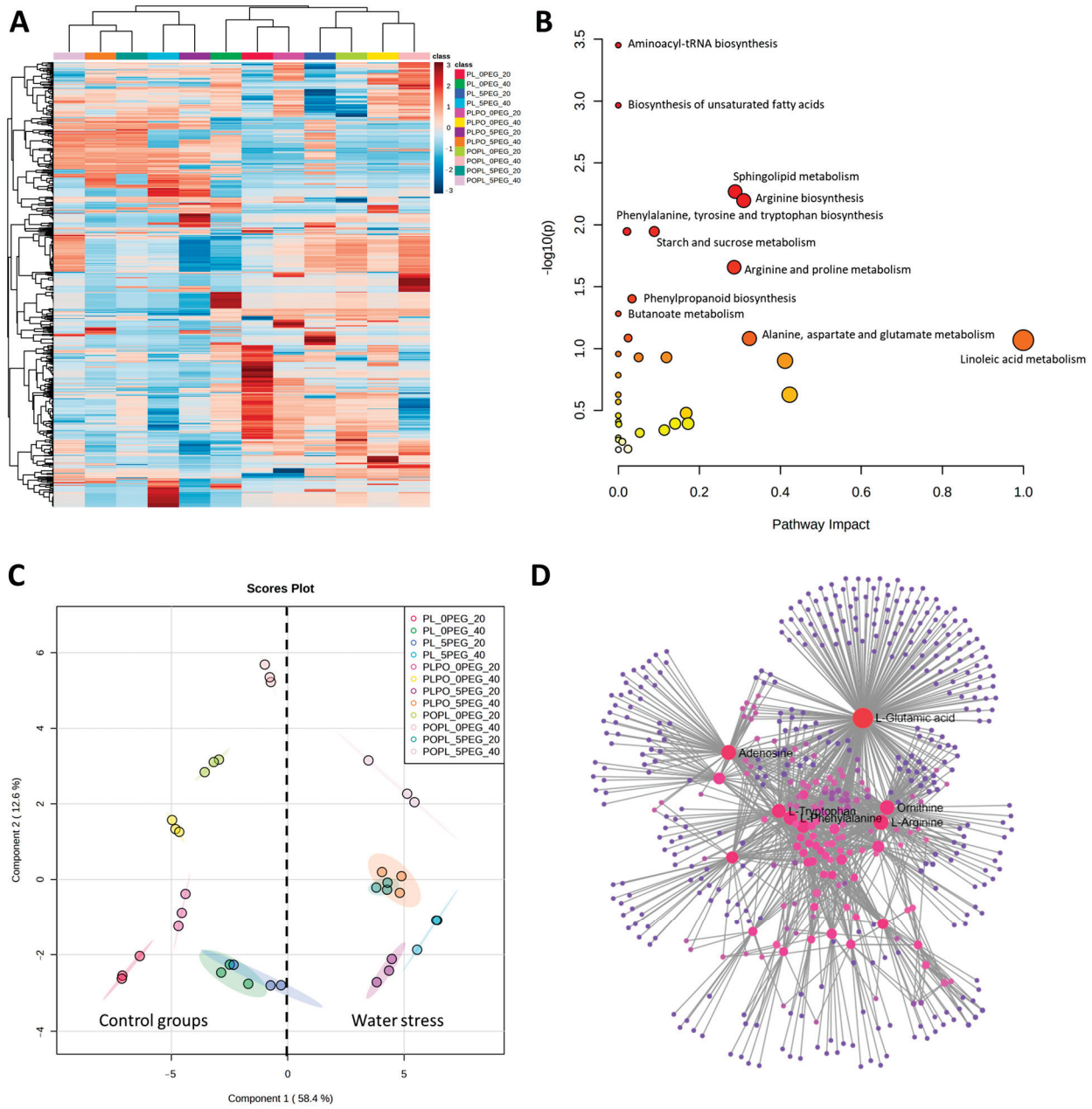


Figure 4. Metabolomics of *Vanilla* plants with and without water stress. (**A**) Hierarchical ordering heatmap, (**B**) pathway analysis, (**C**) principal component analysis (PCA), and (**D**) network analysis of metabolomic data obtained by UPLC-MS-QTOF from vanilla seedlings exposed to 20 and 40 days of PEG-induced water stress. The heatmap was generated using Pearson and Ward for distance measure and clustering algorithm, respectively. The pathway analysis was determined by Mummichog algorithm, KEGG database, and *Oryza sativa* library. PL: *Vanilla planifolia*, PLPO: *V. planifolia* × *V. pompona*, POPL: *V. pompona* × *V. planifolia*, 0 PEG: control group (−0.08 mPa), 5 PEG: hydric stress (−0.49 mPa), 20: 20 days of water stress exposure, and 40: 40 days of water stress exposure.

In order to go deeper into the physio-chemical changes exhibited by PL, PL × PO, and PO × PL plants, we performed a fold change analysis comparing both hybrids against PL at two different times of water stress exposition (Figure 5). At 20 days of water stress exposure, both hybrids accumulated 9,10-epoxystearic acid, butyric acid, gluconic acid, hexadecanedioic acid, sphinganine, adenosine, guanine, biotin, ketoleucine, L-glutamic acid, L-arginine, L-tryptophan, L-tyrosine, L-phenylalanine, *p*-hydroxybenzoic acid, sucrose, and a disaccharide, in comparison to PL. Xanthine, sphingosine, galabiose, and

pheophorbide a were accumulated only in PL × PO; α-linoleic acid, 8,11,14-eicosatrienoic acid, ascorbic acid and L-asparagine were accumulated only in PO × PL (Figure 5A,B; detailed information in Tables S3–S6). At 40 days of water stress exposure, both hybrids accumulated 8,11,14-eicosatrienoic acid, ascorbic acid, butyric acid, gluconic acid, α-linolenic acid, γ-linolenic acid, adenosine, guanine, ketoleucine, L-tryptophan, L-phenylalanine, galabiose, sucrose, and a disaccharide, in comparison to PL. Xanthine was accumulated only in PL × PO, while 9,10-epoxystearic acid, phytosphingosine, sphinganine, sphingosine, deoxyguanosine, L-glutamic acid, L-arginine, L-asparagine, *p*-hydroxybenzoic acid, L-tyrosine, pheophorbide a, and lotaustralin were accumulated only in PO × PL (Figure 5C,D; detailed information in Tables S3–S6).

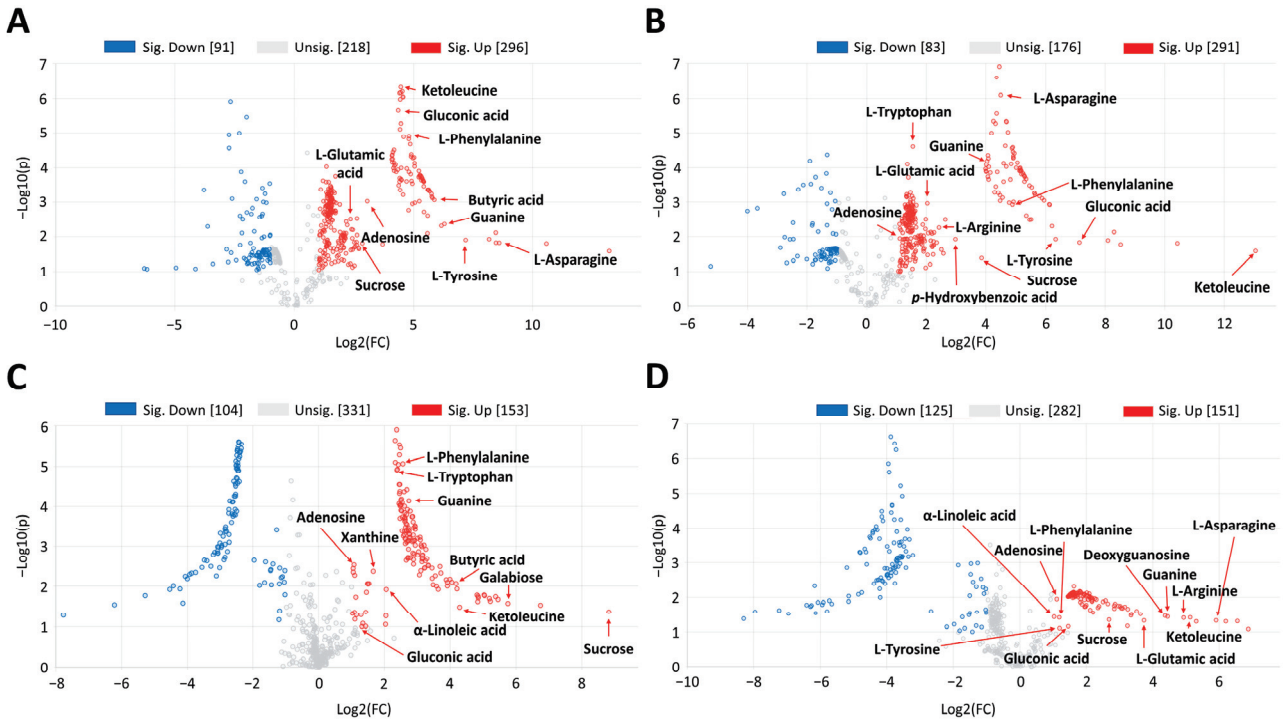


Figure 5. Volcano plots of metabolomic data obtained by UPLC-MS-QTOF comparing *V. planifolia* × *V. pompona* (PL × PO) and *V. pompona* × *V. planifolia* (PO × PL) hybrids versus *V. planifolia* (PL) under water stress exposure. (**A**) PL × PO/PL at 20 days of water stress exposure. (**B**) PO × PL/PL at 20 days of water stress exposure. (**C**) PL × PO/PL at 40 days of water stress exposure. (**D**) PO × PL/PL at 40 days of water stress exposure. Some differential metabolites (fold change >2) that are tentatively identified are shown in each subfigure; the complete list is in Tables S3–S6.

In addition, ten phenolic compounds and one amino acid were identified and quantified through targeted metabolomics in all the vanillas exposed to water stress and full hydration. The phenolics identified in both conditions were salicylic acid, vanillic acid, vanillin, ferulic acid, sinapic acid, 4-coumaric acid, *trans*-cinnamic acid, luteolin, protocatechuic acid, and 4-hydroxybenzoic acid. In addition, phenylalanine, which is a phenolics precursor, was also identified and quantified (Figure 6; the concentration of each phenolic compound is shown in Table S7). Phenylalanine had the highest concentration among all the compounds quantified. Salicylic acid was detected only in PO × PL at 20 days, and a decrease in its concentration was observed at 40 days. Luteolin was identified in all vanillas under full hydration, but under water stress, it could be observed only in PL × PO at 40 days. The content of vanillic acid, vanillin, ferulic acid, *trans*-cinnamic acid, protocatechuic acid, and 4-hydroxybenzoic acid decreased in all the vanillas, and only sinapic acid and 4-coumaric acid exhibited an enrichment under water stress.

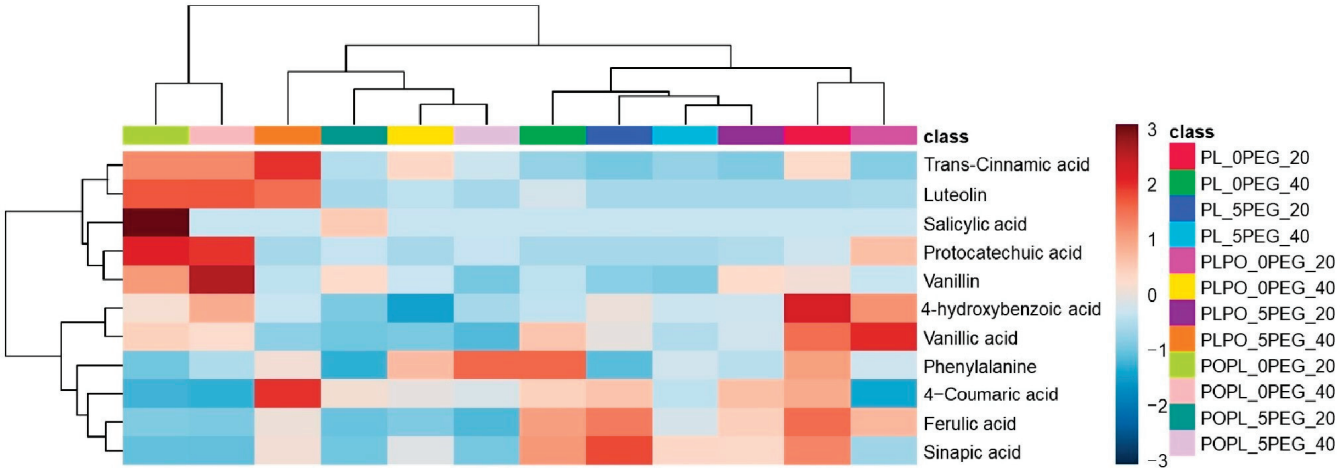


Figure 6. Hierarchical ordering heatmap of phenolic compounds identified by targeted metabolomics as result of exposure of vanilla seedlings to 20 and 40 days of PEG-induced water stress. The heatmap was generated using Pearson and Ward for distance measure and clustering algorithm, respectively. PL: *Vanilla planifolia*, PLPO: *V. planifolia* × *V. pompona*, POPL: *V. pompona* × *V. planifolia*, 0 PEG: hydration status (control group, −0.08 mPa), 5 PEG: hydric stress (−0.49 mPa), 20:20 days of water stress exposure, and 40:40 days of water stress exposure. Complementary information about the phenolic compounds is shown in Table S7.

3. Discussion

3.1. Morphological and Physiological Responses

Plants use various strategies to improve their water management efficiency when water is limited, several of which were observed in the vanilla seedlings exposed to water stress. In this work, we determined the morphological and physio-chemical responses to PEG-induced water stress of *V. planifolia* and *V. pompona* hybrids and the parental *V. planifolia* genotype. PO × PL hybrid increased root length rather than root number, which allowed for an increased water uptake, a strategy observed in water-stress-tolerant orchids [9,38,39], whereas the PL × PO stem length and the number of leaves produced under water stress were reduced, reducing water loss [11,40].

The decrease in LWC is considered a direct indicator of the decrease in water uptake by the plant due to water stress [20]. In order to conserve water, the production of leaves decreases, stomatal closure is induced, and SLA is reduced as well [41–45]. This might explain why PL and PL × PO exhibited the greatest reduction in leaves produced, stomatal closure, and SLA as a possible strategy to conserve water [18]. In contrast, PO × PL exhibited the lowest reduction in LWC, perhaps explaining why it produced more leaves than the other genotypes, and why it was able to maintain the stomata open and not reduce SLA.

Although transpiration is reduced by the morphological responses previously mentioned, they have a negative effect on CO_2 uptake and the energy investment capacity, and the RGR is affected [11,13,18,43]. The RGR is closely and positively associated with the SLA, as it is an indicator of the plant's investment capacity [43,44]. PL (compared to PO × PL) exhibited lower values of both the RGR and SLA and thus, it had the shortest stem length. The SLA is highly plastic in several environmental gradients (light, nutrients, and water stress) and its responses are consistent in different groups of plants, including orchids [46,47]. It is related to the LWC, SC, RGR, and biochemical changes [41–45]. The unchanged SLA in PO × PL may be explained because this genotype exhibited higher LWC, SC, and RGR, as well as an accumulation of several osmolytes and salicylic acid (discussed later), allowing it to conserve water without having to reduce the leaf area [41,43].

3.2. Physio-Chemical Responses

Under water stress, plants often modify primary and secondary metabolism with the aim of accumulating metabolites that act as osmolytes and antioxidants [16,45]. Changes in the metabolic response of both hybrids can be observed either in the heatmap or the PCA (Figure 4A,C). Both hybrids exhibited metabolic changes at 20 days of water stress exposure, as has been seen in other drought-tolerant plants [10,35]. In contrast, PL at 20 days of exposure did not change, suggesting that this species might be slower to modify its metabolism. This could be one of the reasons why PL was more susceptible to water stress [36]. Both hybrids accumulated carbohydrates at 20 and 40 days of water stress exposure (Tables S3–S6). Sucrose is reported to act as an osmolyte and antioxidant, and it could contribute to osmotic maintenance and ROS clearance in the system, and promote root elongation (Figure 7) [19–21]. Amino acids are an important part of osmotic regulation because they are both osmolytes and antioxidants [33,48]. In both hybrids (compared to PL) we found the accumulation of L-asparagine, ketoleucine, L-glutamic acid, L-arginine, L-tryptophan, L-tyrosine, and L-phenylalanine (Tables S3–S6), suggesting they play a role in the response to water stress conditions (Figure 7).

Regarding purines, xanthine was observed in PL × PO at 20 and 40 days of water stress exposure, and deoxyguanosine in PO × PL just at 20 days, while guanine and adenosine were present in both hybrids at both time points (Tables S3–S6). Purines are suggested to play a key role in tolerance to water stress given their antioxidant activity, as it has been reported in orchids of the genus *Dendrobium* [15,16]. In addition, they participated in other metabolic processes, such as the synthesis of nucleic acids, as precursors for the synthesis of primary and secondary metabolites and as a source of energy [14–16,48–51].

In vascular plants, organic acids contribute to osmotic maintenance, in addition to the cell structure, energy storage, and signaling, mitigating the negative impact of environmental stressors [15,33,52]. Linoleic acid, gluconic acid, and hexadecanedioic acid were accumulated at 20 and 40 days of water stress exposure in both hybrids, potentially because of their antioxidant activity [53–55] Ascorbic acid was also found (Tables S3–S6), which may have a role as growth promotor [56].

The change in the concentration of phenols in plants after water stress exposure has been previously reported [16,48]. Phenylalanine (precursor of phenolic compounds in the shikimate pathway), ferulic acid, *trans*-cinnamic acid, sinapic acid, 4-coumaric acid, and protocatechuic acid were found in higher concentrations in both hybrids compared to PL. This may reflect their osmoprotective and antioxidant activity, and because they might have contributed to maintain photosynthesis, as has been reported [10,57–65] (Table S7). The accumulation of these phenolic compounds could also be related to the greater RGR value found in both hybrids (Figure 7). On the other hand, 4-hydroxybenzoic acid, vanillic acid, vanillin, and luteolin decreased in all the plants studied when exposed to water stress (Table S7), which is possibly related to ROS clearance [66–71].

Salicylic acid only appeared in PO × PL. This phenolic compound is commonly called "the stress hormone" since it activates various plant defense mechanisms under biotic and abiotic stress [72]. The mechanisms of action of salicylic acid include the accumulation of osmolytes such as amino acids and soluble sugars in order to maintain osmotic homeostasis [73]; however, it also promotes the production of secondary/specialized metabolites, such as terpenes, other phenols, and nitrogenous compounds, stimulating the antioxidant system [72–74]. In addition to biochemical responses, salicylic acid can play a role in morphological and physiological responses, since it promotes root elongation, as well as stomatal closure [74,75] (Figure 7), which agrees with our results. The presence of salicylic acid only in PO × PL might be related to a better response to water stress, in comparison with the reciprocal hybrid and the parental genotype.

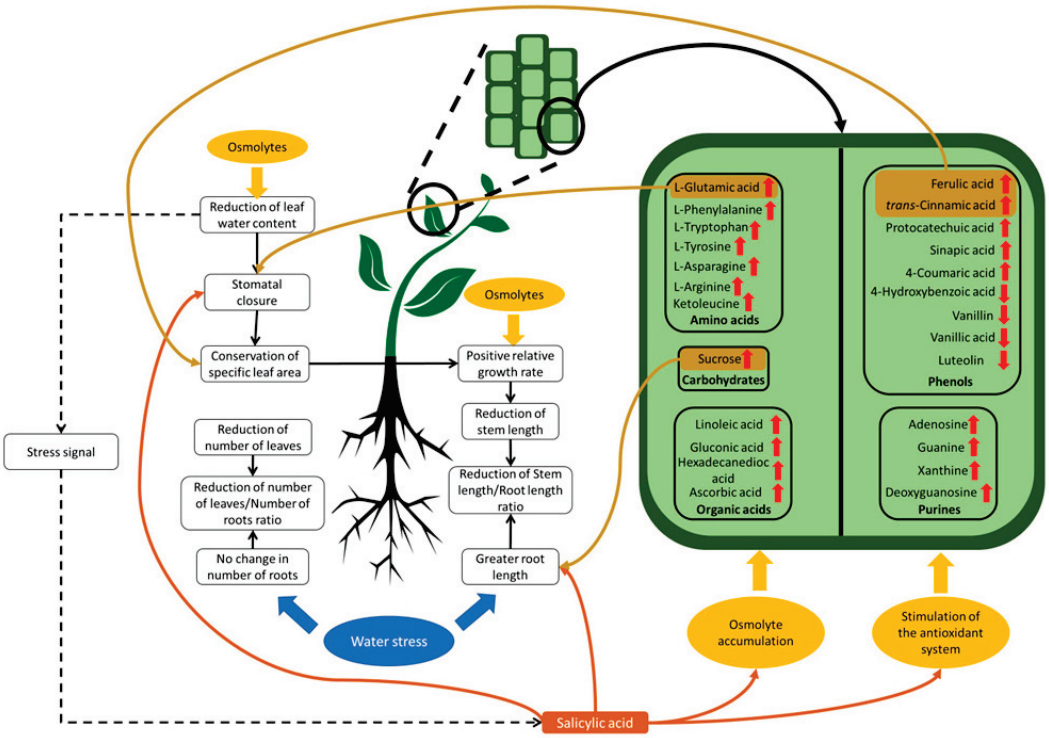


Figure 7. Morphological and physio-chemical responses of vanilla hybrids exposed to PEG-induced water stress. Water-stress-tolerant plants exhibit changes in morphology such as root elongation and reduction in leaf production and size, which increase water uptake. Additionally, stomatal closure is induced. Primary and secondary metabolism is modified in order to promote the accumulation of metabolites that act as osmolytes and antioxidants. Red arrows indicate an increase in the content of certain metabolites, while purple arrows indicate a decrease. Figure includes data from the present study and from the literature [11,14,16–20,40,41,45–49,57–62,73–75].

Overall, both hybrids had multiple responses to tolerate PEG-induced water stress. However, the PO × PL hybrid exhibited better tolerance than the reverse cross. In plant hybrids, there is usually a greater expression of the characteristics of the ovule donor over those of the pollen donor [76,77]. This may partly reflect the expression of maternal traits in chloroplasts and mitochondria DNA [78]. Both organelles are usually inherited from maternal parent [78,79]. Another plausible explanation for the discrepancy between the reciprocal hybrids is epigenetic regulation, since this regulation in plants is more sensitive to environmental stress [80,81]; however, the hypothesis must be tested in the future. Although *V. pompona* has been previously used in hybridization programs [76,82,83], it has rarely been used as a maternal species. The present study suggests that in future work, *V. pompona* should be used as a donor of ovules instead of pollen, and also, future work must provide specific evidence of the advantages of these *V. pompona* × *V. planifolia* hybrids over the parental species.

4. Materials and Methods

4.1. Biological Material

In vitro seedlings of *V. planifolia* (PL), *V. planifolia* × *V. pompona* (PL × PO) and *V. pompona* × *V. planifolia* (PO × PL) were used [84]; the first species mentioned is the ovule donor and the second is the pollen donor. Previous subcultures were performed in full-strength Murashige and Skoog medium (MS) supplemented with 6-benzylaminopurine at 0.2 mg/L [85] for vegetative multiplication to acquire enough plant material. Once the seedlings reached 2 cm in height and generated at least one root of 1 cm in length [86], they were subcultured in MS without PEG (control group, −0.08 mPa), and MS medium

with PEG 5% w/v was added to induce moderate water stress (−0.49 mPa) [31]. The water potential in the medium was calculated using the following formula [87]:

$$\varphi = 0.84[PEG]^2T - 118[PEG]^2 - 0.267[PEG]T - 11.8[PEG]$$

where φ: water potential in megapascals (mPa); [PEG]: grams of PEG per milliliters of distilled water; and T: temperature of the medium (°C). There was one vanilla seedling per bottle, 10 repetitions per treatment, and they were placed in random blocks. The seedlings were placed in glass jars with a capacity of 235 mL (10.51 cm high × 6.5 cm diameter) on 25 mL of culture medium. Exposure to PEG-induced water stress lasted five weeks. The PEG-induced water stress condition was established at the beginning of the experiment and no further evaluations of the culture medium were conducted [33]. However, we performed comparisons at the same time between water-stressed versus full-hydrated (control groups) seedlings and between hybrids versus the parental genotype. In this work, *V. pompona* seedlings were not included because of the difficulty of germinating them in in vitro conditions.

4.2. Morphological and Physiological Responses to Water Stress

The increase in shoot length was recorded once a week during the five weeks. The comparison between treatments was made from the final values. The relative growth rate (RGR) was calculated from the weekly measurements, using the following formula [88]:

$$RGR = \frac{LnM_2 - LnM_1}{\Delta t}$$

where LnM_1: natural logarithm of the initial measurement and LnM_2: natural logarithm of the following measurement. The length of the longest root produced by each plant was recorded at the end of the stress exposure time. The number of leaves and roots produced by the vanillas under the two water conditions and the leaf: root ratio and stem: longest root ratio were calculated. At the end of the period of exposure to water stress, three seedlings for each treatment were randomly selected, and one leaf of each seedling was harvested to be used for the evaluation of stomatal conductance (SC), specific leaf area (*SLA*), and leaf water content (LWC). Stomatal conductance was measured on the abaxial side of the leaves using a porometer (Decagon Devices, model SC-1); the vapor flow over the leaf surface was recorded (μ mol/m^2 s) hourly between 9:00 and 14:00 h. For the determination of SLA, the area of the freshly cut leaves was measured with the *Imagej* 1.8.0 program [89] and was calculated with the SLA formula for in vitro seedlings [90]:

$$SLA = \frac{fresh\ leaf\ area\ (\mathrm{cm}^2)}{dry\ leaf\ weight\ (\mathrm{mg})}$$

For the measurement of the leaf water content, the leaves of each treatment were weighed at the moment of being cut (fresh weight), and later, they were dried in an oven at 60 °C for 24 h to obtain the dry weight [91]. The leaf water content was calculated as the difference between fresh weight and dry weight.

4.3. Physio-Chemical Response to Water Stress

4.3.1. Untargeted Metabolomics

Seedling leaves of each treatment were harvested at 20 and 40 days of exposure to water stress. The leaves were placed in liquid nitrogen immediately after being cut and then stored at −80 °C until subsequent analyses. The samples were lyophilized (Freezone 1, Labconco, Kansas City, MO, USA), and pulverized in a mortar and pestle with liquid nitrogen. To obtain the crude extracts, accelerated solvent extraction system was used (ASE350, Dionex, Thermo Scientific, Waltham, MA, USA) [44]. The solvent used was methanol (HPLC grade) and water (MilliQ) in 80:20 ratio. Excess solvent was removed by evaporation under reduced pressure in rotary evaporator (Rotovap® RII, Büchi,

Newmarket, UK). From the resulting extracts, 500 µL of each sample was taken, and they were then placed in 1.5 mL UPLC tubes. The extract was filtered using a 0.45 µm filter prior to analysis. The samples were injected at a concentration of 20 mg/mL. The samples were evaluated in triplicates.

For the identification of distinctive metabolites associated with water stress, untargeted metabolomic analyses were carried out [92]. It was performed on an ultra-high-performance liquid chromatography system (UPLC, Acquity class I, Waters™, Milford, MA, USA), coupled to a high-resolution quadrupole time-of-flight mass spectrometer (QTOF, HDMI Synapt G2-Si model, Waters™). Chromatographic separation was carried out on an Acquity BEH column. (1.7 µm, 2.1 × 50 mm) at a temperature of 40 °C and 15 °C for the sample. The mobile phase consisted of (A) water and (B) acetonitrile, both with 0.1% formic acid. The conditions of the mobile phase gradient were 0–13 min linear gradient 1–80% B, 13–14 min 80% B isocratic, and 14–15 min linear gradient 80–1% B (total analysis time was 20 min). The flow rate was 0.3 mL/min and 5 µL of each extract was injected.

Mass spectrometry analysis was performed with an electrospray ionization source in positive and negative mode. The sampling cone, capillary, and source offset voltages were 3000, 40, and 80 V, respectively. The source temperature was 100 °C and the desolvation temperature was 20 °C. The desolvation gas flow was 600 L/h and the nebulizer pressure was 6.5 bar. The mass range was from 50 to 1200 Da, function 1 CE, and 6 V; function 2 CER 10–30 V and exploration time 0.5 s. Leucine–enkephalin (556.2771 $[M + H]^+$; 554.2615 $[M-H]^-$) was used as reference solution. Mass/charge ratios (m/z), retention time (RT), and peak intensity (total counts) were obtained and analyzed using Masslynx and MarkerLynx software (version 4.1, Waters™).

4.3.2. Identification of Phenolics Compounds by UPLC-MS-MS

The extracts prepared for untargeted metabolomics were also used for the identification and quantification of individual phenolic compounds in leaves. Phenolic-targeted metabolomics was performed by ultra-high performance liquid chromatography (Agilent 1290 series) coupled with a triple quadrupole mass spectrometer (Agilent 6460), with a dynamic multiple reaction monitoring (dMRM) acquisition method, following the protocol established by [93] that included a total of 60 phenolic compounds. The chromatographic analysis was carried out on a ZORBAX SB-C18 column (1.8 µm, 2.1 × 50 mm; Agilent Technologies) with the column oven temperature at 40 °C. The mobile phase consisted of (A) water and (B) acetonitrile, both containing 0.1% formic acid. The gradient conditions of the mobile phase were as follows: 0 min 1% B, 0.1–40 min linear gradient 1–40% B, 40.1–42 min linear gradient 40–90% B, 42.1–44 min isocratic 90% B, 44.1–46 min linear gradient 90–1% B, and 46.1–47 min 1% B isocratic (total run time 47 min). The flow rate was 0.3 mL/min, and 2 µL was the sample injection volume. The ESI source operated in positive and negative ionization modes. The desolvation temperature was 300 °C, the cone gas (N_2) flow was 5 L/min, the nebulizer pressure was 45 Psi, the sheath gas temperature was 250 °C, the sheath gas flow was 11 L/min, the capillary voltage (positive and negative) was 3500 V, and the nozzle voltage (positive and negative) was 500 V. The fragmentor voltage was 100 V and the cell accelerator voltage was 7 V for all compounds. The identity was confirmed by co-elution with authentic standards under the same analytical conditions described above for each compound. For quantitation of each phenolic compound a calibration curve in a concentration range of 1–9 µM was prepared (r^2 values $\geq$ 0.97 were considered for the linearity range). The data were processed using the MassHunter Workstation Software, version B.06.00 (Agilent Technologies). The results were expressed as µg/g of sample (dry weight).

4.4. Statistical Analysis

For the morphological and physiological responses, two-level factorial arrangement was used: PEG concentration and vanilla genotype. Total length, longest root length, and the relationships between total length/root length and number of leaves/number

of roots were analyzed using generalized linear model (GLM), gamma distribution, and inverse link function ($p < 0.05$). The number of leaves and roots produced was analyzed using GLM, Poisson distribution, logarithm link function ($p < 0.05$). RGR, SLA, stomatal conductance, and leaf water content, which were then analyzed by two-way analysis of variance (ANOVA) and Tukey's post hoc test ($p < 0.05$). R 4.0.3 software was used [94], with packages Rmisc [95], agricolae [96], multcomp [97], and ggplot2 [98].

For the biochemical response, the intensities of the RT_m/z signals were used to elaborate heatmaps, PCA, network analysis, and volcano plots, using the MetaboAnalyst 5.0 software [99]. The main purpose of heatmap is to display grouping based on the abundance of the different spectrometric features (m/z_Rt) detected. The analysis of the enrichment of metabolic pathways was performed through the Mummichog algorithm, using the same MetaboAnalyst 5.0 software. For this, the KEGG database (Kyoto Encyclopedia of Genes and Genomes) was used [100], with *Oryza sativa* library for monocots. The RT-m/z signals that contributed to the discrimination between analyzed groups (fold change (FC) values greater or equal than 2) were considered differential chemical markers between conditions [101–104]. For the tentative identification of the distinctive metabolites (markers) that could be associated with changes in water potential, the KEGG databases were used [105] with a literature review, with a maximum mass error less than or equal to ± 5 parts per million (ppm). The comparison between the concentration of the quantified phenols was carried out by two-way ANOVA and Tukey's post hoc test ($p < 0.05$).

5. Conclusions

This is the first study in which both the morphological and physio-chemical responses to water stress of vanilla are addressed simultaneously. Additionally, it represents one of the few studies in which *V. planifolia* physiology is studied. Tolerance to PEG-induced water stress was achieved through root elongation to increase the water uptake and reduction in the number of leaves, as well as in the stomatal opening and the SLA, in order to reduce water loss. We observed an accumulation of carbohydrates, amino acids, purines, organic acids, and phenols; osmolyte and antioxidant activity was needed to achieve tolerance to water stress. The performance shown by the hybrid organisms suggests their ability to resist PEG-induced water stress under in vitro conditions. The *V. pompona* × *V. planifolia* hybrid exhibited the best PEG-induced water stress performance, possibly due to the maternal inherence of *V. pompona*. Vanilla hybrids are a potential alternative to the traditional cultivation of *V. planifolia*, especially given the difficulties that this crop will face due to the lack of water as a result of climate change; however, future studies should be carried out to determine the composition and content of the aromatic compounds in the fruits of these organisms. The approach of this study should be extended to other hybrids of *V. planifolia* and *V. pompona*, and also, to varieties of both species that are currently in cultivation, to determine which are likely to be best adapted for growth in water-limited conditions. These efforts will help protect *Vanilla* farmers from future extreme weather events and the *Vanilla* industry and consumers from future shortages and price fluctuations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24054690/s1.

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

Funding: This research was funded by CONACYT (Consejo Nacional de Ciencia y Tecnología), grant number 779887 to J.M.B-C., and to the CONACYT project number 316998: "Maintenance of the analytical platforms (UPLC-MS-QqQ and UPLC-MS-QTOF) used in metabolomics and chemical profiling studies, associated with frontier research of regional and national importance between INECOL and its strategic allies from program SUPPORT FOR FRONTIER SCIENCE: STRENGTHENING AND MAINTENANCE OF COMMON-USE RESEARCH INFRASTRUCTURES AND TECHNICAL TRAINING 2021", which allowed the maintenance and optimal operation of the chemical-analytical platforms used in this study.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: To the CITRO orchid garden for the donation of biological material to carry out this research.

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

References

1. Ranadive, A.S. Quality control of vanilla beans and extracts. In *Handbook of Vanilla Science and Technology*, 2nd ed.; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; pp. 239–260. [CrossRef]
2. Hernández-Hernández, J. Mexican Vanilla Production. In *Handbook of Vanilla Science and Technology*; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; Volume 2, pp. 3–26. [CrossRef]
3. Armenta-Montero, S.; Menchaca-García, R.; Pérez-Silva, A.; Velázquez-Rosas, N. Changes in the Potential Distribution of Vanilla planifolia Andrews under Different Climate Change Projections in Mexico. *Sustainability* **2022**, *14*, 2881. [CrossRef]
4. Menchaca-García, R.A. In vitro germination of Vanilla. In *Handbook of Vanilla Science and Technology*; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; Volume 2, pp. 181–190. [CrossRef]
5. Hernández-Hernández, J. Vanilla diseases. In *Handbook of Vanilla Science and Technology*; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; Volume 2, pp. 27–40. [CrossRef]
6. Trenberth, K.E.; Dai, A.; van der Schrier, G.; Jones, P.D.; Barichivich, J.; Briffa, K.R.; Sheffield, J. Global warming and changes in drought. *Nat. Clim. Chang.* **2014**, *4*, 17–22. Available online: http://www.nature.com/doifinder/10.1038/nclimate2067 (accessed on 20 November 2020). [CrossRef]
7. Hatfield, J.L.; Prueger, J.H. Temperature extremes: Effect on plant growth and development. *Weather Clim. Extr.* **2015**, *10*, 4–10. [CrossRef]
8. Frenkel, C.; Ranadive, A.S.; Vázquez, J.T.; Havkin-Frenkel, D. Curing on Vanilla. In *Handbook of Vanilla Science and Technology*; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; Volume 2, pp. 191–222. [CrossRef]
9. Zhang, S.; Yang, Y.; Li, J.; Qin, J.; Zhang, W.; Huang, W.; Hu, H. Physiological diversity of orchids. *Plant Divers.* **2018**, *40*, 196–208. [CrossRef] [PubMed]
10. Khan, N.; Bano, A.; Rahman, M.; Rathinasabapathi, B.; Babar, M. UPLC-HRMS-based untargeted metabolic profiling reveals changes in chickpea (*Cicer arietinum*) metabolome following long-term drought stress. *Plant Cell Environ.* **2018**, *42*, 115–132. [CrossRef]
11. Salehi-Lisar, S.; Bakhshayeshan-Agdam, H. Drought Stress in Plants: Causes, Consequences, and Tolerance. In *Drought Stress Tolerance in Plants*; Hossain, M., Wani, S., Bhattacharjee, S., Burrit, D., Tran, L., Eds.; Springer: Cham, Switzerland, 2016; pp. 1–16. [CrossRef]
12. Flexas, J.; Gallé, A.; Galmés, J.; Ribas-Carbo, M.; Medrano, H. The response of photosynthesis to soil water stress. In *Plant Responses to Drought Stress from Morphological to Molecular Features*; Aroca, R., Ed.; Springer: Granada, Spain, 2012. [CrossRef]
13. Taiz, L.; Zeiger, E. *Plant Physiology*, 5th ed.; Sianuer Associates: Sunderland, MA, USA, 2010; pp. 1–623.
14. Farooq, M.; Hussain, M.; Wahid, A.; Siddique, H. Drought stress in plants: An overview. In *Plant Responses to Drought Stress from Morphological to Molecular Features*; Aroca, R., Ed.; Springer: Granada, Spain, 2012; pp. 1–34. [CrossRef]
15. Zhang, C.; Chen, J.; Huang, W.; Song, X.; Niu, J. Transcriptomics and Metabolomics Reveal Purine and Phenylpropanoid Metabolism Response to Drought Stress in Dendrobium sinense, an Endemic Orchid Species in Hainan Island. *Front. Gen.* **2021**, *12*, 692702. [CrossRef]
16. Zhang, M.; Yu, Z.; Zeng, D.; Si, C.; Zhao, C.; Wang, H.; Li, C.; He, C.; Duan, J. Transcriptome and Metabolome Reveal Salt-Stress Responses of Leaf Tissues from Dendrobium officinale. *Biomolecules* **2021**, *11*, 736. [CrossRef] [PubMed]
17. Zivcak, M.; Brestic, M.; Sytar, O. Osmotic Adjustment and Plant Adaptation to Drought Stress. In *Drought Stress Tolerance in Plants, Vol. I. Physiology and Chemistry*; Hossain, M., Wani, S., Bhattacharjee, S., Burrit, D., Tran, L., Eds.; Springer: Cham, Switzerland, 2016; pp. 104–144.

18. Samarah, N. Understanding How Plants Respond to Drought Stress at the Molecular and Whole Plant Levels. In *Drought Stress Tolerance in Plants, Vol. II.*; Hossain, M., Wani, S., Bhattacharjee, S., Burrit, D., Tran, L., Eds.; Springer: Cham, Switzerland, 2016; pp. 1–38. [CrossRef]
19. Sami, F.; Yusuf, M.; Faizan, M.; Faraz, A.; Hayat, S. Role of sugars under abiotic stress. *Plant. Physiol.* **2016**, *109*, 54–61. [CrossRef] [PubMed]
20. Ozturk, M.; Turkyilmaz Unal, B.; García-Caparrós, P.; Khursheed, A.; Gul, A.; Hassanuzzaman, M. Osmoregulation and its actions during the drought stress in plants. *Physiol. Plant.* **2020**, *172*, 1321–1335. [CrossRef]
21. Yang, X.; Lu, M.; Wang, Y.; Wang, Y.; Liu, Z.; Chen, S. Response Mechanism of Plants to Drought Stress. *Horticulturae* **2021**, *7*, 50. [CrossRef]
22. Gundaraniya, S.; Ambalam, P.; Tomar, R. Metabolomic Profiling of Drought-Tolerant and Susceptible Peanut (*Arachis hypogaea* L.) Genotypes in Response to Drought Stress. *ACS Omega* **2020**, *5*, 31209–31219. [CrossRef] [PubMed]
23. Pectu, E.; Arsintescu, A.; Stanciu, D. The effect of drought stress on fatty acid composition in some Romanian sunflower hybrids. *Rom. Agric. Res.* **2001**, *15*, 39–43.
24. Ashraf, M.; Iqbal, M.; Rasheed, R.; Hussain, I.; Riaz, M.; Arif, M. Environmental Stress and Secondary Metabolites in Plants: An Overview. In *Plant Metabolites and Regulation Under Environmental Stress*; Ahmad, P., Ahanger, M., Singh, V., Tripathi, D., Alam, P., Alyemeni, M., Eds.; Elsevier: London, UK, 2018; pp. 153–167. ISBN 978-0-12-812689-9.
25. Quan, N.T.; Anh, L.H.; Khang, D.T.; Tuyen, P.T.; Toan, N.P.; Minh, T.N.; Bach, D.T.; Ha, P.T.T.; Elzaawely, A.A.; Khanh, T.D. Involvement of secondary metabolites in response to drought stress of rice (*Oryza sativa* L.). *Agriculture* **2016**, *6*, 23. [CrossRef]
26. Wang, X.; Oh, M.; Sakata, K.; Komatsu, S. Gel-free/label-free proteomic analysis of root tip of soybean over time under flooding and drought stresses. *J. Prot.* **2016**, *130*, 42–55. [CrossRef] [PubMed]
27. Rapala-Kozik, M.; Kowalska, E.; Ostrowska, K. Modulation of thiamine metabolism in *Zea mays* seedlings under conditions of abiotic stress. *J. Exp. Bot.* **2008**, *59*, 4133–4143. [CrossRef]
28. Li, J.; Demesyeux, L.; Brym, M.; Chambers, A.H. Development of species-specific molecular markers in Vanilla for seedling selection of hybrids. *Mol. Biol. Rep.* **2020**, *47*, 1905–1920. [CrossRef]
29. Divakaran, M.; Babu, K.; Peter, K. Cryopreservation of vanilla pollen and its utilization in inter specific hybridization between *V. planifolia* and *V. aphylla*. *Eur. J. Biomed. Pharm. Sci.* **2017**, *4*, 441–446.
30. van Esse, H.; Reuber, T.; van der Does, D. Genetic modification to improve disease resistance in crops. *New Phytol.* **2019**, *225*, 70–86. [CrossRef]
31. Flores-Jiménez, A.; Reyes-López, D.; Jiménez-García, D.; Romero-Arenas, O.; Rivera-Tapia, J.A.; Huerta-Lara, J.A.; Pérez-Silva, A. Diversidad de *Vanilla* spp. (Orchidaceae) y sus perfiles bioclimáticos en México. *Rev. Biol. Trop.* **2016**, *65*, 975–987. [CrossRef]
32. Barreda-Castillo, J.M.; Menchaca-García, R.A.; Pérez-Silva, A.; Sánchez-Coello, N.G.; Luna-Rodríguez, M. Influencia de la temperatura en la infectividad de *Fusarium oxysporum* f. sp. vanillae en Vanilla planifolia y en híbridos *V. planifolia* × *V. pompona*. *Biotecnia* **2022**, *25*, 177–183. [CrossRef]
33. Osmolovskaya, N.; Shumilina, J.; Kim, A.; Didio, A.; Grishina, T.; Bilova, T.; Keltsieva, O.; Zhukov, I.; Tarakhovskaya, E.; Frolov, A.; et al. Methodology of drought stress research experimental setup and physiological characterization. *Int. J. Mol. Sci.* **2018**, *19*, 4089. [CrossRef]
34. Van der Weele, C.; Spollen, W.; Sharp, R.; Baskin, T. Growth of Arabidopsis thaliana seedlings under water deficit studied by control of water potential in nutrient-agar media. *J. Exp. Bot.* **2000**, *51*, 1555–1562. [CrossRef] [PubMed]
35. Xian-Pan, B.; Feng-Feng, J.; Feng-Tang, W.; Cai-Zhi, C.; Fu-Sun, Y. Response to Drought Stress Simulated by PEG of Phalaenopsis pulcherrima. *J. Trop. Subtrop. Bot.* **2019**, *28*, 53–61.
36. Martínez-Santos, E.; Cruz-Cruz, C.A.; Spinoso-Castillo, J.L.; Bello-Bello, J.J. In vitro response of vanilla (*Vanilla planifolia* Jacks. ex Andrews) to PEG-induced osmotic stress. *Sci. Rep.* **2021**, *11*, 22611. [CrossRef] [PubMed]
37. Najla, S.; Sanoubar, R.; Murshed, R. Morphological and biochemical changes in two parsley varieties upon water stress. *Physiol. Mol. Biol. Plants* **2012**, *18*, 133–139. [CrossRef] [PubMed]
38. Niechayev, N.A.; Pereira, P.N.; Cushman, J.C. Understanding trait diversity associated with crassulacean acid metabolism (CAM). *Curr. Opin. Plant Biol.* **2019**, *49*, 74–85. [CrossRef] [PubMed]
39. Zotz, G.; Winkler, U. Aerial roots of epiphytic orchids: The velamen radicum and its role in water and nutrient uptake. *Oecologia.* **2013**, *171*, 733–741. [CrossRef] [PubMed]
40. De Micco, V.; Arone, G. Morpho-anatomical traits for plant adaptation to drought. In *Plant Responses to Drought Stress from Morphological to Molecular Features*; Aroca, R., Ed.; Springer: Granada, Spain, 2012; pp. 37–63. [CrossRef]
41. Karimi, S.; Rahemi, M.; Rostami, A.; Sedaghat, S. Drought effects on growth, water content and osmoprotectants in four olive cultivars with different drought tolerance. *Int. J. Fruit Sci.* **2018**, *18*, 254–267. [CrossRef]
42. Gantiva, E.; Díez, C.; Moreno, F. Efecto de la interacción luz-agua sobre la fotosíntesis de la *Vanilla planifolia* (Orchidaceae). *Rev. Biol. Trop.* **2020**, *68*, 1250–1261.
43. Carranza, C.; Lanchero, O.; Miranda, D.; Chaves, B. Growth analysis of 'Batavia' lettuce (*Lactuca sativa* L.) cultivated in a saline soil of the Bogota Plateau. *Agron. Colomb.* **2009**, *27*, 41–48.
44. Arias, R.C.; Reyes, J.J.; Ray, J.V.; Benítez, D.G.; Hernández, L.G.; Ledea, J.L. 2019. Morphometric indicators in new megatermal varieties of cenchrus purpureus tolerant to water stress. *Trop. Subtrop. Agroecos.* **2019**, *22*, 115–125.

45. Mata, A.; Jorge, T.; Pires, M.; Antonio, C. Drought stress tolerance in plants: Insights from metabolomics. In *Drought Stress Tolerance in Plants, Vol. II. Molecular and Genetic Perspectives*; Hossain, M., Wani, S., Bhattacharjee, S., Burrit, D., Tran, L., Eds.; Springer: Cham, Switzerland, 2016; pp. 187–217. [CrossRef]
46. Hietz, P.; Wagner, K.; Nunes Ramos, F.; Cabral, J.S.; Agudelo, C.; Benavides, A.M.; Cach-Pérez, M.J.; Cardelús, C.L.; Chilpa Galván, N.; Erickson Nascimento da Costa, L.; et al. Putting vascular epiphytes on the traits map. *J. Ecol.* **2022**, *110*, 340–358. [CrossRef]
47. Wright, I.J.; Reich, P.B.; Westoby, M.; Ackerly, D.D.; Baruch, Z.; Bongers, F.; Villar, R. The worldwide leaf economics spectrum. *Nature* **2004**, *428*, 821–827. [CrossRef] [PubMed]
48. Leyva, V.; Lopez, J.; Zevallos-Ventura, A.; Cabrera, R.; Cañari-Chumpitaz, C.; Toubiana, D.; Maruenda, H. NMR-based leaf metabolic profiling of *V. planifolia* and three endemic *Vanilla* species from the Peruvian Amazon. *Food Chem.* **2021**, *358*, 129365. [CrossRef]
49. Liu, J.; Kang, R.; Liu, Y.; Wu, K.; Yan, X.; Song, Y.; Pan, L.; Thang, Z. Differential Metabolite Accumulation in Different Tissues of Gleditsia sinensis under Water Stress and Rehydration Conditions. *Forests* **2020**, *11*, 542–556. [CrossRef]
50. Rai, V.K. Role of amino acids in plant responses to stresses. *Biol. Plant* **2002**, *45*, 481–487. [CrossRef]
51. Pirasteh-Anosheh, H.; Emam, Y.; Pessarakli, M. Changes in endogenous hormonal status in corn (*Zea mays*) hybrids under drought stress. *J. Plant Nutr.* **2013**, *36*, 1695–1707. [CrossRef]
52. Shanker, A.; Maheswari, M.; Yadav, S.; Desai, S.; Bhanu, D.; Attal, N.; Venkateswarlu, B. Drought stress responses in crops. *Func. Int. Gen.* **2014**, *14*, 11–22. [CrossRef] [PubMed]
53. Feizabadi, A.; Noormohammadi, G.; Fatehi, F. Changes in growth, physiology, and fatty acid profile of rapeseed cultivars treated with vermicompost under drought stress. *J. Soil Sci. Plant Nutr.* **2021**, *21*, 200–208. [CrossRef]
54. Auler, P.A.; Souza, G.M.; da Silva Engela, M.R.G.; do Amaral, M.N.; Rossatto, T.; da Silva, M.G.Z.; Braga, E.J.B. Stress memory of physiological, biochemical and metabolomic responses in two different rice genotypes under drought stress: The scale matters. *Plant Sci.* **2021**, *311*, 110994. [CrossRef] [PubMed]
55. Mibei, E.K.; Owino, W.O.; Ambuko, J.; Giovannoni, J.J.; Onyango, A.N. Metabolomic analyses to evaluate the effect of drought stress on selected African Eggplant accessions. *J. Sci. Food Agric.* **2018**, *98*, 205–216. [CrossRef]
56. Khazaei, Z.; Esmaielpour, B.; Estaji, A. Ameliorative effects of ascorbic acid on tolerance to drought stress on pepper (*Capsicum annuum* L.) plants. *Phys. Mol. Biol. Plants* **2020**, *26*, 649–1662. [CrossRef]
57. Hura, T.; Grzesiak, S.; Hura, K.; Thiemt, E.; Tokarz, K.; Wędzony, M. Physiological and biochemical tools useful in drought-tolerance detection in genotypes of winter triticale: Accumulation of ferulic acid correlates with drought tolerance. *Ann. Bot.* **2007**, *100*, 767–775. [CrossRef] [PubMed]
58. Piasecka, A.; Sawikowska, A.; Kuczyńska, A.; Ogrodowicz, P.; Mikołajczak, K.; Krystkowiak, K.; Kachlicki, P. Drought-related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their metabolomic quantitative trait loci. *Plant J.* **2017**, *89*, 898–913. [CrossRef]
59. Xuan, T.D.; Khang, D.T. Effects of exogenous application of protocatechuic acid and vanillic acid to chlorophylls, phenolics and antioxidant enzymes of rice (*Oryza sativa* L.) in submergence. *Molecules* **2018**, *23*, 620. [CrossRef] [PubMed]
60. Zhang, S.; Gai, Z.; Gui, T.; Chen, J.; Chen, Q.; Li, Y. Antioxidant Effects of Protocatechuic Acid and Protocatechuic Aldehyde: Old Wine in a New Bottle. *Evid.-Based Complement. Altern. Med.* **2021**, *2021*, 6139308. [CrossRef] [PubMed]
61. Araniti, F.; Lupini, A.; Mauceri, A.; Zumbo, A.; Sunseri, F.; Abenavoli, M.R. The allelochemical trans-cinnamic acid stimulates salicylic acid production and galactose pathway in maize leaves: A potential mechanism of stress tolerance. *Plant Physiol. Biochem.* **2018**, *128*, 32–40. [CrossRef]
62. Sarker, U.; Oba, S. Drought stress enhances nutritional and bioactive compounds, phenolic acids and antioxidant capacity of Amaranthus leafy vegetable. *BMC Plant Biol.* **2018**, *18*, 258. [CrossRef]
63. Juliano, F.F.; Alvarenga, J.F.R.; Lamuela-Raventos, R.M.; Massarioli, A.P.; Lima, L.M.; Santos, R.C.; Alencar, S.M. Polyphenol analysis using high-resolution mass spectrometry allows differentiation of drought tolerant peanut genotypes. *J. Sci. Food Agric.* **2020**, *100*, 721–731. [CrossRef] [PubMed]
64. Shawon, R.A.; Kang, B.S.; Lee, S.G.; Kim, S.K.; Lee, H.J.; Katrich, E.; Ku, Y.G. Influence of drought stress on bioactive compounds, antioxidant enzymes and glucosinolate contents of Chinese cabbage (*Brassica rapa*). *Food Chem.* **2020**, *308*, 125657. [CrossRef] [PubMed]
65. Król, A.; Amarowicz, R.; Weidner, S. Changes in the composition of phenolic compounds and antioxidant properties of grapevine roots and leaves (*Vitis vinifera* L.) under continuous of long-term drought stress. *Acta Physiol. Plant* **2014**, *36*, 1491–1499. [CrossRef]
66. Bhardwaj, R.D.; Kaur, L.; Srivastava, P. Comparative evaluation of different phenolic acids as priming agents for mitigating drought stress in wheat seedlings. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **2017**, *87*, 1133–1142. [CrossRef]
67. Yeloojeh, K.A.; Saeidi, G.; Sabzalian, M.R. Drought stress improves the composition of secondary metabolites in safflower flower at the expense of reduction in seed yield and oil content. *Ind. Crops Prod.* **2020**, *154*, 112496. [CrossRef]
68. Guo, X.; Xin, Z.; Yang, T.; Ma, X.; Zhang, Y.; Wang, Z.; Lin, T. Metabolomics response for drought stress tolerance in Chinese wheat genotypes (*Triticum aestivum*). *Plants* **2020**, *9*, 520. [CrossRef]
69. Jafari, M.; Shahsavar, A. The effect of foliar application of melatonin on changes in secondary metabolite contents in two Citrus species under drought stress conditions. *Front. Plant Sci.* **2021**, *2021*, 1509. [CrossRef]

70. Mechri, B.; Tekaya, M.; Hammami, M.; Chehab, H. Effects of drought stress on phenolic accumulation in greenhouse-grown olive trees (*Olea europaea*). *Biochem. Syst. Ecol.* **2020**, *92*, 104112. [CrossRef]
71. Hodaei, M.; Rahimmalek, M.; Arzani, A.; Talebi, M. The effect of water stress on phytochemical accumulation, bioactive compounds and expression of key genes involved in flavonoid biosynthesis in *Chrysanthemum morifolium* L. *Ind. Crops Prod.* **2018**, *120*, 295–304. [CrossRef]
72. Kumar, M.; Patel, M.K.; Kumar, N.; Bajpai, A.B.; Siddique, K.H. Metabolomics and molecular approaches reveal drought stress tolerance in plants. *Int. J. Mol. Sci.* **2021**, *22*, 9108. [CrossRef] [PubMed]
73. Hein, J.A.; Sherrard, M.E.; Manfredi, K.P.; Abebe, T. The fifth leaf and spike organs of barley (*Hordeum vulgare* L.) display different physiological and metabolic responses to drought stress. *BMC Plant Biol.* **2016**, *16*, 248. [CrossRef]
74. Koo, Y.M.; Heo, A.Y.; Choi, H.W. Salicylic acid as a safe plant protector and growth regulator. *Plant Pat. J.* **2020**, *36*, 1. [CrossRef] [PubMed]
75. Ullah, A.; Manghwar, H.; Shaban, M.; Khan, A.H.; Akbar, A.; Ali, U.; Fahad, S. Phytohormones enhanced drought tolerance in plants: A coping strategy. *Environ. Sci. Pollut. Res.* **2018**, *25*, 33103–33118. [CrossRef]
76. Belanger, F.; Havkin-Frenkel, D. Molecular analysis of a vanilla hybrid cultivated in Costa Rica. In *Handbook of Vanilla Science and Technology*; Havkin-Frenkel, D., Belanger, F., Eds.; Wiley: New York, NY, USA, 2019; Volume 2, pp. 391–402. [CrossRef]
77. Pontes, M. *Orquídeas, Notas Sobre Cultivo y más*; Asociación Mexicana de Orquideología: Mexico City, Mexico, 2019; pp. 1–238.
78. Park, H.S.; Lee, W.K.; Lee, S.C.; Lee, H.O.; Joh, H.J.; Park, J.Y.; Yang, T.J. Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. *Sci. Rep.* **2021**, *11*, 2506. [CrossRef] [PubMed]
79. Li, D.; Gan, G.; Li, W.; Li, W.; Jiang, Y.; Liang, X.; Wang, Y. Inheritance of Solanum chloroplast genomic DNA in interspecific hybrids. *Mitochondrial* **2021**, *6*, 351–357. [CrossRef] [PubMed]
80. Kumar, S.; Singh, A. Epigenetic regulation of abiotic stress tolerance in plants. *Adv. Plants Agric. Res.* **2016**, *5*, 179. [CrossRef]
81. Baulcombe, D.C.; Dean, C. Epigenetic regulation in plant responses to the environment. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a019471. [CrossRef]
82. Chambers, A.H. Vanilla (*Vanilla* spp.) breeding. In *Advances in Plant Breeding Strategies: Industrial and Food Crops*; Al-Khayri, J., Jain, S., Johnson, D., Eds.; Springer: Cincinnati, OH, USA, 2019; pp. 707–734. [CrossRef]
83. Grisoni, M.; Nany, F. The beautiful hills: Half a century of vanilla (*Vanilla planifolia* Jacks. ex Andrews) breeding in Madagascar. *Genet. Resour. Crop Evol.* **2021**, *68*, 1691–1708. [CrossRef]
84. Menchaca-García, R.A. Germinación in Vitro de Híbridos de *Vanilla planifolia* y *V. pompona*. Ph.D. Thesis, Universidad Veracruzana, Xalapa, Veracruz, Mexico, 2012.
85. Lee-Espinosa, H.E.; Murguía-González, B.J.; García-Rosas, A.L.; Cordova-Contreras, A.; Luguna-Cerda, J.O.; Mijangos-Cortés, L.F.; Barahona-Pérez, L.G.; Iglesias-Andreu, L.G.; Santana-Buzzy, N. In vitro clonal propagation of Vanilla (*Vanilla planifolia* 'Andrews'). *Hortscience* **2008**, *43*, 454–458. [CrossRef]
86. Divakaran, M.; Babu, K.; Peter, K. Protocols for biotechnological interventions in improvement of vanilla (*Vanilla planifolia* Andrews). In *Protocols for In Vitro Cultures and Secondary Metabolite Analysis of Aromatic and Medicinal Plants*; Jain, M., Ed.; Springer: New York, NY, USA, 2016; pp. 47–63. [CrossRef]
87. Michel, B.E. Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. *Plant Physiol.* **1983**, *72*, 66–70. [CrossRef] [PubMed]
88. Santos-Castellanos, M.; Segura Abril, M.; Ñústez López, C.E. Growth analysis and source-sink relationship of four potato cultivars (*Solanum tuberosum* L.) in the Zipaquira Town (Cundinamarca, Colombia). *Rev. Fac. Nac. Agron. Medellín* **2010**, *63*, 5253–5266.
89. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Met.* **2012**, *9*, 671–675. [CrossRef] [PubMed]
90. Calvete, E.O.; Kämpf, A.N.; Bergamaschi, H.; Daudt, R. Avaliação do crescimento de plantas de morangueiro, durante a aclimatização ex vitro. *Hort. Bras.* **2000**, *18*, 188–192. [CrossRef]
91. Fernández, G.; Johnston, M. *Fisiología Vegetal Experimental*; Instituto Panamericano para la Agricultura: San José, Costa Rica, 1986; pp. 249–256.
92. Monribot-Villanueva, J.L.; Altúzar-Molina, A.; Aluja, M.; Zamora-Briseño, J.A.; Elizalde-Contreras, J.M.; Bautista-Valle, M.; Arellano, J.; Sánchez-Martínez, D.; Rivera-Reséndiz, F.; Vázquez-Rosas-Landa, M.; et al. Integrating proteomics and metabolomics approaches to elucidate the ripening process in white Psidium guajava. *Food Chem.* **2022**, *367*, 130656. [CrossRef] [PubMed]
93. Juárez-Trujillo, N.; Monribot-Villanueva, J.L.; Alvarado-Olivarez, M.; Luna-Solano, G.; Guerrero-Analco, J.A.; Jiménez-Fernández, M. Phenolic profile and antioxidative properties of Pulp and sedes of *Randia monatha* Benth. *Ind. Crops Prod.* **2018**, *124*, 53–58. [CrossRef]
94. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 12 September 2022).
95. Rmisc: Ryan Miscellaneous. Available online: https://CRAN.R-project.org/package=Rmisc (accessed on 12 September 2022).
96. Agricolae: Statistical Procedures for Agricultural Research. Available online: https://CRAN.R-project.org/package=agricolae (accessed on 12 September 2022).
97. Hothorn, T.; Bretz, F.; Westfall, P. Simultaneous Inference in General Parametric Models. *Biomet. J.* **2008**, *50*, 346–363. [CrossRef]
98. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; pp. 1–182.

99. Xia, J.; Sinelnikov, I.; Han, B.; Wishart, D. MetaboAnalyst 3.0—Making metabolomics more meaningful. *Nucleic Acids Res.* **2015**, *43*, 251–257. [CrossRef]
100. KEGG PATHWAY Database. Wiring Diagrams of Molecular Interactions, Reactions, and Relations. Available online: https://www.genome.jp/kegg/pathway.html (accessed on 12 September 2022).
101. Adler, M.; Alon, U. Fold-change detection in biological systems. *Curr. Opt. Syst. Biol.* **2018**, *8*, 81–89. [CrossRef]
102. Gaffney, I.; Sallach, J.; Wilson, J.; Bergstrom, E.; Thomas-Oates, J. Metabolomic Approaches to Studying the Response to Drought Stress in Corn (*Zea mays*) Cobs. *Metabolites* **2021**, *11*, 438. [CrossRef] [PubMed]
103. Melandri, G.; AbdElgawad, H.; Riewe, D.; Hageman, J.A.; Asard, H.; Beemster, G.; Bouwmeester, H. Biomarkers for grain yield stability in rice under drought stress. *J. Exp. Bot.* **2020**, *71*, 669–683. [CrossRef] [PubMed]
104. Michaletti, A.; Naghavi, M.R.; Toorchi, M.; Zolla, L.; Rinalducci, S. Metabolomics and proteomics reveal drought-stress responses of leaf tissues from spring-wheat. *Sci. Rep.* **2018**, *8*, 5710. [CrossRef] [PubMed]
105. KEGG COMPOUND Database. Chemical Substances Integrated with Genomics. Available online: https://www.genome.jp/kegg/compound/ (accessed on 12 September 2022).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

International Journal of Molecular Sciences Editorial Office
E-mail: ijms@mdpi.com
www.mdpi.com/journal/ijms

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-0599-0