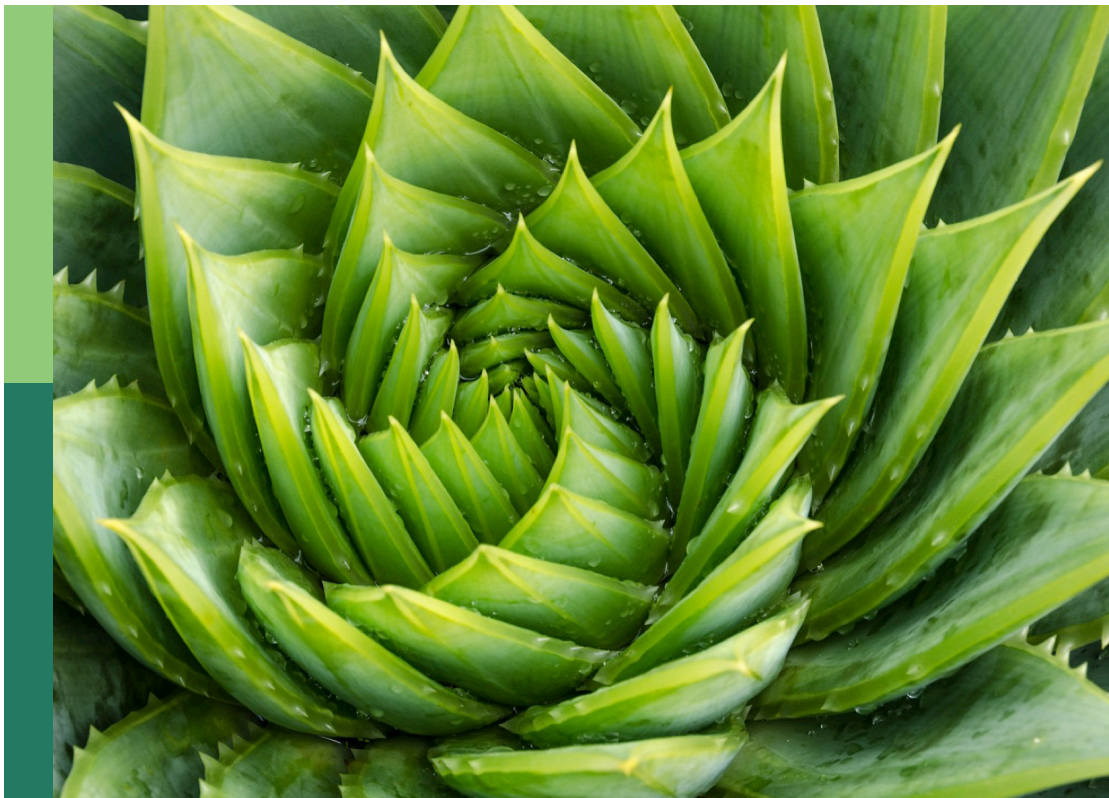
frontiers | Research Topics

Multi-omics and computational biology in horticultural plants: From genotype to phenotype, volume II

Edited by
Yunpeng Cao, Xiaoxu Li, Hui Song, Muhammad Abdullah and Muhammad Aamir Manzoor

Published in
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4470-9
DOI 10.3389/978-2-8325-4470-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Multi-omics and computational biology in horticultural plants: From genotype to phenotype, volume II

Topic editors

Yunpeng Cao — Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China

Xiaoxu Li — Chinese Academy of Agricultural Sciences (CAAS), China

Hui Song — Qingdao Agricultural University, China

Muhammad Abdullah — The University of Queensland, Australia

Muhammad Aamir Manzoor — Shanghai Jiao Tong University, China

Citation

Cao, Y., Li, X., Song, H., Abdullah, M., Manzoor, M. A., eds. (2024). *Multi-omics and computational biology in horticultural plants: From genotype to phenotype, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4470-9

Table of contents

06 Editorial: Multi-omics and computational biology in horticultural plants: from genotype to phenotype, volume II
Yunpeng Cao, Xiaoxu Li, Hui Song, Muhammad Abdullah and Muhammad Aamir Manzoor

11 Comparative time-course transcriptome analysis of two contrasting alfalfa (*Medicago sativa* L.) genotypes reveals tolerance mechanisms to salt stress
Dongmei Ma, Jinjun Cai, Qiaoli Ma, Wenjing Wang, Lijuan Zhao, Jiawen Li and Lina Su

26 Integrated analysis of miRNAome transcriptome and degradome reveals miRNA-target modules governing floral florescence development and senescence across early- and late-flowering genotypes in tree peony
Lili Guo, Yuying Li, Chenjie Zhang, Zhanying Wang, John E. Carlson, Weinlun Yin, Xiuxin Zhang and Xiaogai Hou

47 Genome-Wide Identification of *14-3-3* gene family reveals their diverse responses to abiotic stress by interacting with StABI5 in Potato (*Solanum tuberosum* L.)
Qianqian Wang, Chongchong Yan, Yuying Fu, Yu Wang, Pengfei Jiang, Yinyuan Ding and Huajun Liao

61 Phylogenetic studies of magnoliids: Advances and perspectives
Zhiguo Shen, Xin Ding, Jianming Cheng, Fangfang Wu, Hengfu Yin and Minyan Wang

69 Genome-wide analysis and identification of microRNAs in *Medicago truncatula* under aluminum stress
Zhongjie Lu, Zhengyu Yang, Zheng Tian, Qihui Gui, Rui Dong and Chao Chen

80 Effect of shading on physiological attributes and comparative transcriptome analysis of *Camellia sinensis* cultivar reveals tolerance mechanisms to low temperatures
Shah Zaman, Jiazhi Shen, Shuangshuang Wang, Dapeng Song, Hui Wang, Shibo Ding, Xu Pang, Mengqi Wang, Irfan Ali Sabir, Yu Wang and Zhaotang Ding

99 Salt-responsive bermudagrass microRNAs and insights into light reaction photosynthetic performance
Shugao Fan, Erick Amombo, Sheila Avoga, Yating Li and Yanling Yin

112 Resequencing of durian genomes reveals large genetic variations among different cultivars
Wanapinun Nawae, Chaiwat Naktang, Salisa Charoensri, Sonicha U-thoomporn, Nattapol Narong, Orwintinee Chusri, Sithichoke Tangphatsornruang and Wirulda Pootakham

127 Transcriptomic and metabolomic insights on the molecular mechanisms of flower buds in responses to cold stress in two *Camellia oleifera* cultivars
Ya-Jun Wang, Ling-Li Wu, Min-hong Sun, Ze Li, Xiao-Feng Tan and Jian-An Li

142 Temporal transcriptome analysis reveals several key pathways involve in cadmium stress response in *Nicotiana tabacum* L.
Chenyang Li, Yi Hong, Jinhao Sun, Guoping Wang, Huina Zhou, Liangtao Xu, Long Wang and Guoyun Xu

153 High-quality genome assembly and comparative genomic profiling of yellowhorn (*Xanthoceras sorbifolia*) revealed environmental adaptation footprints and seed oil contents variations
Juan Wang, Haifei Hu, Xizhen Liang, Muhammad Tahir ul Qamar, Yunxiang Zhang, Jianguo Zhao, Hongqian Ren, Xingrong Yan, Baopeng Ding and Jinping Guo

163 The apple 14-3-3 gene *MdGRF6* negatively regulates salt tolerance
Yuqing Zhu, Wei Kuang, Jun Leng, Xue Wang, Linlin Qiu, Xiangyue Kong, Yongzhang Wang and Qiang Zhao

175 The creation of autotetraploid provides insights into critical features of DNA methylome changes after genome doubling in water spinach (*Ipomoea aquatica* Forsk)
Yuanyuan Hao, Xiao Su, Wen Li, Lin Li, Yu Zhang, Muhammad Ali Mumtaz, Huangying Shu, Shanhan Cheng, Guopeng Zhu and Zhiwei Wang

190 The spatiotemporal profile of *Dendrobium huoshanense* and functional identification of *bHLH* genes under exogenous MeJA using comparative transcriptomics and genomics
Xiaomei He, Wenwu Zhang, Irfan Ali Sabir, Chunyan Jiao, Guohui Li, Yan Wang, Fucheng Zhu, Jun Dai, Longyun Liu, Cunwu Chen, Yingyu Zhang and Cheng Song

208 Low-coverage whole genome sequencing of eleven species/subspecies in *Dioscorea* sect. *Stenophora* (Dioscoreaceae): comparative plastome analyses, molecular markers development and phylogenetic inference
Ke Hu, Xiao-Qin Sun, Min Chen and Rui-Sen Lu

219 Research progress on iron absorption, transport, and molecular regulation strategy in plants
Xinyi Ning, Mengfei Lin, Guohua Huang, Jipeng Mao, Zhu Gao and Xiaoling Wang

230 QTL mapping and analysis for drought tolerance in rice by genome-wide association study
Yueming Yi, Muhammad A. Hassan, Xinxin Cheng, Yiru Li, Huan Liu, Wuyun Fang, Qian Zhu and Shimei Wang

241 Phylotranscriptomics supports numerous polyploidization events and phylogenetic relationships in *Nicotiana*
Shuaibin Wang, Junping Gao, Zhaowu Li, Kai Chen, Wenxuan Pu and Chen Feng

257 Potential regulatory genes of light induced anthocyanin accumulation in sweet cherry identified by combining transcriptome and metabolome analysis
Yao Zhang, Chaoqun Chen, Yiling Cui, Qinsong Du, Wenjing Tang, Wenlong Yang, Guanqiong Kou, Wanjia Tang, Hongxu Chen and Ronggao Gong

275 Physiological and lipidomic response of exogenous choline chloride alleviating salt stress injury in Kentucky bluegrass (*Poa pratensis*)
Zhi-Fang Zuo, Yan Li, Xin-Feng Mi, Yong-Long Li, Chen-Yuan Zhai, Guo-Feng Yang, Zeng-Yu Wang and Kun Zhang

287 Recent genome resequencing paraded COBRA-*Like* gene family roles in abiotic stress and wood formation in Poplar
Muhammad Sajjad, Adeel Ahmad, Muhammad Waheed Riaz, Quaid Hussain, Muhammad Yasir and Meng-Zhu Lu

307 Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties
Abhisek Chakraborty, Shruti Mahajan, Manohar S. Bisht and Vineet K. Sharma

324 Identification of candidate genes associated with less-photosensitive anthocyanin phenotype using an EMS mutant (*pind*) in eggplant (*Solanum melongena* L.)
Lei Luo, Jos Molthoff, Qiang Li, Ying Liu, Shuangxia Luo, Na Li, Shuxin Xuan, Yanhua Wang, Shuxing Shen, Arnaud G. Bovy, Jianjun Zhao and Xueping Chen

338 KASP-IEva: an intelligent typing evaluation model for KASP primers
Xiaojing Chen, Longyu Huang, Jingchao Fan, Shen Yan, Guomin Zhou and Jianhua Zhang

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 02 February 2024
DOI 10.3389/fpls.2024.1368909

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Nunzio D'Agostino,
University of Naples Federico II, Italy

*CORRESPONDENCE
Yunpeng Cao
xfcypeng@126.com
Xiaoxu Li
lixx@blsa.com.cn

RECEIVED 11 January 2024
ACCEPTED 24 January 2024
PUBLISHED 02 February 2024

CITATION
Cao Y, Li X, Song H, Abdullah M and Manzoor MA (2024) Editorial: Multi-omics and computational biology in horticultural plants: from genotype to phenotype, volume II.
Front. Plant Sci. 15:1368909.
doi: 10.3389/fpls.2024.1368909

COPYRIGHT
© 2024 Cao, Li, Song, Abdullah and Manzoor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Multi-omics and computational biology in horticultural plants: from genotype to phenotype, volume II

Yunpeng Cao[1]*, Xiaoxu Li[2,3]*, Hui Song[4], Muhammad Abdullah[5] and Muhammad Aamir Manzoor[6]

[1]School of Health and Nursing, Wuchang University of Technology, Wuhan, China, [2]Beijing Life Science Academy, Beijing, China, [3]Technology Center, China Tobacco Hunan Industrial Co., Ltd., Changsha, China, [4]Key Laboratory of National Forestry and Grassland Administration on Grassland Resources and Ecology in the Yellow River Delta, College of Grassland Science, Qingdao Agricultural University, Qingdao, China, [5]Queensland Alliance of Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia, [6]Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

KEYWORDS

multi-omics, computational biology, horticultural plants, genotype, phenotype

Editorial on the Research Topic
Multi-omics and computational biology in horticultural plants: from genotype to phenotype, volume II

Horticulture, an integral component of the broader field of agriculture, has played a pivotal role in the development of human civilization. The shift from nomadic lifestyles to settled farming communities was greatly enabled by advancements in horticultural practices. This domain encompasses the scientific, technological, and artistic aspects of growing, breeding, processing, and commercializing various plant types, such as ornamental species, flowers, fruits, vegetables, nuts, seeds, and herbs. In recent years, there has been a surge in the sequencing of numerous horticultural plant genomes (Marks et al., 2021). The field of multi-omics and computational biology, particularly as they relate to horticultural plants and transition from genotype to phenotype, have experienced significant growth and diversification (Cao et al., 2022a). This progress has been driven by the incorporation of high-throughput technologies and innovative computational methods, yielding profound insights into plant physiological adaptation and biological mechanisms. The current Research Topic is focused on merging advanced omics and computational biology techniques to associate genotypes with phenotypes and link genetic markers to traits in various horticultural crops (Figure 1).

This Research Topic showcases a collection of 24 scholarly articles. Within this assemblage, two pieces are comprehensive reviews, while the other 22 constitute original research papers. Among these, a pair delves into the genome sequencing of horticultural crops. Three articles focus on the study of fruit crops, another trio sheds light on vegetable research, and a single paper explores the realm of Chinese herbal medicine. Additionally,

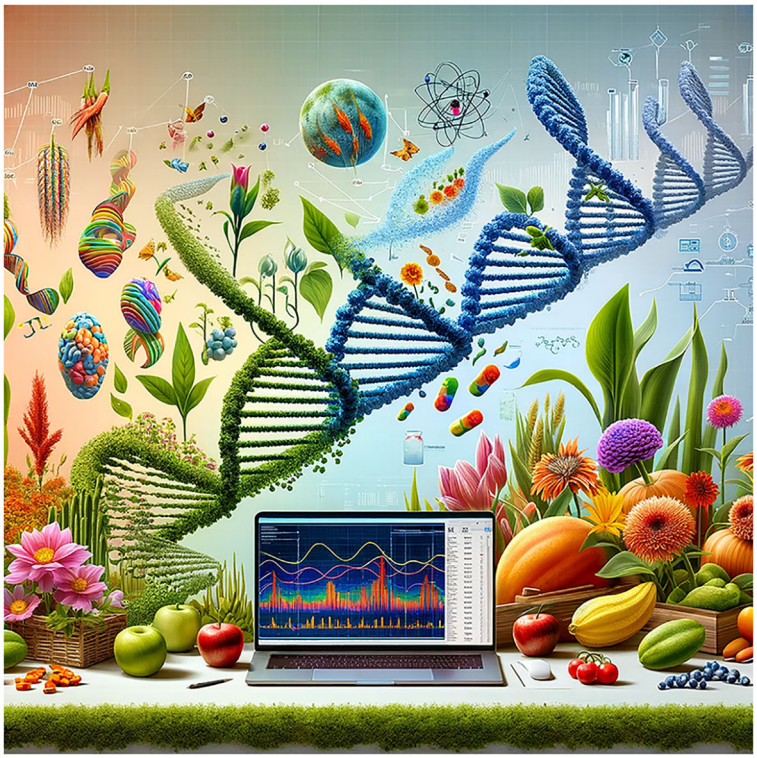


FIGURE 1
Multi-omics and computational biology in horticultural plants: From genotype to phenotype.

one article is dedicated to ornamental crops, three examine various industrial crops, and the final nine provide insightful research on model crop species.

Multi-omics data in horticulture plants

The integration of multi-omics data in horticulture significantly enhances our understanding of plant biology, leading to accelerated advancements in both research and breeding efforts within the field. This approach underscores the critical importance of combining various multi-omics datasets, such as genomics, transcriptomics, proteomics, and metabolomics, in horticultural plants. By combining various multi-omics datasets, offers a comprehensive view of complex biological processes, enhancing our understanding of plant growth, development, and environmental responses, and facilitates the identification of key genetic markers and traits for crop improvement, supporting innovative breeding strategies (Mondal et al., 2022). Moreover, it aids in more effective management of plant health and productivity, contributing to the overall advancement of horticultural science.

Holistic insights

Integrating multiple omics approaches yields a holistic and comprehensive perspective of biological systems, enabling researchers to concurrently examine various biological layers, thereby deepening the understanding of plant biology. By analyzing genetic variations and their impact on gene expression, protein synthesis, and metabolite levels helps shed light on how these molecular changes translate into observable plant traits. This integrative analysis is crucial for unraveling the complex biological processes in plants, leading to significant advancements in fields such as plant biology, agriculture, and horticulture.

Unraveling plant phenotypes

The integration of multi-omics data is crucial in revealing the intricate interactions and regulatory processes that shape plant phenotypes. For example, by correlating these omics data, researchers can discern the impacts of DNA sequence alterations on gene expression, protein synthesis, and metabolite concentrations (Cao et al., 2022a). Such insights shed light on how genetic variances contribute to the observable characteristics of plants. Yang et al. employed an Eigen Genome-Wide Association Study (EigenGWAS) methodology on a collection of 331 tomato varieties to uncover critical genetic determinants that affect metabolite variation throughout the tomato's domestication and enhancement processes, thereby emphasizing the significance of polygenic selection in the developing of tomatoes with superior fruit quality (Yang et al., 2022).

Elevating plant trait predictions

Enhancing the predictive capabilities of models that forecast plant traits can be significantly achieved through the integration of

multi-omics data, and these models can more accurately account for the interactions among different biological processes. This comprehensive approach leads to more precise and accurate predictions of plant traits, acknowledging the intricate and interconnected nature of biological systems. For example, Jiang et al. elucidated the mechanism of polysaccharide biosynthesis of *Bletilla striata* by combining genomic and transcriptomic data (Jiang et al., 2022a). Furthermore, multi-omics analysis of potential R2R3-MYB transcription factors within the Euphorbiaceae family has successfully identified a specific MYB gene significantly involved in the biosynthesis of seed oil (Cao et al., 2023).

Precision breeding advancements

Understanding the relationship between genotype and phenotype is fundamental for precise breeding design, as it lays the groundwork for predicting how genetic variations influence observable traits. The integration of multi-omics data significantly enhances precision breeding by identifying molecular markers linked to desirable traits in plants across various biological levels (Mahmood et al., 2022). This approach enables breeders to select traits with increased accuracy, thereby contributing to the development of superior plant varieties. By leveraging multi-omics data, breeders can gain a deeper understanding of the genetic basis of desired traits (Cao et al., 2022a). This holistic view allows for more informed selection processes and efficient breeding strategies, paving the way for the development of plants with optimized qualities.

Unveiling plant-environment dynamics

The integration of multi-omics data enables researchers to gain a more comprehensive understanding of plant-environment interactions, revealing the effects of various environmental factors on gene expression, protein production, and metabolic processes. Such insights are crucial for understanding how plants grow, develop, and respond to stress. By analyzing these data, scientists can identify how external conditions such as temperature, moisture, and soil composition influence plant biology at multiple levels, which leads to a better understanding of plant adaptation and resilience (Großkinsky et al., 2018).

Holistic insights for plant disease management

Integrating multi-omics data offers a holistic perspective on plant biology, significantly aiding in the diagnosis and management of plant diseases (Wang et al., 2021; Rathnasamy et al., 2023). This comprehensive approach can facilitate the identification of molecular markers associated with disease resistance, which are pivotal in guiding the development of disease-resistant plant varieties. At the same time, insights from multi-omics data also can inform and refine disease management strategies. By understanding the complex interactions at the genetic, transcriptomic, proteomic, and metabolomic levels, researchers can develop more effective methods to prevent and control plant diseases, leading to healthier crops and improved agricultural outcomes (Wang et al., 2021).

From genotype to phenotype in horticulture plants

Integrating multi-omics data within horticultural studies offers a robust method for deepening our understanding of plant biology. In this context, we covered the crucial importance of multi-omics data integration in horticulture with a specific focus on tracing the journey from genotype to phenotype. This approach significantly enriches our comprehension of plant biology, ranging from the genetic foundation to observable characteristics, thereby enhancing the efficiency and depth of research and breeding programs in horticulture.

Plant growth and development

To elucidate the genetic foundations of complex traits in plants, identifying candidate genes is a crucial step in modern genomic research. Previously, this task faced significant challenges due to the scarcity of high-quality genomic resources for various crops. However, the last twenty years have witnessed a revolutionary increase in the availability of detailed crop genomes and pan-genomes, which has greatly facilitated the process of connecting genotypes with phenotypes, a key aspect in understanding plant biology. Central to this understanding is the intricate network of gene interactions that orchestrate plant growth and development. Modern multi-omics techniques have become invaluable tools in deciphering this complexity.

Genomic studies, provide a comprehensive overview of a plant's genetic makeup, facilitating the identification of key genes, such as those involved in the auxin signaling pathway. They establish links between genetic variations, including single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs), and observable traits (Zhang et al., 2014; Cao et al., 2022a). Complementary to genomic data, transcriptomics offers insights into gene expression patterns across various development stages of development and in response to environmental factors. Technologies such as RNA sequencing (RNA-seq) are instrumental in monitoring the expression dynamics of genes, including those associated with the auxin pathway, under different conditions (Cao et al., 2022b). Furthermore, metabolomics adds another layer to our understanding by analyzing the metabolic changes that occur during plant development and focusing on identifying active biochemical pathways and tracking the fluctuation of metabolites, including hormones like auxin. Together, these multi-omics approaches provide a holistic view of plant biology, linking genotype to phenotype and paving the way for advanced breeding strategies and crop improvement.

Disease resistance pathways

In the realm of plant biology, a key area of study is the diverse and intricate pathways plants have evolved for disease resistance. This biological defense system is orchestrated through a complex network of genes, proteins, and metabolites, with each component playing a role in combating a wide array of pathogens. Multi-omics approaches are indispensable for dissecting and understanding these defense mechanisms at a molecular level.

Genomic studies in this area predominantly concentrate on *resistance* (*R*) genes and analyze genetic variations, such as SNPs and INDELs, to assess a plant's innate disease resistance potential, as detailed by (Jiang et al., 2022b), with this genotypic analysis being essential for understanding the baseline resistance of different plant species or varieties. Complementing genomic data, transcriptomics provides a dynamic view of how plants respond to pathogen attacks at the gene expression level (Li et al., 2022). Utilizing RNA-seq, researchers track the expression patterns of R genes under pathogenic stress, which reveals the activation of the disease resistance pathway during an infection, offering insights into the temporal response of plants to pathogenic threats (Cao et al., 2021). Proteomics further enhances our understanding by examining the post-transcriptional changes in R proteins during pathogen attacks, elucidating the functional roles these proteins play in the plant's defense mechanisms, and uncovering the biochemical processes involved in resisting pathogen invasion. Lastly, metabolomics completes the multi-omics picture by identifying the specific metabolites produced by plants in response to pathogen attacks, such as phytoalexins, thereby offering a comprehensive view of plants chemical defenses at the molecular level (Obata and Fernie, 2012). Collectively, these multi-omics approaches form a cohesive and detailed picture of plant disease resistance, providing vital insights for the development of more resilient crop varieties and innovative disease management strategies in agriculture.

Stress response pathways

Plants, inherently stationary organisms, are subjected to a myriad of environmental stresses including drought, salinity, and extreme temperatures. The study of their molecular responses to such stresses is pivotal in advancing our understanding of plant resilience, particularly in the context of crop improvement. Multi-omics approaches offer an integrated and comprehensive set of tools for dissecting these complex stress response pathways.

At the genomic level, studies focus on identifying key genes that are activated or suppressed in response to environmental stresses, encompassing the discovery of diverse transcription factors that significantly influence the regulation of other stress-responsive gene expressions(Zhang et al., 2022). For example, brassinazole resistant 1 (BZR1) regulates the target gene *ethylene response factors 49* (*ERF49*) to enhance plant sensitivity to heat stress (Chen et al., 2022). Transcriptomics takes this analysis a step further by tracking the changes in gene expression under various stress conditions (Nye et al., 2023), which revealed how plants modulate their gene expression in real-time to adapt to and survive under adverse conditions. Proteomics adds another layer to our understanding of plant stress responses by examining proteins involved in these processes, including their abundance, modifications, interactions, and elucidating their regulatory roles and functional significance in the adaptation to stress (Kirk et al., 2022). Lastly, metabolomics complements these approaches by elucidating the metabolic alterations in plants under stress, pinpointing critical metabolites produced or modified in the plant's defense, and serving as key indicators of stress response and adaptation (Obata and Fernie, 2012). Taken together, these multi-omics data offer a comprehensive perspective on plant responses to environmental stress, enhancing our understanding of stress tolerance mechanisms and facilitating the development of environmentally resilient crops.

Future prospects of integrating multi-omics in horticulture plants

The integration of multi-omics into plant phenotyping marks the onset of a transformative epoch in horticultural research and practices, heralding a wealth of groundbreaking opportunities. For example, precision horticulture, increasingly practical through intricate multi-omics data, allows for the customization of each cultivation phase, from sowing to harvest, based on the unique genetic and environmental profiles of individual plants, thereby enhancing productivity, sustainability, and yield quality. Another is predictive breeding, where the amalgamation of multi-omics accelerates and refines the breeding process through predictive modeling that incorporates a wide array of genetic and phenotypic data. This methodology also significantly enhances the management of disease and stress responses in plants by establishing advanced early warning systems and pioneering new management strategies. Sustainable crop management, enabled by multi-omics, incorporates the complex interactions among plants, soil, and climate, fostering ecologically sustainable practices, reducing environmental impact, and enhancing plant biodiversity exploration, potentially uncovering new breeding and conservation resources. Collectively, the advancement of multi-omics democratizes access to sophisticated plant phenotyping techniques, thereby expanding the horizon of horticultural research globally and marking a significant advancement in optimizing horticultural practices and methodologies.

Author contributions

YC: Validation, Writing – original draft, Writing – review & editing. XL: Writing – review & editing. HS: Writing – review & editing. MA: Writing – review & editing. MM: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (Grant No. 32201602), the Natural Science Foundation of Hubei Province (Grant No. 2023AFB1036), and the Funding of Beijing Life Science Academy (Grant No. 2023200CC0270).

Acknowledgments

Editors thank all the contributing authors in this Research Topic.

Conflict of interest

Author XL was employed by the company China Tobacco Hunan Industrial Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cao, Y., Fan, T., Wang, L., Zhang, L., and Li, Y. (2023). Large-scale analysis of putative Euphorbiaceae R2R3-MYB transcription factors identifies a MYB involved in seed oil biosynthesis. *BMC Plant Biol.* 23, 145. doi: 10.1186/s12870-023-04163-5

Cao, Y., Li, Y., Wang, L., Zhang, L., and Jiang, L. (2022b). Evolution and function of ubiquitin-specific proteases (UBPs): Insight into seed development roles in plants. *Int. J. Biol. Macromolecules* 221, 796–805. doi: 10.1016/j.ijbiomac.2022.08.163

Cao, Y., Mo, W., Li, Y., Li, W., Dong, X., Liu, M., et al. (2021). Deciphering the roles of leucine-rich repeat receptor-like protein kinases (LRR-RLKs) in response to Fusarium wilt in the Vernicia fordii (Tung tree). *Phytochemistry* 185, 112686. doi: 10.1016/j.phytochem.2021.112686

Cao, K., Peng, Z., Zhao, X., Li, Y., Liu, K., Arus, P., et al. (2022a). Chromosome-level genome assemblies of four wild peach species provide insights into genome evolution and genetic basis of stress resistance. *BMC Biol.* 20, 139. doi: 10.1186/s12915-022-01342-y

Chen, X., Xue, H., Zhu, L., Wang, H., Long, H., Zhao, J., et al. (2022). ERF49 mediates brassinosteroid regulation of heat stress tolerance in Arabidopsis thaliana. *BMC Biol.* 20, 1–17. doi: 10.1186/s12915-022-01455-4

Großkinsky, D. K., Syaifullah, S. J., and Roitsch, T. (2018). Integration of multi-omics techniques and physiological phenotyping within a holistic phenomics approach to study senescence in model and crop plants. *J. Exp. Bot.* 69, 825–844. doi: 10.1093/jxb/erx333

Jiang, L., Lin, M., Wang, H., Song, H., Zhang, L., Huang, Q., et al. (2022a). Haplotype-resolved genome assembly of Bletilla striata (Thunb.) Reichb.f. to elucidate medicinal value. *Plant J.* 111, 1340–1353. doi: 10.1111/tpj.15892

Jiang, X., Su, Y., and Wang, M. (2022b). Mapping of a novel clubroot disease resistance locus in Brassica napus and related functional identification. *Front. Plant Sci.* 13, 1014376. doi: 10.3389/fpls.2022.1014376

Kirk, P., Amsbury, S., German, L., Gaudioso-Pedraza, R., and Benitez-Alfonso, Y. (2022). A comparative meta-proteomic pipeline for the identification of plasmodesmata proteins and regulatory conditions in diverse plant species. *BMC Biol.* 20, 1–21. doi: 10.1186/s12915-022-01331-1

Li, Y., Jiang, L., Mo, W., Wang, L., Zhang, L., and Cao, Y. (2022). AHLs' life in plants: Especially their potential roles in responding to Fusarium wilt and repressing the seed oil accumulation. *Int. J. Biol. Macromolecules* 208, 509–519. doi: 10.1016/j.ijbiomac.2022.03.130

Mahmood, U., Li, X., Fan, Y., Chang, W., Niu, Y., Li, J., et al. (2022). Multi-omics revolution to promote plant breeding efficiency. *Front. Plant Sci.* 13, 1062952. doi: 10.3389/fpls.2022.1062952

Marks, R. A., Hotaling, S., Frandsen, P. B., and Vanburen, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nat. Plants* 7, 1571–1578. doi: 10.1038/s41477-021-01031-8

Mondal, S., Song, H., Zhang, L., and Cao, Y. (2022). Multi-omics and computational biology in horticultural plants: From genotype to phenotype. *Front. Plant Sci.* 13, 1073266. doi: 10.3389/fpls.2022.1073266

Nye, D. G., Irigoyen, M. L., Perez-Fons, L., Bohorquez-Chaux, A., Hur, M., Medina-Yerena, D., et al. (2023). Integrative transcriptomics reveals association of abscisic acid and lignin pathways with cassava whitefly resistance. *BMC Plant Biol.* 23, 657. doi: 10.1186/s12870-023-04607-y

Obata, T., and Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cell. Mol. Life Sci.* 69, 3225–3243. doi: 10.1007/s00018-012-1091-5

Rathnasamy, S. A., Gothandapani, S., Chellamuthu, S., Chakraborty, A., Gurusamy, D., and Roy, A. (2023). "Omics technologies unravelling the plant–pathogen interaction and stress response," in *Genomics of plant–pathogen interaction and the stress response* (CRC Press), 74–110.

Wang, S., Liu, L., Mi, X., Zhao, S., An, Y., Xia, X., et al. (2021). Multi-omics analysis to visualize the dynamic roles of defense genes in the response of tea plants to gray blight. *Plant J.* 106, 862–875. doi: 10.1111/tpj.15203

Yang, J., Liang, B., Zhang, Y., Liu, Y., Wang, S., Yang, Q., et al. (2022). Genome-wide association study of eigenvectors provides genetic insights into selective breeding for tomato metabolites. *BMC Biol.* 20, 120. doi: 10.1186/s12915-022-01327-x

Zhang, S., Chen, W., Xin, L., Gao, Z., Hou, Y., Yu, X., et al. (2014). Genomic variants of genes associated with three horticultural traits in apple revealed by genome re-sequencing. *Horticulture Res.* 1, 14045. doi: 10.1038/hortres.2014.45

Zhang, H., Zhu, J., Gong, Z., and Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nat. Rev. Genet.* 23, 104–119. doi: 10.1038/s41576-021-00413-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fpls.2022.1070846

Check for updates

OPEN ACCESS

EDITED BY
Hui Song,
Qingdao Agricultural University, China

REVIEWED BY
Qiang Zhou,
Lanzhou University, China
Peng Jiao,
Jilin Agricultural University, China

*CORRESPONDENCE
Dongmei Ma
madm88@126.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 15 October 2022
ACCEPTED 16 November 2022
PUBLISHED 08 December 2022

CITATION
Ma D, Cai J, Ma Q, Wang W, Zhao L, Li J and Su L (2022) Comparative time-course transcriptome analysis of two contrasting alfalfa (*Medicago sativa* L.) genotypes reveals tolerance mechanisms to salt stress. *Front. Plant Sci.* 13:1070846. doi: 10.3389/fpls.2022.1070846

COPYRIGHT
© 2022 Ma, Cai, Ma, Wang, Zhao, Li and Su. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparative time-course transcriptome analysis of two contrasting alfalfa (*Medicago sativa* L.) genotypes reveals tolerance mechanisms to salt stress

Dongmei Ma[1,2,3]*, Jinjun Cai[4], Qiaoli Ma[5], Wenjing Wang[1,2,3], Lijuan Zhao[1,2,3], Jiawen Li[1,2,3] and Lina Su[1,2,3]

[1]Breeding Base for State Key Laboratory of Land Degradation and Ecological Restoration in Northwest China, Ningxia University, Yinchuan, China, [2]Ministry of Education Key Laboratory for Restoration and Reconstruction of Degraded Ecosystems in Northwest China, Ningxia University, Yinchuan, China, [3]Key Laboratory of Modern Molecular Breeding for Dominant and Special Crops in Ningxia, Ningxia University, Yinchuan, China, [4]Institute of Agricultural Resources and Environment, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan, China, [5]Agricultural College, Ningxia University, Yinchuan, China

Salt stress is a major abiotic stress affecting plant growth and crop yield. For the successful cultivation of alfalfa (*Medicago sativa* L.), a key legume forage, in saline-affected areas, it's essential to explore genetic modifications to improve salt-tolerance.Transcriptome assay of two comparative alfalfa genotypes, Adina and Zhaodong, following a 4 h and 8 h's 300 mM NaCl treatment was conducted in this study in order to investigate the molecular mechanism in alfalfa under salt stress conditions. Results showed that we obtained 875,023,571 transcripts and 662,765,594 unigenes were abtained from the sequenced libraries, and 520,091 assembled unigenes were annotated in at least one database. Among them, we identified 1,636 differentially expression genes (DEGs) in Adina, of which 1,426 were up-regulated and 210 down-regulated, and 1,295 DEGs in Zhaodong, of which 565 were up-regulated and 730 down-regulated. GO annotations and KEGG pathway enrichments of the DEGs based on RNA-seq data indicated that DEGs were involved in (1) ion and membrane homeostasis, including ABC transporter, *CLC*, *NCX*, and *NHX*; (2) Ca^{2+} sensing and transduction, including BK channel, EF-hand domain, and calmodulin binding protein; (3) phytohormone signaling and regulation, including *TPR, FBP, LRR*, and *PP2C*; (4) transcription factors, including zinc finger proteins, *YABBY*, and SBP-box; (5) antioxidation process, including *GST*, PYROX, and *ALDH*; (6) post-translational modification, including *UCH*, ubiquitin

family, *GT, MT* and *SOT*. The functional roles of DEGs could explain the variations in salt tolerance performance observed between the two alfalfa genotypes Adina and Zhaodong. Our study widens the understanding of the sophisticated molecular response and tolerance mechanism to salt stress, providing novel insights on candidate genes and pathways for genetic modification involved in salt stress adaptation in alfalfa.

KEYWORDS

alfalfa, transcriptome, salt tolerance, post-translational modification, signal sensing and transduction

Introduction

Salt stress is a key abiotic stress threatening germination, growth, development, and seed set formation of plants (Munns and Tester, 2008). It's crucial to improve plants' endurance to salinity in order to support plant growth and crop yield. Plants undergo a number of morphological, cellular, physiological, biochemical, and molecular changes when coping with salt stress, and they have evolved sophisticated tolerance arrangements to defend themselves from salt stress conditions (Parihar et al., 2015; Haak et al., 2017). These adjustments are regulated by multiple genes and molecular mechanisms. Therefore, it is imperative to identify and understand the genes and molecular mechanisms involved in salt stress responses in order to improve salt resistance capacity in plants, which would subsequently make great contributions for the genetic modification of crops and food production in saline affected areas.

During the past decades, key molecular mechanisms mediated plant salt response and tolerance have been identified. We know that the hyperosmotic component and the ionic Na^+ component caused by salinity are perceived by the salt stress sensing and signaling components (e.g. reactive oxygen species (ROS), abscisic acid (ABA), Ca^{2+} response, and kinases (calcium-dependent protein kinases (CDPKs), and calcineurin B-like proteins with CBL-interacting protein kinases (CIPKs), which would transduce the signal of hyperosmotic pressure (Deinlein et al., 2014; Dinneny, 2015; You and Chan, 2015). The salt stress sensory and signaling mechanisms to achieve tolerance can be integrated by the linkage of transcription factors in plants. These transcription factors include calmodulin-binding transcription activators (CAMTAs), GT element-binding like proteins (GTLs), basic leucine zipper (bZIP), WRKY, MYB, APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF), NAC, and basic helix–loop–helix (bHLH) families (Deinlein et al., 2014; Rasul et al., 2017; Li et al., 2018; Sun et al., 2018; Wang et al., 2018; Büyük et al., 2019), which can regulate the transcript expression patterns of genes involved in salt response and tolerance in plants. In addition, salt overly sensitive (SOS) in maintaining low Na^+ in the cytoplasm, HKT in Na^+ partitioning, and some other additional regulators have been reported to be intimately associated with salt tolerance in plants (Deinlein et al., 2014). Yet, salt tolerance mechanisms remained elusive due to the sophisticated gene regulation network and the genetic variability among plant species.

In recent years, next-generation sequencing (NGS) technologies have been broadly adopted for their high precision and throughput to explore the molecular mechanism of salt response and tolerance in plants. It is worth clarifying that transcriptome studies have used RNA sequencing approaches not only in model plants and important crops like rice (*Oryza sativa*) (Shankar et al., 2016), maize (*Zea mays*) (Du et al., 2017), wild cotton species (*Gossypium klotzschianum*) (Wei et al., 2017), and *Brachypodium distachyon* (Priest et al., 2014), but also in plant species for which no reference genome is available, for instance ryegrass (*Lolium perenne*) (Hu et al., 2016), *Carex rigescens* (Li et al., 2017; Zhang et al., 2020a), and radish (*Raphanus sativus* L.) (Sun et al., 2016). A number of transcripts that play critical roles in salt stress response and tolerance regulation have been identified, largely contributing to the understanding of plant salt tolerance mechanisms. The usage of the RNA sequencing approaches makes the dynamic detection of transcripts in plants under salt stress accessible. Multi-transcriptomes at different time points can be used as comparisons to identify more genes and mechanisms intimately linked to plant salt stress tolerance, as studies have done in *Carex rigescens* (Zhang et al., 2020a), potato (*Solanum tuberosum* L.) (Li et al., 2020), sorghum (*Sorghum bicolor* L.) (Cui et al., 2018), and soybean (*Glycine max* L.) (Liu et al., 2019a). Yet the regulatory network involved in dynamic salt response is still not fully understood. Transcriptome analyses of different plant species should be considered the preferred way to reveal salt response and tolerance systems in plants.

Alfalfa (*Medicago sativa* L.), an important perennial forage legume, is widely cultivated across the world and it's valued for its high protein content, nutritional value, stress resistance

capabilities, and biomass production (Luo et al., 2014). The mainly areas where alfalfa is cultivated in the northwestern, northeastern, and northern coastal regions of China, are unfortunately affected by increased soil salinization, which hinders the growth and production of alfalfa plants (Ashrafi et al., 2014). As such it is vital to understand salt tolerance mechanisms in alfalfa. Research on the molecular mechanisms involved in salt stress adaptation and tolerance in alfalfa has been progressing: (1) large-scale dissection of transcripts, proteins, metabolites, and genetic loci have been identified *via* transcriptome (Long et al., 2015), proteome (Long et al., 2016; Long et al., 2018) and genome-wide association analyses (Yu et al., 2016); (2) a number of genes and regulators involved in salt tolerance in alfalfa, such as *MsGRP* (Long et al., 2013), *MsZEP* (Zhang et al., 2016), miR393 (Long et al., 2017) and miR156 (Arshad et al., 2017), have been identified by molecular function assays; (3) it has been observed that the overexpression of salt tolerance genes such as *AgcodA* (Li et al., 2014) and *AtNDPK2* (Wang et al., 2014), and the co-overexpression of *ZxNHX* and *ZxVP1-1* (Kang et al., 2016) improve salt stress adaptation in alfalfa. Nevertheless, research on the integrated signaling pathways and regulatory networks involved in the response to salt stress in alfalfa is still limited, and understanding of the mechanisms and the genetic modifications of alfalfa necessary to salt stress tolerance is still a challenge.

In this study, we hypothesize that the salt response and tolerance mechanisms can be revealed by time-course transcriptional comparison between two *Medicago sativa* genotypes (Adina and Zhaodong), which were previously screened as contrast salt responsive genotypes by physiological tests. We perform gene expression analysis and identify differentially expressed genes in salt stress conditions, subsequently we carry out GO annotation and KEGG pathway enrichment analysis. Our aim is to achieve a wide view of the transcriptional expression of genes in salt stress response and to reveal salt tolerance genes and mechanisms in alfalfa. In addition, we explore the genetic variation at transcriptional level of two salt tolerance comparable alfalfa genotypes.

Materials and methods

Plant materials and treatment conditions

Two alfalfa (*Medicago sativa*) cultivars (cvs. Adina, Zhaodong) were used in this study. Our growth and physiological assessments noted that Adina is salt-tolerant, while Zhaodong is salt-sensitive (Figure S1). Alfalfa seeds were immersed in 75% ethanol for 30 seconds and then sterilized by 0.1% $HgCl_2$ solution for 8 minutes. After rinsing with double distilled water 5 times, the seeds were let germinate on filter paper on sterilized petri dishes. Alfalfa seed germination was achieved in a growth chamber under conditions of 16-h-light (1200 μmol m^{-2} s^{-1})/8-h-dark, 25°C temperature, and 85% humidity. Five alfalfa seedlings in similar growth conditions were transferred in a tube (13 cm×5 cm) containing half-strength Hoagland's nutrient solution for hydroponic culture. The nutrient solution was replaced every two days to keep it fresh.

Thirty-day-old seedlings were subjected to salt treatment by transferring them to 300 mM NaCl nutrient solution. The root was chosen for transcriptome analysis as it's the first organ to experience salt stress (Postnikova et al., 2013). The roots of Adina and Zhaodong were harvested at 4 h and 8 h after salt treatment and rapidly rinsed with double distilled water. The roots of Adina and Zhaodong without NaCl addition were sampled as control. Three individual plants were collected as one sample, and three biological replicates were conducted in this study. The harvested samples were flash-frozen with liquid nitrogen and preserved at -80°C. N0, N4, and N8, were used to indicate the Adina root samples respectively at 0, 4 h, and 8 h after salt treatments, and M0, M4, and M8 for Zhaodong samples at 0, 4 h, and 8 h after salt treatments.

Total RNA isolation and sequencing

Total RNA in the alfalfa root samples was isolated using Trizol Reagent (Invitrogen, Beijing, China) following the manufacturer's instruction. The RNA quality was examined on 1% RNase free agarose gel, the purity was confirmed by a Nanodrop-2000 spectrophotometer (Thermo Scientific, DE, USA), the integrity was determined by an Agilent Bioanalyzer 2100 System (Agilent Technologies, CA, USA), and the precise concentration was determined by a Qubit 2.0 Fluorimeter (Life Technologies, CA, USA). The sequencing library was constructed using a NEBNext® Ultra™ RNA Library Prep Kit (NEB, USA) according to the manufacturer's recommendations and index codes. After verifying the quality of the cDNA library obtained, the samples were sequenced with the Illumina HiSeq/MiSeq sequencer.

De novo assembly and functional annotation

Raw reads, which have been submitted to the SRA public database (accession: PRJNA821982), were cleaned removing adapter sequences, low-quality sequences, reads with ambiguous bases 'N', and reads with more than 10% bases with Q< 20. The high-quality clean data was used for the *de novo* assembly by Trinity assembler (http://trinityrnaseq.sourceforge.net/) (Grabherr et al., 2011). The functional annotations of the assembled unigenes were compared to the databases: non-redundant protein sequences (Nr); non-redundant nucleotide sequences (Nt); protein family (Pfam);

eukaryotic Ortholog Groups/Clusters of Orthologous Groups of proteins (KOG/COG) and a manually annotated and reviewed protein sequence database (Swiss-Prot). The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analysis annotation was obtained with the KEGG Automatic Annotation Server (KAAS). The Gene Ontology (GO) annotation of unigenes was accessed by Blast2GO program (v2.5) according to the Nr and Pfam annotation results.

Differential unigene expression analysis

The clean reads were mapped back to the assembled sequences with Bowtie2 in RSEM software (Li and Dewey, 2011). The read counts were converted to FPKM (Fragments Per Kilobase Million) to represent the gene expression level. The differential expression of the unigenes, which was generated with the Benjamini and Hochberg's approach, was determined by DESeq R package with adjusted *P* value< 0.05 and the criteria of |log2FC| > 1. The GO enrichment analysis was conducted with the GO-seq and topGO R packages. The KEGG analysis was carried out with the KOBAS software.

Quantitative real-time PCR validation

Seven genes were randomly selected for quantitative RT-PCR analysis in order to validate the RNA-seq data. RNA samples of alfalfa roots were reverse transcribed into cDNA using a PrimeScript^{TM} RT reagent kit with gDNA Eraser (Takara, China). The amplification of the candidate genes was conducted on an ABI qPCR System (Applied biosystems, United States) with SYBR Premix Ex Taq (TliRNaseH Plus) (Takara, China). The PCR cycling reaction procedure was 95°C for 30 s, followed by 42 cycles of 95°C for 30 s, and 55°C for 15 s. After each PCR run a dissociation curve was generated to confirm the specificity of the product. The primers for qRT-PCR validation were designed using Beacon Designer software (version 7.8) and are listed in Table S1. The *β-ACTIN* gene of *M. sativa* was used as the internal control for gene expression normalization. The expression level of the candidate genes was calculated with the $2^{-\Delta\Delta Ct}$ method.

Results

Transcriptome sequencing and assembly

We sequenced eighteen cDNA libraries of two alfalfa cultivars. We acquired raw reads for each library by PCR and subsequent Illumina sequencing, we then obtained the clean reads removing ambiguous nucleotides, adapter sequences, and low-quality sequences. The raw reads, clean reads, and the quality assessments of the eighteen libraries are shown in Table S2. From the 18 sequenced libraries we obtained in total 875,023,571 transcripts with a mean length of 623 bp, an N50 of 946 bp, and an N90 of 256 bp, and 662,765,594 unigenes with a mean length of 1,014 bp, an N50 of 1,353 bp, and an N90 of 491 bp (Table 1). The transcripts and unigenes were grouped by length as shown in Table 1.

Unigenes annotations and pathway assignments

In order to explore their integrated function, we blasted the unigenes against seven databases (NR, NT, KO, SwissProt, PFAM, GO, KOG) for similarity searches. The numbers of the annotated unigenes and the percentage of the total are listed in Table 2. 520,091 assembled unigenes, 79.58% of the total, were annotated in at least one database.

Among them, 356,281 (54.52%) assembled unigenes showed identity in the GO database. The matched unigenes could be sorted into three categories (biological processes, cellular components, and molecular functions), and further classified in 56 subcategories (Figure S2A). The three main subcategories within the category of biological processes were cellular process, metabolic process, and single-organism process. Cell and cell part were the two dominant subcategories in cellular components, while binding and catalytic activity were the major subcategories within molecular functions.

When analyzed in the KEGG database, 159,597 unigenes were found to belong to five classifications and to be involved in 19 pathways: 6.6% (10,576) in cellular processes, 3.5% (5,524) in environmental information processing, 26.5% (42,290) in genetic information processing, 45.1% (71,943) in metabolism, and 2.8%

TABLE 1 A Summary of assembled transcripts and unigenes for all samples.

	Transcripts	Unigenes
Min Length	201	201
Mean Length	623	1,014
Median Length	348	718
Max Length	20,230	20,230
Number of N50	946	1,353
Number of N90	256	491
Total Nucleotides	875,023,571	662,765,594
Number of length (bp) <301	45,822	581,806
Number of length (bp): 301-500	144,183	341,594
Number of length (bp): 501-1000	241,797	258,808
Number of length (bp): 1001-2000	152,105	152,272
Number of length (bp): >2000	69,556	69,556

N50, the size of the transcripts or the unigenes above which the assembly contains at least 50% of the total length of all the contigs; N90, the size of the transcripts or the unigenes above which the assembly contains at least 50% of the total length of all the contigs.

10.3389/fpls.2022.1070846

TABLE 2 Blast analysis of assembled unigenes.

	Number of Genes	Percentage (%)
Annotated in NR	356,535	54.56
Annotated in NT	278,059	42.55
Annotated in KO	159,597	24.42
Annotated in SwissProt	311,429	47.65
Annotated in PFAM	347,366	53.15
Annotated in GO	356,281	54.52
Annotated in KOG	164,855	25.22
Annotated in all Databases	48,407	7.4
Annotated in at least one Database	520,091	79.58
Total Unigenes Annotated	653,466	100

(4,462) in cellular processes, while translation (18,770), carbohydrate metabolism (14,340), folding, sorting and degradation (13,063), overview (11,543), and transport and catabolism (10,576) were the dominant pathways (Figure S2B).

Expression and cluster analysis of the genes

In order to analyze the expression level of genes, we converted the read counts of alfalfa transcriptome to FPKM, as its density distribution can reflect the integrated gene expression pattern of each sequenced sample. Some of the gene expression levels in the salt-tolerant samples (N0, N4, N8) were higher than in the salt-sensitive samples (M0, M4, M8) (Figure 1). To estimate the gene expression at different times in Adina and Zhaodong, we hierarchically clustered the genes in the six samples with their log2(ratios) values. Results showed that most of the genes were differentially expressed after salt treatment in Adina and Zhaodong and at 4 h and 8 h after salt treatment (Figure 1C).

Differential expression analysis of the genes

The numbers, changes, overlaps, and the ratios to total, of the DEGs identified in the four comparison groups (N4, N8, M4, M8) were exhibited in the Venn gram (Figure 2). After salt treatment we found 583 DEGs in N4 and 1,219 in N8, with an overlap of 166 DEGs. We identified 1,153 and 204 DEGs, with an overlap of 62, respectively in M4 and M8. In addition, we found a total of 1,636 DEGs, 1,426 up-regulated and 210 down-regulated, in the salt-tolerant cultivar Adina, and a total of 1,295 DEGs, 565 up-regulated and 730 down-regulated, in the salt-tolerant cultivar Zhaodong. The fact that more DEGs are upregulated in Adina compared to Zhaodong could indicate the induction of the transcriptional regulation involved in molecular salt-tolerance in Adina.

GO annotations of the DEGs

The GO annotation classification of the DEGs in Adina and Zhaodong were displayed in Figure 3. The DEGs were categorized into three metabolic categories: biological processes, cellular components, and molecular functions. In biological processes, the subcategory of oxidation-reduction was highly enriched in both Adina and Zhaodong, while the metabolic process was significantly enriched only in Adina. In terms of cellular components, the transcription factor complex was enriched only in Adina, while plastid and chloroplast were only significantly enriched in Zhaodong. In the category of molecular functions, the most abundant subcategories in both Adina and Zhaodong were catalytic activity and oxidoreductase activity in. Annotations such as transcription factor activity, cofactor binding, coenzyme binding, nucleic acid binding transcription, were significantly enriched in Adina, whereas processes such as enzyme inhibitor activity and peptidase inhibitor activity were enriched in Zhaodong.

KEGG pathway enrichment of the DEGs

We conducted KEGG pathway analysis of the DEGs in Adina and Zhaodong (Figure 4). We found that DEGs in Adina were mainly enriched in photosynthesis-antenna proteins, protein processing in endoplasmic reticulum, glyoxylate and dicarboxylate metabolism, and pentose and glucuronate interconversions pathways. In Zhaodong DEGs were enriched in plant hormone signal transduction, photosynthesis, starch and sucrose metabolism, pentose phosphate pathway, glycolysis/gluconeogenesis, and glutathione metabolism pathways. The KEGG enrichments in Adina and Zhaodong subjected to salt stress indicate that the biological processes of photosynthesis, energy generation and usage, signal sensing and transduction, and protein synthesis are modified in salt stress conditions.

RNA-seq data validation with qRT-PCR

We randomly chose seven DEGs to validate the RNA-seq data by qRT-PCR assay. The correlation analysis between RNA-seq data and qRT-PCR are presented in Table 3. The correlation coefficients are greater than 0.6, showing a high correlation between the data of RNA-seq and qRT-PCR. The results of the correlation analysis indicate that RNA-seq data is reliable and can be used for further functional analysis.

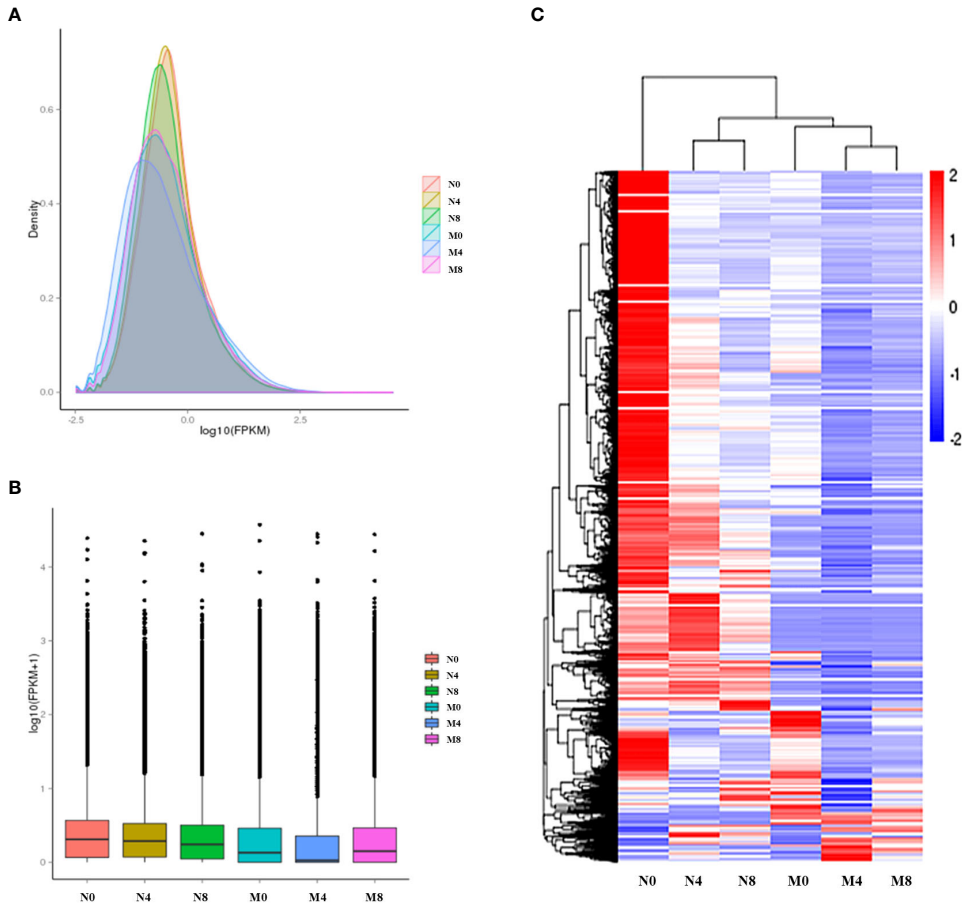


FIGURE 1
Expression analysis of the genes. N0, N4, and N8, represent the alfalfa root samples of Zhaodong respectively after 0 h, 4 h, and 8 h of salt treatment. M0, M4, and M8, represent the alfalfa root samples of Adina respectively after 0 h, 4 h, and 8 h of salt treatment. **(A)** FPKM density distribution. **(B)** FPKM distribution. **(C)** Hierarchical cluster of genes in the six alfalfa samples. The colors represent log2(FPKM value) in each grid, the color key is on the right.

Discussions

In this study we conducted a dynamic transcriptome survey of two alfalfa genotypes to investigate the transcriptional changes happening in alfalfa in response to salt stress conditions. From the sequenced samples we obtained more than 87 million transcripts and 66 unigenes. In addition, we performed differential gene expression analysis, GO annotation and KEGG enrichment analysis of the unigenes and DEGs, generating clues about the salt response mechanism of alfalfa. We revealed a dynamic and wide perspective of salt adaptation and tolerance changes at the transcriptional level through the comparison of DEGs expression in two alfalfa genotypes with significant salt tolerance genetic variations. We will discuss specific DEGs involved in ion and membrane homeostasis, Ca^{2+} sensing and transduction, phytohormone signaling and regulation, transcription factors, antioxidation process, post-translational modification (Figures 5, 6, Table S3) contributing to salt stress adaptation in alfalfa in the next sections of this paper.

DEGs involved in ion and membrane homeostasis

In plants the negative electrical potential caused by salt stress can occur through Na^+ influx in ion channels or other membrane transport proteins, which can enhance the passive diffusion of Na^+ across the plasma membrane. As such, sodium and chloride channels and transporters are essential for salt stress defense processes (Amtmann and Beilby, 2010). In our study, we identified as DEGs after salt treatment 12 ATP-binding cassette (ABC) transporter genes, which are involved in the ATP-powered export or import of various substrates

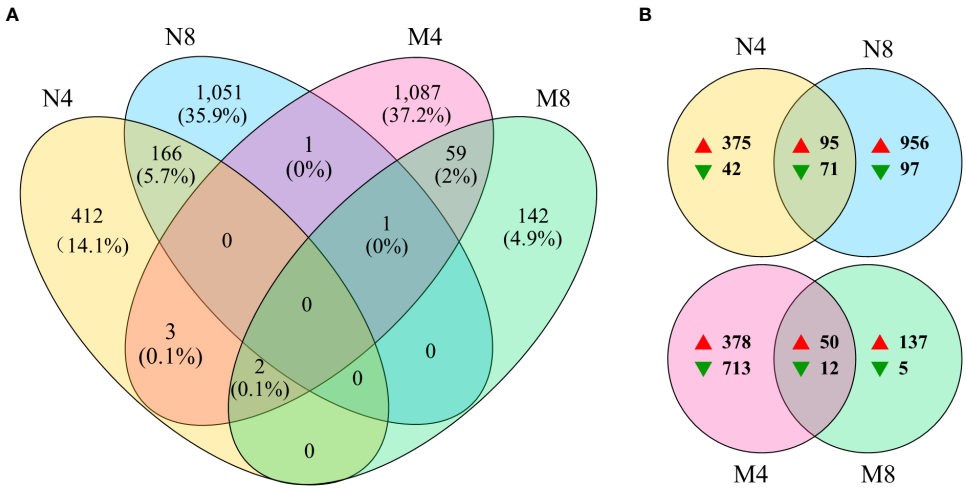


FIGURE 2
Venn diagram of the four comparisons. **(A)** The overlapped numbers of genes in the four groups. **(B)** The number of up-regulated (on the right of the red arrows) and down-regulated (on the right of the green arrows) genes in the four groups. N4 and N8 represent the comparison groups of the genes identified at 4 h and 8 h after treatment to 0 h in Adina. M4 and M8 represent the comparison groups of the genes identified at 4 h and 8 h after treatment to 0 h in Zhaodong.

across biological membranes and act as important membrane transporters in plants (Kang et al., 2010; Moon and Jung, 2014). 11 of the 12 ABC genes were up-regulated, consistently with a prior transcriptome study (Wu et al., 2017). The enhanced expression of the ABC transporters could contribute to salt resistance reducing the passage and transport of sodium (Kang et al., 2010; Kim et al., 2010; Wu et al., 2017). We also identified as DEGs three ABC-2 type transporters and two ABC transporter C-terminal domain, suggesting ABC-transporters could have a role in enabling alfalfa to cope with salt stress through ion signaling and transportation. In addition, we noticed two voltage gated chloride channel (*CLC*) genes, three sodium/calcium exchanger (*NCX*) genes, and a Na^+/H^+ antiporter family (*NHX*) gene were upregulated in Adina, while one *CLC* was downregulated in Zhaodong. *CLC*, *NCX*, and *NHX* are known to play important roles in intracellular Na^+ and Cl^- transportation and homeostasis under salt stress conditions and are considered important gene targets for salt tolerance improvement in plants (Wu et al., 2004; Wang et al., 2012; Wei et al., 2013; Liu et al., 2020). The up-regulation changes observed in Adina indicate the role of the genes in salt tolerance in alfalfa.

In salt stress conditions plants uptake macronutrients such as K^+, NO_3^- and Ca^{2+} to mitigate Na^+ and Cl^- assimilation (Amtmann and Beilby, 2010). It has been proposed that *CLC* could regulate the transport, interaction and homeostasis of NO_3^- in cotton (*Gossypium hirsutum* L.), in addition to its Cl^- transportation role (Liu et al., 2020). We therefore suggest that the ion and membrane homeostasis in alfalfa in salt stress conditions could be maintained by the identified DEGs, especially in Adina, which shows upregulated sodium and chloride channels and transporters genes.

DEGs involved in Ca^{2+} sensing and transduction

Ca^{2+} is a ubiquitous second messenger that plays a critical role in plant responses to salinity (Zeng et al., 2017). We identified three upregulated calcium-activated potassium channel genes (BK channel), two in Adina and one Zhaodong. It has been proposed that BK channels are activated by Ca^{2+} waves and can be coupled by the Ca^{2+} spark (Jonathan Ledoux et al., 2006), therefore the up-regulation of BK channels in Adina could help to hyperpolarizing of the cell membrane for Ca^{2+} sensing and control (Jonathan Ledoux et al., 2006; Hadi and Karimi, 2012). In our study we identified twelve EF-hand domain genes as DEGs, with five upregulated in Adina and none upregulated in Zhaodong. The structural properties of the EF-hand domain allow a rapid response to cytosolic Ca^{2+} concentration and on/off Ca^{2+} binding, and the majority of the key Ca^{2+} sensors and transducers contain the EF-hand domain (Zeng et al., 2017). In rice overexpression of *OsCCD1*, a small calcium-binding protein with one EF-hand motif, can significantly enhance tolerance to salt stress (Jing et al., 2016), while a different study showed how Ca^{2+} binding to the EF-hand domain of *ZmNSA1* can improve salt-tolerance in maize by increasing SOS1 Na^+/H^+ antiporter-mediated root Na^+ efflux (Cao et al., 2020). Consistently with these findings, we also found that the SOS1 Na^+/H^+ antiporter family (*NHX*) was up-

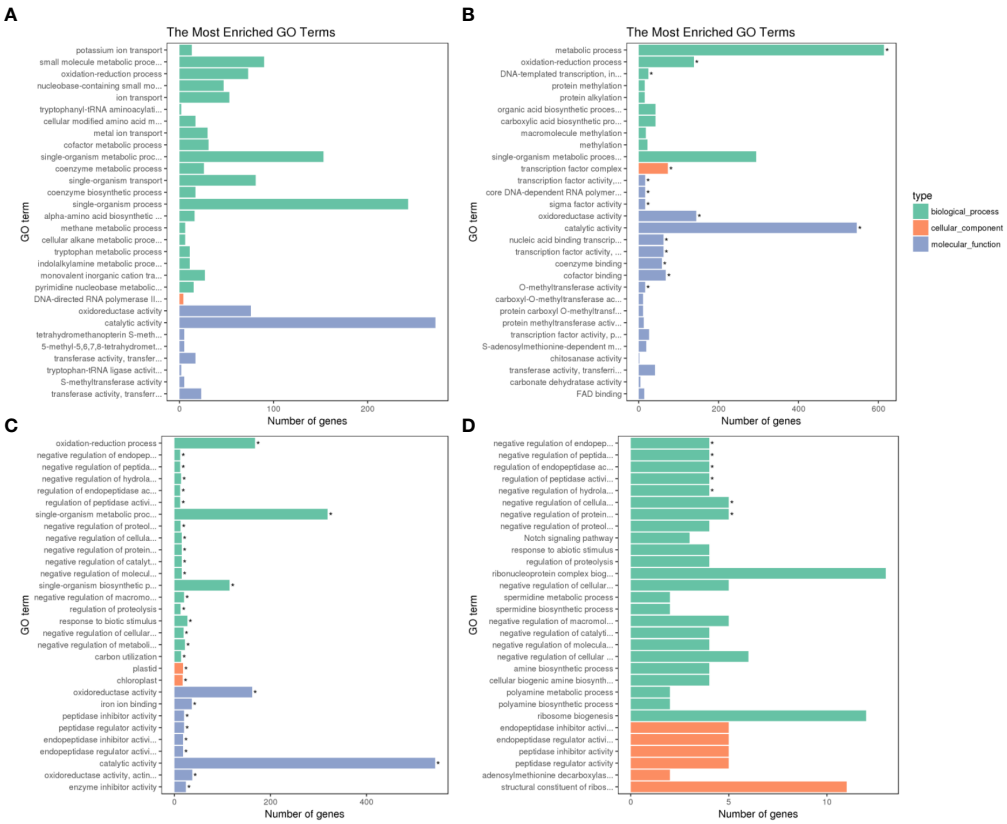


FIGURE 3
GO categorization of the DEGs. DEGs were separated into metabolic categories based on biological processes, cellular components and molecular functions. GO categorization of the DEGs in N4 **(A)**, N8 **(B)**, M4 **(C)**, and M8 **(D)**. N4 and N8 represent the comparison groups of the genes identified at 4 h and 8 h after treatment in Adina. M4 and M8 represent the comparison groups of the genes identified at 4 h and 8 h after treatment to 0 h in Zhaodong. *Asterisk means DEGs were significantly enriched in this GO term.

regulated and modified at the EF-hand domain in Adina, supporting the evidence that Ca^{2+} sensing and signaling play an important role in salt-tolerance in Adina (Cao et al., 2020). At the same time the downregulation of the calmodulin binding protein observed in Zhaodong, together with the down-regulation of the five genes encoding EF-hand domain in Zhaodong, suggests that the Ca^{2+} sensing and transduction processes were not enhanced as in Adina.

DEGs involved in phytohormone signaling and regulation

A number of phytohormonal signals work together in cellular physiology to enable the plant to adapt to the salinity of the environment (Liu et al., 2012). In our study, we identified 21 tetratricopeptide repeats (*TPR*) genes known to be involved in plant salt stress response *via* hormone signaling, 13 genes were up-regulated in Adina and 2 were up-regulated in Zhaodong (Sharma and Pandey, 2016). *TPR*s expression have been proved to be induced under salt stress in rice (Wang et al., 2016b) and Hulless barley (*Hordeum vulgare* L.) (Lai et al., 2020). TPRs regulate salt stress signals *via* gibberellin and cytokinin cross-talk and ABA biosynthesis and transport. In accord with these results, in our study we observed that the induction of *TPR* in Adina in response to salt stress could upregulate plant hormones, such as ABA, GA, cytokinin, and ethylene signaling pathways (Sharma and Pandey, 2016; Wang et al., 2016b; Lai et al., 2020). In addition, we identified ten F-box protein (*FBP*) genes as DEGs after salt stress treatment, with five upregulated in Adina, and none upregulated in Zhaodong. *FBPs*, for example *ARKP1* in *Arabidopsis* (Li et al., 2016), or *TaFBA1* in wheat (*Triticum aestivum*) (Kong et al., 2016), have been documented to be positive regulators for ABA signaling, and plants overexpressing *FBP* present a higher stress-tolerance and a higher antioxidant ability (Abd-Hamid et al., 2020). All of the twelve leucine rich repeat (*LRR*) genes identified in Adina were up-regulated after salt stress treatment, whereas seven of the eight identified in Zhaodong were downregulated. Leucine-rich repeats receptor-like kinases (LRR-RLKs) are known to play

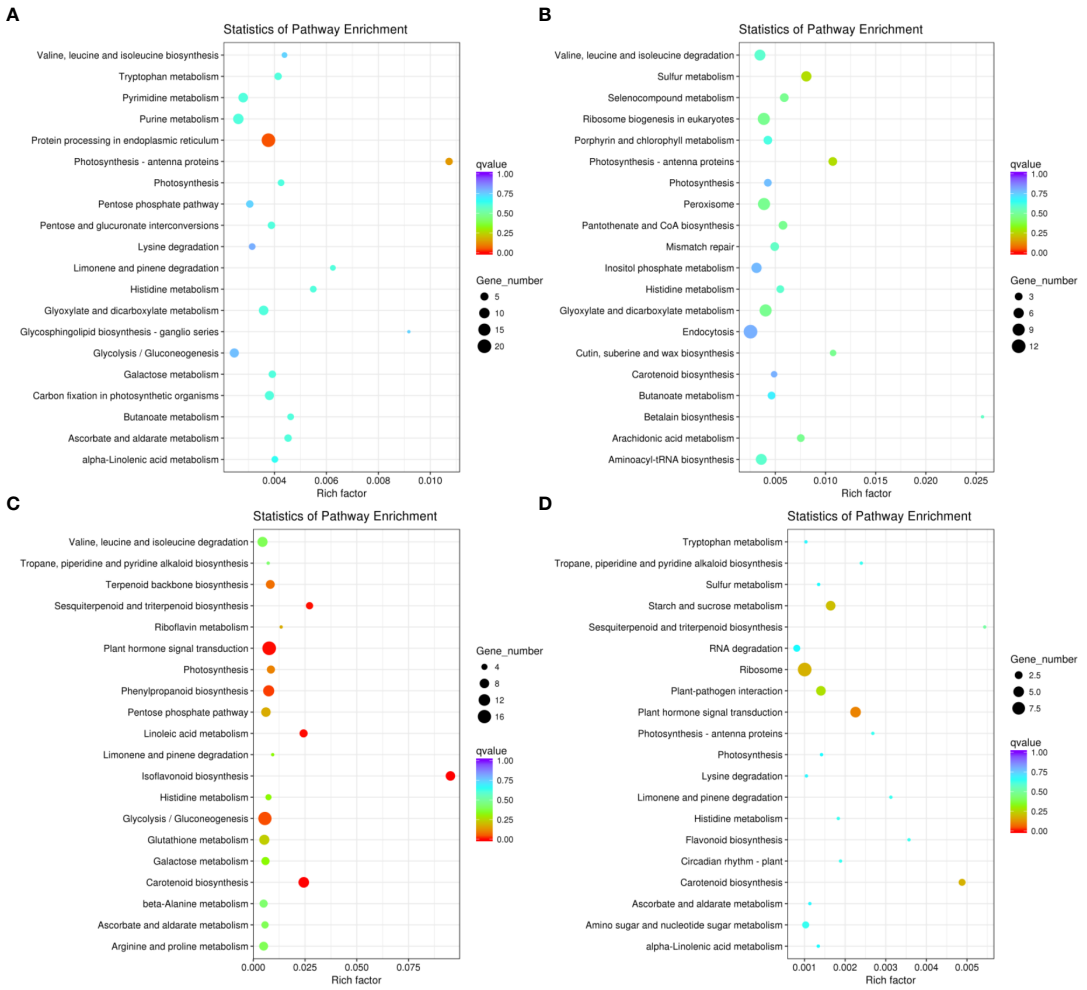


FIGURE 4
KEGG functional classification of the DEGs. The sizes of the spots represent the number of genes, and the colors represent the q-value. KEGG functional classification of the DEGs in N4 **(A)**, N8 **(B)**, M4 **(C)**, and M8 **(D)**. N4 and N8 represent the comparison groups of the genes identified at 4 h and 8 h after treatment to 0 h in Adina. M4 and M8 represent the comparison groups of the genes identified at 4 h and 8 h after treatments to 0 h in Zhaodong.

TABLE 3 Correlation coefficient between RNA-seq data and qRT-PCR.

Gene ID	Correlation coefficient
Cluster-171808.189526	0.8
Cluster-171808.190107	0.731
Cluster-171808.201554	0.819
Cluster-171808.198991	0.886
Cluster-171808.81350	0.933
Cluster-171808.262653	0.616
Cluster-171808.116680	0.782

important roles in plant defense to salt stress through the regulation of salt stress-related and ABA-depending signaling pathways (Wang et al., 2016a). The increased expression of *FBP* and *LRR* could indicate that the ABA-depending signaling pathway is enhanced to improve salt stress sensing and signaling in Adina. We identified eight protein phosphatase 2C (PP2C) genes, which are regulators of ABA early signal transduction, as DEGs with seven increased and one decreased in Zhaodong. Over-expression of *AtPP2CG1* can improve salt tolerance, enabling *PP2C* to move from a repressor-associated suppression status to an activator-mediated transcription status in response to salt stress (Liu et al., 2012; Nguyen et al., 2019), suggesting PP2C has a regulation role in salt stress adaptation in Zhaodong.



FIGURE 5
Heat map of candidate DEGs involved in ion and membrane homeostasis **(A)**, antioxidation process **(B)**, and phytohormone signaling and regulation **(C)**. The four columns, from left to right, represent N4, N8, M4, and M8. Color key is shown on the right. Red represents up-regulation, green down-regulation, and white no significant change.

DEGs encoding transcription factors

Transcription factors (TFs) play important roles in plant adaptation to salt stress regulating downstream genes responsible for salt-tolerance (Sun et al., 2010). In this study, we found that 25 zinc finger protein genes, part of one of the largest transcription factor families in plants, were differentially expressed after salt stress treatments in two alfalfa cultivars. Previous studies have reported these genes (for instance ZFP179 encoding C2H2-type zinc finger protein in rice (Sun et al., 2010), *BrRZFP1* encoding C3HC4-type in *Brassica rapa* (Jung et al., 2013),*GhZFP1* encoding CCCH-type in *Gossypium hirsutum* (Guo et al., 2009), *OsZFP6* encoding CCHC-type in rice (Guan et al., 2014)) play vital roles in the resistance to salt stress by regulating salt-tolerance genes, suggesting the 19 gene encoding these four types of zinc finger protein in alfalfa, with 17 of them up-regulated and two down-regulated, have an important role in salt adaption role. In addition, we observed three genes encoding the B-box zinc finger proteins, and two genes encoding GATA zinc finger proteins were upregulated in Adina. B-box zinc finger

10.3389/fpls.2022.1070846

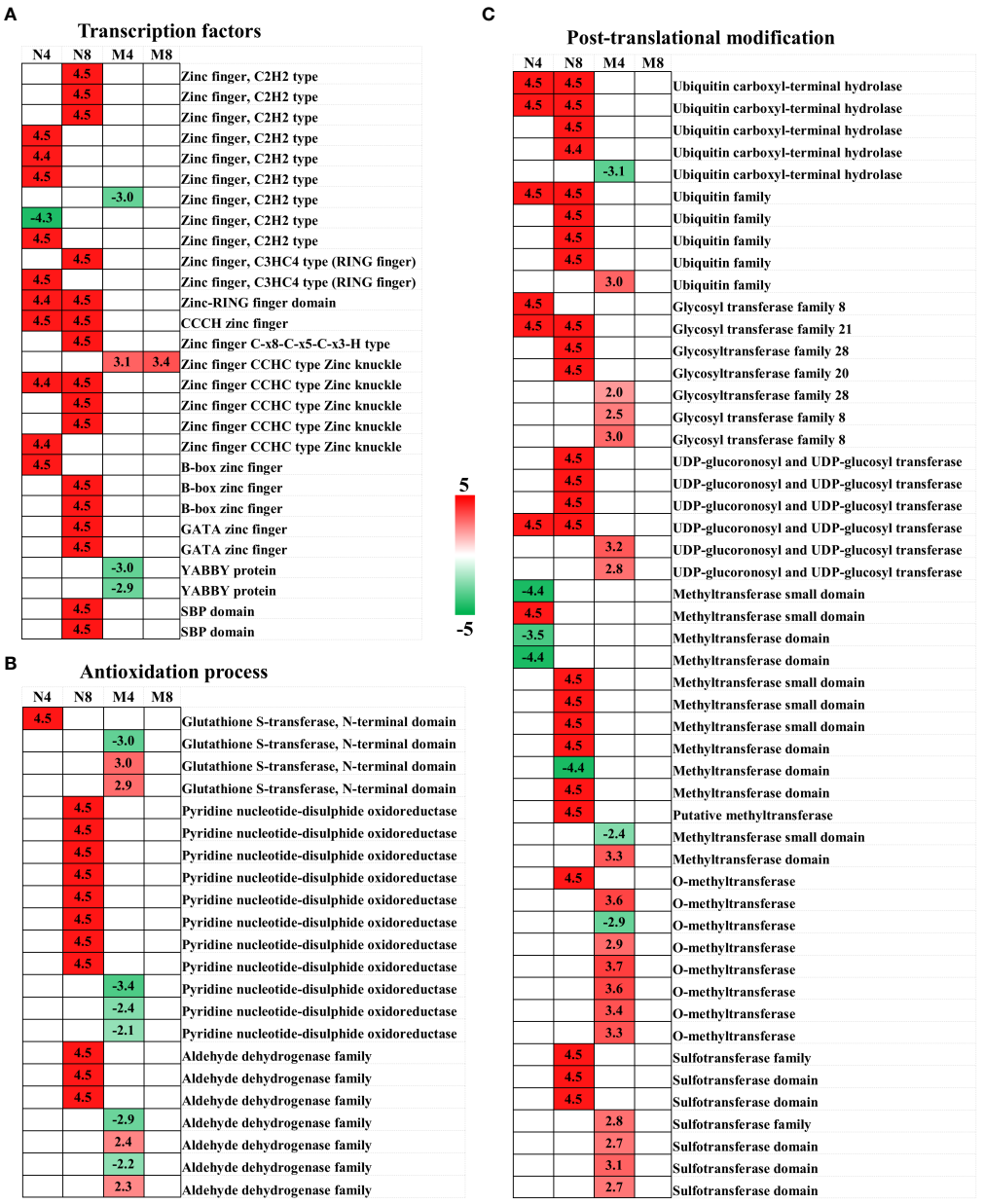


FIGURE 6
Heat map of candidate DEGs involved in transcription factors **(A)**, Ca^{2+} sensing and transduction **(B)**, and post-translational modification **(C)**. Samples values used are $\log_2$(fold change). The four columns, from left to right, represent N4, N8, M4, and M8. Color key is shown on the right. Red represents up-regulation, green down-regulation, and white no significant change.

proteins and GATA zinc finger proteins have been recently discovered to be tightly associated with ABA signaling and ROS scavenging in the defense from salt stress (Gupta et al., 2017; Liu et al., 2019b), indicating the role these five genes play in Adina. We detected two YABBY transcription factors, part of the subfamily of zinc finger protein, differentially expressed in Zhaodong. *YABBY*s were previously identified as the DEGs in cotton (*Gossypium klotzschianum*) in salt stress condition in a study carried out on the transcriptomic data (Wei et al., 2017). *YABBY*s could be negative regulatory factors in salt stress response in Zhaodong, as *AcYABBY4* does in *Arabidopsis* and *GmYABBY10* in *Glycine max* (Zhao et al., 2017; Li et al., 2019). We identified two up-regulated Squamosa promoter binding protein (SBP)-box genes in Adina, which encode transcription factors exclusive to plants. In previous research it was demonstrated how *Arabidopsis* plants over-expressing an SBP-

box transcription factor (*VpSBP16*) from a Chinese wild vitis species had improved salt-tolerance because of the better regulation of SOS and ROS signaling (Hou et al., 2018), indicating *SBPs* may have a role in salt stress defense in Adina.

DEGs involved in the antioxidation process

Exposure of plants to salt stress leads to oxidative stress and damage (Munns and Tester, 2008). Four glutathione S-transferases (*GSTs*), ubiquitous enzymes playing an important role in cellular detoxification under abiotic stress, were identified as DEGs in our study. Overexpressing *GST*s, as *OsGSTU4* (Sharma et al., 2014) and *LbGST1* (Diao et al., 2011), can significantly reduce the production of the reactive oxygen species and oxidative damage. Three of the four GSTs found in our study were upregulated, indicating their roles in salt stress defense *via* enhancing the antioxidation process in alfalfa. In addition, we found eight pyridine nucleotide-disulphide oxidoreductase genes (*PYROX*), which encode the conserved domain of glutathione reductase (GR), a major antioxidase in plants (Trivedi et al., 2013), were up-regulated in Adina, and three down-regulated in Zhaodong. The resulting up-regulation of the pyridine nucleotide-disulphide oxidoreductase gene could indicate a higher expression of the *GR* transcripts, resulting in more GR activity and therefore oxidation resistance in Adina (Trivedi et al., 2013; Mudalkar et al., 2017). The accumulation of aldehydes under salt stress is another factor that leads to the oxidative stress, and the major detoxification pathways of aldehyde is the oxidation of their carbonyl groups into carboxylic acids by aldehyde dehydrogenase (ALDH) (Xu et al., 2013). We identified seven *ALDH* genes as DEGs after salt stress treatment, three in Adina and four in Zhaodong. The 3 *ALDH* genes in Adina were up-regulated, suggesting the detoxification process in Adina by ALDH may be enhanced in these conditions (Gautam et al., 2020).

DEGs involved in post-translational modifications

Post-translational modifications play critical roles for plants' response to salt stress (Zhou et al., 2012). We found four ubiquitin carboxyl-terminal hydrolases (*UCH*), which are involved in the plant ubiquitin C-terminal hydrolases, were increased in Adina, and one decreased in Zhaodong. Genes with UCH domain, such as *UBP16* in *Arabidopsis* (Zhou et al., 2012), and *ZmUBP15*, *ZmUBP16* and *ZmUBP19* in maize (Kong et al., 2019), have been reported to encoding ubiquitin-specific protease for salt-tolerance. In addition, in this study we observed five ubiquitin families were upregulated. The up-regulation of these ubiquitin-specific protease genes is consistent with the enhancement of ubiquitination in protein involved in salt tolerance in Adina. Furthermore, all of the 13 glycosyl transferases (*GTs*), enzymes in charge of glycosylation, were upregulated after salt stress treatment. GTs, as well the UDP-sugar glycosyltransferases (UGTs), have been illustrated to have crucial roles in improving plant resistance to stress conditions (Oussama et al., 2015; Pasquet et al., 2016). The enhanced expression of the *GT*s in our study suggests an advanced glycosylation process in alfalfa, in both Adina and Zhaodong, in order for the plants to achieve stronger salt stress adaptation, which is consistent with the reported salt-tolerance roles of *UGT85A5* (Sun et al., 2013), *UGT76E11* (Qin et al., 2018), and *CrUGT87A1* (Zhang et al., 2020b). We also identified 21 methyltransferases (*MT*), the enzyme that catalyze the methylation reaction. Recent studies suggested that the two Caffeic acid O-methyltransferase (*COMT*) genes, *CrCOMT* (Zhang et al., 2019) and *SlCOMT1* (Sun et al., 2020), had the higher transformation of N-acetylserotonin into melatonin, leading to higher salt tolerance. The regulated methylation processes under salt stress conditions might be an important salt tolerance mechanism in alfalfa. Seven sulfotransferases (*SOT*), enzymes catalyzing the sulfation process, were also increased after salt stress both in Adina and Zhaodong. The changes of the *SOT*s in our study are consistent with what is documented in transcriptional assays, such as the expression of *AtSOT12* (Hirschmann et al., 2014), *BrSOT*s (Jin et al., 2019) and *SjSOT*s (Lu et al., 2020) after exposure to salt stress. A leading hypothesis is that the sulfation of polysaccharides that causes the modification of the composition of cell walls is the main reason for increased salt tolerance in plants (Hirschmann et al., 2014). The specific salt adaptation roles of *SOT*s in alfalfa need to be further explored.

Conclusions

In this study, we conducted a time-course transcriptome assay at four and eight hours after exposure to high salinity to explore the integrative molecular salt stress adaptation mechanisms in alfalfa. DEGs identified from the RNA-seq were analyzed with GO annotation and KEGG pathway enrichment database, in order to highlight the functional roles of transcription factor activity, signal sensing and transduction, and catalytic function. DEGs involved in or encoding ion and membrane homeostasis, Ca^{2+} sensing and transduction, phytohormone signaling and regulation, transcription factors, antioxidation process, and post-translational modification, are shown to contributing to salt stress adaptation in alfalfa. Our transcriptome results widen the knowledge of the molecular regulatory mechanisms in alfalfa and provide a number of

candidate functional genes for salt tolerance breeding and improvements in alfalfa.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA821982.

Author contributions

DM, JC, and QM were responsible for the experimental design. DM, WW, and LZ performed the experiments. DM, JL, and LS prepared the manuscript and coordinated its revision. DM and JC read and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Project of National Natural Science Foundation of China (31760698), the Ningxia Agricultural Breeding Project (2019NYYZ0401), the Project of First-Class Discipline Construction (Ecology) for Ningxia Higher Education (NXYLXK2017B06), the Ningxia Research and Development key project (2019BBF02022-04), and the Project of First-Class University of Western China for Key Laboratory of Ningxia University (GZXM2017001).

Acknowledgments

We want to acknowledge Shanghai LuMing Biological Technology Co. Ltd for performing LC–MS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1070846/full#supplementary-material

References

Abd-Hamid, N.-A., Ahmad-Fauzi, M.-I., Zainal, Z., and Ismail, I. (2020). Diverse and dynamic roles of f-box proteins in plant biology. *Planta* 251, 1–31. doi: 10.1007/s00425-020-03356-8

Amtmann, A., and Beilby, M. J. (2010). "The role of ion channels in plant salt tolerance," in *Signaling and communication in plants* (Springer), 23–46.

Arshad, M., Gruber, M. Y., Wall, K., and Hannoufa, A. (2017). An insight into microRNA156 role in salinity stress responses of alfalfa. *Front. Plant Sci.* 8, 356. doi: 10.3389/fpls.2017.00356

Ashrafi, E., Razmjoo, J., Zahedi, M., and Pessarakli, M. (2014). Selecting alfalfa cultivars for salt tolerance based on some physiochemical traits. *Agron. J.* 106, 1758–1764. doi: 10.2134/agronj13.0569

Büyük, İ., İlhan, E., Şener, D., Özsoy, A. U., and Aras, S. (2019). Genome-wide identification of *CAMTA* gene family members in phaseolus vulgaris l. and their expression profiling during salt stress. *Mol. Biol. Rep.* 46, 2721–2732. doi: 10.1007/s11033-019-04716-8

Cao, Y., Zhang, M., Liang, X., Li, F., Shi, Y., Yang, X., et al. (2020). Natural variation of an EF-hand Ca^{2+}-binding-protein coding gene confers saline-alkaline tolerance in maize. *Nat. Commun.* 11, 1–14. doi: 10.1038/s41467-019-14027-y

Cui, J., Ren, G., Qiao, H., Xiang, X., Huang, L., and Chang, J. (2018). Comparative transcriptome analysis of seedling stage of two sorghum cultivars under salt stress. *J. Plant Growth Regul.* 37, 986–998. doi: 10.1007/s00344-018-9796-9

Deinlein, U., Stephan, A. B., Horie, T., Luo, W., Xu, G., and Schroeder, J. I. (2014). Plant salt-tolerance mechanisms. *Trends Plant Sci.* 19, 371–379. doi: 10.1016/j.tplants.2014.02.001

Diao, G., Wang, Y., Wang, C., and Yang, C. (2011). Cloning and functional characterization of a novel glutathione s-transferase gene from limonium bicolor. *Plant Mol. Biol. Rep.* 29, 77–87. doi: 10.1007/s11105-010-0212-2

Dinneny, J. R. (2015). Traversing organizational scales in plant salt-stress responses. *Curr. Opin. Plant Biol.* 23, 70–75. doi: 10.1016/j.pbi.2014.10.009

Du, X., Wang, G., Ji, J., Shi, L., Guan, C., and Jin, C. (2017). Comparative transcriptome analysis of transcription factors in different maize varieties under salt stress conditions. *Plant Growth Regul.* 81, 183–195. doi: 10.1007/s10725-016-0192-9

Gautam, R., Meena, R. K., Woch, N., and Kirti, P. (2020). Ectopic expression of *BrALDH7B2* gene encoding an antiquitin from *Brassica rapa* confers tolerance to abiotic stresses and improves photosynthetic performance under salt stress in tobacco. *Environ. Exp. Bot.* 180, 104223. doi: 10.1016/j.envexpbot.2020.104223

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Guan, Q., Wang, L.-F., Bu, Q., and Wang, Z. (2014). The rice gene *OsZFP6* functions in multiple stress tolerance responses in yeast and arabidopsis. *Plant Physiol. Biochem.* 82, 1–8. doi: 10.1016/j.plaphy.2014.04.021

Guo, Y., Yu, Y., Wang, D., Wu, C., Yang, G., Huang, J., et al. (2009). GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5. *New Phytol.* 183, 62–75. doi: 10.1111/j.1469-8137.2009.02838.x

Gupta, P., Nutan, K. K., Singla-Pareek, S. L., and Pareek, A. (2017). Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. *Front. Plant Sci.* 8, 1944. doi: 10.3389/fpls.2017.01944

Haak, D. C., Fukao, T., Grene, R., Hua, Z., Ivanov, R., Perrella, G., et al. (2017). Multilevel regulation of abiotic stress responses in plants. *Front. Plant Sci.* 8, 1564. doi: 10.3389/fpls.2017.01564

Hadi, M. R., and Karimi, N. (2012). The role of calcium in plants' salt tolerance. *J. Plant Nutr.* 35, 2037–2054. doi: 10.1080/01904167.2012.717158

Hirschmann, F., Krause, F., and Papenbrock, J. (2014). The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* 5, 556. doi: 10.3389/fpls.2014.00556

Hou, H., Jia, H., Yan, Q., and Wang, X. (2018). Overexpression of a SBP-box gene (*VpSBP16*) from chinese wild vitis species in arabidopsis improves salinity and drought stress tolerance. *J. Int. J. Mol. Sci.* 19, 940. doi: 10.3390/ijms19040940

Hu, T., Jin, Y., Li, H., Amombo, E., and Fu, J. (2016). Stress memory induced transcriptional and metabolic changes of perennial ryegrass (*Lolium perenne*) in response to salt stress. *Physiologia Plantarum* 156, 54–69. doi: 10.1111/ppl.12342

Jing, P., Zou, J., Kong, L., Hu, S., Wang, B., Yang, J., et al. (2016). OsCCD1, a novel small calcium-binding protein with one EF-hand motif, positively regulates osmotic and salt tolerance in rice. *Plant Sci.* 247, 104–114. doi: 10.1016/j.plantsci.2016.03.011

Jin, L., Ouyang, N., Huang, Y., Liu, C., and Ruan, Y. J. P. O. (2019). Genome-wide analysis of sulfotransferase genes and their responses to abiotic stresses in Chinese cabbage (*Brassica rapa* l.). *PloS One* 14, e0221422. doi: 10.1371/journal.pone.0221422

Jonathan Ledoux,, Werner, M. E., Brayden, J. E., and Nelson, M. T. (2006). Calcium-activated potassium channels and the regulation of vascular tone. *Physiology* 21, 69–79. doi: 10.1152/physiol.00040.2005

Jung, Y., Lee, I.-H., Nou, I.-S., Lee, K.-D., Rashotte, A. M., and Kang, K.-K. (2013). BrRZFP1 a brassica rapa C3HC4-type RING zinc finger protein involved in cold, salt and dehydration stress. *Plant Biol.* 15, 274–283. doi: 10.1111/j.1438-8677.2012.00631.x

Kang, P., Bao, A.-K., Kumar, T., Pan, Y.-Q., Bao, Z., Wang, F., et al. (2016). Assessment of stress tolerance, productivity, and forage quality in T1 transgenic alfalfa co-overexpressing *ZxNHX* and *ZxVP1-1* from *Zygophyllum xanthoxylum*. *Front. Plant Sci.* 7, 1598. doi: 10.3389/fpls.2016.01598

Kang, J., Hwang, J.-U., Lee, M., Kim, Y.-Y., Assmann, S. M., Martinoia, E., et al. (2010). PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc. Natl. Acad. Sci.* 107, 2355–2360. doi: 10.1073/pnas.0909222107

Kim, D., Jin, J. Y., Alejandro, S., Martinoia, E., and Lee, Y. (2010). Overexpression of *AtABCG36* improves drought and salt stress resistance in arabidopsis. *Physiologia plantarum* 139, 170–180. doi: 10.1111/j.1399-3054.2010.01353.x

Kong, J., Jin, J., Dong, Q., Qiu, J., Li, Y., Yang, Y., et al. (2019). Maize factors ZmUBP15, ZmUBP16 and ZmUBP19 play important roles for plants to tolerance the cadmium stress and salt stress. *Plant Sci.* 280, 77–89. doi: 10.1016/j.plantsci.2018.11.014

Kong, X., Zhou, S., Yin, S., Zhao, Z., Han, Y., and Wang, W. (2016). Stress-inducible expression of an f-box gene *TaFBA1* from wheat enhanced the drought tolerance in transgenic tobacco plants without impacting growth and development. *Front. Plant Sci.* 7, 1295. doi: 10.3389/fpls.2016.01295

Lai, Y., Zhang, D., Wang, J., Wang, J., Ren, P., Yao, L., et al. (2020). Integrative transcriptomic and proteomic analyses of molecular mechanism responding to salt stress during seed germination in hulless barley. *Int. J. Mol. Sci.* 21, 359. doi: 10.3390/ijms21010359

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, Z., Li, G., Cai, M., Priyadarshani, S. V., Aslam, M., Zhou, Q., et al. (2019). Genome-wide analysis of the YABBY transcription factor family in pineapple and functional identification of AcYABBY4 involvement in salt stress. *Int. J. Mol. Sci.* 20, 5863. doi: 10.3390/ijms20235863

Li, M.-N., Long, R.-C., Feng, Z.-R., Liu, F.-Q., Sun, Y., Zhang, K., et al. (2018). Transcriptome analysis of salt-responsive genes and SSR marker exploration in *Carex rigescens* using RNA-seq. *J. Integr. Agric.* 17, 184–196. doi: 10.1016/S2095-3119(17)61749-0

Li, Q., Qin, Y., Hu, X., Li, G., Ding, H., Xiong, X., et al. (2020). Transcriptome analysis uncovers the gene expression profile of salt-stressed potato (*Solanum tuberosum* l.). *Sci. Rep.* 10, 1–19. doi: 10.1038/s41598-020-62057-0

Liu, X., Li, R., Dai, Y., Yuan, L., Sun, Q., Zhang, S., et al. (2019b). A b-box zinc finger protein, MdBBX10, enhanced salt and drought stresses tolerance in arabidopsis. *Plant Mol. Biol.* 99, 437–447. doi: 10.1007/s11103-019-00828-8

Liu, X., Pi, B., Pu, J., Cheng, C., Fang, J., and Yu, B. (2020). Genome-wide analysis of chloride channel-encoding gene family members and identification of *CLC* genes that respond to Cl^-/salt stress in upland cotton. *Mol. Biol. Rep.* 47 (12), 9361–9371. doi: 10.1007/s11033-020-06023-z

Liu, A., Xiao, Z., Li, M. W., Wong, F. L., Yung, W. S., Ku, Y. S., et al. (2019a). Transcriptomic reprogramming in soybean seedlings under salt stress. *Plant Cell Environ.* 42, 98–114. doi: 10.1111/pce.13186

Liu, X., Zhu, Y., Zhai, H., Cai, H., Ji, W., Luo, X., et al. (2012). AtPP2CG1, a protein phosphatase 2C, positively regulates salt tolerance of arabidopsis in abscisic acid-dependent manner. *Biochem. Biophys. Res. Commun.* 422, 710–715. doi: 10.1016/j.bbrc.2012.05.064

Li, H., Wang, Z., Ke, Q., Ji, C. Y., Jeong, J. C., Lee, H.-S., et al. (2014). Overexpression of *codA* gene confers enhanced tolerance to abiotic stresses in alfalfa. *Plant Physiol. Biochem.* 85, 31–40. doi: 10.1016/j.plaphy.2014.10.010

Li, Y., Zhang, L., Li, D., Liu, Z., Wang, J., Li, X., et al. (2016). The arabidopsis f-box E3 ligase RIFP1 plays a negative role in abscisic acid signalling by facilitating ABA receptor RCAR3 degradation. *Plant Cell Environ.* 39, 571–582. doi: 10.1111/pce.12639

Li, M., Zhang, K., Long, R., Sun, Y., Kang, J., Zhang, T., et al. (2017). iTRAQ-based comparative proteomic analysis reveals tissue-specific and novel early-stage molecular mechanisms of salt stress response in *Carex rigescens*. *Environ. Exp. Bot.* 143, 99–114. doi: 10.1016/j.envexpbot.2017.08.010

Long, R., Gao, Y., Sun, H., Zhang, T., Li, X., Li, M., et al. (2018). Quantitative proteomic analysis using iTRAQ to identify salt-responsive proteins during the germination stage of two medicago species. *Sci. Rep.* 8, 9553. doi: 10.1038/s41598-018-27935-8

Long, R. C., Li, M. N., Kang, J. M., Zhang, T. J., Sun, Y., and Yang, Q. C. (2015). Small RNA deep sequencing identifies novel and salt-stress-regulated microRNAs from roots of *Medicago sativa* and *Medicago truncatula*. *Physiologia Plantarum* 154, 13–27. doi: 10.1111/ppl.12266

Long, R., Li, M., Li, X., Gao, Y., Zhang, T., Sun, Y., et al. (2017). A novel miRNA sponge form efficiently inhibits the activity of miR393 and enhances the salt tolerance and ABA insensitivity in arabidopsis thaliana. *Plant Mol. Biol. Rep.* 35, 409–415. doi: 10.1007/s11105-017-1033-3

Long, R., Li, M., Zhang, T., Kang, J., Sun, Y., Cong, L., et al. (2016). Comparative proteomic analysis reveals differential root proteins in *Medicago sativa* and *Medicago truncatula* in response to salt stress. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00424

Long, R., Yang, Q., Kang, J., Zhang, T., Wang, H., Li, M., et al. (2013). Overexpression of a novel salt stress-induced glycine-rich protein gene from alfalfa causes salt and ABA sensitivity in arabidopsis. *Plant Cell Rep.* 32, 1289–1298. doi: 10.1007/s00299-013-1443-0

Luo, D., Liu, W., Wang, Y., Zhang, J., and Liu, Z. (2014). Development of a rapid one-step PCR protocol to distinguish between alfalfa (*Medicago sativa*) and sweet clover (*Melilotus* spp.) seeds. *Seed Sci. Technol.* 42, 237–246. doi: 10.15258/sst.2014.42.2.12

Lu, C., Shao, Z., Zhang, P., and Duan, D. (2020). Genome-wide analysis of the saccharina japonica sulfotransferase genes and their transcriptional profiles during whole developmental periods and under abiotic stresses. *BMC Plant Biol.* 20, 1–20. doi: 10.1186/s12870-020-02422-3

Moon, S., and Jung, K.-H. (2014). Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses. *J. Plant Physiol.* 171, 1276–1288. doi: 10.1016/j.jplph.2014.05.006

Mudalkar, S., Sreeharsha, R. V., and Reddy, A. R. (2017). Involvement of glyoxalases and glutathione reductase in conferring abiotic stress tolerance to *Jatropha curcas* l. *Environ. Exp. Bot.* 134, 141–150. doi: 10.1016/j.envexpbot.2016.11.011

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Nguyen, N. H., Jung, C., and Cheong, J.-J. (2019). Chromatin remodeling for the transcription of type 2C protein phosphatase genes in response to salt stress. *Plant Physiol. Biochem.* 141, 325–331. doi: 10.1016/j.plaphy.2019.06.012

Oussama, A., Angela, R.-M., Almudena, T.-M., María, F. L. C., Aurelio, G.-C., and Lourdes, G.-G. (2015). Ectopic expression of a stress-inducible glycosyltransferase from saffron enhances salt and oxidative stress tolerance in arabidopsis while alters anchor root formation. *Plant Sci.* 02, 60–73. doi: 10.1016/j.plantsci.2015.02.004

Parihar, P., Singh, S., Singh, R., Singh, V. P., and Prasad, S. M. (2015). Effect of salinity stress on plants and its tolerance strategies: a review. *Environ. Sci. pollut. Res.* 22, 4056–4075. doi: 10.1007/s11356-014-3739-1

Pasquet, J. C., Changenet, V., Macadre, C., Boex-Fontvieille, E., Soulhat, C., Bouchabke-Coussa, O., et al. (2016). A brachypodium UDP-glycosyltransferase confers root tolerance to deoxynivalenol and resistance to fusarium infection. *Plant Physiol. Biochem.* 172, 559–574. doi: 10.1104/pp.16.00371

Postnikova, O. A., Shao, J., and Nemchinov, L. G. (2013). Analysis of the alfalfa root transcriptome in response to salinity stress. *Plant Cell Physiol.* 54, 1041–1055. doi: 10.1093/pcp/pct056

Priest, H. D., Fox, S. E., Rowley, E. R., Murray, J. R., Michael, T. P., and Mockler, T. C. (2014). Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress. *PloS One* 9, e87499. doi: 10.1371/journal.pone.0087499

Qin, L., Hui-Min, Y., Xia-Fei, M., Ji-Shan, L., Yan-Jie, L., and Bing-Kai, H. (2018). Ectopic expression of glycosyltransferase UGT76E11 increases flavonoid accumulation and enhances abiotic stress tolerance in arabidopsis. *Plant Biol.* 20, 10–19. doi: 10.1111/plb.12627

Rasul, I., Nadeem, H., Siddique, M., Atif, R., Ali, M., Umer, A., et al. (2017). Plants sensory-response mechanisms for salinity and heat stress. *JAPS: J. Anim. Plant Sci.* 27 (2).

Shankar, R., Bhattacharjee, A., and Jain, M. (2016). Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep.* 6, 23719. doi: 10.1038/srep23719

Sharma, M., and Pandey, G. K. (2016). Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* 6, 1218. doi: 10.3389/fpls.2015.01218

Sharma, R., Sahoo, A., Devendran, R., and Jain, M. (2014). Over-expression of a rice tau class glutathione s-transferase gene improves tolerance to salinity and oxidative stresses in arabidopsis. *PloS One* 9, e92900. doi: 10.1371/journal.pone.0092900

Sun, S., Guo, S., Yang, X., Bao, Y., Tang, H., Sun, H., et al. (2010). Functional analysis of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice. *J. Exp. Bot.* 61, 2807–2818. doi: 10.1093/jxb/erq120

Sun, Y.-G., Wang, B., Jin, S.-H., Qu, X.-X., Li, Y.-J., and Hou, B.-K. (2013). Ectopic expression of arabidopsis glycosyltransferase *UGT85A5* enhances salt stress tolerance in tobacco. *PloS One* 8, e59924. doi: 10.1371/journal.pone.0059924

Sun, X., Wang, Y., and Sui, N. (2018). Transcriptional regulation of bHLH during plant response to stress. *Biochem. Biophys. Res. Commun.* 503, 397–401. doi: 10.1016/j.bbrc.2018.07.123

Sun, S., Wen, D., Yang, W., Meng, Q., Shi, Q., and Gong, B. (2020). Overexpression of caffeic acid O-methyltransferase 1 (*COMT1*) increases melatonin level and salt stress tolerance in tomato plant. *J. Plant Growth Regul.* 39 (3), 1221–1235. doi: 10.1007/s00344-019-10058-3

Sun, X., Xu, L., Wang, Y., Luo, X., Zhu, X., Kinuthia, K. B., et al. (2016). Transcriptome-based gene expression profiling identifies differentially expressed genes critical for salt stress response in radish (*Raphanus sativus* l.). *Plant Cell Rep.* 35, 329–346. doi: 10.1007/s00299-015-1887-5

Trivedi, D. K., Gill, S. S., Yadav, S., and Tuteja, N. (2013). Genome-wide analysis of glutathione reductase (GR) genes from rice and arabidopsis. *Plant Signaling Behav.* 8, e23021. doi: 10.4161/psb.23021

Wang, Y., Jiang, L., Chen, J., Tao, L., An, Y., Cai, H., et al. (2018). Overexpression of the alfalfa *WRKY11* gene enhances salt tolerance in soybean. *PloS One* 13, e0192382. doi: 10.1371/journal.pone.0192382

Wang, Z., Li, H., Ke, Q., Jeong, J. C., Lee, H.-S., Xu, B., et al. (2014). Transgenic alfalfa plants expressing *AtNDPK2* exhibit increased growth and tolerance to abiotic stresses. *Plant Physiol. Biochem.* 84, 67–77. doi: 10.1016/j.plaphy.2014.08.025

Wang, P., Li, Z., Wei, J., Zhao, Z., Sun, D., and Cui, S. (2012). A Na^+/Ca^{2+} exchanger-like protein (AtNCL) involved in salt stress in arabidopsis. *J. Biol. Chem.* 287, 44062–44070. doi: 10.1074/jbc.M112.351643

Wang, J., Zhang, P., Liu, S., Cong, B., and Chen, K. (2016a). A leucine-rich repeat receptor-like kinase from the Antarctic moss pohlia nutans confers salinity and ABA stress tolerance. *Plant Mol. Biol. Rep.* 34, 1136–1145. doi: 10.1007/s11105-016-0994-y

Wang, W., Zhao, X., Li, M., Huang, L., Xu, J., Zhang, F., et al. (2016b). Complex molecular mechanisms underlying seedling salt tolerance in rice revealed by comparative transcriptome and metabolomic profiling. *J. Exp. Bot.* 67, 405–419. doi: 10.1093/jxb/erv476

Wei, Q., Liu, Y., Zhou, G., Li, Q., Yang, C., and Peng, S.-A. (2013). Overexpression of *CsCLCc*, a chloride channel gene from *Poncirus trifoliata*, enhances salt tolerance in arabidopsis. *Plant Mol. Biol. Rep.* 31, 1548–1557. doi: 10.1007/s11105-013-0592-1

Wei, Y., Xu, Y., Lu, P., Wang, X., Li, Z., Cai, X., et al. (2017). Salt stress responsiveness of a wild cotton species (*Gossypium klotzschianum*) based on transcriptomic analysis. *PloS One* 12, e0178313. doi: 10.1371/journal.pone.0178313

Wu, B., Hu, Y., Huo, P., Zhang, Q., Chen, X., and Zhang, Z. (2017). Transcriptome analysis of hexaploid hulless oat in response to salinity stress. *PloS One* 12, e0171451. doi: 10.1371/journal.pone.0171451

Wu, C., Yang, G., Meng, Q., and Zheng, C. (2004). The cotton *GhNHX1* gene encoding a novel putative tonoplast Na+/H+ antiporter plays an important role in salt stress. *Plant Cell Physiol.* 45, 600–607. doi: 10.1093/pcp/pch071

Xu, X., Guo, R., Cheng, C., Zhang, H., Zhang, Y., and Wang, X. (2013). Overexpression of *ALDH2B8*, an aldehyde dehydrogenase gene from grapevine, sustains arabidopsis growth upon salt stress and protects plants against oxidative stress. *Plant Cell Tissue Organ Culture* 114, 187–196. doi: 10.1007/s11240-013-0314-2

You, J., and Chan, Z. (2015). ROS regulation during abiotic stress responses in crop plants. *Front. Plant Sci.* 6, 1092. doi: 10.3389/fpls.2015.01092

Yu, L.-X., Liu, X., Boge, W., and Liu, X.-P. (2016). Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa (*Medicago sativa* l.) using genotyping-by-sequencing. *Front. Plant Sci.* 7, 956. doi: 10.3389/fpls.2016.00956

Zeng, H., Zhang, Y., Zhang, X., Pi, E., and Zhu, Y. (2017). Analysis of EF-hand proteins in soybean genome suggests their potential roles in environmental and nutritional stress signaling. *Front. Plant Sci.* 8, 877. doi: 10.3389/fpls.2017.00877

Zhang, K., Cui, H., Cao, S., Yan, L., Li, M., and Sun, Y. (2019). Overexpression of *CrCOMT* from *Carex rigescens* increases salt stress and modulates melatonin synthesis in *Arabidopsis thaliana*. *Plant Cell Rep.* 38, 1501–1514. doi: 10.1007/s00299-019-02461-7

Zhang, K., Cui, H., Li, M., Xu, Y., Cao, S., Long, R., et al. (2020a). Comparative time-course transcriptome analysis in contrasting *Carex rigescens* genotypes in response to high environmental salinity. *Ecotoxicol Environ. Saf.* 194, 110435. doi: 10.1016/j.ecoenv.2020.110435

Zhang, K., Sun, Y., Li, M., and Long, R. (2020b). *CrUGT87A1*, a UDP-sugar glycosyltransferases (UGTs) gene from *Carex rigescens*, increases salt tolerance by accumulating flavonoids for antioxidation in arabidopsis thaliana. *Plant Physiol. Biochem.* 159, 28–36. doi: 10.1016/j.plaphy.2020.12.006

Zhang, Z., Wang, Y., Chang, L., Zhang, T., An, J., Liu, Y., et al. (2016). *MsZEP*, a novel zeaxanthin epoxidase gene from alfalfa (*Medicago sativa*), confers drought and salt tolerance in transgenic tobacco. *Plant Cell Rep.* 35, 439–453. doi: 10.1007/s00299-015-1895-5

Zhao, S., Lu, D., Yu, T., Ji, Y., Zheng, W., Zhang, S., et al. (2017). Genome-wide analysis of the YABBY family in soybean and functional identification of GmYABBY10 involvement in high salt and drought stresses. *Plant Physiol. Biochem.* 119, 132–146. doi: 10.1016/j.plaphy.2017.08.026

Zhou, H., Zhao, J., Yang, Y., Chen, C., Liu, Y., Jin, X., et al. (2012). Ubiquitin-specific protease16 modulates salt tolerance in arabidopsis by regulating Na^+/H^+ antiport activity and serine hydroxymethyltransferase stability. *Plant Cell* 24, 5106–5122. doi: 10.1105/tpc.112.106393

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 December 2022
DOI 10.3389/fpls.2022.1082415

Check for updates

OPEN ACCESS

EDITED BY
Hui Song,
Qingdao Agricultural University, China

REVIEWED BY
Qianqian Shi,
Northwest A&F University, China
Shiping Cheng,
Pingdingshan University, China

*CORRESPONDENCE
Xiaogai Hou
hkdhxg@haust.edu.cn
Xiuxin Zhang
zhangxiuxin@caas.cn

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 28 October 2022
ACCEPTED 17 November 2022
PUBLISHED 14 December 2022

CITATION
Guo L, Li Y, Zhang C, Wang Z, Carlson JE, Yin W, Zhang X and Hou X (2022) Integrated analysis of miRNAome transcriptome and degradome reveals miRNA-target modules governing floral florescence development and senescence across early- and late-flowering genotypes in tree peony.
Front. Plant Sci. 13:1082415.
doi: 10.3389/fpls.2022.1082415

COPYRIGHT
© 2022 Guo, Li, Zhang, Wang, Carlson, Yin, Zhang and Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Integrated analysis of miRNAome transcriptome and degradome reveals miRNA-target modules governing floral florescence development and senescence across early- and late-flowering genotypes in tree peony

Lili Guo[1], Yuying Li[1], Chenjie Zhang[1], Zhanying Wang[2], John E. Carlson[3], Weinlun Yin[4], Xiuxin Zhang[5]* and Xiaogai Hou[5]*

[1]College of Tree Peony, Henan University of Science and Technology, Luoyang, Henan, China, [2]Department of Horticulture, Luoyang Academy of Agricultural and Forestry Sciences, Luoyang, Henan, China, [3]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA, United States, [4]College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [5]Center of Peony, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Science, Beijing, China

As a candidate national flower of China, tree peony has extremely high ornamental, medicinal and oil value. However, the short florescence and rarity of early-flowering and late-flowering varieties restrict further improvement of the economic value of tree peony. Specific miRNAs and their target genes engaged in tree peony floral florescence, development and senescence remain unknown. This report presents the integrated analysis of the miRNAome, transcriptome and degradome of tree peony petals collected from blooming, initial flowering, full blooming and decay stages in early-flowering variety *Paeonia ostii* 'Fengdan', an early-flowering mutant line of *Paeonia ostii* 'Fengdan' and late-flowering variety *Paeonia suffruticosa* 'Lianhe'. Transcriptome analysis revealed a transcript (*'psu.G.00014095'*) which was annotated as a xyloglucan endotransglycosylase/hydrolase precursor *XTH-25* and found to be differentially expressed across flower developmental stages in *Paeonia ostii* 'Fengdan' and *Paeonia suffruticosa* 'Lianhe'. The miRNA-mRNA modules were presented significant enrichment in various pathways such as plant hormone signal transduction, indole alkaloid biosynthesis, arachidonic acid metabolism, folate biosynthesis, fatty acid elongation, and the MAPK signaling pathway. Multiple miRNA-mRNA-TF modules demonstrated the potential functions of *MYB*-related, *bHLH*, *Trihelix*, *NAC*, *GRAS* and *HD-ZIP* TF families in floral florescence, development, and senescence of tree peony. Comparative spatio-temporal expression investigation of eight floral-favored miRNA-target

modules suggested that transcript '*psu.T.00024044*' and microRNA *mtr-miR166g-5p* are involved in the floral florescence, development and senescence associated agronomic traits of tree peony. The results might accelerate the understanding of the potential regulation mechanism in regards to floral florescence, development and abscission, and supply guidance for tree peony breeding of varieties with later and longer florescence characteristics.

KEYWORDS

tree peony, miRNA-target modules, florescence, floral development, senescence

1 Introduction

Tree peony (*Paeonia suffruticosa* Andr.) is a perennial deciduous shrub of *Paeoniaceae*. All kinds of tree peony species are endemic to China (Wang et al., 2019). Tree peony, famous for 'the king of flowers', is big, colorful, fragrant, graceful and elegant, symbolizing wealth and prosperity, and is renowned as a symbol of Chinese civilization (Yang et al., 2020). The poem 'Only peony is the true color of the country, which moves the capital when they are blossoming', vividly describes the people's deep love for peony flowers in the Tang Dynasty. These days, tree peonies are cultivated all over the world, and people's enthusiasm to cultivate and plant tree peonies is still increasing (Hong et al., 2017).

Research on ornamental characters of tree peony has made continuous progress from identification and cultivation of varieties to improved cultivation technology. There are already more than 2,000 ornamental varieties of tree peony in China (Luo et al., 2021). However, the flowering time of tree peony is still quite short and convergent. Under natural conditions, it takes only 50-60 days from budding to fading; the flowering period is 3-5 days for a single flower and 10-15 days for a colony of plants. Most tree peony varieties are middle-flowering varieties, with the proportion of early-flowering and late-flowering varieties being quite few (Li et al., 2011). In addition, due to the lack of research on the genetic basis and molecular regulation mechanism of flowering in tree peony, it is difficult to improve the breeding and cultivation techniques to meet the demand for prolonging of flowering time of tree peony. These are all important factors restricting the improvement of tree peony ornamental value and the further development of international markets (Kamenetsky et al., 2003). Florescence has thus been one of the key ornamental traits that limit the improving of the economic value of tree peony, and the theoretical and technical research needed to solve this problem has become a key focus in both the public and scientific horticulture communities.

In efforts to prolong the ornamental period of tree peony, studies have been carried out from such aspects as early and late flowering hybrid breeding and growth regulator regulation. However, conventional breeding of woody perennials in tree peony is time-consuming and labor-intensive making it difficult to quickly meet the market demand for new varieties. Although high-density genetic maps and QTL identification are now being reported, the low breeding efficiency greatly currently remains a limiting factor in the selection and breeding process of tree peony varieties (Cai et al., 2015; Li et al., 2019; Zhang et al., 2019). It should thus be of great value to construct the transcriptional regulatory network for blooming, to reveal the regulation basis of early and late blooming, and explore the application of genetic regulation factors for accelerating the breeding of tree peony varieties with extra-early, late, and long-lasting flowers.

MicroRNAs (miRNA) are short (21-24 nucleotide) RNAs originated from noncoding RNAs root in the expression of miRNA genes (MIR genes) (Basso et al., 2019). MiRNAs have become crucial modulator of gene expression, primarily by means of the cleavage/inhibit of target genes translation during or after transcription (Xie et al., 2020). MiRNAs regulate almost all the crucial biological processes of the plants' life cycle, such as growth and development (Li et al., 2019), flowering (Spanudakis and Jackson, 2014), ripening (Guo et al., 2018), postharvest senescence (Pei et al., 2013; Chen et al., 2020), and plant-environment interactions (Basso et al., 2019). MiRNAs can rapidly reprogram the expression patterns of downstream genes that strictly regulate agronomic trait, for instance, florescence (Wang et al., 2009). Previous studies have emphasized the significance of miRNAs involved in floral transition and flowering regulation (Waheed and Zeng, 2020). Studies of *miR156*, *miR172*, *miR390*, *miR159*, *miR169*, and *miR399* have shown that they are key factors affecting flowering time (Waheed and Zeng, 2020).

Integrated analysis of the miRNAome, transcriptome, and degradome analyses can enhance the understanding of the genome wide co-expression patterns of miRNA-mRNA pairs and links the biological interactions of miRNA-target modules

(Liu et al., 2020a; Wang et al., 2021). Integrated miRNA, transcriptome, and degradome-seq analysis of miRNA-mRNA involved in flowering of pepper (Shu et al., 2021), floral development and abscission of yellow lupine (Glazinska et al., 2019), female sterility of pomegranate (Chen et al., 2020), male sterility of rice (Sun et al., 2021), flower development across capsicum species (Lopez-Ortiz et al., 2021), stamen development in moso bamboo (Cheng et al., 2019), flowering induction in *Lilium×formolongi* (Zhang et al., 2021), floral transition in *Magnolia×soulangeana* 'Changchun' (Sun et al., 2021) has provided evidence demonstrated of regulatory pathways and gene networks of miRNAs and their targets associated with flowering. Moreover, studies have shown that *miR319-TCP*, *miR156-SPL*, *miR159-MYB*, *miR172-AP2* and *miR399-PHO2* nodes play important roles in floral transition (Waheed and Zeng, 2020).

Prediction of miRNAs in tree peony have been reported, including miRNAs in response to copper stress (Jin et al., 2015), involved in bud dormancy release (Zhang et al., 2018), seed fatty acid synthesis (Yin et al., 2018), flower spot formation (Zhao et al., 2019), petal variegation (Shi et al., 2019), flower development (Han et al., 2020), and brassinolide treatment on flowering (Zhang et al., 2022). In herbaceous peony, miRNAs involved in response to stress from high temperature (Hao et al., 2017) and *Botrytis cinerea* infection (Zhao et al., 2015), as well as lateral branch formation (Liu et al., 2020) have been predicted. In addition, *miR156e-3p* of herbaceous peony has been proved to enhance anthocyanin accumulation in lateral branches of transgenic *Arabidopsis thaliana* (Zhao et al., 2017). However, currently, miRNAome analysis during reproductive growth, has not yet been applied to the elucidation of miRNA-mRNA module regulatory networks specific to the trait of flowering among tree peony varieties with contrasting flowering times.

In this study, we identified and determined the critical miRNA and their MIR genes using miRNAome analysis combined with transcriptome, degradome and qRT-PCR verification during flower development stages in three tree peony varieties with different flowering times. This research might enlighten the composition of post-transcriptional networks in tree peony floral florescence, development and abscission, and facilitate innovations for breeding programs aiming to prolong the flowering period.

2 Materials and methods

2.1 Materials preparation

Early-flowering variety *Paeonia ostii* 'Fengdan' (FD), an early-flowering mutant line of *Paeonia ostii* 'Fengdan' (MU), and late-flowering variety *Paeonia suffruticosa* 'Lianhe' (LH) were selected as the experiment materials. FD, MU, and LH used in this study were 13-years-old plants with single and white flowers. Fresh petals were collected at 9:00-10:00 am on different days at blooming stage (BS), initial flowering stage (IF), full bloom stage (FB), and decay stage (DE), respectively. Abbreviations of sample and library names were presented in Table 1. Three biological replicates are different flowers on different stems of the same plant at each developmental stage for each variety respectively. For each flower, all the petals were sampled and pooled prior to freezing by liquid nitrogen. The sampled petals were stored in the freezer (-80°C) for RNA extraction.

2.2 RNA isolation, library construction and sequencing

In total, 36 libraries were prepared for miRNAome and transcriptome analysis separately (Table S1). Approximately 200 mg petals were used for total RNA extraction. Total RNA integrity was initially assessed by denaturing agarose gel electrophoresis, then confirmed by Bioanalyzer 2100 (Agilent, CA, USA). Total RNA amount and purity quantification were performed on NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA). Total RNA with a concentration>100 ng/μL, RNA integrity number

TABLE 1 Summary of the abbreviations used for sample names and library names.

Type of Terms	Abbreviations	Description
Genotype	*MU*	*Mutant of Paeonia ostii 'Fengdan', opening earlier than Paeonia ostii 'Fengdan'*
	FD	*Paeonia ostii 'Fengdan', early flowering cultivar*
	LH	*Paeonia suffruticosa 'Lianhe', late flowering cultivar*
Developmental stage	*CE*	*Color Exposure Stage*
	BS	*Blooming Stage*
	IF	*Initial Flowering Stage*
	HO	*Half Opening Stage*
	FB	*Full Blooming Stage*
	ID	*Initial Decay Stage*
	DE	*Decay Stage*

>7.0, OD260/280>1.8 and amount>50 μg were used for library construction. Petals from four flower development stages of each variety were mixed prior to the degradome library construction. That is, a total of 3 degradome libraries were constructed for degradome sequencing. Kits and reagents used for RNA isolation, purification, quantification, and libraries construction are listed in Table S2. Transcriptome sequencing was performed by the 2×150 bp paired-end sequencing (PE150) on Illumina Novaseq™ 6000. miRNAome and degradome sequencing were performed by the 1×50 bp single-end sequencing on Illumina Hiseq2500. Libraries construction and sequencing were performed at LC-BIO (Hangzhou, China) according to the vendor's recommended protocol.

2.3 Data processing of miRNAome, transcriptome, and degradome

2.3.1 Data processing of miRNAome

Raw data filtering was processed using ACGT101-miR (LC Sciences, Houston, Texas, USA). Sequences with a length of 18-26 nt were mapped to the genome of tree peony (https://ftp.cngb.org/pub/CNSA/data1/CNP0000281/CNS0044072/CNA0002540/) (Lv et al., 2020) and miRBase 22.0 (http://www.mirbase.org/). Sequences mapped to miRBase 22.0 were characterized as known miRNA. Sequences unmapped to miRBase 22.0 and matched to tree peony genome were identified as candidate novel miRNAs. Secondary structure prediction of miRNA was performed by RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). Sequences possess stem-loop structure and satisfy the requirement of miRNA prediction (Axtell and Meyers, 2018) were considered as real miRNAs of tree peony. Differentially expressed miRNA (DEM) analysis based on normalized (Li et al., 2016) deep-sequencing counts was performed by ANOVA with the criterion of $P \leq 0.05$. DEM target genes prediction was performed by PsRobot 1.2 to characterize the miRNA binding sites.

2.3.2 Data processing of transcriptome

Raw data filtering was conducted by FASTP (https://github.com/OpenGene/fastp) to remove reads containing adaptor contamination, low quality bases and undetermined bases. FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 0.11.9) and FASTP was used for sequences quality verification. HISAT2 (https://ccb.jhu.edu/software/hisat2) was used for the reads mapping to the reference genome of tree peony (Lv et al., 2020). Reads were assembled by StringTie (http://ccb.jhu.edu/software/stringtie/) (Version: stringtie-1.3.4d.Linux_x86_64). Data merging was conducted using gffcompare http://ccb.jhu.edu/software/stringtie/gffcompare.shtml). Transcripts' expression levels were estimated by StringTie according to FPKM method (FPKM=[total_exon_fragments/mapped_reads (millions)×exon_length(kB)]). Identification of DEGs with fold change>2 or <0.5 and P value<0.05 were performed using edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html). GO and KEGG pathway investigation were performed by DAVID (https://david.ncifcrf.gov/). TFs were investigated by iTALK (v1.2) software596. WGCNA was performed according to Langfelder and Horvath (2008).

2.3.3 Data processing of degradome

Degradome data processing was performed by program ACGT10-DEG (LC Sciences, Houston, Texas, USA) using software package CleaveLand4 according the following command: degradome and transcriptome data alignment and generate a degradome density file, miRNAs and transcriptome alignment to parse miRNA-mRNA potential target site, cross-referencing to the degradome data to demonstrate the slicing site.

2.4 miRNAs and targets expression assay

In total, eight miRNA-target pairs associated with floral florescence, development and senescence were randomly selected for quantitative real-time PCR (qRT-PCR) analysis. Fresh petals of FD, MU and LH were collected at developmental stages (BS, IF, FB, DE) respectively. For each developmental stage, petals from three different flowers on different stems of the same plant were sampled individually. Total RNA extraction, miRNA extraction, cDNA synthesis for mRNA and miRNA were performed according to the instructions of manufacturers. Kits information for total RNA/miRNA extraction and cDNA synthesis are shown in Table S2. The *EF1-α* and *U6* were used as the reference for mRNA and tailing reaction miRNA analysis separately. Primer sequences for qRT-PCR assay are listed in Table S3. qRT-PCR analysis for miRNA and targets were both performed on a BIORAD CFX96 machine. Three technical replicates per reaction were conducted in the qRT-PCR analyses to ensure statistical validity. The relative quantity was calculated on the basis of $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

3 Results

3.1 Morphological comparison of flowering time in three tree peony varieties

The date when 80% of flowers reached color exposure (CE), blooming stage (BS), initial flowering (IF), half opening (HO), full blooming (FB), initial decay (ID), and decay (DE) stages were investigated in FD, MU, and LH in 2020 and 2021 respectively (Figure 1A). The flower duration time (date from CE to DE) of 80% flowers of FD, MU, LH were also investigated.

FD is an early flower variety, while LH is a late flowering variety. MU was a mutant of FD, which blossomed seven days earlier than FD in 2020, and nine days earlier than FD in 2021. Flowering time of FD was 16 days earlier than LH in 2020, and 10 days earlier than LH in 2021. Flowering time of MU was 23 and 19 days earlier than LH in 2020 and 2021, respectively. In addition, floral florescence per plant was 16-17 days for FD, 13-15 days for MU, and 11-12 days for LH, which demonstrated that FD possesses the longest blooming time (Figures 1B, C).

3.2 MiRNAome analysis

3.2.1 Expression miRNAs of tree peony revealed by miRNAome

Petal samples were collected at BS, IF, FB, and DE stages from three replicates of each from the three varieties, respectively. An overview of miRNAome sequencing data is presented in Table S4. Over 650.34 million raw reads were produced, therein to, 123.55 million filtered miRNA reads were acquired. A total of 252, 297, and 263 miRNAs were determined to be expressed in common across the four developmental stages (BS, IF, FB, and DE) in FD, MU, and LH, respectively (Figures S1A–C). There are altogether 290, 311, 252, and 226 miRNAs expressed in common across the varieties (FD, MU, and LH) at developmental stages BS, IF, FB, and DE, respectively (Figures S1E–H). Finally, a total of 164 miRNA were expressed in common across all four developmental stages and three varieties (Figures S1D, I).

3.2.2 Known and predicted miRNAs of tree peony revealed by sequencing the miRNAome

In total, 2,444 pre-miRNA and 2,617 unique miRNAs were identified by miRNAome analysis (Table 2). The miRNAs were classified into known and predicted groups (Table S5). Four types of known miRNAs were included in group1 (gp1), group2a (gp2a), group1 (gp2b), and group1 (gp3), while only predicted miRNAs were included in group4 (gp4). In 'gp1' were placed the reads that mapped to specific miRNAs/pre-miRNAs in miRbase for which the pre-miRNAs mapped to the tree peony genome and to the ESTs. In 'gp2a' were placed reads that mapped to selected miRNAs/pre-miRNAs in miRbase, but for which the pre-miRNAs did not map to the tree peony genome. However, the reads (the miRNAs of the pre-miRNAs) in this group did

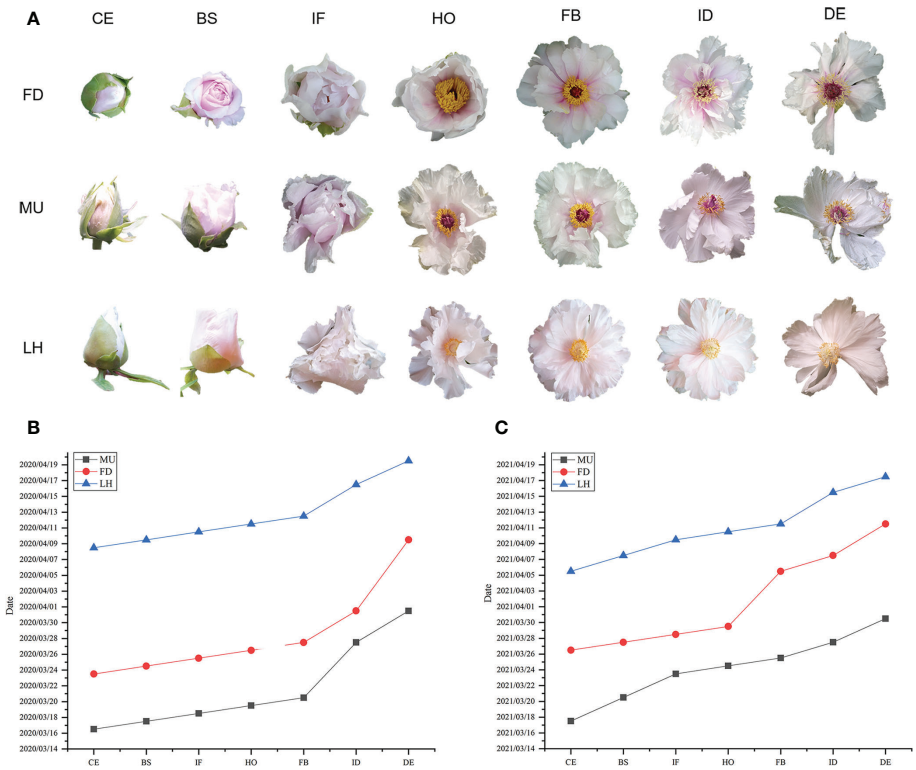


FIGURE 1
Phenotype and flowering time investigation of FD, MU, and LH. **(A)** Flower phenotype at different developmental stages of FD, MU, and LH. **(B)** The date when 80% flowers reached different development stages in FD, MU, and LH investigated in 2020. **(C)** The date when 80% flowers reached different development stages in FD, MU, and LH investigated in 2021.

TABLE 2 Summary of known and predicted miRNA.

	groups	pre-miRNA	total	unique miRNA	total
known miRNA	group1	10	735	15	796
	group2a	75		106	
	group2b	611		632	
	group3	39		43	
predicted miRNA	group4	1709	1709	1812	1821
Total		2444		2617	

map to the tree peony genome. Also, the extended sequences from the corresponding loci in the tree peony genome which could form hairpins. In 'gp2b' were placed reads that mapped to miRNAs/pre-miRNAs of selected species in miRbase. The pre-miRNAs did not map to the tree peony genome, however, the reads (the miRNAs from the pre-miRNAs) did map to the tree peony genome. Also, the extended sequences at the genome loci could not form hairpins in this case. In 'gp3' were placed reads that mapped to the selected miRNAs/pre-miRNAs in miRbase but for which the pre-miRNAs do not map to the tree peony genome, and also the reads did not map to the tree peony genome. In 'gp4' were placed reads that did not map to the selected pre-miRNAs in miRbase. However, these reads did map to the tree peony genome and the extended sequences from the genome loci which could form hairpins. In summary, for pre-miRNA, a total of 735 known miRNAs including gp1 (10), gp2a (75), gp2b (611), gp3 (39) were identified, and a total of 1,709 predicted miRNAs which only include gp4 were characterized. For unique miRNA, a total of 796 known miRNAs including gp1 (15), gp2a (106), gp2b (632), gp3 (43) were defined, and a total of 1,812 predicted miRNAs which only include gp4 were detected (Table 2).

3.2.3 Differentially expressed miRNAs among stages in floral florescence development

To identify DEMs engaged in floral florescence and senescence, significantly differential expressed miRNAs ($P<0.01$, $P<0.05$, $P<0.1$) were analyzed across four developmental stages and three varieties. In total, 146, 313, and 201 miRNAs showed signifcant differential expression ($P<0.05$) across four developmental stages in FD, MU, and LH, respectively. A total of 253, 227, 285, 270 miRNAs revealed signifcant differential expression ($P<0.05$) across three varieties (FD, MU, LH) at stage of BS, IF, FB, and DE separately. The numbers of differentially expressed up- and down-regulated miRNAs in different groups are shown in Figure S2. Up-regulated DEMs refer to the miRNAs having signifcantly higher expression, while down-regulated DEMs refer to the miRNAs which present prominently lower expression.

A wide variety of miRNAs showed differential expression specific to genotype and developmental stage. However, no DEMs were found to be co-expressed across the four developmental stages even in a single variety, which meant the analysis of co-expressed DEMs between varieties based on intersection of DEMs across developmental stages could not be performed. In fact, higher numbers of differentially expressed miRNAs were presented in DEvsBS, DEvsFB, and DEvsIF when comparing DEMs across developmental stages irrespective of variety. MU constantly had a higher number of DEMs commonly expressed when compared with FD and LH (Figures 2A–C). When comparing across varieties (FD, MU and LH) at specific developmental stages, 18, 16, 24 and 9 intersecting DEMs were obtained at developmental stages BS, IF, FB and DE respectively (Figures 3A–D). Meanwhile, there were more DEMs exclusively expressed in the FB stage than BS, IF and DE stages in all three of the varieties (Figures 3A–D). Interestingly however, three DEMs were identified that simultaneously expressed in the BS, IF, and FB stages. Finally, the expression of 6, 8, 16, and 8 stage-specific DEMs, at BS, IF, FB and DE stages, respectively, were not affected by variety (Figure 3E).

3.3 Transcriptome analysis

3.3.1 Expression genes revealed by transcriptome sequencing

Overview of raw reads, valid reads, Q30 and GC content data are shown in Table S6. Over 227 million valid reads, after filtering of 245 million raw reads, were obtained. In total, 500,378 contigs, 35,687 genes (G), 35,687 unique Transcripts (T) were identified separately. Of these, 29,271 GO annotated genes and 10,556 KEGG annotated genes were obtained separately. A total of 15,099, 14,906, and 14,544 intersecting annotated genes were detected across developmental stages (BS, IF, FB, and DE) in FD, MU, and LH varieties respectively (Figures S3A–C). A total of 14,672, 14,844, 14,841, and 14,779 intersecting annotated genes were detected across all varieties (FD, MU, and LH) at all developmental stages BS, IF, FB, and DE respectively (Figures S4A–D). Finally, a total of 13,203 intersecting annotated genes were detected across four developmental stages and three varieties (Figure S3D, Figure S4E).

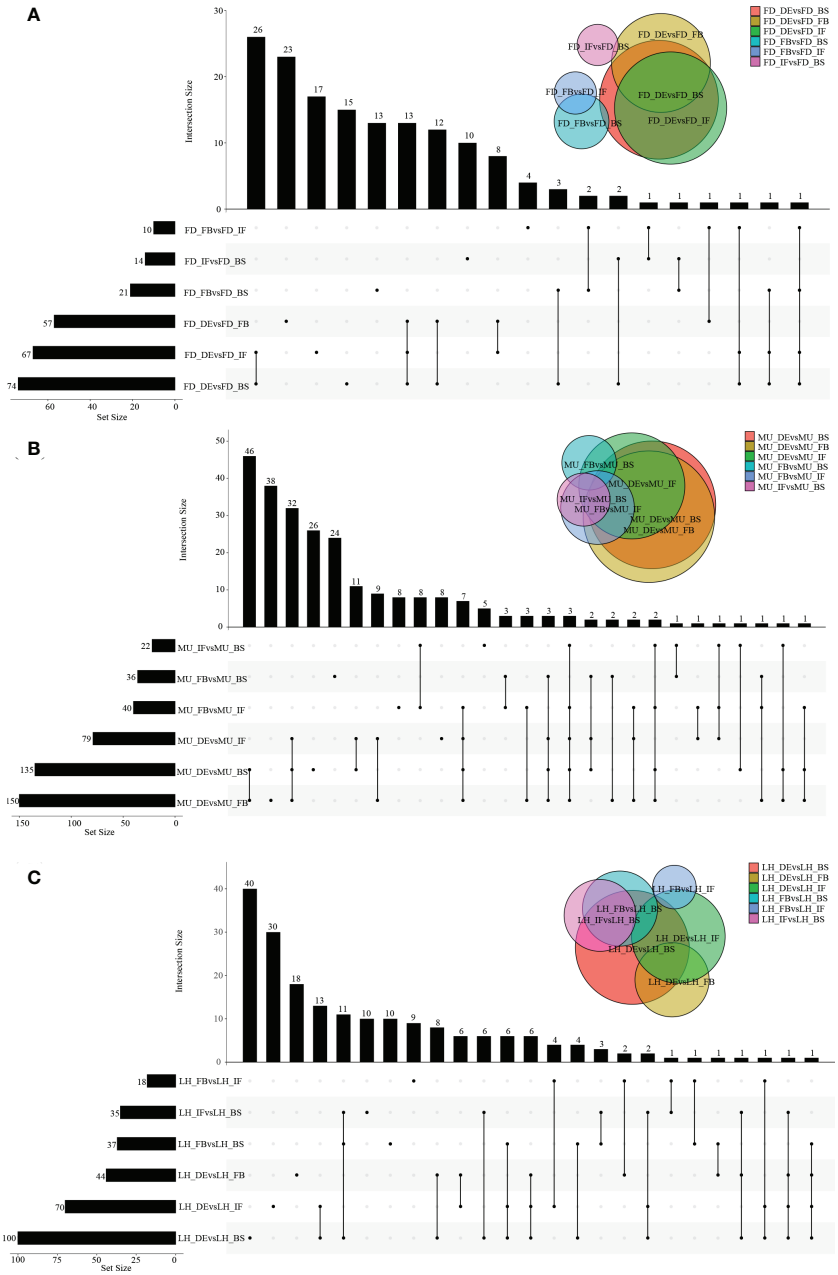


FIGURE 2
The distribution of DEMs across flower developmental stages in FD, MU, and LH. **(A)** The distribution of DEMs across developmental stages (BS, IF, FB, DE) in FD libraries. **(B)** The distribution of DEMs across developmental stages (BS, IF, FB, DE) in MU libraries. **(C)** The distribution of DEMs across developmental stages (BS, IF, FB, DE) in LH libraries.

3.3.2 Differentially expressed genes identified in floral florescence development stages

To identify differentially expressed genes (DEGs) with floral florescence patterns, gene expression was compared across developmental stages in each of the varieties (Figures 4A–C). Numbers of differentially expressed DEGs including up-regulated and down-regulated DEGs in different groups are shown in Figure S5. A total of 12, 36, and 69 genes showed significant differential expression among flower developmental stages (BS, IF, FB, DE) in FD, MU and LH varieties, respectively (Figure 4D). Interestingly, after determining the intersection of differentially expressed genes in the three varieties, we identified only one co-expressed DEG (psu.G.00014095), in varieties FD and LH, that was expressed across all four development stages, which indicates that expression

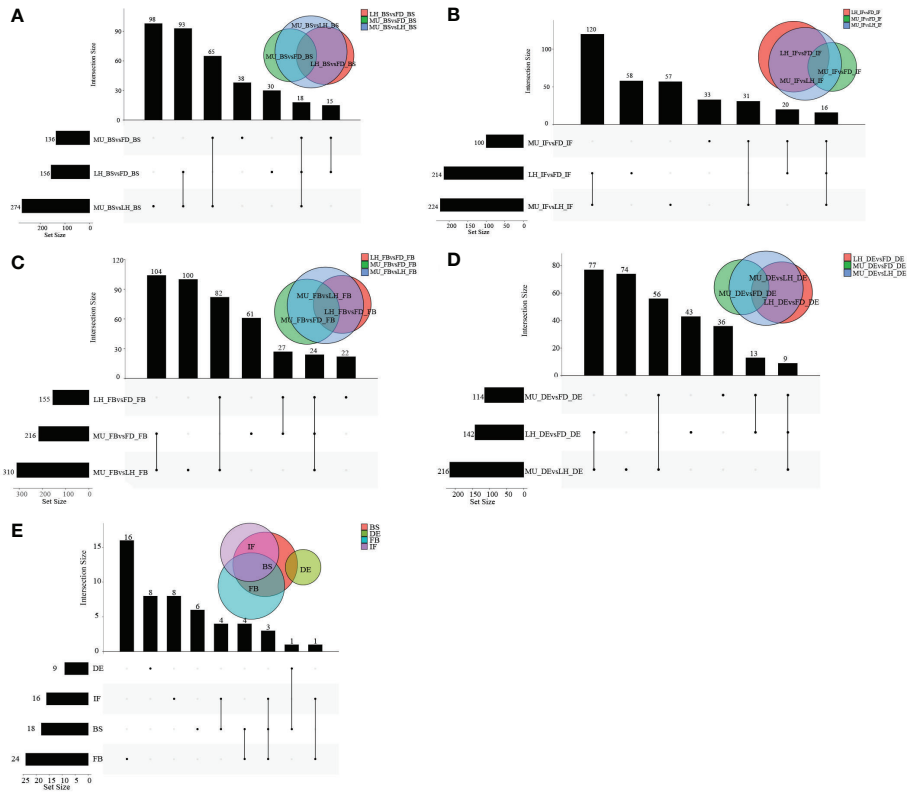


FIGURE 3
The distribution of DEMs across varieties at BS, IF, FB, and DE stages. **(A)** The distribution of DEMs across tree peony varieties (FD, MU and LH) at BS libraries. **(B)** The distribution of DEMs across tree peony varieties (FD, MU and LH) at IF libraries. **(C)** The distribution of DEMs across tree peony varieties (FD, MU and LH) at FB libraries. **(D)** The distribution of DEMs across tree peony varieties (FD, MU and LH) at DE libraries. **(E)** The distribution of intersection DEMs across tree peony varieties (FD, MU and LH) and across four flower developmental stages (BS, IF, FB, DE).

of DEG psu.G.00014095 was dependent both on developmental stage and genotype. The function of this gene was annotated as xyloglucan endotransglycosylase/hydrolase precursor *XTH-25*, which may regulate the floral florescence, development and senescence of tree peony. In addition, genes differentially expressed DEGs at different developmental stages across all varieties were identified to discover genes that may regulate flowering (Figures 5A–D). In total, 437, 232, 446, and 343 co-expressed genes showed significant differential expression among stages BS, IF, FB, and DE, respectively, across varieties FD, MU and LH (Figure 5E). Examining intersections of the DEGs further revealed 16 DEGs that simultaneously expressed at the four developmental stages and in the three varieties (Figure 5E). Besides, 257, 104, 257, and 240 specific DEGs showed significant differential expression across in stages in FB, IF, FB, and DE, respectively, in all the three varieties (Figure 5E).

3.3.3 Gene ontology and kyoto encyclopedia of genes and genomes analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) based functional enrichment analysis of DEGs were conducted to uncover the biological roles in governing floral florescence, development and abscission across early- and late-flowering genotypes in tree peony.

Some GO terms (chloroplast, chloroplast envelope, plasmodesma, and molecular function) for DEGs were common across both varieties and developmental stages, while some were specific to either the variety (e.g., defense response) or to developmental stages (flavonoid biosynthetic process, flavonoid glucuronidation, quercetin 7-o-glucosyltransferase activity, quercetin 3-o-glucosyltransferase activity, chloroplast stroma, kinase activity, response to cold, chloroplast thylakoid membrane, thylakoid, chloroplast thylakoid). It was worth noting that GO terms like abscisic acid-activated signaling pathway and protein serine/threonine kinase activity processes were significantly enriched across developmental stages in FD and MU which were early-flowering genotype, while biological process, plant type cell wall, vacuole, hydrolase activity, hydrolyzing o-glucosyl compounds, golgi apparatus, transferase activity, transferring glycosyl groups, fatty acid biosynthetic process, chloroplast inner membrane processes were only enriched across developmental stages in LH which

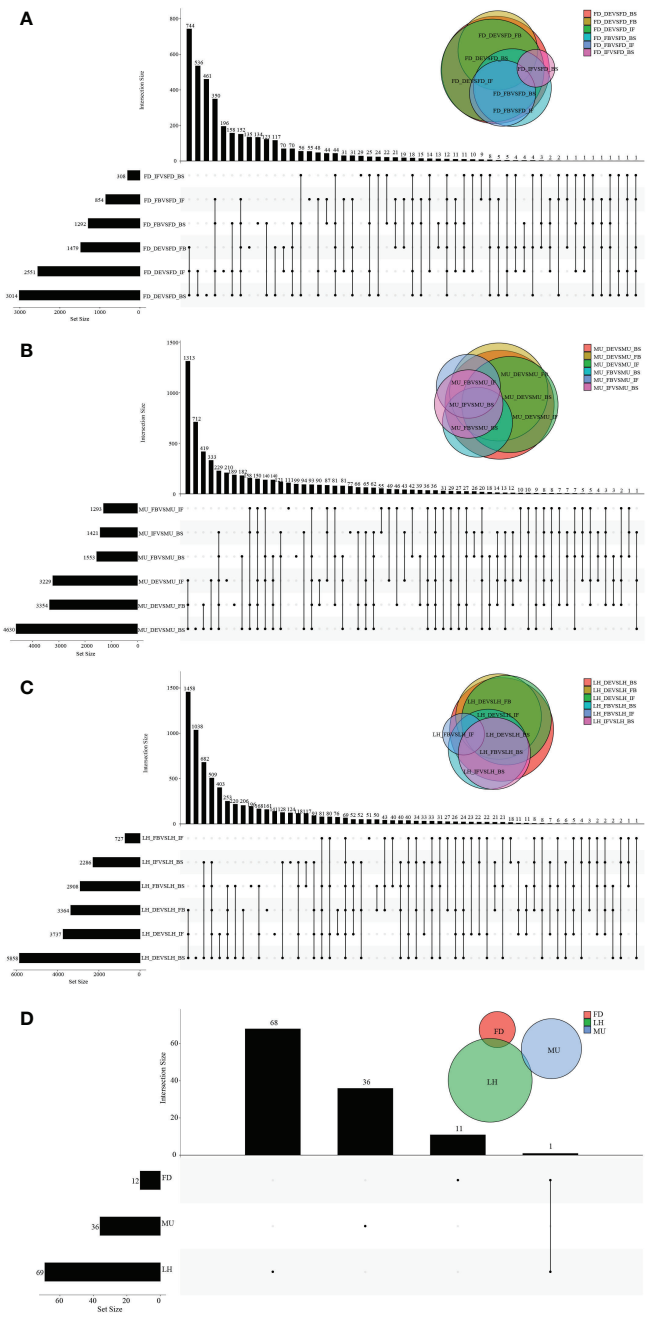


FIGURE 4
The distribution of DEGs across flower developmental stages in FD, MU, and LH. **(A)** The distribution of DEGs across flower developmental stages (BS, IF, FB, DE) in FD libraries. **(B)** The distribution of DEGs across flower developmental stages (BS, IF, FB, DE) in MU libraries. **(C)** The distribution of DEGs across flower developmental stages (BS, IF, FB, DE) in LH libraries. **(D)** The distribution of intersection DEGs across four flower developmental stages (BS, IF, FB, DE) and across the three tree peony varieties (LH, MU and LH).

was late-flowering genotype. The GO terms for response to abscisic acid, wounding, chitin, and protein phosphorylation were only enriched in MU, which was an early flowering mutant genotype of FD (Table 3).

KEGG enrichment analysis of DEGs revealed that plant hormone signal transduction, starch and sucrose metabolism, MAPK signaling pathway-plant, phenylpropanoid biosynthesis, and carotenoid biosynthesis pathway were common across

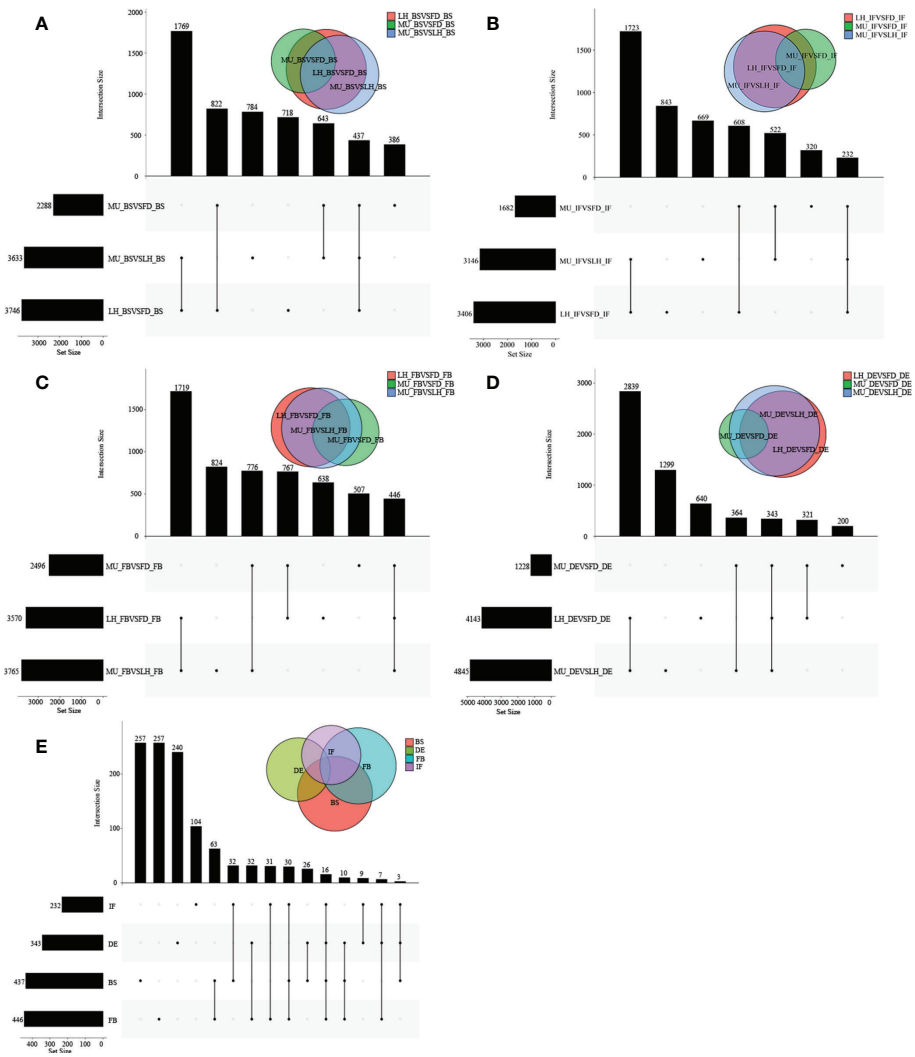


FIGURE 5
The distribution of DEGs across varieties at BS, IF, FB, and DE stages. **(A)** The distribution of DEGs across tree peony varieties (LH, MU and LH) at BS libraries. **(B)** The distribution of DEGs across tree peony varieties (LH, MU and LH) at IF libraries. **(C)** The distribution of DEGs across tree peony varieties (LH, MU and LH) at FB libraries. **(D)** The distribution of DEGs across tree peony varieties (LH, MU and LH) at DE libraries. **(E)** The distribution of intersection DEGs across the three tree peony varieties (LH, MU and LH) and across the four flower developmental stages (BS, IF, FB, DE).

developmental stages in both early- and late-flowering varieties, which suggests a possible role in the floral florescence, development and abscission in tree peony. Here, it was interesting that KEGG pathways like cyanoamino acid metabolism, galactose metabolism, other glycan degradation, fatty acid elongation, and amino sugar and nucleotide sugar metabolism were specific to varieties, while some other pathways were specific to developmental stages (for example, porphyrin and chlorophyll metabolism and fructose and mannose metabolism). Thus, these KEGG pathways might possibly have roles in variety-specific or developmental stage-specific responses. It is worth mentioning that pathways for stilbenoid, diarylheptanoid and gingerol biosynthesis, glycosphingolipid biosynthesis-ganglio series, arachidonic acid metabolism, terpenoid backbone biosynthesis, and zeatin biosynthesis were common to the early-flowering varieties FD and MU, while fructose and mannose metabolism, sphingolipid metabolism, fatty acid biosynthesis, inositol phosphate metabolism, pyruvate metabolism, phosphatidylinositol signaling system were specific to the late-flowering genotype LH, which may contribute to the late-flowering phenotype (Table 4).

3.3.4 Identification of DEGs encoding transcription factors

Transcription factors (TFs), a kind of DNA-binding proteins which play important roles in transcription, perform a number of function in flowering (Kumar et al., 2012; Shibuya et al., 2015).

TABLE 3 GO enrichment analysis of DEGs across genotypes and developmental stages in tree peony.

GO term	Sample
Chloroplast	A, B, C, D, E, F, G
Chloroplast envelope	A, B, C, D, E, F, G
Plasmodesma	A, B, C, D, E, F, G
Molecular function	A, B, C, D, E, F, G
Defense response	A, B, C, E, F
Flavonoid biosynthetic process	A, B, D, E, F, G
Flavonoid glucuronidation	A, B, D, E, F, G
Quercetin 7-o-glucosyltransferase activity	A, B, D, E, F, G
Quercetin 3-o-glucosyltransferase activity	A, B, D, E, F, G
Chloroplast stroma	B, C, D, E, F, G
Kinase activity	B, C, D, E, F, G
Response to cold	B, C, D, E, F, G
Chloroplast thylakoid membrane	D, E, F, G
Thylakoid	D, E, F, G
Chloroplast thylakoid	D, E, F, G
Abscisic acid-activated signaling pathway	A, B, G
Protein serine/threonine kinase activity	A, B
Extracellular region	A
Plasma membrane	A
Regulation of organ growth	A
Monooxygenase activity	A, G
Protein phosphatase inhibitor activity	A, G
Abscisic acid binding	A, G
Protein serine/threonine phosphatase activity	A, F, G
Response to wounding	B, E, F
Response to abscisic acid	B
Protein phosphorylation	B
Response to chitin	B
Biological process	C, D, F, G
Plant type cell wall	C, D
Vacuole	C, E
Hydrolase activity, hydrolyzing O-glucosyl compounds	C
Golgi apparatus	C
Transferase activity, transferring glycosyl groups	C
Fatty acid biosynthetic process	C
Chloroplast inner membrane	C
Oxidation-reduction process	A, C, F
Anchored component of plasma membrane	A, C
Apoplast	B, C, D, F
Cell wall	B, C
Response to water deprivation	D, E
Response to heat	D
Response to salt stress	D
Response to light stimulus	E
Plastoglobule	E

A: Samples collected from FD across four developmental stages. B: Samples collected from MU across four developmental stages. C: Samples collected from LH across four developmental stages. D: Samples collected from BS stage across three genotypes. E: Samples collected from IF stage across three genotypes. F: Samples collected from FB stage across three genotypes. G: Samples collected from DE stage across three genotypes.

In total, 116 differentially expressed TFs, belonging to 26 TF families including *bHLH*, *C2H2*, *ERF*, *B3*, *MYB-related*, *NAC*, *HD-ZIP*, *bZIP*, *GRAS*, *HSF*, *MYB*, *Trihelix*, etc., were identified across the four flower developmental stages in the tree peony varieties. Among the above identified TFs, the *bHLH* family, *C2H2* family and *ERF* family accounted for the largest proportion, the following are *B3* family, *MYB*-related family protein and *NAC* family (Table S7).

In addition, 1,117 differentially expressed TFs belonging to 53 TF families, covering *bHLH*, *ERF*, *NAC*, *C2H2*, *B3*, *MYB*-related, *Trihelix*, *MYB*, *FAR1*, *GRAS*, *C3H*, and *WRKY*, etc., were identified across the three varieties at BS, IF, FB, and DE stages. Among these TFs, the *bHLH* family accounted for the largest proportion, followed by *ERF*, *NAC*, *C2H2*, *B3*, and *MYB*-related family (Table S8). It was worth mention that *bHLH* family, *C2H2* family, *ERF* family, *B3* family, *MYB*-related family, and *NAC* family showed high dominance both across genotype and across developmental stages, which demonstrated that these TFs might play crucial functions in regulating floral florescence, development and senescence in tree peony.

3.4 Degradome sequencing revealed miRNA-regulated mRNAs

Degradome, used for miRNA and siRNA targets characterization by the 5'-ends of uncapped RNAs (German et al., 2008; Gregory et al., 2008). Petals of the four flower developmental stages (BS, IF, FB and DE) for each genotype (FD, MU and LH) were pooled prior the degradome libraries construction (FD, LH, and MU) and used for degradome analysis. Around 10.9 million raw reads were gained from the degradome libraries. Detailed information of total raw reads, unique raw reads, mappable reads, unique mappable reads, mapped reads, unique mapped reads, number of input transcripts, and number of target transcripts are shown in Table S9. A total of 7,571, 9,250 and 6,457 mRNAs were silenced by miRNAs in FD, MU and LH respectively. The alignment information of miRNAs-mRNA pairs can be found in Table S10. In total, 11,391 targets showed a differential degradation pattern between FD and LH. In addition, 11,928 targets presented a differential degradation characteristic between MU and FD. Moreover 11,605 targets demonstrated a differential degradation characteristic between MU and LH.

3.5 Integrated analysis of floral florescence, development, and senescence dependent miRNA-mRNA modules

DEGs identified across flower developmental stages (BS, IF, FB, DE) in FD, MU and LH were assembled into a unified set.

DEGs identified across varieties (FD, MU, LH) at developmental stages BS, IF, FB and DE were then assembled into another unified set. DEGs from the union sets were then used together for further miRNA-mRNA target pairing confirmation by multi-omics analysis. Subsequently, miRNA-mRNA pairs with opposite regulatory patterns were characterized in terms of the gene-silencing function of miRNAs. With this method, miRNA-mRNA pairs associated with floral florescence, development and senescence were identified.

In total, 32 miRNA targets showed antagonistic regulatory patterns during developmental stages in FD, MU and LH, which might thus be candidates for regulating the floral development and senescence of tree peony (Table S11). Another 191 miRNA targets were differentially expressed across tree peony varieties and developmental stages, which suggests important roles for those in floral florescence regulation (i.e., the timing of flower opening and senescence) (Table S11). Of these, ten miRNA targets showed significantly different expression patterns both across development and across varieties, which suggests possible dual roles in floral florescence regulation in tree peony. Expression patterns of these floral florescence, development and senescence dependent miRNA targets are presented in Figures 6A–E. GO (Figure 7A) and KEGG (Figure 7B) pathway analysis revealed that these floral florescence-, development- and senescence-dependent miRNA targets were enriched in pathways like plant hormone signal transduction, indole alkaloid biosynthesis, arachidonic acid metabolism, folate biosynthesis, fatty acid elongation, MAPK signaling pathway, etc.

The miRNA-guided floral florescence, development and senescence regulatory networks were complicated, which imply that one specific miRNA might be able to adjust and control many mRNAs, and one specific mRNA also might be targeted by divers miRNAs simultaneously (Liu et al., 2020a). Multiple-to-multiple inter-associations between miRNAs and their target genes which enriched in KEGG pathways and encode TFs were constructed by Cytoscape (Figure 8). The multiple-to-multiple miRNA-mRNA-TF modules were enriched in pyruvate metabolism, carbon fixation in photosynthetic organisms, pentose and glucuronate interconversions, sesquiterpenoid and triterpenoid biosynthesis, aminoacyl-tRNA biosynthesis (Figure 8A). The result also displayed that the miRNA-mRNA-TF modules mainly consisted of *MYB*-related, *bHLH*, *Trihelix*, *NAC*, *GRAS* and *HD-ZIP* TF families, demonstrating their potential functions in tree peony floral florescence, development and senescence (Figure 8B).

3.6 Weighted gene co-expression network analysis reveals candidate hub genes

In order to uncover the regulation mechanism of flowering time based on transcriptome data across varieties and flower development stages, weighted gene co-expression network analysis (WGCNA) was performed to detect co-expressed genes to disclose the hub gene which might regulate floral florescence, development and senescence. WGCNA analysis resulted in 43 distinct co-expressed gene modules were exhibited by distinctive colors and shown by a heatmap (Figure 9A). Each heatmap represented an expression cluster, which straightly elucidated the relationship between the clusters of three tree peony varieties and four development stages (Figure 9B). Then, correlation analysis was performed between modules and samples to find modules with the highest correlation. The candidate hub genes were confirmed by taking the intersection of gene in modules with the highest correlation and the genes used for integrated analysis of miRNA-mRNA-TF. Furthermore, the intersected genes were selected for the hub genes network construction (Figures 9C, D). Finally, hub genes *psu.G.00014449*, *psu.G.00003047* and *psu.G.00009129* in LH, *psu.G.00032165* and *psu.G.00007421* in MU may play crucial roles in the floral florescence, development and senescence in tree peony. Regrettably, none intersected hub genes were detected in FD since the lower gene connectivity.

3.7 qRT-PCR analysis for the DEMs and DEGs verification

A total of eight DEM and DEG pairs were selected for the qRT-PCR analysis to verify the expression pattern of miRNA and mRNA data obtained from miRNAome, transcriptome, and degradome (Figure 10). The selected miRNA-target pairs were presented as follows: *seu-MIR11025-p5* and *psu.T.00024044*, *miR166-5p* and *psu.T.00024044*, *PC-5p-564_43386* and *psu.T.00034433*, mtr-*miR396b-5p* and *psu.T.00010381*, *PC-3p-602268_25_S* and *psu.T.00020538*, *PC-5p-429002_51_S* and *psu.T.00018467*, *mtr-MIR2592bj-p3* and *psu.T.00015108*, *PC-5p-143784_277_S* and *psu.T.00016751*. Target t-plots (Figure S6) show the cleavage sites of target genes silenced by miRNAs during developmental stages. In general, the expression levels of DEM-target pairs were consistent with miRNA-guided mRNA cleavage signatures validated by degradome.

As expected, majority of the expression patterns of the examined miRNA-target pairs were similar to the RNA-seq data, with the exception of transcripts *psu.T.00010381*, *psu.T.00016751* and *mtr-miR396b-5p*, confirming the accuracy and reliability of the sequencing data in general (Figure 10). Our study confirmed an exactly negative correlation for miRNAs and target genes *PC-3p-602268_25* and *psu.T.00020538* expressed at the four developmental stages in FD and LH, target genes *PC-5p-429002_51* and *psu.T.00018467* at all four developmental stages in FD, and target genes *mtr-miR166g-5p* and *psu.T.00024044* at the four developmental stages in MU. While the remaining tested miRNA-target pairs did not always show a negative relationship across the four developmental stages, this is

TABLE 4 KEGG enrichment analysis of DEGs across genotypes and developmental stages in tree peony.

Pathway name	Sample
Starch and sucrose metabolism	A, B, C, D, E, F, G
Carotenoid biosynthesis	A, B, C, D, E, F, G
MAPK signaling pathway-plant	A, B, C, D, E, F, G
Phenylpropanoid biosynthesis	A, B, C, D, E, F, G
Plant hormone signal transduction	A, B, C, D, E, F, G
Cyanoamino acid metabolism	A, B, C, E, F, G
Galactose metabolism	A, B, C, D, E, F
Other glycan degradation	A, B, C, D, E, F
Fatty acid elongation	A, B, C, D, E
Amino sugar and nucleotide sugar metabolism	A, B, C
Porphyrin and chlorophyll metabolism	A, B, D, E, F, G
Fructose and mannose metabolism	C, D, E, F, G
Stilbenoid, diarylheptanoid and gingerol biosynthesis	A, B, D, E, F
Glycosphingolipid biosynthesis-ganglio series	A, B, D
Arachidonic acid metabolism	A, B
Terpenoid backbone biosynthesis	A, B
Zeatin biosynthesis	A, B
Cutin, suberine and wax biosynthesis	A, D, E, F
Photosynthesis-antenna proteins	A, E, F, G
Phenylalanine metabolism	A, G
Prodigiosin biosynthesis	B, F, G
Anthocyanin biosynthesis	B
Fructose and mannose metabolism	C, D, E, F, G
Sphingolipid metabolism	C, D, F
Fatty acid biosynthesis	C, F, G
Inositol phosphate metabolism	C, D
Pyruvate metabolism	C, G
Phosphatidylinositol signaling system	C
Glycosaminoglycan degradation	A, C, D
Plant-pathogen interaction	B, C, E, F, G
Circadian rhythm-plant	B, C, D
Thiamine metabolism	D, E, F.
Glycerophospholipid metabolism	D, E
Ubiquinone and other terpenoid–quinone biosynthesis	D, F
Flavonoid biosynthesis	E
Linoleic acid metabolism	E
Indole alkaloid biosynthesis	E, G
Other types of O–glycan biosynthesis	F
Ascorbate and aldarate metabolism	G
Protein export	G
Riboflavin metabolism	G
Lysine biosynthesis	G
Valine, leucine and isoleucine biosynthesis	G

A: Samples collected from FD across four developmental stages. B: Samples collected from MU across four developmental stages. C: Samples collected from LH across four developmental stages. D: Samples collected from BS stage across three genotypes. E: Samples collected from IF stage across three genotypes. F: Samples collected from FB stage across three genotypes. G: Samples collected from DE stage across three genotypes.

consistent with previous research (Liu et al., 2017; Zuluaga et al., 2017).

Previous research revealed that miRNA and their target genes were not always presented a specific one-to-one regulatory relationship, which is due to that a specific miRNA can regulate several target mRNAs simultaneously and a specific mRNA also could be targeted by multiple miRNAs (Liu et al., 2020a; Liu et al., 2020b). Thus, the expression pattern of miRNA and their target genes does not emerge a negative correlation all the time (Liu et al., 2017; Zuluaga et al., 2017; Zuluaga et al., 2018). Our study showed that *psu.T.00024044* could be targeted by *mtr-miR166g-5p* and *seu-MIR11025-p5_2ss4CA17CA* simultaneously which was consistent with data elucidated previously. Furthermore, changes in expression levels of miRNA and mRNA could have derived from the differences arising from the biological replicates (Pradervand et al., 2009; Xu et al., 2018; Zuluaga et al., 2018; Liu et al., 2020b). Correlation analysis of miRNA-target pairs expression profiles between sequencing and qRT-PCR are shown in Figure S7.

4 Discussion

Florescence plays a crucial role in ornamental value of tree peony. However, current understanding of the regulatory mechanism underlying florescence in tree peony is still far beyond understanding. Here, we reported the combined analysis of the miRNAome, transcriptome and degradome to reveal the potential regulatory mechanism of florescence, using petals from four developmental stages in three tree peony varieties-FD (an early-flowering variety), LH (a late-flowering variety), and MU (a natural mutant line of FD) that flowers 2-3 days earlier than FD. Previous research on tree peony miRNA discovery sampled bud (Zhang et al., 2018; Han et al., 2020) or root, stem and leaf (Jin et al., 2015) tissues of FD, as well as seeds of high and low-ALA-content varieties *Paeonia rokii* 'Sai gui fei' and 'Jing shen huan fa' (Hao et al., 2017). This is also the first report of miRNA-target identification in varieties with contrasting flowering time phenotypic trait. Knowledge obtained on the miRNA-mRNA modules in tree peony varieties will provide crucial preliminary data for the further researches on floral florescence, development and abscission in tree peony.

4.1 Pathways associated with floral florescence, development, and senescence

Floral-associated pathways usually involved the vernalization, autonomous, ambient temperature, photoperiod,

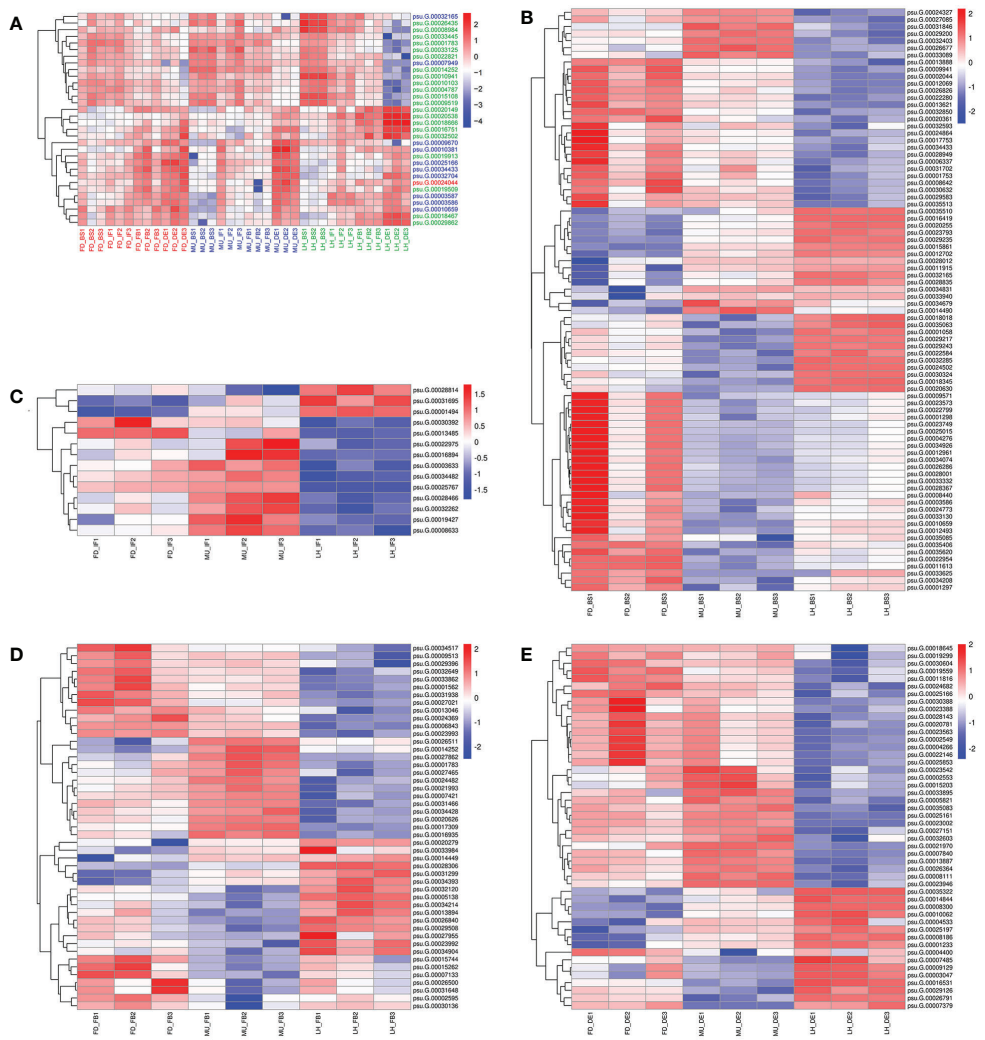


FIGURE 6
Expression pattern of miRNA targets specific to flower developmental stages and varieties by miRNAome, transcriptome and degradome integrated analysis of tree peony. **(A)** Expression pattern of miRNA targets identified across flower developmental stages in FD, MU, and LH. **(B)** Expression pattern of miRNA targets identified across tree peony varieties at flower developmental stage BS. **(C)** Expression pattern of miRNA targets identified across tree peony varieties at flower developmental stage IF. **(D)** Expression pattern of miRNA targets identified across tree peony varieties at flower developmental stage FB. **(E)** Expression pattern of miRNA targets identified across tree peony varieties at flower developmental stage DE.

gibberellic acid, aging and sugar pathways (Han et al., 2021). Genes involved flowering-time regulation in tree peony (Wang et al., 2019) that were identified including genes involved in floral organ and meristem, vernalization pathway, age pathway, GA pathway, autonomous pathway, photoperiod pathway. High intensity light promotes flowering through the photoperiod pathway which requires the cooperation of chloroplast retrograde signals and silencing transcription of *Flowering Locus C* (*FLC*). To response high light induction, transcription factor *PTM* localized at chloroplast envelope suppresses *FLC* transcription. It is also known that an intracellular signaling pathway originated from chloroplasts regulates the flowering transition (Feng et al., 2016; Susila et al., 2016). Furthermore, it was discovered that the expression of chloroplast protein *CEBP* changed during flower development and senescence (Iordachescu et al., 2009). Our results showing that chloroplast and chloroplast envelope genes were expressed in common among both varieties and developmental stages assayed, suggests important roles in floral florescence, development and senescence in tree peony. We also observed a general role for the interaction of turgor pressure and plasmodesmata affecting floral development, perhaps through regulation of plasmodesmata aperture, for which an association with transition to flowering was previously shown (Hernández-Hernández et al., 2019).

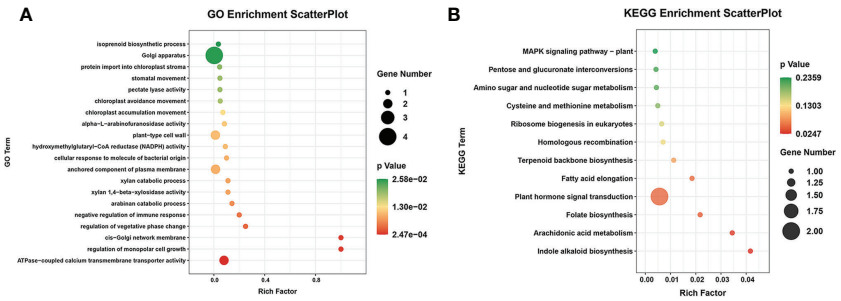


FIGURE 7
GO and KEGG analysis of miRNA targets specific to flower developmental stages and varieties by miRNAome, transcriptome and degradome integrated analysis of tree peony. **(A)** GO analysis of target genes identified by integrated analysis of miRNAome, transcriptome and degradome. **(B)** KEGG pathway analysis of target genes identified by integrated analysis of miRNAome, transcriptome and degradome.

4.2 Plant hormone signal transduction contributing to floral florescence, development, and senescence

Cell division, expansion, differentiation and stress response in organisms were proved to be regulated by hormones in the earlier report (Artur, 2022). ABA signaling presents multiple connections with the photoperiodic pathway (Martignago et al., 2020). Exogenous spraying of ABA result in changes of flowering time, indicating that ABA possibly is an internal factor regulating the floral transition (Conti et al., 2014). *GIGANTEA* (*GI*) is a key flowering gene required for photoperiod perception (Mishra and Panigrahi, 2015). ABA signaling integration through *GI* operates *via* up-regulation of *FT* (Riboni et al., 2016). Overexpression of the chrysanthemum *R2R3-MYB* delays flowering in Arabidopsis (Shan et al., 2012). ABA hypersensitive 1 suppresses frigida-mediated delayed flowering in Arabidopsis (Bezerra et al., 2004). *FD* and *FD-like bZIPs* protein complexes play a significant role in modulating ABA signaling (Martignago et al., 2020). ABA activates an intricate regulatory network of signals including TFs that have contrary effects on florescence (Conti et al., 2014). While the role of ABA in flowering in model plants is emerging, the ABA molecular control of flowering still poorly revealed in tree peony. This research found that many of the most highly differentially expressed genes were relevant to plant hormone signal transduction especially ABA. These results might provide a potential basis for further research on mechanisms parsing of ABA regulating floral florescence, development and senescence in tree peony.

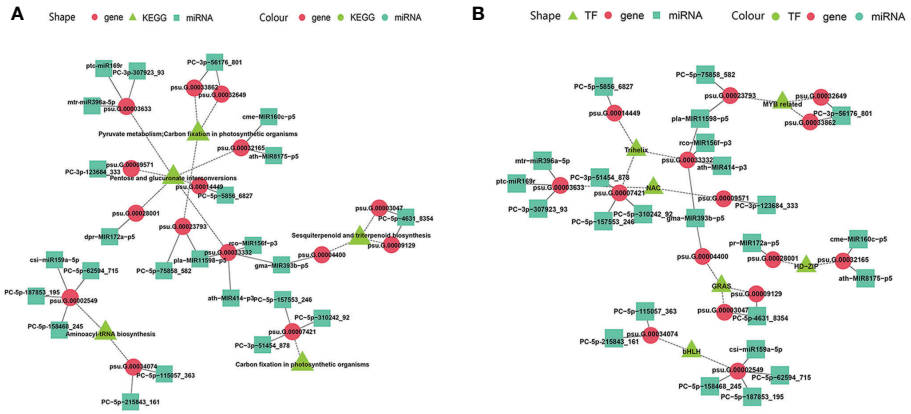


FIGURE 8
MiRNA-mRNA-TF modules regulatory network identified by miRNAome, transcriptome and degradome integrated analysis across flower developmental stages and varieties of tree peony **(A)** Regulatory network mediated by miRNA-mRNA-TF modules identified across flower developmental stages in FD, MU, and LH. **(B)** Regulating network mediated by miRNA-mRNA-TF modules identified across tree peony varieties at flower developmental stages BS, IF, FB, DE.

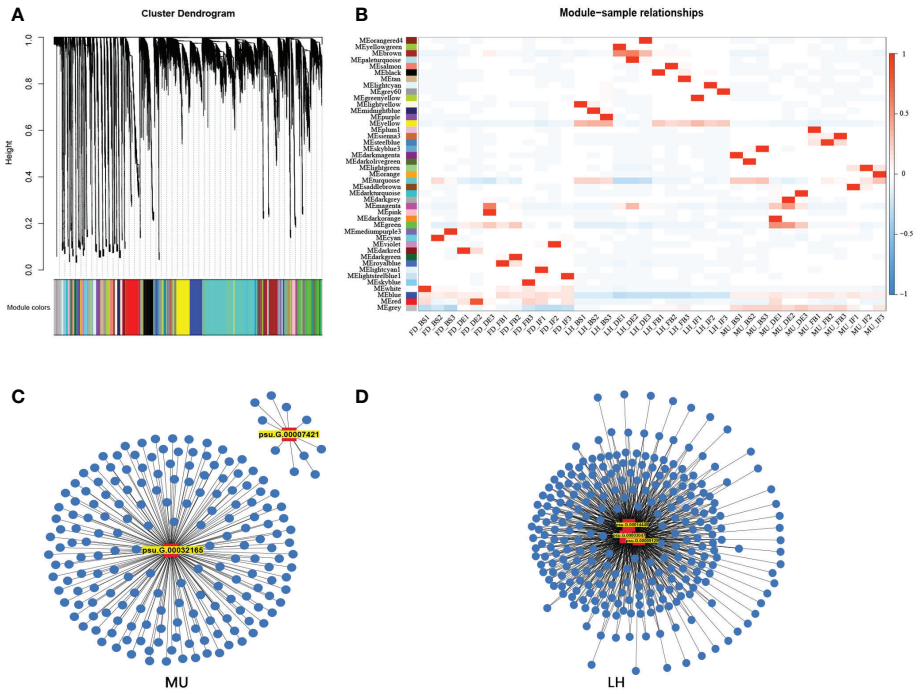


FIGURE 9
WGCNA analysis of differentially expressed genes regulating floral florescence, development and senescence. **(A)** Hierarchical cluster tree showing co-expression modules. **(B)** Module-samples relationships. **(C)** Hub gene regulating floral florescence, development and senescence in LH. **(D)** Hub gene regulating floral florescence, development and senescence in MU.

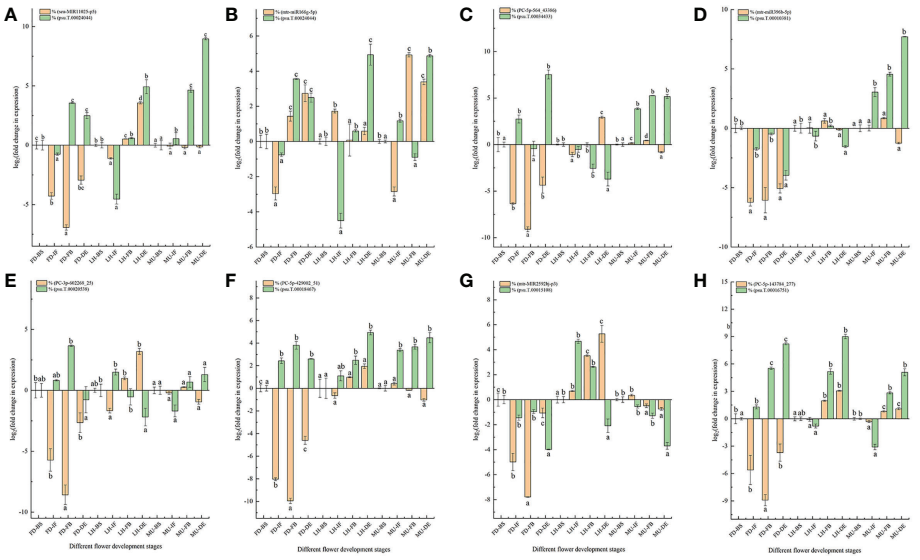


FIGURE 10
qRT-PCR analysis of miRNA-mRNA pairs. **(A)** qRT-PCR analysis of *MIR11025-p5* and *psu.T.00024044*. **(B)** qRT-PCR analysis of *miR166-5p* and *psu.T.00024044*. **(C)** qRT-PCR analysis of *PC-5p-564_43386* and *psu.T.00034433*. **(D)** qRT-PCR analysis of *mtr-miR396b-5p* and *psu.T.00010381*. **(E)** qRT-PCR analysis of *PC-3p-602268_25* and *psu.T.00020538*. **(F)** qRT-PCR analysis of *PC-5p-429002_51* and *psu.T.00018467*. **(G)** qRT-PCR analysis of *mtr-MIR2592bj-p3* and *psu.T.00015108*. **(H)** qRT-PCR analysis of *PC-5p-143784_277* and *psu.T.00016751*. MU, Mutant of Paeonia ostii 'Fengdan'; FD, Paeonia ostii 'Fengdan'; LH, Paeonia suffruticosa 'Lianhe'; BS, Blooming Stage; IF, Initial Flowering Stage; FB, Full Blooming Stage; DE, Decay Stage.

4.3 miRNA-mRNA-TF regulate floral florescence, development, and senescence

MiRNAs have been proved to be of great importance in regulation of gene expression, defense responses, and cell function in plants (Achkar et al., 2016; Choudhary et al., 2021). Recent research have shown that miRNAs play crucial roles in regulating gene expression associated with flowering (Spanudakis and Jackson, 2014). According to the latest report, *miR156* and *miR172* possibly participate in flowering *via* an aging pathway (Waheed and Zeng, 2020). *MiR167* was reported to be involved in governing floral/fiber-associated agronomic traits in cotton (Arora and Chaudhary, 2021). Inhibition of *miR168* in rice could improve yield, prolong flowering and enhance immunity (Wang et al., 2021). While overexpression of *miR159* resulted in late flowering, whereas suppression of *miR159* leaded to the acceleration of flowering in the ornamental flowering plant gloxinia (Li et al., 2013; Millar et al., 2019).

Researches have demonstrated that TFs play vital roles in floral transition as miRNA targets. MiRNAs and their TF targets regulate gene expression at post-transcriptional level and transcriptional level respectively (Waheed and Zeng, 2020). Previously, the miRNA-mRNA-TF modules like *pos-miR319a-3p.2–3p/TCP2*, *pos-miR159/GAMYB*, *pos-miR169/nuclear transcription factor Y subunit A*, and *pos-miR828/WER* were identified in variety FD (Han et al., 2020). TF target genes (*AP2* and *SPL*) might have splice sites for *PsmiR172a* and *PsmiR156a*, suggesting that *miR156* and *miR172* probably play important roles during dormancy transition in FD (Zhang et al., 2018). Additionally, studies have revealed that *miR156-SPL* (Xie et al., 2020; Rao et al., 2021), *miR172-AP2* (Ó'Maoiléidigh et al., 2021), *miR319-TCP* (Zhu et al., 2021), *miR159-MYB* (Spanudakis and Jackson, 2014) and *miR399-PHO2* (Kim et al., 2011) play important roles in floral transition.

Transcription factors *MYC2*, *MYC3*, and *MYC4* in *bHLH* family were involved in jasmonate-mediated flowering inhibition in Arabidopsis (Wang et al., 2017). Previous studies showed that the *bZIP* transcription factors were functionally required for flower development (Strathmann et al., 2001). The *TBZF* gene encoding *bZIP* were reported to be abundant in senescing flower buds (Yang et al., 2002). The *C2H2* zinc finger family perform functions in pollen development regulation in grapevine (Arrey-Salas et al., 2021). Overexpression of the *CcNAC1* gene promotes early flowering in jute (Zhang et al., 2021). Acting as a *B3* domain transcription factor, *AtREM16* prolongs flowering by coupling on the promoters of *SOC1* and *FT* (Yu et al., 2020). Expression of chrysanthemum *Trihelix* transcription factors showed that they played important roles in chrysanthemum inflorescences (Song et al., 2016). It has also been shown that expression of chrysanthemum transcription factor *ERF* can influence flowering time in Arabidopsis (Xing et al., 2019). Additionally, it was shown that *CmERF110* interacts with *CmFLK* to promote flowering by regulating the circadian clock (Huang et al., 2022).

In this study, we identified 16 miRNA-mRNA-TF modules across flower developmental stages (Table S12). The TFs belongs to 16 families. Among these TF families, the *bHLH*, *bZIP*, and *C2H2* accounted for the largest proportion, followed by *B3*, *Trihelix*, *ERF*, etc. In addition, 71 miRNA-mRNA-TF modules were identified across the tree peony varieties at the flower developmental stages (Table S12). The TFs consisted of 37 TF families, of which the *NAC*, *Trihelix* and *bHLH* families accounted for the largest proportion, followed by *ERF*, *B3*, *MYB* family etc. In this study, the *bHLH*, *NAC*, *C2H2*, *bZIP*, and *Trihelix* displayed the most highly differential expression, suggesting that these miRNA-mRNA-TF modules may be crucial factors in floral florescence, development and senescence in tree peony. Their specific functional contributions in tree peony remains to be further explored.

4.4 Limiting factors in the study and future research on tree peony

The first draft genome assembly (~13.79 GB) of tree peony variety 'Luo shen xiao chun' reported recently represents the largest sequenced genome in dicotyledon to date (Lv et al., 2020). However, due to the unusually large and complex genome, the draft genome assembly of tree peony is still on scaffold level, which hinders miRNA-mRNA pairs identification by transcriptome and sRNAome analysis using the reference genome. A high-quality reference genome is expected to provide substantial fundamental resources for further research in tree peony in the future. Furthermore, lack of a homologous genetic transformation system has hindered functional genomics research in tree peony (Wen et al., 2020). Breakthroughs in virus-induced gene silencing (VIGS) in rose (Cheng et al., 2018; Liang et al., 2020; Cheng et al., 2021; Zhang et al., 2021) and tree peony (Zhao et al., 2017; Xie et al., 2019; Wang et al., 2020; Han et al., 2022), provide hope for the functional characterization and identification of miRNA-mRNA modules for floral florescence, development, and senescence identified in this study in tree peony.

5 Conclusion

Prolonging blooming period has been an important target for tree peony breeding. The present study provides the integrated analysis of tree peony miRNA-mRNA modules regulated at the transcriptional level, with the purpose of illustrating the regulatory network of floral florescence, development, and senescence. The expression profiles described include developing flowers at multiple developmental time-points, in three varieties with contrasting flowering time (early- and late-flowering), including

an early-flowering mutant line found previously. A total of 2,444 tree peony miRNAs were identified, with 1,709 of them being novel. A transcriptome analysis resulted in discovery of 35,687 genes, with a significant number of floral florescence, development, and senescence DEGs associated with chloroplast, chloroplast envelope, plasmodesma, and molecular function process, and involved in plant hormone signal transduction, starch and sucrose metabolism, MAPK signaling, phenylpropanoid and carotenoid biosynthesis pathway. Multi-omics analysis of flowering time regulation networks identified key miRNA-target pairs including transcription factors, protein kinases, and hormone regulators that were antagonistically regulated. Newly discovered functional miRNA-mRNA-TF modules provide molecular resources for further interpretation of the florescence mechanism, and for germplasm resource innovation aimed at prolonging flowering time of tree peony.

Data availability statement

The data presented in the study are deposited in the China National GenBank repository, accession number CNP0002984, https://db.cngb.org/search/?q=CNP0002984.

Author contributions

LG performed the experiments, analyzed the data and wrote the manuscript. YL, CZ, and ZW participated in the sample collection and experimental assay. JC and WY contributed to the manuscript review and editing. XZ contributed to the experimental design, review and editing. XH contributed to the conceptualization, review, editing, and supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China [Grant number U1804233], the Innovation Scientists and Technicians Troop Construction Projects of Henan Province [Grant number 202101510003], the Outstanding Youth Fund of Natural Science Foundation of Henan Province [Grant number 202300410119], and the Science and Technology Innovation Talents in Universities of Henan Province [22HASTIT036].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1082415/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Expressed miRNA detected across flower developmental stages and varieties in tree peony. **(A)** The distribution of expressed miRNAs across the four flower developmental stages (BS, IF, FB, DE) in FD. **(B)** The distribution of expressed miRNAs across the four flower developmental stages (BS, IF, FB, DE) in MU. **(C)** The distribution of expressed miRNAs across the four flower developmental stages (BS, IF, FB, DE) in LH. **(D)** Intersection of expressed miRNAs across flower developmental stages (BS, IF, FB, DE) and tree peony varieties (LH, MU and LH). **(E)** The distribution of expressed miRNAs across varieties (FD, MU and LH) at flower developmental stage BS. **(F)** The distribution of expressed miRNAs across varieties (FD, MU and LH) at flower developmental stage IF. **(G)** The distribution of expressed miRNAs across varieties (FD, MU and LH) at flower developmental stage FB. **(H)** The distribution of expressed miRNAs across varieties (FD, MU and LH) at flower developmental stage DE. **(I)** Intersection of expressed miRNAs across tree peony varieties (FD, MU and LH) and flower developmental stages (BS, IF, FB, DE).

SUPPLEMENTARY FIGURE 2
Summary of the number of miRNAs showing significant differential expression at different flower developmental stages in FD, MU, and LH. **(A)** Number of miRNAs showing significant differential expression ($P<0.01$, $P<0.05$, and $P<0.1$) at different flower developmental stages in FD. **(B)** Number of miRNAs showing significant differential expression ($P<0.01$, $P<0.05$, and $P<0.1$) at different flower developmental stages in MU. **(C)** Number of miRNAs showing significant differential expression ($P<0.01$, $P<0.05$, and $P<0.1$) at different flower developmental stages in LH. The blue colored bars represent the number of upregulated miRNAs. The red colored bars represent the number of downregulated miRNAs.

SUPPLEMENTARY FIGURE 3
Annotated genes detected across flower developmental stages in FD, MU, and LH. **(A)** The distribution of annotated genes across developmental stages (BS, IF, FB, DE) in tree peony genotype FD. **(B)** The distribution of annotated genes across developmental stages (BS, IF, FB, DE) in tree peony genotype MU. **(C)** The distribution of annotated genes across developmental stages (BS, IF, FB, DE) in tree peony genotype LH. **(D)** Intersection of annotated genes across developmental stages (BS, IF, FB, DE) and tree peony varieties (LH, MU and LH).

SUPPLEMENTARY FIGURE 4
Annotated genes detected across tree peony varieties at different flower developmental stages. **(A)** The distribution of annotated genes across varieties (FD, MU and LH) at developmental stage BS. **(B)** The distribution of annotated genes across varieties (FD, MU and LH) at developmental stage IF. **(C)** The distribution of annotated genes across varieties (FD, MU and LH) at developmental stage FB. **(D)** The distribution of annotated genes across varieties (FD, MU and LH) at developmental stage DE. **(E)**

Intersection of annotated genes across varieties (FD, MU and LH) and flower developmental stages (BS, IF, FB, DE).

SUPPLEMENTARY FIGURE 5
Summary of the number of genes showing significant differential expression at different flower developmental stages in FD, MU, and LH. **(A)** Number of genes showing significant differential expression ($P<0.05$) across different flower developmental stages (BS, IF, FB, DE) in tree peony variety FD, MU, and LH, respectively. **(B)** Number of genes showing significant differential expression ($P<0.05$) across varieties (FD, MU, LH) at different flower developmental stages (BS, IF, FB, DE), respectively. The blue colored bars represent the number of upregulated genes. The red colored bars represent the number of downregulated genes.

SUPPLEMENTARY FIGURE 6
T plot of mRNA-miRNA pairs selected and verified by qRT-PCR. **(A)** T plot of *MIR11025-p5* and *psu.T.00024044*. **(B)** T plot of *miR166-5p* and *psu.T.00024044*. **(C)** T plot of *PC-5p-564_43386* and *psu.T.00034433*. **(D)** T plot of *mtr-miR396b-5p* and *psu.T.00010381*. **(E)** T plot of *PC-3p-602268_25_S* and *psu.T.00020538*. **(F)** T plot of *PC-5p-429002_51_S* and *psu.T.00018467*. **(G)** T plot of *mtr-MIR2592bj-p3* and *psu.T.00015108*. **(H)** T plot of *PC-5p-143784_277_S* and *psu.T.00016751*.

SUPPLEMENTARY FIGURE 7
Correlation analysis of miRNA and target expression profiles between high-throughput sequencing and qRT-PCR. **(A)** Correlation analysis of *psu.T.00024044* with *MIR11025-p5* and *miR166-5p*. **(B)** Correlation analysis of *psu.T.00034433* with *PC-5p-564_43386*. **(C)** Correlation analysis of *psu.T.00010381* with *mtr-miR396b*-5p. **(D)** Correlation analysis of *psu.T.00020538* with *PC-3p-602268_25*. **(E)** Correlation analysis of *psu.T.00018467* with *PC-5p-429002_51*. **(F)** Correlation analysis of *psu.T.00015108* with *mtr-MIR2592bj-p3*. **(G)** Correlation analysis of *psu.T.00016751* with *PC-5p-143784_277*.

References

Achkar, N. P., Cambiagno, D. A., and Manavella, P. A. (2016). miRNA biogenesis: a dynamic pathway. *Trends Plant Sci.* 21, 1034–1044. doi: 10.1016/j.tplants.2016.09.003

Arora, S., and Chaudhary, B. (2021). Global expression dynamics and miRNA evolution profile govern floral/fiber architecture in the modern cotton (Gossypium). *Planta.* 254, 1–20. doi: 10.1007/s00425-021-03711-3

Arrey-Salas, O., Caris-Maldonado, J. C., Hernández-Rojas, B., and Gonzalez, E. (2021). Comprehensive genome-wide exploration of C2H2 zinc finger family in grapevine (Vitis vinifera l.): Insights into the roles in the pollen development regulation. *Genes.* 12, 302. doi: 10.3390/genes12020302

Artur, M. (2022). Ethylene and gibberellins at the crossroads of rice root elongation. *Plant Cell.* 34, 1157–1158. doi: 10.1093/plcell/koac020

Axtell, M. J., and Meyers, B. C. (2018). Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell.* 30, 272–284. doi: 10.1105/tpc.17.00851

Basso, M. F., Ferreira, P., Kobayashi, A. K., Harmon, F. G., and Grossi-De-Sá, M. F. (2019). MicroRNAs and new biotechnological tools for its modulation and improving stress tolerance in plants. *Plant Biotechnol. J.* 17, 1482–1500. doi: 10.1111/pbi.13116

Bezerra, I. C., Michaels, S. D., Schomburg, F. M., and Amasino, R. M. (2004). Lesions in the mRNA cap-binding gene ABA HYPERSENSITIVE 1 suppress FRIGIDA-mediated delayed flowering in arabidopsis. *Plant J.* 40, 112–119. doi: 10.1111/j.1365-313X.2004.02194.x

Cai, C. F., Cheng, F. Y., Wu, J., Zhong, Y., and Liu, G. X. (2015).). the first high-density genetic map construction in tree peony (Paeonia sect. moutan) using genotyping by specific-locus amplified fragment sequencing. *PLos One* 10, e0128584. doi: 10.1371/journal.pone.0128584

Cheng, C. X., Gao, J. P., and Ma, N. (2018). Investigation of petal senescence by TRV-mediated virus-induced gene silencing in rose. *Plant Sci. Humana Press.* 38, 49–63. doi: 10.1007/978-1-4939-7672-0_4

Cheng, Z. C., Hou, D., Ge, W., Li, X. Y., and Gao, J. (2019). Integrated mRNA, microRNA transcriptome and degradome analyses provide insights into stamen development in moso bamboo. *Plant Cell Physiol.* 61, 76–87. doi: 10.1093/pcp/pcz179

Cheng, C. X., Yu, Q., Wang, Y. R., Wang, H., Dong, Y. H., Ji, Y. Q., et al. (2021). Ethylene-regulated asymmetric growth of the petal base promotes flower opening in rose (Rosa hybrida). *Plant Cell.* 33, 1229–1251. doi: 10.1093/plcell/koab031

Chen, C., Liu, C. H., Jiang, A. L., Zhao, Q. Q., Zhang, Y. H., and Hu, W. Z. (2020). MiRNA and degradome sequencing identify miRNAs and their target genes involved in the browning inhibition of fresh-cut apples by hydrogen sulfide. *J. Agr. Food Chem.* 68, 8462–8470. doi: 10.1021/acs.jafc.0c02473

Chen, L. N., Luo, X., Yang, X. W., Jing, D., Xia, X. C., Li, H. X., et al. (2020). Small RNA and mRNA sequencing reveal the roles of microRNAs involved in pomegranate female sterility. *Int. J. Mol. Sci.* 21, 558. doi: 10.3390/ijms21020558

Choudhary, A., Kumar, A., Kaur, H., and Kaur, N. (2021). MiRNA: the taskmaster of plant world. *Biologia.* 76, 1551–1567. doi: 10.1007/s11756-021-00720-1

Conti, L., Galbiati, M., and Tonelli, C. (2014). *ABA and the floral transition* (Dordrecht: Springer). doi: 10.1007/978-94-017-9424-4_18

Feng, P. Q., Guo, H. L., Chi, W., and Zhang, L. X. (2016). Chloroplast retrograde signal regulates flowering. *P Natl. Acad. Sci. U.S.A.* 113 (38), 10708–10713. doi: 10.1073/pnas.152159911

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* 26, 941–946. doi: 10.1038/nbt1417

Glazinska, P., Kulasek, M., Glinkowski, W., and Wojciechowski, W. (2019). Integrated analysis of small RNA, transcriptome and degradome sequencing provides new insights into floral development and abscission in yellow lupine (Lupines luteus l.). *Int. J. Mol. Sci.* 20, 5122. doi: 10.3390/ijms20205122

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., et al. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* 10, 593–601. doi: 10.1038/ncb1722

Guo, D. L., Li, Q., Lv, W. Q., Zhang, G. H., and Yu, Y. H. (2018). MicroRNA profiling analysis of developing berries for 'Kyoho' and its early-ripening mutant during berry ripening. *BMC Plant Biol.* 18, 1–16. doi: 10.1186/s12870-018-1516-x

Han, R., Truco, M. J., and Lavelle, D. O. (2021). Michelmore, R.W. a composite analysis of flowering time regulation in lettuce. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.632708

Han, J. G., Zhang, T. R., Li, J., and Hu, Y. H. (2020). Identification of miRNA responsive to early flowering in tree peony (Paeonia ostii) by high-throughput sequencing. *J. Hortic. Sci. Biotech.* 96, 1–14. doi: 10.1080/14620316.2020.1846466

Han, L., Zhou, L., Zou, H., Yuan, M., and Wang, Y. (2022). PsGSTF3, an anthocyanin-related glutathione s-transferase gene, is essential for petal coloration in tree peony. *Int. J. Mol. Sci.* 23, 1423. doi: 10.3390/ijms23031423

Hao, Z. J., Liu, D., Gong, S. J., Zhao, D. Q., and Tao, J. (2017). High throughput sequencing of herbaceous peony small RNAs to screen thermo-tolerance related microRNAs. *Genes Genom.* 39, 1–12. doi: 10.1007/s13258-016-0505-4

Hernández-Hernández, V., Benítez, M., and Boudaoud, A. (2019). Interplay between turgor pressure and plasmodesmata during plant development. *J. Exp. Bot.* 71, 768–777. doi: 10.1093/jxb/erz434

Hong, D. Y., Zhou, S. L., He, X. J., Yuan, J. H., Zhang, Y. L., Cheng, F. Y., et al. (2017). Current status of wild tree peony species with special reference to conservation. *Biodivers Sci.* 25, 781–793. doi: 10.17520/biods.2017129

Huang, Y. Y., Xing, X. J., Tang, Y., Jin, J. Y., Ding, L., Song, A. P., et al. (2022). An ERF transcription factor and a FLK homologue jointly modulate photoperiodic flowering in chrysanthemum. *Plant Cell Environ.* 45, 1442–1456. doi: 10.1111/pce.14261

Iordachescu, M., Bowman, H., Sasaki, K., Imai, R., Satoh, S., and Verlinden, S. (2009). Subcellular localization and changes in mRNA abundance of CEBP, a nuclear-encoded chloroplast protein, during flower development and senescence. *J. Plant Biol.* 52, 365–373. doi: 10.1007/s12374-009-9047-5

Jin, Q. J., Xue, Z. Y., Dong, C. L., Wang, Y. J., and Chu, L. L. (2015). Identification and characterization of microRNAs from tree peony (Paeonia ostii) and their response to copper stress. *PLos One* 10, e0117584. doi: 10.1371/journal.pone.0117584

Kamenetsky, R., Barzilay, A., Erez, A., and Halevy, A. H. (2003). Temperature requirements for floral development of herbaceous peony cv. *'Sarah Bernhardt'*. *Sci. Hortic.* 97 (3-4), 309–320. doi: 10.1016/S0304-4238(02)00153-X

Kim, W., Ahn, H. J., Chiou, T. J., and Ahn, J. H. (2011). The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in arabidopsis thaliana. *Mol. Cells* 32, 83–88. doi: 10.1007/s10059-011-1043-1

Kumar, S. V., Lucyshyn, D., Jaeger, K. E., Alós, E., Alvey, E., Alvey, E., et al. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature.* 484, 242–245. doi: 10.1038/nature10928

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi: 10.1186/1471-2105-9-559

Liang, Y., Jiang, C., Liu, Y., Gao, Y., and Gao, J. (2020). Auxin regulates sucrose transport to repress petal abscission in rose (Rosa hybrida). *Plant Cell.* 32, 3485–3499. doi: 10.1105/tpc.19.00695

Li, X. Y., Bian, H. W., Song, D. F., Ma, S. Y., Han, N., Wang, J. H., et al. (2013). Flowering time control in ornamental gloxinia (Sinningia speciosa) by manipulation of miR159 expression. *Ann. Bot.* 111, 791–799. doi: 10.1093/aob/mct034

Li, S. M., Lv, S. Z., Yu, K., Wang, Z. Y., and Huang, J. T. (2019). Construction of a high-density genetic map of tree peony (Paeonia suffruticosa andr. moutan) using restriction site associated DNA sequencing (RADseq) approach. *Tree Genet. Genomes.* 15, 63. doi: 10.1007/s11295-019-1367-0

Li, X., Shahid, M., Wu, J. W., Wang, L., Liu, X. D., and Lu., Y. G. (2016). Comparative small RNA analysis of pollen development in autotetraploid and diploid rice. *Int. J. Mol. Sci.* 17, 499. doi: 10.3390/ijms17040499

Liu, H., Able, A. J., and Able, J. A. (2017). Water-deficit stress-responsive microRNAs and their targets in four durum wheat genotypes. *Funct. Integr. Genomics* 17, 237–251. doi: 10.1007/s10142-016-0515-y

Liu, H., Able, A. J., and Able, J. A. (2020a). Integrated analysis of small RNA, transcriptome, and degradome sequencing reveals the water-deficit and heat stress response network in durum wheat. *Int. J. Mol. Sci.* 21, 6017. doi: 10.3390/ijms21176017

Liu, H., Able, A. J., and Able, J. A. (2020b). Multi-omics analysis of small RNA, transcriptome, and degradome in t. turgidum-regulatory networks of grain development and abiotic stress response. *Int. J. Mol. Sci.* 21, 7772. doi: 10.3390/ijms21207772

Liu, L., Wu, Y. Q., Zhao, D. Q., and Tao, J. (2020). Integrated mRNA and microRNA transcriptome analyses provide insights into paclobutrazol inhibition of lateral branching in herbaceous peony. *3 Biotech.* 10, 1–9. doi: 10.1007/s13205-020-02489-7

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods.* 25, 402–408. doi: 10.1006/meth.2001.1262

Li, J. J., Zhang, X. F., and Zhao, X. Q. (2011). *Tree peony in China* (Beijing, China).

Lopez-Ortiz, C., Pea-Garcia, Y., Bhandari, M., Abburi, V. L., and Reddy, U. K. (2021). Identification of miRNAs and their targets involved in flower and fruit development across domesticated and wild capsicum species. *Int. J. Mol. Sci.* 22, 4866. doi: 10.3390/ijms22094866

Luo, X. N., Sun, D. Y., Wang, S., Luo, S., Fu, Y. Q., Niu, L. X., et al. (2021). Integrating full-length transcriptomics and metabolomics reveals the regulatory mechanisms underlying yellow pigmentation in tree peony (Paeonia suffruticosa andr.) flowers. *Hort. Res.* 8 (1), 15. doi: 10.1038/s41438-021-00666-0

Lv, S. Z., Cheng., S., Wang, Z. Y., Li, S. M., Jin, X., Lan, L., et al. (2020). Draft genome of the famous ornamental plant paeonia suffruticosa. *Ecol. Evol.* 10, 4518–4530. doi: 10.1002/ece3.5965

Martignago, D., Siemiatkowska, B., Lombardi, A., and Conti, L. (2020). Abscisic acid and flowering regulation: many targets, different places. *Int. J. Mol. Sci.* 21, 9700. doi: 10.3390/ijms21249700

Millar, A. A., Lohe, A., and Wong, G. (2019). Biology and function of miR159 in plants. *Plants.* 8, 255. doi: 10.3390/plants8080255

Mishra, P., and Panigrahi, K. C. S. (2015). GIGANTEA-an emerging story. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00008

Ó'Maoiléidigh, D. S., van Driel, A. D., Singh, A., Sang, Q., Le Bec, N., Vincent, C., et al. (2021). Systematic analyses of the MIR172 family members of arabidopsis define their distinct roles in regulation of APETALA2 during floral transition. *PLos Biol.* 19, e3001043. doi: 10.1371/journal.pbio.3001043

Pei, H. X., Ma, N., Chen, J. W., Zheng, Y., Tian, J., Li, J., et al. (2013). Integrative analysis of miRNA and mRNA profiles in response to ethylene in rose petals during flower opening. *PLos One* 8, e64290. doi: 10.1371/journal.pone.0064290

Pradervand, S., Weber, J., Thomas, J., Bueno, M., Wirapati, P., Lefort, K., et al. (2009). Impact of normalization on miRNA microarray expression profiling. *RNA.* 15, 493–501. doi: 10.1261/rna.1295509

Rao, S. P., Li, Y., and Chen, J. H. (2021). Combined analysis of microRNAs and target genes revealed miR156-SPLs and miR172-AP2 are involved in a delayed flowering phenomenon after chromosome doubling in black goji (Lycium ruthencium). *Front. Genet.* 12. doi: 10.3389/fgene.2021.706930

Riboni, M., Robustelli Test, A., Galbiati, M., Tonelli, C., and Conti, L. (2016). ABA-dependent control of GIGANTEA signalling enables drought escape *via* up-regulation of FLOWERING LOCUS T in arabidopsis thaliana. *J. Exp. Bot.* 67, 6309–6322. doi: 10.1093/jxb/erw384

Shan, H., Chen, S. M., Jiang, J. F., Chen, Y., Gu, C. S., Li, P. L., et al. (2012). Heterologous expression of the chrysanthemum R2R3-MYB transcription factor cmMYB2 enhances drought and salinity tolerance, increases hypersensitivity to ABA and delays flowering in arabidopsis thaliana. *Mol. Biotechnol.* 51, 160–173. doi: 10.1007/s12033-011-9451-1

Shibuya, K., Shimizu, K., Niki, T., and Ichimura, K. (2015). Identification of a NAC transcription factor, EPHEMERAL1, that controls petal senescence in Japanese morning glory. *Plant J.* 79, 1044–1051. doi: 10.1111/tpj.12605

Shi, Q. Q., Zhang, X. X., Li, X., Zhai, L. J., Luo, X. N., Luo, J. R., et al. (2019). Identification of microRNAs and their targets involved in paeonia rockii petal variegation using high-throughput sequencing. *J. Am. Soc Hortic. Sci.* 144, 118–129. doi: 10.21273/JASHS04395-18

Shu, H. Y., Zhou, H., Mu, H. L., Wu, S. H., Jiang, Y. L., Yang, Z., et al. (2021). Integrated analysis of mRNA and non-coding RNA transcriptome in pepper (Capsicum chinense) hybrid at seedling and flowering stages. *Front. Genet.* 1497. doi: 10.3389/fgene.2021.685788

Song, A. P., Wu, D., Fan, Q. Q., Tian, C., Chen, S. M., Guan, Z. Y., et al. (2016). Transcriptome-wide identification and expression profiling analysis of chrysanthemum trihelix transcription factors. *Int. J. Mol. Sci.* 17, 198. doi: 10.3390/ijms17020198

Spanudakis, E., and Jackson, S. (2014). The role of microRNAs in the control of flowering time. *J. Exp. Bot.* 65, 365–380. doi: 10.1093/jxb/ert453

Strathmann, A., Kuhlmann, M., Heinekamp, T., and Dröge-Laser, W. (2001). BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development. *Plant J.* 28, 397–408. doi: 10.1046/j.1365-313X.2001.01164.x

Sun, L. Y., Jiang, Z., Ju, Y., Zou, X., Wan, X., Chen, Y., et al. (2021). A potential endogenous gibberellin-mediated signaling cascade regulated floral transition in magnolia×soulangeana 'Changchun'. *Mol. Genet. Genomics* 296, 207–222. doi: 10.1007/s00438-020-01740-3

Sun, Y. J., Xiong, X. G., Wang, Q., Zhu, L., Wang, L., He, Y., et al. (2021). Integrated analysis of small RNA, transcriptome, and degradome sequencing reveals the MiR156, MiR5488 and MiR399 are involved in the regulation of male sterility in PTGMS rice. *Int. J. Mol. Sci.* 22, 2260. doi: 10.3390/ijms22052260

Susila, H., Jin, S., and Ahn, J. H. (2016). Light intensity and floral transition: chloroplast says "time to flower". *Mol. Plant* 9, 1551–1553. doi: 10.1016/j.molp.2016.10.013

Waheed, S., and Zeng, L. H. (2020). The critical role of miRNAs in regulation of flowering time and flower development. *Genes.* 11, 319. doi: 10.3390/genes11030319

Wang, J. W., Czech, B., and Weigel, D. (2009). MiR156 regulated SPL transcription factors define an endogenous flowering pathway in arabidopsis thaliana. *Cell.* 138, 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, S. L., Gao, J., Xue, J. Q., Xue, Y. Q., Li, D. D., Guan, Y. R., et al. (2019). *De novo* sequencing of tree peony (Paeonia suffruticosa) transcriptome to identify critical genes involved in flowering and floral organ development. *BMC Genomics* 20, 1–22. doi: 10.1186/s12864-019-5857-0

Wang, X. J., Liang, H. Y., Guo, D. L., Guo, L. L., Duan, X. G., Jia, Q. S., et al. (2019). Integrated analysis of transcriptomic and proteomic data from tree peony (P. ostii) seeds reveals key developmental stages and candidate genes related to oil biosynthesis and fatty acid metabolism. *Hort. Res.* 6 (1), 1–19. doi: 10.1038/s41438-019-0194-7

Wang, H., Li, Y., Chern, M. S., Zhu, Y., Zhang, L. L., Lu, J. H., et al. (2021). Suppression of rice miR168 improves yield, flowering time and immunity. *Nat. Plants.* 7, 129–136. doi: 10.1038/s41477-021-00852-x

Wang, X., Li, J., Guo, J., Qiao, Q., and Ma, Y. (2020). The WRKY transcription factor PlWRKY65 enhances the resistance of paeonia lactiflora (herbaceous peony) to alternaria tenuissima. *Hort. Res.* 7, 12. doi: 10.1038/s41438-020-0267-7

Wang, H. P., Li, Y., Pan, P. P., Lou, D. J., and Hu, Y. R. (2017). The bHLH transcription factors MYC2, MYC3, and MYC4 are required for jasmonate-mediated inhibition of flowering in arabidopsis. *Mol. Plant* 10, 1461–1464. doi: 10.1016/j.molp.2017.08.007

Wang, S., Liu, L., Mi, X., Zhao, S., and Wei, C. (2021). Multi-omics analysis to visualize the dynamic roles of defense genes in the response of tea plants to gray blight. *Plant J.* 106, 862–875. doi: 10.1111/tpj.15203

Wen, S. S., Chen, L., and Tian, R. N. (2020). Micropropagation of tree peony (Paeonia sect. moutan): A review. *Plant Cell Tiss Org.* 141, 1–14. doi: 10.1007/s11240-019-01747-8

Xie, L., Yan, Z., Li, M., Tian, Y., and Zhang, Y. (2020). Identification of phytochemical markers for quality evaluation of tree peony stamen using comprehensive HPLC-based analysis. *Ind. Crops Prod.* 154, 112711. doi: 10.1016/j.indcrop.2020.112711

Xie, L., Zhang, Q., Sun, D., Yang, W., and Zhang, Y. (2019). Virus-induced gene silencing in the perennial woody paeonia ostii. *PEERJ.* 7, e7001. doi: 10.7717/peerj.700

Xie, Y. R., Zhou, Q., Zhao, Y. P., Li, Q. Q., Liu, Y., Ma, M. D., et al. (2020). FHY3 and FAR1 integrate light signals with the miR156-SPL module-mediated aging pathway to regulate arabidopsis flowering. *Mol. Plant* 13, 483–498. doi: 10.1016/j.molp.2020.01.013

Xing, X. J., Jiang, J. F., Huang, Y. Y., Zhang, Z. X., Song, A. P., Ding, L., et al. (2019). The constitutive expression of a chrysanthemum ERF transcription factor influences flowering time in arabidopsis thaliana. *Mol. Biotechnol.* 61, 20–31. doi: 10.1007/s12033-018-0134-z

Xu, Y. J., Zhu, S. W., Liu, F., Wang, W., Wang, X. W., Han, G. M., et al. (2018). Identification of arbuscular mycorrhiza fungi responsive microRNAs and their regulatory network in maize. *Int. J. Mol. Sci.* 19, 3201. doi: 10.3390/ijms19103201

Yang, Y., Sun, M., Li, S. S., Chen, Q. H., Silav, J. A. T. D., Wang, A. J., et al. (2020). Germplasm resources and genetic breeding of paeonia: a systematic review. *Hort. Res.* 7 (1), 19. doi: 10.1038/s41438-020-0332-2

Yang, S. H., Yamaguchi, Y., Koizumi, N., Kusano, T., and Sano, H. (2002). Promoter analysis of tbzF, a gene encoding a bZIP type transcription factor, reveals distinct variation in cis-regions responsible for transcriptional activation between senescing leaves and flower buds in tobacco plants. *Plant Sci.* 162, 973–980. doi: 10.1016/S0168-9452(02)00043-2

Yin, D. D., Li, S. S., Shu, Q. Y., Gu, Z. Y., and Wang, L. S. (2018). Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. *Gene.* 666, 72–82. doi: 10.1016/j.gene.2018.05.011

Yu, Y. C., Qiao, L. F., Chen, J. C., Rong, Y., Zhao, Y. H., Cui, X. K., et al. (2020). Arabidopsis REM16 acts as a B3 domain transcription factor to promote flowering time *via* directly binding to the promoters of SOC1 and FT. *Plant J.* 103, 1386–1398. doi: 10.1111/tpj.14807

Zhang, L., Guo, D. L., Guo, L. L., Guo, Q., Wang, H. F., and Hou, X. G. (2019). Construction of a high-density genetic map and QTLs mapping with GBS from the interspecific F1 population of p. ostii 'Fengdan bai' and p. suffruticosa 'Xin riyuejin'. *Sci. Hortic.* 246, 190–200. doi: 10.1016/j.scienta.2018.10.039

Zhang, G. Y., Huang, S. Q., Zhang, C., Li, D. F., Wu, Y. B., Deng, J. L., et al. (2021). Overexpression of CcNAC1 gene promotes early flowering and enhances drought tolerance of jute (Corchorus capsularis l.). *Protoplasma.* 258, 337–345. doi: 10.1007/s00709-020-01569-y

Zhang, L., Song, C. W., Guo, D. L., Guo, L. L., Hou, X. G., and Wang, H. F. (2022). Identification of differentially expressed miRNAs and their target genes in response to brassinolide treatment on flowering of tree peony (Paeonia ostii). *Plant Signal Behav.* 17, 2056364. doi: 10.1080/15592324.2022.2056364

Zhang, Y. X., Wang, Y. Y., Gao, X. K., Liu, C. Y., and Gai, S. P. (2018). Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Sci. Rep.* 8, 1–14. doi: 10.1038/s41598-018-22415-5

Zhang, Y., Wu, Z., Feng, M., Chen, J., Qin, M., Wang, W., et al. (2021). The circadian-controlled PIF8–BBX28 module regulates petal senescence in rose flowers by governing mitochondrial ROS homeostasis at night. *Plant Cell.* 33, 2716–2735. doi: 10.1093/plcell/koab152

Zhang, Q., Zhao, Y. Q., Gao, X., and Jia, G. X. (2021). Analysis of miRNA-mediated regulation of flowering induction in lilium×formolongi. *BMC Plant Biol.* 21, 1–16. doi: 10.1186/s12870-021-02961-3

Zhao, A. J., Cui, Z., Li, T. G., Pei, H. Q., Sheng, Y. Q., Li, X. Q., et al. (2019). mRNA and miRNA expression analysis reveal the regulation for flower spot patterning in phalaenopsis 'Panda'. *Int. J. Mol. Sci.* 20, 4250. doi: 10.3390/ijms20174250

Zhao, D. Q., Gong, S. J., Hao, Z. J., and Tao, J. (2015). Identification of miRNAs responsive to botrytis cinerea in herbaceous peony (Paeonia lactiflora pall.) by high-throughput sequencing. *Genes.* 6, 918–934. doi: 10.3390/genes6030918

Zhao, D., Xia, X., Wei, M. G., Sun, J., Meng, J. S., and Tao, J. (2017). Overexpression of herbaceous peony miR156e-3p improves anthocyanin accumulation in transgenic arabidopsis thaliana lateral branches. *3 Biotech.* 7, 1–19. doi: 10.1007/s13205-017-1011-3

Zhu, L., Li, S. S., Ma, Q. Y., Wen, J., Yan, K. Y., and Li, Q. Z. (2021). The acer palmatum TCP transcription factor ApTCP2 controls leaf morphogenesis, accelerates senescence, and affects flowering *via* miR319 in arabidopsis thaliana. *J. Plant Growth Regul.* 41, 244–256. doi: 10.1007/s00344-021-10299-1

Zuluaga, D. L., De Paola, D., Janni, M., Curci, P. L., and Sonnante, G. (2017). Durum wheat miRNAs in response to nitrogen starvation at the grain filling stage. *PLos One* 12, e0183253. doi: 10.1371/journal.pone.0183253

Zuluaga, D. L., Vittoria, L., Luca, C. P., and Gabriella, S. (2018). MicroRNAs in durum wheat seedlings under chronic and short-term nitrogen stress. *Funct. Integr. Genomics* 18, 645–657. doi: 10.1007/s10142-018-0619-7

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fpls.2022.1090571

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Yexiong Quian,
Anhui Normal University, China
Asif Ali,
Nanjing Forestry University, China

*CORRESPONDENCE
Qianqian Wang
qqwangahau@163.com
Huajun Liao
lhj76814@163.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 05 November 2022
ACCEPTED 13 December 2022
PUBLISHED 09 January 2023

CITATION
Wang Q, Yan C, Fu Y, Wang Y, Jiang P, Ding Y and Liao H (2023) Genome-Wide Identification of *14-3-3* gene family reveals their diverse responses to abiotic stress by interacting with StABI5 in Potato (*Solanum tuberosum L.*). *Front. Plant Sci.* 13:1090571. doi: 10.3389/fpls.2022.1090571

COPYRIGHT
© 2023 Wang, Yan, Fu, Wang, Jiang, Ding and Liao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-Wide Identification of *14-3-3* gene family reveals their diverse responses to abiotic stress by interacting with StABI5 in Potato (*Solanum tuberosum L.*)

Qianqian Wang[1]*, Chongchong Yan[1], Yuying Fu[1], Yu Wang[2], Pengfei Jiang[2], Yinyuan Ding[1] and Huajun Liao[1]*

[1]Key Laboratory of Horticultural Crop Germplasm innovation and Utilization, Institute of Horticulture, Anhui Academy of Agricultural Sciences, Hefei, China, [2]National Engineering Laboratory of Crop Stress Resistance, School of Life Sciences, Anhui Agricultural University, Hefei, China

The *14-3-3* genes are widely present in plants and participate in a wide range of cellular and physiological processes. In the current study, twelve *14-3-3s* were identified from potato genome. According to phylogenetic evolutionary analysis, potato *14-3-3s* were divided into ϵ and non-ϵ groups. Conserved motif and gene structure analysis displayed a distinct class-specific divergence between the ϵ group and non-ϵ group. Multiple sequence alignments and three-dimensional structure analysis of 14-3-3 proteins indicated all the members contained nine conservative antiparallel α-helices. The majority of *14-3-3s* had transcript accumulation in each detected potato tissue, implying their regulatory roles across all stages of potato growth and development. Numerous cis-acting elements related to plant hormones and abiotic stress response were identified in the promoter region of potato *14-3-3s*, and the transcription levels of these genes fluctuated to different degrees under exogenous ABA, salt and drought stress, indicating that potato *14-3-3s* may be involved in different hormone signaling pathways and abiotic stress responses. In addition, eight potato 14-3-3s were shown to interact with StABI5, which further demonstrated that potato 14-3-3s were involved in the ABA-dependent signaling pathway. This study provides a reference for the identification of the 14-3-3 gene family in other plants, and provides important clues for cloning potential candidates in response to abiotic stresses in potato.

KEYWORDS

14-3-3* gene family, abscisic acid, abiotic stress, potato, *StABI5

Introduction

The 14-3-3 proteins have been demonstrated to be ubiquitously present in eukaryotes and is recognized as one group of the most critical phospholipid binding proteins (Sehnke et al., 2000; Sehnke et al., 2002a; Paul et al., 2012). They were first identified from cow brain tissue by Moor and Perez, accounting for about 1% of all soluble proteins in brain tissue so that they are considered as brain specific for a long time (Moore and Perez, 1967)). Subsequent studies found that the 14-3-3 proteins present in various tissues of all multicellular eukaryotic organisms, and they were extremely conserved in evolution from unicellular eukaryotes to high plants and animals (Ferl, 1996; Sehnke et al., 2002a). The identified 14-3-3 proteins are relatively tiny acid proteins (25~32 kDa) consisting of nine typically antiparallel α-helices that result in forming homo/hetero dimers (Dong et al., 1995; Yang et al., 2006). It is intriguing that both of the 14-3-3 proteins in the dimer are capable of binding to two different target proteins or two binding sites of one protein respectively, which makes them function as scaffolding proteins or adaptor proteins to involved in various biological pathways, such as enhancing or inhibiting the catalytic activity of target proteins, regulation of nuclear transport and subcellular localization of target proteins, changing the binding ability of target proteins to other proteins and avoiding target protein degradation (Sehnke et al., 2002a; Aitken, 2006; Ryu et al., 2007; Latz et al., 2010; Paul et al., 2012; Cotelle et al., 2014). It is reported that 14-3-3 proteins bind to target protein through 3 types of phosphorylation modes, including R(S/Ar)(+)p(S/T)XP, RX(Ar)(+)p(S/T)XP and PSX1-2-COOH, of which P (S/T) represents phosphorylated serine/threonine, Ar represents aromatic amino acids, + represents basic amino acids, and X after p(S/T) represents any types of amino acids, but it is usually leucine, glutamic, alanine or methionine (Yaffe et al., 1997; Ganguly et al., 2005).

Compared with 14-3-3 proteins in animals, plant 14-3-3 proteins were found relatively late, and they were first identified from model plant *Arabidopsis thaliana* as well as *Spinacia oleracea*, *Oenothera hookeri* and *Hordeum vulgare* (Brandt et al., 1992; Hirsch et al., 1992; Lu et al., 1992). With the development of molecular biology techniques, an increasing number of 14-3-3 proteins have been identified in various plants, such as rice (8), tomato (12), common bean (18), banana (25), apple (18) and so on (Sehnke et al., 2002b; Chen et al., 2006; Xu and Shi, 2006; Li et al., 2021; Lyu et al., 2021; Shao et al., 2021). As a research hotspot, 14-3-3 proteins have received more and more attention from researchers in plant researches. Studies have shown that 14-3-3s can bind to the plant plasma membrane H+-ATPase enzymes involved in phosphorylation to enhance its activity (Wilson et al., 2016). Besides, plant 14-3-3 proteins are also able to be modified through conformation and combined with the signal proteins to regulate the physiological and biochemical reactions. For example, 14-3-3s can interact with signal proteins through phosphorylation, acetylation or ubiquitination to regulate starch synthesis metabolism and protein transport during maize grain development (Yu et al., 2016). In addition, a plenty of previous studies have demonstrated that plant 14-3-3s are also involved in other biological processes, including primary metabolism, signal transduction, ion transport, transcriptional regulation, biotic and abiotic stress response (Ganguly et al., 2005; Schoonheim et al., 2007; Denison et al., 2011; Chen et al., 2017; Huang et al., 2022).

In the long-term evolution, plants have formed a set of complex mechanisms to respond to various abiotic stresses, such as drought, high temperature, cold and salt stress (Zhu, 2016; Gong et al., 2020). As extremely conserved regulatory proteins, 14-3-3s are at the intersection of the plant's response to abiotic stress signal network and respond to a variety of stresses (Fu et al., 2000). For instance, the expression of four rice 14-3-3 genes, OsGF14b, OsGF14c, OsGF14e and OsGF14f are induced by drought, low temperature and salt stress, and overexpression of OsGF14c in rice enhances the drought tolerance of transgenic seedlings (Chen et al., 2006). Additional analysis shows that this process depends on the regulation of OsCDPK1 by OsGF14c. In Arabidopsis, two 14-3-3 proteins λ and κ have been confirmed to negatively regulate salt tolerance by interacting with the salt-sensitive protein SOS2 to inhibit its kinase activity (Zhou et al., 2014). Overexpressing of GsGF14o in Arabidopsis significantly reduce drought tolerance during seed germination and seedling growth. Further analysis showed that transgenic plants exhibited lower stomata opening, incomplete root hair development, reduced transpiration rate and water intake, which indicates that GsGF14o plays a negative regulatory role in drought tolerance (Sun et al., 2013).

Potato (*Solanum tuberosum* L.) is used as crop grain, vegetable or industrial raw material, which is grown all over the world and plays an essential role in solving hunger and ensuring food (Halterman et al., 2016). As the characteristics of comprehensive nutrition and wide adaptability, in 2015, potato was listed as the fourth staple food crop after maize, wheat and rice in China. Potato cultivation is exposed to the external environment, hence, it is frequently affected by various stress factors during the whole growth cycle, such as abiotic stress (eg, drought, salinity, freezing damage, flood, and elevated temperature) and biotic stress (virus, fungal pathogens, bacterial pathogens). These environmental stress factors result in reduced potato yield and quality. Plants can transmit identified stress signals through signal transduction pathways, thereby activating the expression of stress-related genes that regulate the plant's tolerance to the corresponding stress (Gong et al., 2020). Therefore, it is increasingly important and urgent for potato production and molecular breeding to explore stress response genes and elucidate their molecular mechanisms. The 14-3-3 proteins have been confirmed to play momentous roles in

various stress responses in kinds of plant (Roberts et al., 2002). Although some individual *14-3-3s* were reported in potato (Żuk et al., 2003; Aksamit et al., 2005), the specific information and molecular mechanisms of potato14-3-3 gene family members were still poorly understood. In this study, we performed a comprehensive analysis of potato 14-3-3 gene family through bioinformatics methods at genome-wide level. A total of 12 potato *14-3-3s* were identified, then we analyzed their phylogenic relationships, chromosomal locations, gene structures, cis-elements, conserved motifs, and subcellular locations, as well as characterized their expression in responses to ABA, drought and salt stress. Furthermore, eight 14-3-3s were shown to be able to interact with StABI5, a key member in ABA-dependent pathways. These results provide a foundation for further functional studies of the *14-3-3s* in potato.

Materials and methods

Plant materials and stress treatments

The sequenced doubled monoploid potato plants DM1-3 (*S.phureja*) were used in this study. Potato seedlings *in vitro* were cultured in an artificial climate chamber with a photoperiod of 16 h light/8 h dark at temperature 24 ± 2 °C and 55% relative humidity. The stress treatment method for potato seedlings was according to previously described protocols and made some modifications (Charfeddine et al., 2015). Four-weeks-old seedlings were sprayed with 100 μM ABA (abscisic acid) or submitted to 15% polyethylene glycol 6000 (PEG 6000) and 100 mM sodium chloride (NaCl), which simulated the drought and salt stresses, respectively. Then the treated materials were harvested at the indicated time points, 0, 1, 6 and 12 h, and immediately frozen in liquid nitrogen for further analysis. Seedlings grown in normal condition without treatment were served as control. All the samples were conducted three biological replicates.

Identification of 14-3-3s in potato and other plants

The potato genome database version 4.03 of the doubled monoploid S. tuberosum Phureja DM1-3 was downloaded from Spud DB Potato Genomics Resources website (http://spuddb.uga.edu/), then two approaches were used to identify potato *14-3-3s*. Firstly, we constructed a potato protein local database and searched against it using the query of 14-3-3 proteins Hidden Markov Model (HMM) sequence (PF00244.20) through BLASTP program (E-value set as 0.001) (Tang et al., 2016). Secondly, we downloaded the 14-3-3 family protein sequences of *Arabidopsis thaliana*, *Solanum lycopersicum*, *Zea mays* and *Brachypodium distachyon* from Phytozome v13 (https://phytozome-next.jgi.doe.gov/) according to reported protein ID (Table S2) (Cao et al., 2016). All the downloaded sequences were employed to search against the potato protein database to identify all potential 14-3-3s. The Pfam (http://pfam.xfam.org/) and SMART (http://smart.embl-heidelberg.de/) databases were used to further check the candidates on the basis of presence of conserved 14-3-3 domain (Schultz et al., 1998; Finn et al., 2013). Finally, repeated and incomplete sequences were manually removed. The ExPASy database (http://www.expasy.org/tools/) was employed to identify isoelectric points (PI) and molecular weights (MW) of potato 14-3-3 proteins.

Characterization of 14-3-3s sequences, phylogenetic tree construction and cis-element analysis

The identified potato full-length 14-3-3 protein sequences were employed to perform multiple sequence alignment with an inner Muscle program of MEGA version 7.0 software (http://www.megasoftware.net/) (Kumar et al., 2016), then an un-root phylogenetic tree was constructed through maximum likelihood method (ML)with 1000 bootstrap replicates. To investigate the phylogenetic relationships among different plants, the same method was used to generate the other phylogenetic tree with 14-3-3 proteins from potato, *Arabidopsis*, tomato, maize and *Brachypodium*. The online Gene Structure Display Server 2.0 program (GSDS) (http://gsds.cbi.pku.edu.cn/) was applied to determine the exon–intron structure of potato *14-3-3s* based on their CDS sequences and corresponding DNA sequences (Hu et al., 2014). The Multiple Expectation maximization for Motif Elicitation (MEME) program (http://meme-suite.org/tools/meme) was employed to analyze distribution of conserved motifs in potato 14-3-3s with parameters of optimum width ranges from 6 to 200, maximum number of motifs being 15 and any number of repetitions (Bailey et al., 2009). Then the identified motif sequences were further annotated using Pfam and SMART databases. Based on the location information of potato *14-3-3s* in the genome database, the physical chromosome location image of each gene was generated using MapInspect software (https://mapinspect.software.informer.com/). The DNAMAN version 6.0 software was adopted to construct the multiple alignments of potato 14-3-3s, and three-dimensional structure of each protein was predicted using the online Phyre^2 program with normal mode (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley et al., 2015). The upstream 1.5 kb genomic sequence from the start codon of potato 14-3-3 genes were derived, which was further submitted to PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to detect the putative cis-elements in promoter region of each gene (Lescot et al., 2002).

Expression pattern analysis using transcriptome data

Genome-wide transcriptome data of potato DM1-3 (*S.phureja*) was downloaded from Spud DB Potato Genomics Resources website (http://spuddb.uga.edu/) (Massa et al., 2011). The fragments per kilobase of exon model per million mapped reads (FPKM) values of each *14-3-3s* were retrieved. Referring to previous study (Kumar et al., 2013; Wang et al., 2015), the FPKM data was further processed. For the spatial and temporal expression, FPKM valves were transformed by taking $\log_2$ (FPKM +1), thereafter, the processed data was loaded into R software to perform cluster analysis and generate a heatmap.

RNA extraction and quantitative real-time PCR analysis

The Trizol reagent (Invitrogen, USA) was adopted for total RNA isolation of all collected samples referring to manufacturer's protocol. 2 μg of extracted RNA was measured for reverse transcription into cDNA using PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa). Potato 14-3-3s specific primers for this experiment were designed through Primer Express 3.0 software (Applied Biosystems), and the specificity of all primers were further verified in NCBI database with Primer Blast program (Table S3). The *EF1α* (Elongation factor 1*α*) gene was served as an internal control (Charfeddine et al., 2015). The qPCR was performed on an ABI Quant studio 3 Real-Time system (Applied Biosystems) using SYBR® Green Realtime PCR Master Mix (TOYOBO, Japan) with the reaction program: The program as follows: denaturation (95°C for 5min), amplification and quantification (40 cycles of 95°C for 15 s and 60°C for 1 min), melting curve analysis (60–95°C, with a heating rate of 0.3°C/s). The comparative delta cycle threshold (DDCT) method was adopted to calculate relative transcript levels of 14-3-3s (Wang et al., 2017). Statistical analysis was performed by one-way analysis of variance (ANOVA) test using SPSS 19.0 software (http://www.spss.com.cn/). Each qPCR assay was established with three technical replicates.

Determination of subcellular localization of potato 14-3-3s

The stop codon-free cDNA sequences of potato *14-3-3s* were inserted into pCAMBIA1305-GFP vector by homologous recombination method, which were transformed into *Agrobacterium tumefaciens GV3101*. The *Agrobacterium tumefaciens* containing fusion-expression and empty vectors infected one-month-old tobacco (*Nicotiana benthamiana*) seedlings leaves, respectively. StABI5, a reported potato *bZIP* transcription factor localized in the nucleus in tobacco, was also transformed as a control (Zhu et al., 2020). 48 h later, infected leaves were sampled for fluorescence observation under a Zeiss LSM7800 (Zeiss, Germany) confocal microscope. The DNA dye 4, 6-diamidino-2-phenylindole (DAPI) was applied to visualize the nucleus localization. The primers used in this experiment are listed in Table S3.

Bimolecular fluorescence complementation assay

Full-length coding sequences without termination codon of *StABI5* and 11 potato *14-3-3*s were respectively cloned into the pUC-SPYNE and pUC-SPYCE vectors for protein-protein interaction assays, respectively. The fusion plasmids *StABI5*-cYFP and *14-3-3s*-nYFP were transformed jointly into tobacco leaves through *Agrobacterium tumefaciens* (strain GV3101) mediated methods. Then the next steps follow the subcellular localization approach described above. The primers used in this experiment are listed in Table S3.

Results

Genome-wide identification and characterization of 14-3-3s in potato

The downloaded Hidden Markov Model sequence (PF00244.20) and reported *14-3-3s* of *Arabidopsis* and tomato were served as two queries to search against potato genome database to identify the putative potato *14-3-3s*. A total of twelve nonredundant sequences were eventually retained, which were verified to contain the complete domain of 14-3-3 gene family. Refer to the naming methods in previous studies (Wang et al., 2017; Shao et al., 2021), these sequences were named as *StGF14a* to *StGF14l* based on their order in potato chromosomes (Table S1; Figure S1). The relevant information for all *14-3-3s* was collected and listed in Table S1. Twelve 14-3-3 members located on 7 of the 12 chromosomes and contained 249 to 285 amino acid residues with molecular weight (MW) of 28.2-32.2 kDa (Table S1; Figure S1). The predicted isoelectric point (pI) of these proteins were between 4.64 and 4.96, which was consistent with reported values of the 14-3-3 family members in other plants.

Phylogenetic analysis of 14-3-3s in potato and other plants

To gain insight into the evolutionary relationships among members of the potato *14-3-3* family, we constructed an unrooted phylogenetic tree using maximum likelihood method. It was shown that potato 14-3-3s exhibited close

phylogenetic relationships with high bootstrap values (>53%) support (Figure 1A). Like the classification scheme of 14-3-3 families in other plant species, potato 14-3-3 members were also able to be divided into two subgroups (ϵ subgroup and non-ϵ subgroup) based on their phylogenetic relationships. The ϵ subgroup simply contains four members, *StGF14f*, *-14h*, *-14g*, *-14l*, nevertheless, the non-ϵ subgroup contains the rest eight members, *StGF14a*, *-14b*, *-14c*, *-14d*, *-14e*, *-14i*, *-14j* and-*14k* (Figure 1A).

In order to understand the relationship of 14-3-3 family members among different plant species, a total of 57 sequences from maize, *Arabidopsis*, tomato, *Brachypodium distachyon* and potato were extracted to construct an unrooted phylogenetic tree. Phylogenetic analysis showed that 14-3-3s also clustered into two groups (ϵ-group and non-ϵ-group) (Figure 2), the ϵ-group contains 16 members, while the non-ϵ-group contains 41 members, which was consistent with the clustering characteristics of 14-3-3s in potato (4 belong to ϵ-group and 8 belong to non-ϵ-group) and other plant species. In addition, the 14-3-3 members of potato were more closely related to these of tomato according to the phylogenetic tree. For instance, most tomato and potato 14-3-3 members were clustered together in pairs at the terminal branch of the phylogenetic tree with strong bootstrap support (Figure 2), which was in accordance with the evolutionary relationships among these five species.

Gene structure and motif composition analysis of potato 14-3-3s

Gene organization analysis showed that the number of introns of potato*14-3-3s* was distributed between 3 to 6, and members in the same subgroup possessed the same number of introns as well as similar arrangement patterns of introns and exons (Figure 1C). For example, non-ϵ subgroup members contain three introns, while ϵ subgroup contains five introns with the exception of *StGF14g* having six introns. Although intron/exon distributions were similar among potato *14-3-3s* in the same subgroup, the intron length of each gene varies considerably, which may be one of the reasons for the functional differentiation of the same subgroup 14-3-3s (Figure 1C). To understand the diversity in motif composition of potato 14-3-3s, amino acid sequences were submitted to the MEME database for analysis. A total of 15 conservative motifs were identified and named as motif 1to motif 15 (Figure 1B; Table S4). Motif 1 and 2 were annotated as typical 14-3-3 family domains in the Pfam and SMART databases, which were present in all potato 14-3-3 proteins, indicating that these proteins are all genuine14-3-3 family members. However, the remaining 12 motifs had not been documented in the two current databases to date and require further study. Motif 3 and motif 4 were present in each potato 14-3-3s, suggesting these proteins may play similar roles in some pathway (Figure 1B). In contrast, the distribution of additional motifs was extremely variable. For example, Motif 5 only was present only in StGF14a and StGF14b, motif 6 was present only in StGF14d, StGF14e and StGF14j, and motif 9 was only present in StGF14f, suggesting that these motifs might play important roles in gene-specific functions. In addition, StGF14g contains the largest number of motifs (11), which indicated that StGF14g might play more diversified functions in potato growth and development. Similar to gene structure analysis, we found that 14-3-3s in the same subgroups showed highly similar motif arrangement, and different subgroups exhibited certain differences (Figure 1B). Additionally, multiple sequences alignment analysis showed that the protein sequences of the 12 potato 14-3-3s were relatively

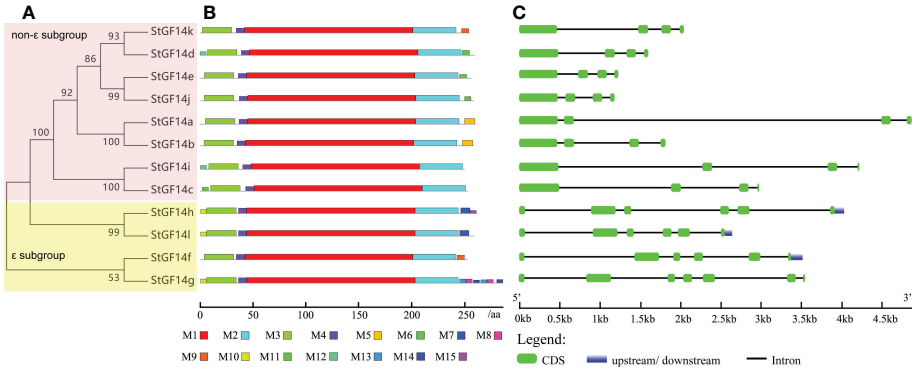


FIGURE 1
Phylogenetic relationships, conserved motifs distribution and gene structure of potato 14-3-3s. **(A)** Phylogenetic tree of twelve potato 14-3-3 proteins. **(B)** Distribution of conserved motifs within potato 14-3-3 protein sequences. The differently coloured boxes represent fifteen different conserved motifs. **(C)** Exon/intron structures of potato *14-3-3s*. The green boxes represent exons, the black lines represent introns, and the blue boxes indicates untranslated regions, and the lengths of exons can be inferred from the scale at the bottom.

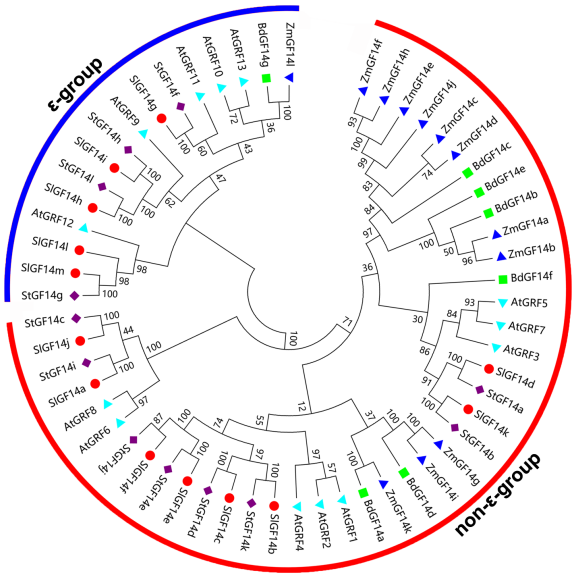


FIGURE 2
Phylogenetic analysis of 14-3-3s from five species. The phylogenetic tree ia constructed by MEGA 7.0 software using 57 14-3-3 protein sequences from *Arabidopsis* (*At*), tomato (*Sl*), maize (*Zm*), *Brachypodium distachyon* (*Bd*) and potato (*St*). The rootless tree is divided into two subfamilies, the ϵ-group is marked with blue line, and the non-ϵ-group is marked by red line.

conservative and contained nine antiparallel α-helices (α1-α9) (Figure 3); the amino acid sequences of α- helices regions were greatly similar, while the sequences of the N-terminal and C-terminal regions exhibit a high degree of variability. Meanwhile, the three-dimensional structure predicted analysis also indicated that each potato 14-3-3 protein contained nine antiparallel α-helices, which was consistent with the result of multiple protein sequence alignment (Figure S2). These results also support the credibility of phylogenetic tree classification.

Analysis of the cis-elements of potato 14-3-3s

It has been reported that the cis-acting elements play an influential role for gene function and regulatory patterns. The plant *14-3-3s* have been demonstrated to participate in various abiotic stress and hormonal regulation pathways. To explore the potential function of potato *14-3-3s*, the 1,500 bp upstream genomic sequence from start codon of each gene were submitted to the PlantCARE online tool for cis-acting regulatory element analysis. Eight types of *cis*-elements were detected in current study, including three hormone-related responsive elements: ABRE (ABA responsive element), GARE (gibberellin response elements), AuxRE (auxin response element), and five stress responses elements: HSE (heat stress element), MBS (drought-inducible response), LTR (low temperature response), TC-rich (defense and stress response), TCA-element (salicylic acid-related element) (Figure 4).

The distribution of these cis-elements in the promoter region of *14-3-3s* exhibited considerable variation, e.g., the number and types of cis-elements varied from 2 to 10 and 2 to 6, respectively. The *StGF14i* contained the maximum number (10) and types (6) of cis-elements, on the contrary, the *StGF14b* and *StGF14d* had Minimum number (2) and types (2). In addition, the HSE element was found in almost all gene promoter regions except for *StGF14d* and *StGF14e*, followed by TC-rich element in nine gene promoters (*StGF14a, -4c, -14d, -14e, -14f, -14h, -14i, -14k, -14l*), while the LTR element was only existed in *StGF14j*. These results suggest that potato *14-3-3s* are presumably regulated by multiple hormones and environmental stresses, and that different individuals may be involved in distinct signaling pathways.

Expression profiles of potato 14-3-3s in different tissues

To explore the temporal and spatial expression patterns of *14-3-3s* in different development stages of potato, the publicly-available transcriptome data was adopted to detect the expression level of each potato *14-3-3s* in 12 different tissues, including leaf, shoot, root, callus, tuber, sepal, stamen, stolon, flower, petiole and carpel (Figure 5; Table S5). As shown in the heatmap, different colors represent the transcript levels of *14-3-3s* in different tissues, which indicated that all genes were expressed in at least two tissues, but the expression patterns of different genes showed some degree of difference. For example, eleven *14-3-3s* (*StGF14a*, *-14b*, *-14c*, *-14d*, *-14e*, *-14f*, *-14h*, *-14i*, *-14j*, *-14k*, *-14l*) were consistently expressed in examined potato tissues, which suggested that these genes played roles in all stages of potato growth and development. On the contrary, the expression of *StGF14g* was only detected in flower and stamen at relatively low level, implying *StGF14g* may be essential for the development of flowering (Figure 5). In addition, we also observed that the expression level of *14-3-3s* was various in different tissues. For instance, the *StGF14a* had the highest expression level in root, *StGF14e* in stamen, *StGF14f* in callus, *StGF14h* and *StGF14i* in petiole, suggesting that these genes may play a vital role in certain specific tissues or developmental stages of potato.

Expression profiles of potato 14-3-3s in response to drought, salt and ABA stress

The 14-3-3 family genes have been reported to be involved in multiple abiotic stress response in numerous plant species (Chen et al., 2006; Li et al., 2021; Shao et al., 2021). Previous

10.3389/fpls.2022.1090571

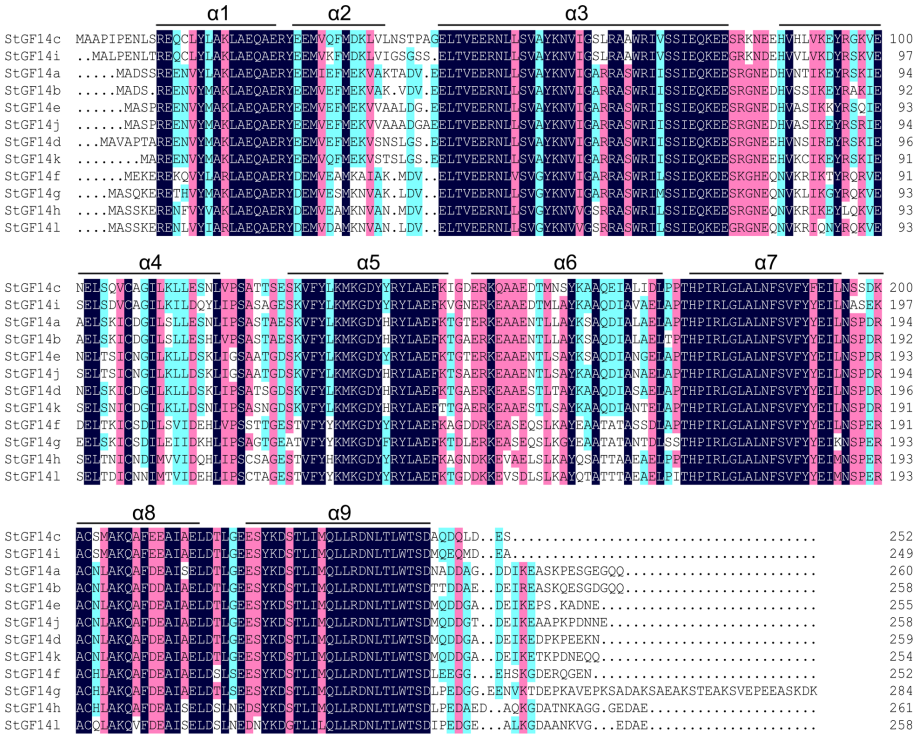


FIGURE 3
Sequences alignment of candidate potato 14-3-3 proteins. Amino acid sequences alignment of the twelve potato 14-3-3 proteins is performed using DNAMAN software. Nine α-helices were marked as α1-α9, respectively.

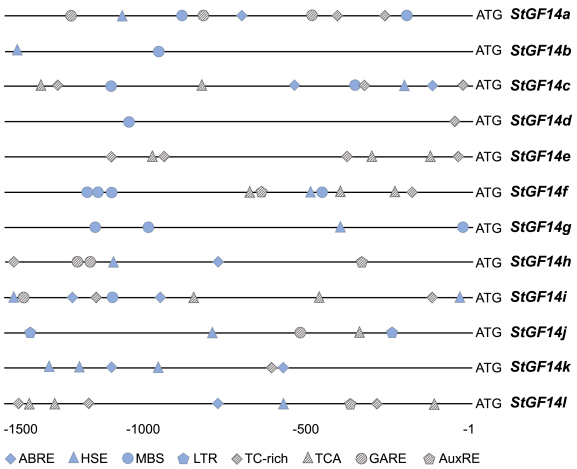


FIGURE 4
Cis-elements in the promoter regions of potato *14-3-3s*. Boxes of different colors and shapes represent eight types of cis-elements. The position of each cis-element in the promoters referring to the ruler at the bottom. ABRE (ABA responsive element), GARE (gibberellin response elements), AuxRE (auxin response element), HSE (heat stress element), MBS (drought-inducible response), LTR (low temperature response), TC-rich (defense and stress response), TCA-element (salicylic acid-related element).

phylogenetic analysis showed that 14-3-3 gene family were relatively conserved among potato, rice, maize and *Arabidopsis*, meanwhile, cis-element analysis indicated potato 14-3-3 family members containing abundant abiotic stress

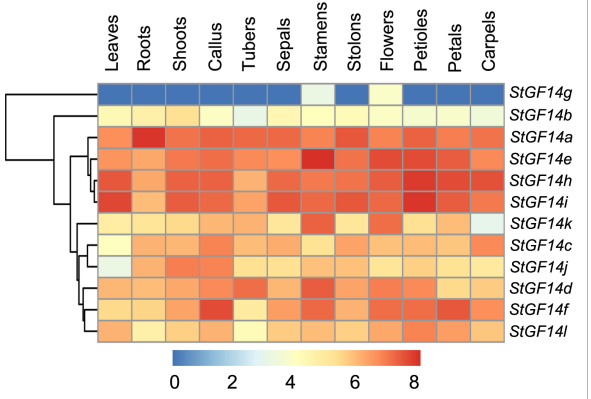


FIGURE 5
Expression profiles of potato *14-3-3s* across different tissues. Blue and red indicate low and high levels of transcript of tewlve potato *14-3-3s*, respectively. Tissues from different developmental stages are shown at the bottom of the heatmap.

related cis-elements (Figures 2, 4), which implied that potato 14-3-3 family members might also participate in abiotic stress response. Then we examined the transcriptional level of 12 potato 14-3-3 genes under drought (PEG), salt (NaCl) and exogenous ABA treatment. The results showed that almost all tested genes were up-regulated or down-regulated to some extent, except for *StGF14g*, whose expression was not detected during the whole treatment stage under three treatments (Figure 6), which was consistent with tissue expression patterns and further demonstrated the *StGF14g* gene only functioning during the reproductive stage of flower development. Eight genes were up-regulated under ABA treatment, including *StGF14a*, *-14b*, *-14c*, *-14d*, *-14e*, *-14f*, *-14j* and *-14k*, of which *StGF14d* and *StGF14e* displayed great amplitude of variation of expression level (>5-fold), while the *StGF14i* and *StGF14h* were slightly down-regulated (Figure 6A).

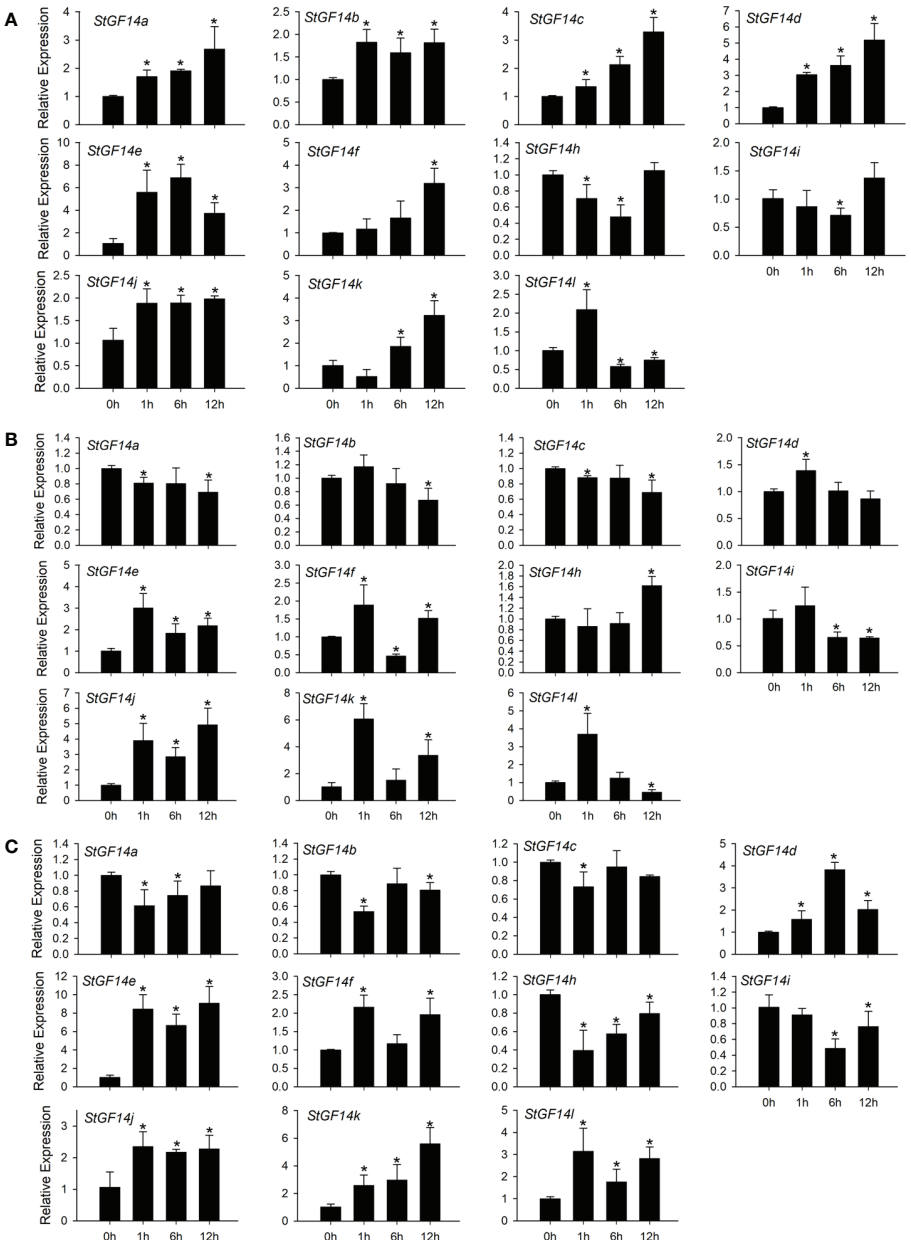


FIGURE 6
Expression analysis of potato14-3-3s in response to drought, salt and ABA treatments. The potato seedlings were treated with 100 μM ABA **(A)**, 100 mM NaCl **(B)**, and 15% PEG 6000 **(C)**. The significant differences between data were calculated using Student's t test, and indicated with an asterisks (*), $P < 0.05$..

In addition, the *StGF14l* was up-regulated at 1 h and then down-regulated at 6 h to 12 h. Under salt stress, four genes *StGF14a*, *-14b*, *-14c* and-*14i* were slightly down-regulated, while *StGF14d*, *-14e*, *-14h*, *-14j* and *-14k* were up-regulated to varying degrees (Figure 6B). Overall, *StGF14f* was up-regulated under salt stress treatment, but down-regulated at 6 h, whereas *StGF14l* was up-regulated at 1 h and then showed a downward trend (Figure 6B). Six genes, *StGF14d*, *-14e*, *-14f*, *-14j*, *-14K* and *-14l* were detected significantly up-regulated under drought treatment, by contrast, *StGF14a*, *-14b*, *-14c*, *-14h* and *-14i* were down-regulated (Figure 6C). It was worth noting that among the twelve genes, ten were inducible simultaneously by three types of treatment, and four genes, *StGF14e*, *-14j*, *-14k* and *-14l*, were dramatically up-regulated (Figure 6), which suggested that these *14-3-3s* might be involved in drought and salt stress through ABA-mediated signaling pathway.

Subcellular localization of potato 14-3-3 proteins

Previous studies have demonstrated that plant 14-3-3 proteins were mostly localized in both the nucleus and cytoplasm, which might be related to their extensive involvement in different biological pathway (Chen et al., 2006; Li and Sangeeta, 2010). To explore the subcellular localization of 14-3-3s in potato, the coding sequences of eleven *14-3-3s* without stop codons were cloned. On account of *StGF14g* had no expression in many potato tissues, its coding sequence was not available (Figure 5). Then the fusion vectors were successfully constructed through linked coding sequences of potato 14-3-3s with GFP reporter gene driven by CaMV 35S promoter, which were transiently expressed in tobacco epidermal cells to observe fluorescent signals. *StABI5*, a reported potato *bZIP* transcription factor localized in the nucleus in tobacco, was also transformed as a control. According to the result, the GFP signal of empty vector control was observed in both the nucleus and cytoplasm, and the GFP signal of *StABI5* was only present in the nucleus (Figure 7). The GFP signal accumulation of 11 potato14-3-3s were the same as that of empty vector control, which implicated these proteins were also located in the nucleus and cytoplasm, consistent with reported results (Zuo et al., 2021).

Potato 14-3-3 proteins can interact with StABI5

Plant 14-3-3 proteins have been certified to interact with transcription factors of ABA-dependent signaling pathway to

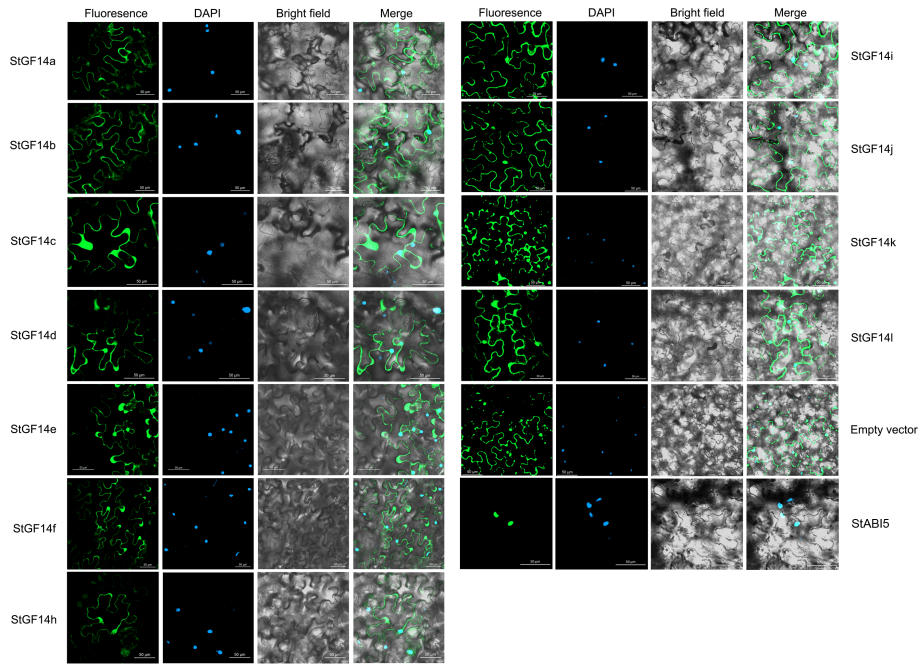


FIGURE 7
Subcellular localization analysis of potato 14-3-3s. Fusion vectors of each potato *14-3-3s* and *StABI5* are transiently expressed under control of the CaMV35S promoter in tobacco leaves and observed under a laser scanning confocal microscope. The green reprents the GFP signal, and blue represents the nucleus are stained by dapi Bars = 50 μm.

regulate the expression of downstream genes, thus participating in the regulation of plant growth and development and abiotic stress response (Casaretto and Ho, 2003). In a separate study, we identified a potato *StABI5* gene that functional analysis suggests plays a negative regulatory role in drought stress response through the ABA-dependent pathway. Therefore, a BIFC assay was performed to verify whether potato 14-3-3 protein could interact with StABI5. Our result indicated that eight 14-3-3s (StGF14a, -14c, -14e, -14f, -14h, -14i, -14j and -14k) were capable of binding to StABI5 and localized in the nucleus, while the rest three proteins could not interact with StABI5 (Figure 8). These results suggested that potato14-3-3s might also participate in ABA-dependent signaling pathway by interacting with transcription factors.

Discussion

As global temperatures continue to rise, the frequency and extent of subsequent droughts will continue to increase. Mining influential functional genes and cultivating new crop varieties is one of the most effective means to improvie plant environmental adaptability. The 14-3-3s have been demonstrated to play crucial roles in response to biotic and abiotic stress (Oecking and Jaspert, 2009; Denison et al., 2011). However, compared to wheat (Shao et al., 2021), Citrus sinensis (Lyu et al., 2021), rice (Chen et al., 2006), apple (Zuo et al., 2021) and common bean (Ruihua et al., 2015), the information on potato 14-3-3s is still limited. Therefore, we performed a genome-wide identification and comprehensive analysis of the potato 14-3-3 gene family, which will provide a reference for future functional studies of potato 14-3-3s.Twelve potato *14-3-3s* from the whole potato genome were identified, which were further divided into two subgroups: ϵ group and non-ϵ group, according to their phylogenetic relationship (Figure 1 and Table S1). Previous research had shown that plant 14-3-3 gene families were relatively conservative during the evolution, and they shared some similar characteristics, such as containing two subgroups, more members in non-ϵ group than in ϵ group, containing nine antiparallel α-helices, etc. (Dong et al., 1995; Yang et al., 2006). Further comprehensive analysis of potato 14-3-3 family members, we also found the more members in non-ϵ group (8) than in ϵ group (4) (Figures 1, 2). On the contrary, the number of introns and extrons of ϵ group members was more than that of non-ϵ group members, which was consistent with 14-3-3 families in apple and mango, suggesting that evolution had driven this diversity and stability (Zuo et al., 2021; Xia et al., 2022). In addition, multiple sequence alignment and three-dimensional structure predicted analysis of potato 14-3-3s demonstrated that each member contained nine relatively conservative antiparallel α-helices (Figures 3, S2), further suggesting the conservative evolutionary relationship of 14-3-3 family members within and among species. However, the sequences in the N-terminal and C-terminal regions of each 14-3-3s showed a large degree of variability. This characteristic was also reflected from motif analysis where N-terminal and C-terminal motifs were extremely variable, which was regarded as

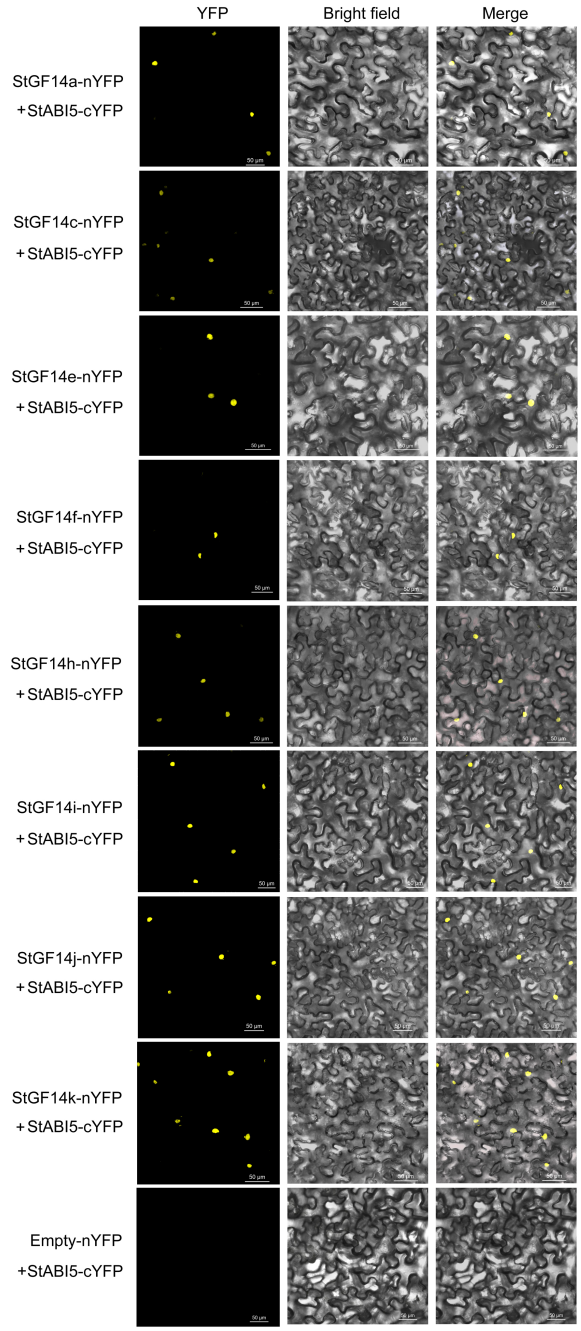


FIGURE 8
BiFC assay verifies the interactions between candidate potato 14-3-3s and StABI5. The potato 14-3-3s-nYFP and StABI5-cYFP vectors are transient transformed in tobacco leaves determined. Yellow fluorescent protein (YFP) images are detected in nulceus of tobacco leaf epidermal cells, indicating the interaction between potato 14-3-3s and StABI5. Bars = 50 μm.

the key factor for *14-3-3s* to bind various proteins and participate in different biological pathways (Coblitz et al., 2006). Potato and tomato belong to the *solanaceae* family, which diverged from common ancestry at around 7.3 Ma (Sato et al., 2012). Multispecies phylogenetic analysis showed that most potato and tomato *14-3-3*s were clustered together in pairs at the terminal branch of the phylogenetic tree with strong bootstrap support (Figure 2), which indicated the closer relationship between potato and tomato, and consisted with the evolutionary history (Tang et al., 2022). Taken together, these results demonstrate that potato *14-3-3s* share common properties with the plant *14-3-3* gene family, and provide further evidence for the conservation of the *14-3-3* gene family in plants.

The involvement of *14-3-3s* in a wide range of biological processes in plants is closely related to their specific expression patterns in different tissues throughout the plant growth cycle (Wilson et al., 2016). Our data displayed that potato *14-3-3s* were expressed in every tested tissue with the exception of *StGF14g* (Figure 5). In particular, *StGF14a*, *-14e*, *-14h* and *-14i* were consistently and strongly expressed, suggesting that these four genes played a critical role throughout growth and development in potato. However, the transcription of *StGF14g* was detected only in flower and stamen, implying that *StGF14g* may function during the reproductive phase of potato development. The involvement of plant *14-3-3s* in the regulation of flowering has been thoroughly demonstrated. For example, *Arabidopsis 14-3-3-ω*, as a bridge linker, formed a complex with the zinc finger transcription factor OXS2 (oxidative stress 2) and FT (flower-forming factor), which changed the subcellular location of FT, thereby affecting the flowering period of *Arabidopsis* (Liang and Ow, 2019). In addition, compared with wild plants, knockout of 14-3-3ν and 14-3-3μ in *Arabidopsis* also resulted in a delayed flowering phenotype (Mayfield et al., 2007). *OsGF14c*, a 14-3-3 gene in rice, has been shown to act as a negative regulator of flowering. Overexpression of *OsGF14c* resulted in a delayed flowering phenotype, while knockout it resulted in an early flowering phenotype (Asih et al., 2009). The transcript of all potato *14-3-3s* were detected in flower and stamen, indicating that these genes were also essential for potato flower development (Figure 5).

Increasing evidence shows that *14-3-3s* are widely involved in the signaling pathways of abiotic stress response in a variety of plants. *TFT4*, a tomato 14-3-3 gene, was significantly upregulated under salt or alkali stress (Xu and Shi, 2006). Further functional mechanism analysis indicated that *TFT4* was involved in the regulation of PKS5-J3 signaling pathway and effectively regulated the concentration of H+ in cells, thereby alleviating alkali stress (Xu et al., 2013). Similarly, the *14-3-3* gene *OsGF14e* in rice was phosphorylated by *OsCPK21* under salt stress, which promoted ABA signal transduction, thus enhancing plant salt tolerance (Chen et al., 2017). GsGF14o was demonstrated to play a negative regulatory role in drought tolerance (Sun et al., 2013). Overexpression of GsGF14o in Arabidopsis significantly reduced drought tolerance during seed germination and seedling growth with lower stomata opening, incomplete root hair development, reduced transpiration rate and water intake. In addition, recent studies have shown that the transcription of 14-3-3 family members of bananas and *Brachypodium distachyon* were regulated by a variety of abiotic stress (Cao et al., 2016; Li et al., 2016). In this study, the expression analysis of potato *14-3-3s* under drought and salt stress implied that the transcript levels of *14-3-3* family members were up- or down-regulated to varying degrees, with the exception of *StGF14g* that was not expressed (Figure 6). The majority of these 12 genes respond to both types of stress treatments, especially four of them, *StGF14e*, *-14j*, *-14k* and *-14l*, were significantly up-regulated. Additionally, four abiotic stress-responsive elements, HSE, MBS, TC-rich and LTR, were widely present in promoter regions of potato *14-3-3s* (Figure 4), suggesting the potato *14-3-3s* were potential candidates for stress-inducible gene under suitable environmental conditions. These results provide strong evidence for the role of *14-3-3s* in response to abiotic stress.

Plant hormones, such as auxin, gibberellin, abscisic acid and ethylene, play essential roles in all stages of plant growth and development and adaptation to biotic and abiotic stresses (Verma et al., 2016). The *14-3-3s* are considered to be key components of plant hormones signaling pathway, which participates in the regulation of signaling pathway by binding to relevant functional proteins. For example, the tobacco 14-3-3 proteins could bind with phosphorylated RSG (REPRESSION OF SHOOT GROWTH) and alter its subcellular localization, so that RSG could be retained in the cytoplasm instead of entering the nucleus to regulate the expression of GA biosynthesis related genes, which controlled the GA contents (Ishida et al., 2004). Similarly, 14-3-3 proteins in rice and *Arabidopsis* were capable of regulating the activity of ACS (1-aminocyclopropane carboxylic acid synthase) by binding to the phosphorylated C-terminus of ACS, thereby protecting ACS from degradation during ethylene biosynthesis (Yao et al., 2007). In addition, a recent study showed that the expression of 14-3-3 genes in citrus were induced under IAA, SA and ABA treatments (Lyu et al., 2021). In the current study, three cis-elements related to abscisic acid, gibberellin and AuxRE response were identified from the promoter regions of twelve potato *14-3-3s* (Figure 4). Among them, ABRE element existed in 6 *14-3-3* members' promoters (*StGF14a*, *-14c*, *-4h*, *-14i*, *-14k*, *-14l*), GARE element existed in 4 (*StGF14a*, *-14h*, *-14i*, *-14j*), and AuxRe exists in 3 (*StGF14f*, *-14h*, *-14l*), respectively, suggesting that potato 14-3-3 members may also involve in hormone signaling pathways.

Abscisic acid (ABA), an essential plant hormone, plays a critical regulatory role in plant response to various adverse environmental stimuli (Raghavendra et al., 2010). Under

drought stress, ABA content is rapidly accumulated in plant, which activates or inhibits the expression of stress-related genes in ABA- dependent signaling pathway to improve plant stress tolerance through regulating stomatal closure to reduce transpiration rate, regulating cell osmotic pressure, and changing root morphology to obtain additional soil water (Zhang et al., 2006). The *14-3-3s* have been shown to be crucial members of ABA-dependent signaling pathways, which modulates the expression of downstream ABA-responsive genes by interacting with transcription factor genes in plants. Under the exogenous ABA treatment, two potato 14-3-3 genes, *StGF14h* and *StGF14i* were slightly down regulated, while other nine genes (*StGF14a*, *-14b*, *-14c*, *-14d*, *-14e*, *-14f*, *-14j*, *-14k*, *-14l*) were up regulated to varying degrees (Figure 6), which suggested that potato 14-3-3 genes might also participate in ABA-dependent signaling pathway.

The ABA insensitive 5 (*ABI5*) is a transcription factor with basic leucine zinc finger structure (*bZIP*), which is a key target gene in ABA-dependent signaling pathway (Miura et al., 2019). In barley, the 14-3-3 proteins acted as a linker to associate the ABA effector VP1 (viviparous 1) with HvABI5 to jointly regulate the expression of downstream ABA response genes (Casaretto and Ho, 2003). In the current study, The BIFC assay showed that eight potato 14-3-3 proteins, StGF14a, -14c, -14e, -14f, -14h, -14i, -14j and -14k could interact with StABI5 (Figure 8). In addition, our separate study showed that the *StABI5* played a negative regulatory role in potato drought stress response (data not shown), indicting these 8 14-3-3s might jointly participate in drought stress response through the interaction with StABI5. These results further confirm that potato 14-3-3 members participate in ABA-dependent signaling pathway through interactions with downstream TFs.

In conclusion, we performed a comprehensive analysis of potato 14-3-3s in this study. Twelve potato 14-3-3 proteins were identified and they exhibited distinct transcript accumulation in different tissues during the growth and development process as well as in response to a variety of abiotic stresses. In addition, the key target member in ABA-dependent signaling pathway StABI5 was demonstrated to act as binding partners of potato 14-3-3s. These results further certified the correlation between 14-3-3s and adversity stress response, which also provided new clues to discover important candidate genes for potato resistance breeding. Nevertheless, future studies on the functional mechanism of potato14-3-3s regulation of stress tolerance should be strengthened.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

QW and HL conceived and designed the study. CY, YF, YW and YD performed the experiments. QW and PJ analyzed data. QW wrote the manuscript. All authors contributed to the article and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (31901439), Natural Science Foundation of Anhui Province (1908085QC134), The Grants from the Earmarked Fund for Anhui Science and Technology Major Project (202003b06020016), The Modern Agro-Industry Technology Research System of China (CARS-09-ES10).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1090571/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Chromosomal locations of potato 14-3-3s.

SUPPLEMENTARY FIGURE 2
The predicted 3D models of twelve 14-3-3 proteins in potato. The α1-α9 represent nine α-helices in the protein structures, respectively, and marked with different colour curves.

SUPPLEMENTARY FIGURE 3
BiFC assay verifies the interactions between candidate potato 14-3-3s and StABI5.

SUPPLEMENTARY TABLE 1
Characteristics of 14-3-3 genes in potato.

SUPPLEMENTARY TABLE 2
The nomenclature and characteristics of *14-3-3s* in four representative species.

SUPPLEMENTARY TABLE 3
List of gene-specific primers used in the present study.

SUPPLEMENTARY TABLE 4
Sequences of conserved motifs identified in potato 14-3-3 proteins.

SUPPLEMENTARY TABLE 5
FPKM of potato *14-3-3s* in different tissues.

References

Żuk, M., Skała, J., Biernat, J., and Szopa, J. (2003). Repression of six 14-3-3 protein isoforms resulting in the activation of nitrate and carbon fixation key enzymes from transgenic potato plants. *Plant Sci.* 165 (4), 731–741. doi: 10.1016/S0168-9452(03)00231-0

Aitken, A. (2006). 14-3-3 proteins: a historic overview. *Semin. Cancer Biol.* 16, 162–172. doi: 10.1016/j.semcancer.2006.03.005

Aksamit, A., Korobczak, A., Skala, J., Lukaszewicz, M., and Szopa, J. (2005). The *14-3-3* gene expression specificity in response to stress is promoter-dependent. *Plant Cell Physiol.* 46, 1635–1645. doi: 10.1093/pcp/pci179

Asih, P. Y., Yuka, O., Shojiro, T., Hiroyuki, T., and Ko, S. (2009). The 14-3-3 protein GF14c acts as a negative regulator of flowering in rice by interacting with the florigen Hd3a. *Plant Cell Physiol.* . 50, 429–438. doi: 10.1093/pcp/pcp012

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37 (suppl_2), W202–W208. doi: 10.1093/nar/gkp335

Brandt, J., Thordal-Christensen, H., Vad, K., Gregersen, P. L., and Collinge, D. B. (1992). A pathogen-induced gene of barley encodes a protein showing high similarity to a protein kinase regulator. *Plant J.* 2 (5), 815–820. doi: 10.1111/j.1365-313X.1992.tb00151.x

Cao, H., Xu, Y., Yuan, L., Bian, Y., Wang, L., Zhen, S., et al. (2016). Molecular characterization of the 14-3-3 gene family in *Brachypodium distachyon l.* reveals high evolutionary conservation and diverse responses to abiotic stresses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01099

Casaretto, J., and Ho, T.-h. (2003). The transcription factors HvABI5 and HvVP1 are required for the abscisic acid induction of gene expression in barley aleurone cells. *Plant Cell.* 15 (1), 271–284. doi: 10.1105/tpc.007096

Charfeddine, S., Saïdi, M. N., Charfeddine, M., and Gargouri-Bouzid, R. (2015). Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. *Mol. Biol. Rep.* 42 (7), 1163–1174. doi: 10.1007/s11033-015-3853-2

Chen, F., Li, Q., Sun, L., and He, Z. (2006). The rice *14-3-3* gene family and its involvement in responses to biotic and abiotic stress. *DNA Res.* 13 (2), 53–63. doi: 10.1093/dnares/dsl001

Chen, Y., Zhou, X., Chang, S., Chu, Z., Wang, H., Han, S., et al. (2017). Calcium-dependent protein kinase 21 phosphorylates 14-3-3 proteins in response to ABA signaling and salt stress in rice. *Biochem. Biophys. Res. Commun.* 493 (4), 1450–1456. doi: 10.1016/j.bbrc.2017.09.166

Coblitz, B., Wu, M., Shikano, S., and Min, L. (2006). C-terminal binding: An expanded repertoire and function of 14-3-3 proteins. *FEBS Lett.* 580 (6), 1531–1535. doi: 10.1016/j.febslet.2006.02.014

Cotelle, V., Meek, S. E., Provan, F., Milne, F. C., Morrice, N., and Mackintosh, C. (2014). 14-3-3s regulate global cleavage of their diverse binding partners in sugar-starved *Arabidopsis* cells. *EMBO J.* 19 (12), 2869–2876. doi: 10.1093/emboj/19.12.2869

Denison, F. C., Paul, A. L., Zupanska, A. K., and Ferl, R. J. (2011). 14-3-3 proteins in plant physiology. *Semin. Cell Dev. Biol.* 22 (7), 720–727. doi: 10.1016/j.semcdb.2011.08.006

Dong, L., Bienkowska, J., Petosa, C., Collier, R. J., and Liddington, R. (1995). Crystal structure of the zeta isoform of the 14-3-3 protein. *Nature.* 376, 191–194. doi: 10.1038/376191a0

Ferl, R. J. (1996). 14-3-3 PROTEINS AND SIGNAL TRANSDUCTION. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47 (1), 49–73. doi: 10.1146/annurev.arplant.47.1.49

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: the protein families database. *Nucleic Acids Res.* 42 (D1), D222–D230. doi: 10.1093/nar/gkt1223

Fu, H., Subramanian, R. R., and Masters, S. C. (2000). 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.* 40 (1), 617–647. doi: 10.1146/annurev.pharmtox.40.1.617

Ganguly, S., Weller, J. L., Ho, A., Chemineau, P., Malpaux, B., and Klein, D. C. (2005). Melatonin synthesis: 14-3-3-dependent activation and inhibition of arylalkylamine n-acetyltransferase mediated by phosphoserine-205. *Proc. Natl. Acad. Sci. U.S.A.* 102 (4), 1222–1227. doi: 10.1073/pnas.0406871102

Gong, Z., Xiong, L., Shi, H., Yang, S., and Zhu, J. K. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci. China. Life Sci.* 63 (5), 635–674. doi: 10.1007/s11427-020-1683-x

Halterman, D., Guenthner, J., Collinge, S., Butler, N., and Douches, D. (2016). Biotech potatoes in the 21st century: 20 years since the first biotech potato. *Am. J. Potato Res.* 93, 1–20. doi: 10.1007/s12230-015-9485-1

Hirsch, S., Aitken, A., Bertsch, U., and Soll, J. (1992). A plant homologue to mammalian brain 14-3-3 protein and protein kinase c inhibitor. *FEBS Lett.* 296 (2), 222–224. doi: 10.1016/0014-5793(92)80384-s

Huang, Y., Wang, W., Yu, H., Peng, J., Hu, Z., and Chen, L. (2022). The role of 14-3-3 proteins in plant growth and response to abiotic stress. *Plant Cell Rep.* 41, 833–852. doi: 10.1007/s00299-021-02803-4

Hu, B., Jin, J., Guo, A. Y., He, Z., and Ge, G. (2014). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 31 (8), 1296–1297. doi: 10.1093/bioinformatics/btu817

Ishida, S., Fukazawa, J., and Takahashi, Y. Y. (2004). Involvement of 14-3-3 signaling protein binding in the functional regulation of the transcriptional activator REPRESSION OF SHOOT GROWTH by gibberellins. *Plant Cell.* 16 (10), 2641–2651. doi: 10.1105/tpc.104.024604

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10 (6), 845–858. doi: 10.1038/nprot.2015.053

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi: 10.1093/molbev/msw054

Kumar, S. A., Vishal, S., Kumar, P. A., Vishal, A., and Singh, A. P. (2013). Genome-wide organization and expression profiling of the NAC transcription factor family in potato (*Solanum tuberosum l.*). *DNA Res.* 20 (4), 403–423. doi: 10.1093/dnares/dst019

Latz, A., Becker, D., Hekman, M., Mueller, T., Beyhl, D., Marten, I., et al. (2010). TPK1, a Ca2+-regulated *Arabidopsis* vacuole two-pore k+ channel is activated by 14-3-3 proteins. *Plant J.* 52 (3), 449–459. doi: 10.1111/j.1365-313X.2007.03255.x

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Liang, M., and Ow, D. W. (2019). Nucleocytoplasmic OXIDATIVE STRESS 2 can relocate FLOWERING LOCUS T. *Biochem. Biophys. Res. Commun.* 517 (4), 735–740. doi: 10.1016/j.bbrc.2019.07.124

Li, M., Ren, L., Xu, B., Yang, X., Xia, Q., He, P., et al. (2016). Genome-wide identification, phylogeny, and expression analyses of the *14-3-3* family reveal their involvement in the development, ripening, and abiotic stress response in banana. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01442

Li, M., Ren, L., Zou, Z., Hu, W., Xiao, S., Yang, X., et al. (2021). Identification and expression analyses of the special 14–3-3 gene family in papaya and its involvement in fruit development, ripening, and abiotic stress responses. *Biochem. Genet.* 59 (6), 1599–1616. doi: 10.1007/s10528-021-10077-4

Li, X. Y., and Sangeeta, D. (2010). Soybean 14-3-3 gene family: identification and molecular characterization. *Planta.* 233 (3), 569–582. doi: 10.1007/s00425-010-1315-6

Lu, G., Delisle, A. J., de Vetten, N. C., and Ferl, R. J. (1992). Brain proteins in plants: an *Arabidopsis* homolog to neurotransmitter pathway activators is part of a DNA binding complex. *Proc. Natl. Acad. Sci. U.S.A.* 89 (23), 11490–11494. doi: 10.1073/pnas.89.23.11490

Lyu, S., Chen, G., Pan, D., Chen, J., and She, W. (2021). Molecular analysis of 14-3-3 genes in citrus sinensis and their responses to different stresses. *Int. J. Mol. Sci.* 22 (2), 568. doi: 10.3390/ijms22020568

Massa, A. N., Childs, K. L., Lin, H., Bryan, G. J., Giuliano, G., and Buell, C. R. (2011). The transcriptome of the reference potato genome solanum tuberosum

group phureja clone DM1-3 516R44. *PloS One* 6 (10), e26801. doi: 10.1371/journal.pone.0026801

Mayfield, J. D., Folta, K. M., Paul, A.-L., and Ferl, R. J. (2007). The 14-3-3 proteins mu and upsilon influence transition to flowering and early phytochrome response. *Plant Physiol.* 145 (4), 1692–1702. doi: 10.1104/pp.107.108654

Miura, K., Lee, J., Jin, J. B., Yoo, C. Y., Miura, T., and Hasegawa, P. M. (2019). Sumoylation of ABI5 by the arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. *Proc. Natl. Acad. Sci. U.S.A.* 106 (13), 5418–5423. doi: 10.1073/pnas.0811088106

Moore, B. W., and and Perez, V. J. (1967). . Specific acidic proteins of the nervous system. FD Carlson *Physiological and Biochemical Aspects of Nervous Integration*, 343–359.

Oecking, C., and Jaspert, N. (2009). Plant 14-3-3 proteins catch up with their mammalian orthologs. *Curr. Opin. Plant Biol.* 12 (6), 760–765. doi: 10.1016/j.pbi.2009.08.003

Paul, A. L., Denison, F. C., Schultz, E. R., Zupanska, A. K., and Ferl, R. J. (2012). 14-3-3 phosphoprotein interaction networks–does isoform diversity present functional interaction specification. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00190. eCollection 2012.

Raghavendra, A. S., Gonugunta, V. K., Christmann, A., and Grill, E. (2010). ABA perception and signalling. *Trends Plant Sci.* 15 (7), 395–401. doi: 10.1016/j.tplants.2010.04.006

Roberts, M. R., Salinas, J., and Collinge, D. B. (2002). 14-3-3 proteins and the response to abiotic and biotic stress. *Plant Mol. Biol.* 50 (6), 1031–1039. doi: 10.1023/a:1021261614491

Ruihua, L., Xiaotong, J., Donghao, J., Sangeeta, D., Shaomin, B., Xuyan, L., et al. (2015). Identification of 14-3-3 family in common bean and their response to abiotic stress. *PloS One* 10 (11), e0143280. doi: 10.1371/journal.pone.0143280

Ryu, H., Kim, K., Cho, H., Park, J., Choe, S., and Hwang, I. (2007). Nucleocytoplasmic shuttling of BZR1 mediated by phosphorylation is essential in *Arabidopsis* brassinosteroid signaling. *Plant Cell* 19 (9), 2749–2762. doi: 10.1105/tpc.107.053728

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., and Gianese, G. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature.* 485 (7400), 635–641. doi: 10.1038/nature11119

Schoonheim, P. J., Sinnige, M. P., Casaretto, J. A., Veiga, H., Bunney, T. D., Quatrano, R. S., et al. (2007). 14-3-3 adaptor proteins are intermediates in ABA signal transduction during barley seed germination. *Plant J.* 49 (2), 289–301. doi: 10.1111/j.1365-313X.2006.02955.x

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95 (11), 5857–5864. doi: 10.1073/pnas.95.11.5857

Sehnke, P. C., Delille, J. M., and Ferl, R. J. (2002a). Consummating signal transduction: the role of 14-3-3 proteins in the completion of signal-induced transitions in protein activity. *Plant Cell.* 14 (81), S339–S354. doi: 10.1105/tpc.010430

Sehnke, P. C., Henry, R., Cline, K., and Ferl, R. J. (2000). Interaction of a plant 14-3-3 protein with the signal peptide of a thylakoid-targeted chloroplast precursor protein and the presence of 14-3-3 isoforms in the chloroplast stroma. *Plant Physio.* 122 (1), 235–241. doi: 10.1104/pp.122.1.235

Sehnke, P. C., Rosenquist, M., Alsterfjord, M., Delille, J., Sommarin, M., Larsson, C., et al. (2002b). Evolution and isoform specificity of plant 14-3-3 proteins. *Plant Mol. Biol.* 50 (6), 1011–1018. doi: 10.1023/a:1021289127519

Shao, W., Chen, W., Zhu, X., Zhou, X., and Qiao, Y. (2021). Genome-wide identification and characterization of wheat 14-3-3 genes unravels the role of TaGRF6-a in salt stress tolerance by binding MYB transcription factor. *Int. J. Mol. Sci.* 22 (4), 1904. doi: 10.3390/ijms22041904

Sun, X., Luo, X., Sun, M., Chen, C., Ding, X., Wang, X., et al. (2013). A glycine soja 14-3-3 protein GsGF14o participates in stomatal and root hair development and drought tolerance in *Arabidopsis thaliana*. *Plant Cell Physio.* 55 (1), 99–118. doi: 10.1093/pcp/pct161

Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nature.* 606 (7914), 535–541. doi: 10.1038/s41586-022-04822-x

Tang, R., Zhu, W., Song, X., Lin, X., Cai, J., Wang, M., et al. (2016). Genome-wide identification and function analyses of heat shock transcription factors in potato. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00490

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16 (1), 86. doi: 10.1186/s12870-016-0771-y

Wang, Q., Liu, J., Yu, W., Zhao, Y., and Cheng, B. (2015). Systematic analysis of the maize PHD-finger gene family reveals a subfamily involved in abiotic stress response. *Int. J. Mol. Sci.* 16 (10), 23517–23544. doi: 10.3390/ijms161023517

Wang, Q., Wang, Y., Chai, W., Song, N., Wang, J., Cao, L., et al. (2017). Systematic analysis of the maize cyclophilin gene family reveals ZmCYP15 involved in abiotic stress response. *Plant Cell Tiss Org.* 128 (3), 543–561. doi: 10.1007/s11240-016-1132-0

Wilson, R. S., Swatek, K. N., and Thelen, J. J. (2016). Regulation of the regulators: Post-translational modifications, subcellular, and spatiotemporal distribution of plant 14-3-3 proteins. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00611

Xia, L., He, X., Huang, X., Yu, H., and Luo, C. (2022). Genome-wide identification and expression analysis of the 14-3-3 gene family in mango (*Mangifera indica l.*). *Int. J. Mol. Sci.* 23 (23), 1593. doi: 10.3390/ijms23031593

Xu, W., Jia, L., Shi, W., Baluˌka, F.e., Kronzucker, H. J., Liang, J., et al. (2013). The tomato 14-3-3 protein TFT4 modulates h+ efflux, basipetal auxin transport, and the PKS5-J3 pathway in the root growth response to alkaline stress. *Plant Physio.* 163 (4), 1817–1828. doi: 10.1104/pp.113.224758

Xu, W. F., and Shi, W. M. (2006). Expression profiling of the 14-3-3 gene family in response to salt stress and potassium and iron deficiencies in young tomato (*Solanum lycopersicum)* roots: analysis by real-time RT-PCR. *Ann. Bot.* 98 (5), 965–974. doi: 10.1093/aob/mcl189

Yaffe, M. B., Rittinger, K., Volinia, S., Caron, P. R., Aitken, A., Leffers, H., et al. (1997). The structural basis for 14-3-3:Phosphopeptide binding specificity. *Cell.* 91 (7), 961–971. doi: 10.1016/S0092-8674(00)80487-0

Yang, X., Wen, H. L., Sobott, F., Papagrigoriou, E., and Elkins, J. M. (2006). Structural basis for protein-protein interactions in the 14-3-3 protein family. *Proc. Natl. Acad. Sci. U.S.A.* 103 (46), 17237–17242. doi: 10.1073/pnas.0605779103

Yao, Y., Ying, D., Lin, J., and Liu, J. Y. (2007). Interaction between ACC synthase 1 and 14-3-3 proteins in rice: a new insight. *Biochemistry.* 72 (9), 1003–1007. doi: 10.1134/s000629790709012x

Yu, T., Li, G., Dong, S., Liu, P., Zhang, J., and Zhao, B. (2016). Proteomic analysis of maize grain development using iTRAQ reveals temporal programs of diverse metabolic processes. *BMC Plant Biol.* 16 (1), 241. doi: 10.1186/s12870-016-0878-1

Zhang, J., Jia, W., Ismail, Y. A. M., and Ismail, Y. A. M. (2006). Role of ABA in integrating plant responses to drought and salt stresses. *Field Crop Res.* 97 (1), 111–119. doi: 10.1016/j.fcr.2005.08.018

Zhou, H., Lin, H., Chen, S., Becker, K., Yang, Y., Zhao, J., et al. (2014). Inhibition of the arabidopsis salt overly sensitive pathway by 14-3-3 proteins. *Plant Cell.* 26 (3), 1166–1182. doi: 10.1105/tpc.113.117069

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell.* 167 (2), 313–324. doi: 10.1016/j.cell.2016.08.029

Zhu, T., Li, L., Feng, L., and Ren, M. (2020). *StABI5* involved in the regulation of chloroplast development and photosynthesis in potato. *Int. J. Mol. Sci.* 21 (3), 1068. doi: 10.3390/ijms21031068

Zuo, X., Wang, S., Xiang, W., Yang, H., Tahir, M. M., Zheng, S., et al. (2021). Genome-wide identification of the 14–3-3 gene family and its participation in floral transition by interacting with TFL1/FT in apple. *BMC Genomics* 22 (1), 41. doi: 10.1186/s12864-020-07330-2

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 16 January 2023
DOI 10.3389/fpls.2022.1100302

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Songlin He,
Henan Agricultural University, China
Wenfang Gong,
Central South University Forestry and Technology, China
Yonghua Li,
Henan Agricultural University, China

*CORRESPONDENCE
Zhiguo Shen
hnslky668@163.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 16 November 2022
ACCEPTED 28 December 2022
PUBLISHED 16 January 2023

CITATION
Shen Z, Ding X, Cheng J, Wu F, Yin H and Wang M (2023) Phylogenetic studies of magnoliids: Advances and perspectives. *Front. Plant Sci.* 13:1100302. doi: 10.3389/fpls.2022.1100302

COPYRIGHT
© 2023 Shen, Ding, Cheng, Wu, Yin and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Phylogenetic studies of magnoliids: Advances and perspectives

Zhiguo Shen[1]*, Xin Ding[1], Jianming Cheng[2], Fangfang Wu[2], Hengfu Yin[3] and Minyan Wang[3]

[1]National Innovation Alliance of Wintersweet, Henan Academy of Forestry, Zhengzhou, China, [2]Scientific Research Department, Scientific Research Department, Henan Colorful Horticulture Co., Ltd, Zhengzhou, China, [3]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, Zhejiang, China

Magnoliids are the largest flowering plant clades outside of the eudicots and monocots, which are distributed worldwide and have high economic, ornamental and ecological values. Eudicots, monocots and magnoliids are the three major clades of Mesangiospermae, and their phylogenetic relationship is one of the most interesting issues. In recent years, with the continuous accumulation of genomic information, the evolutionary status of magnoliids has become a hot spot in plant phylogenetic research. Although great efforts have been made to study the evolution of magnoliids using molecular data from several representative species such as nuclear genome, plastid genome, mitochondrial genome, and transcriptome, the results of current studies on the phylogenetic status of magnoliids are inconsistent. Here, we systematically describe the current understanding of the molecular research on magnoliid phylogeny and review the differences in the evolutionary state of magnoliids. Understanding the research approaches and limitations of magnoliid phylogeny can guide research strategies to further improve the study of the phylogenetic evolution of magnoliids.

KEYWORDS

magnoliids, phylogeny, monocots, eudicots, genome

Introduction

Angiosperms, also known as flowering plants, are the highest and most diverse category of the plant kingdom and have a significant dominance on Earth (Tang et al., 2014; Yang L. et al., 2020; Yang Y. et al., 2020). It has been reported that there are over 35,2000 species of angiosperms (http://www.theplantlist.org/), which are essential sources of oxygen, food, fiber, medicines and other materials for humans and animals (Judd et al., 1999; Tilman et al., 2002). Darwin referred to the phenomenon of rapid origin and species diversity of angiosperms in a relatively short geological period as an "abominable mystery" (Davies et al., 2004; Crepet and Niklas, 2009; Friedman, 2009; Buggs, 2017; Chen et al., 2017). Phylogenetic relationships among organisms are fundamental to evolutionary biology and many other disciplines (Zhang et al., 2012). The establishment of a classification system that truly reflects plant

phylogeny has been an important goal of botanical and evolutionary biology research for the largest group in the plant kingdom, angiosperms (Hsu, 1984; Cronquist, 1988; Takhtajan, 1997).

Angiosperms have long been classified into two major groups: monocotyledons and dicotyledons, according to the four major classification systems of Cronquist, Takhtajan, Engler, and Hutchinson (Engler, 1964; Hutchinson, 1973; Cronquist, 1988; Takhtajan, 1997). With the development of molecular biology, the phylogenetic studies of angiosperms have made amazing progress and the taxonomic perspective of angiosperms has undergone revolutionary changes. The Angiosperm Phylogeny Group (APG), composed of several scholars, proposed the APG system based on molecular data in 1998 (Bremer et al., 1998), which is a new classification system of angiosperms based on cladistics and molecular systematics in three revised versions (Bremer et al., 2003; The APG, 2009; Chase et al., 2016). The APG system has changed the traditional phylogenetic research based on fossil records, species morphology and physiological characteristics, and has had a significant impact on the phylogenetic study of angiosperms. Today, the APG system has become a widely used classification system for angiosperms. In the most recent APG IV classification system, angiosperms are classified as basal angiosperms and Mesangiospermae (Chase et al., 2016). The basal angiosperms (ANA clade) include Amborellales, Nymphaeales and Austrobaileyales (Qiu et al., 1999), and Mesangiospermae include five branches: eudicots, monocots, magnoliids, ceratophyllales, and chloranthales. Among them, eudicots and monocots are the two most abundant groups, accounting for about 75% and 20% of angiosperm species, respectively (Zeng et al., 2014). Magnoliids are the third largest branch with more than 10,000 species, accounting for less than 3% of angiosperm species (http://www.theplantlist.org/). Chloranthales and ceratophyllales are few in number and rare in morphology, with only 77 and 6 species, respectively (Zeng et al., 2014).

Many species of magnoliids are early diverging lineages and play an important role in the study of plant evolution and phylogeny, which can be used to better understand the evolution of extant angiosperms (Massoni et al., 2015; Chen et al., 2019; Shang et al., 2020; Liu et al., 2020; Wu et al., 2021). Moreover, many species have high economic, ornamental and ecological values and are widely distributed worldwide (Massoni et al., 2014; Zeng et al., 2014; Massoni et al., 2015; Dong et al., 2021; Shen et al., 2021). Therefore, magnoliids are of great interest to botanists and plant breeders. Nevertheless, the evolutionary relationships between eudicots, monocots, and magnoliids remain unclear, and differences in topology may reveal the phylogenetic complexity behind the rapid radiation of angiosperms (Soltis and Soltis, 2019). In this paper, the phylogenetic research of magnoliids is reviewed and the potential reasons for differences in the evolutionary state of magnoliids are summarized and discussed, with a view to providing guidance for future research.

Overview of magnoliids

The majority of phylogenetic findings support Magnoliales, Laurales, Canellales and Piperales as a branch of Mesangiospermae with early and rapid differentiation (Soltis et al., 1999; Soltis et al., 2007, Cai et al., 2006; Cantino et al., 2007; Moore et al., 2010; Qiu et al., 2010; Moore et al., 2011; Soltis et al., 2011; Ruhfel et al., 2014), but this branch differs from the Magnoliidae, as defined in the Takhtajan classification system (Takhtajan, 1997) or the Cronquist classification system (Cronquist, 1988). Giulietti et al. (2005) and Cantino et al. (2007) associated the name of Magnoliidae with this branch, while in the APG system (Bremer et al., 2003; The APG, 2009; Chase et al., 2016), with this branch being referred to as magnoliids. This paper follows the name of magnoliids in the APG system, which is equivalent to Magnoliidae in some of the literature (Cantino et al., 2007; Massoni et al., 2014; Massoni et al., 2015).

Magnoliids are the next clades of angiosperms after eudicots and monocots, including some of the "earliest angiosperms" defined in earlier studies (Zeng et al., 2014). Magnoliids have played an important role in the development of human society, with species such as black pepper (*Piper nigrum*), avocado (*Persea americana*), *Litsea cubeba* and *Chimonanthus salicifolius* having high economic value, while other species such as *C. praecox*, *Yulania denudata*, *Magnolia grandiflora*, and *Liriodendron chinense* have high ornamental value. Besides, many organisms (including various butterfly and beetle groups) are highly dependent on this group for feeding or reproduction, which is an important part of the forest ecosystem (Massoni et al., 2015). Magnoliids have morphological characteristics of both eudicots and monocots (Tang et al., 2014). For example, the *L. chinense* flower has three cardinal numbers with single pore pollen grains and exhibits typical monocot characteristics, while the cotyledons and roots show typical eudicot characteristics (Chen et al., 2019).

The phylogeny of magnoliids, monocots and eudicots is related to the early origin and evolution of angiosperms. Clarifying the phylogenetic status of magnoliids will provide a new direction for phytogenetic studies and promote the interpretation of evolutionary mysteries and the disclosure of earth history (Zhang et al., 2022). In recent years, the evolutionary status of magnoliids has become a hot spot for plant phylogenetic studies. Among the existing published studies on the phylogeny of magnoliids, most researchers have tried to explain the phylogenetic status of magnoliids using nuclear genomes and other molecular data (plastid genomes, mitochondrial genomes, transcriptomes, etc.) of several representative species, but the conclusions are inconsistent. In general, the topological structure of the evolutionary relationships between the three clades include the following three types: (1) magnoliids + (eudicot + monocot); (2) monocot + (eudicot + magnoliids); (3) eudicot + (magnoliids + monocot).

Phylogeny based on the sequencing of magnoliids own nuclear genome

Plant cells contain three sets of genomes: the nuclear genome, the plastid genome and the mitochondrial genome. The nuclear whole-genome sequences contain rich genetic information and have greater application potential in phylogenetic research, with potent capabilities to decipher complex phylogenetic models and evolutionary processes, which can deepen our understanding of plant phylogeny and evolution. The reported species of magnoliids with sequenced nuclear genomes include Magnoliales, Laurales, and

Piperales (17 species, 23 references in total). Examples include Laurales: *Cinnamomum kanehirae* (Chaw et al., 2019), avocado (*Persea americana*) (Rendón-Anaya et al., 2019; Nath et al., 2022), *Litsea cubeba* (Chen Y. C. et al., 2020), *Phoebe bournei* (Chen S. P. et al., 2020; Han et al., 2022), *Cinnamomum camphora* (Jiang et al., 2022; Shen et al., 2022; Sun et al., 2022; Wang et al., 2022), *C. burmannii* (Li et al., 2022), *Litsea coreana* (Zhang et al., 2022) of Lauraceae, *Chimonanthus praecox* (Shang et al., 2020; Shen et al., 2021) and *C. salicifolius* (Lv et al., 2020) of Calycanthaceae; Magnoliales: *L. chinense* (Chen et al., 2019), *Magnolia biondii* (Dong et al., 2021), *M. officinalis* (Yin et al., 2021); *Annona muricata* (Strijk et al., 2021), *A. glabra* (He et al., 2022); Pepperales: *Piper nigrum* (Hu et al., 2019), *Aristolochia fimbriata* (Qin et al., 2021) and *A. contorta* (Cui et al., 2022). Twenty of the 23 references in the genome analysis magnoliids discussed the evolutionary status of magnoliids, providing new insights into the early evolution of angiosperms, but the results of the analysis are inconsistent.

Magnoliids form a sister clade to monocots and eudicots

The diversity of rapid formation or differentiation of common ancestors of magnoliids, monocots and eudicots leads to differences in the topological results of phylogenetic trees (incomplete lineage sorting, ILS). Chen et al. (2019) sequenced the nuclear genome of *L. chinense* and deduced three topological structures based on 502 low-copy (no more than two per species) gene trees from 17 species. The species tree was further constructed by the amino acid coalescence approach, and further topological analysis of 78 chloroplast genes and gene families specific to monocots and eudicots in the Liriodendron genome was performed, respectively, which all supported that magnoliids are sister plants to monocots and eudicots. Hu et al. (2019), based on nuclear genome sequencing of *P. nigrum* and 82 single-copy genes identified in 21 species, used the amino acid concatenation approach to support that magnoliids form a sister clade to monocots and eudicots. Rendón-Anaya et al. (2019) sequenced the nuclear genome of avocado and determined 176 single-copy genes from 19 species. Based on protein sequences, avocado is considered to be sister to monocots and eudicots; based on CDS sequences, avocado is sister to monocots; also, based on 4,694 low-copy genes, avocado is considered to be sister to eudicots. Besides, a neighbor-joining tree was generated based on modal dissimilarity scores from thousands of syntenically validated ortholog pairs, indicating that avocado is sister plants to monocots and eudicots. By evaluating the three positions of avocado, the authors concluded that the three different positions of avocado in angiosperms may be indistinguishable for purely biological reasons. However, according to the Akaike information criterion (AIC) comparison based on the free rate (FR) model, avocado is preferred as sister plants to monocots and eudicots. Chen S. P. et al. (2020) sequenced the nuclear genome of *P. bournei* and constructed five evolutionary trees based on 292 single-copy genes from 18 species. Among them, three trees (Bayesian tree, coalescent and concatenation trees based on amino acid sequences) support that magnoliids form a sister clade to monocots and eudicots; two trees (coalescent and concatenation trees based on nucleotide sequences) support that magnoliids are sister to monocots. The authors support the Bayesian tree. Zhang et al. (2022) sequenced the nuclear genome of *L. coreana*. Using 71 single-copy genes of the nuclear genome from 13 species (*Amborella trichopoda* as an outgroup), nucleic acid sequence-based and amino acid sequence-based concatenation trees were constructed to support that magnoliids are sister to eudicots; the constructed amino acid sequence-based coalescent tree supports that magnoliids form a sister clade to monocots and eudicots and that black pepper is closer to monocots and eudicots. The authors concluded that *Magnoliids* are more likely to be the basal species of angiosperms due to the possibility of ILS. Consistently, the results of nuclear genome sequencing analysis of *M. officinalis*, *M. biondii*, *C. camphora*, etc. support that magnoliids form a sister clade to monocots and eudicots (Dong et al., 2021; Yin et al., 2021; Jiang et al., 2022).

Magnoliids are sister to eudicots

Chaw et al. (2019) sequenced the nuclear genome of *C. kanehirae*. Using 211 single-copy genes determined from 13 species, amino acid sequence-based coalescent and concatenation trees support the idea that magnoliids are sister to eudicots. Meanwhile, this topology is also supported by transcriptome data from 22 magnoliids species (although the BS is somewhat low). Similarly, in the genome sequencing analysis of *C. salicifolius*, Lv et al. (2020) constructed an amino acid sequence-based concatenation tree of 103 single-copy gene sets and coalescent tree of 1,420 low-copy gene sets from 17 species, all of which support that magnoliids are sister plants to eudicots. In addition, Shang et al. (2020) used two software (OrthoMCL and SonicParanoid) to identify two single-copy gene sets and construct amino acid sequence-based concatenation and coalescent trees, respectively. A total of four trees support that magnoliids are sister plants to eudicots. However, based on 38 chloroplast single-copy genes from 26 species, the results of amino acid sequence-based concatenation tree support that magnoliids form a sister clade to monocots and eudicots. The authors suggest that the phylogenetic inconsistency between chloroplast genomes and nuclear genomes may be caused by the ILS effect. Furthermore, a concatenation phylogenetic tree was constructed using nucleic acid sequences of 2,420 gene sets from 29 plants (including transcriptome data), again demonstrating that magnoliids are sister plants to eudicots. Therefore, the authors believe that it is relatively accurate that magnoliids are the sister plants to eudicots in the current data set. In the nuclear genomic analysis of the red flower wintersweet, taking into full consideration various factors that may affect the evolutionary position of magnoliids, Shen et al. (2021) constructed concatenation and coalescent trees using nucleic acid and amino acid sequences of 70 single-copy gene families from 25 genomes, as well as phylogeny trees of 123 plants (47 transcripts, 76 genomes) based on the nucleotide sequences of selected low-copy nuclear ortholog groups. The results suggest that magnoliids are more likely to form a sister clade to eudicots, which is supported by more phylogenetic trees. Consistently, the results of nuclear genome sequencing analysis of *C. camphora*, *A. glabra, P. bournei, A. contorta*, *C. burmannii*, etc. support that magnoliids are sister plants to eudicots (Cui et al., 2022; Han et al., 2022; He et al., 2022; Li et al., 2022; Shen et al., 2022; Sun et al., 2022; Wang et al., 2022).

Magnoliids are sister to monocots or the evolutionary relationships remain unresolved

Qin et al. (2021) compared the genome structure of *A. fimbriata* and representative species of major angiosperm groups and placed magnoliids as sister groups of monocots. Chen Y. C. et al. (2020) sequenced the *L. cubeba* nuclear genome, obtained 160 common single-copy gene families of 34 angiosperms from the BUSCO database, and constructed concatenation and coalescent trees using nucleic acid and amino acid sequences. Among them, the amino acid sequence-based coalescent tree supports that magnoliids are sister plants to monocots, and the other three trees support that magnoliids are sister plants to eudicots. Analysis by ASTRAL software suggested that a possible ILS effect on the rapid differentiation of early Mesangiospermae. Based on this, the authors conclude that the evolutionary relationships between magnoliids, monocots, and eudicots remain unresolved.

To sum up, among the 23 references for phylogenetic analysis based on sequencing of magnoliids own nuclear genome, 8 references supported magnoliids as a sister clade to monocots and eudicots, 11 references supported magnoliids as a sister clade to eudicots, and 1 reference supported magnoliids as a sister clade to monocots. In addition, the authors of one reference considered that the evolutionary relationships between magnoliids, monocots and eudicots remain unresolved (Table 1).

Phylogeny of magnoliids based on other molecular data

Over the years, researchers have also integrated plastid, mitochondrial, nuclear genome and transcriptome molecular data from multiple species to analyze the phylogeny of magnoliids and the early diversification of angiosperms. Different phylogenetic relationships have also emerged regarding the status of magnoliids.

Magnoliids form a sister clade to monocots and eudicots

In terms of phylogenetic analysis using plastid genomes, Cai et al. (2006); Moore et al. (2007), and Ruhfel et al. (2014) conducted phylogenetic analysis of 61 plastid protein-coding genes from 35 taxa, 61 plastid protein-coding genes from 45 species, and 78 plastid protein-coding data from 360 species, respectively; Moore et al. (2010) conducted a phylogenetic analysis of 83 protein-coding and rRNA genes from 86 seed plant plastid genomes. Gitzendanner et al. (2018) analyzed the phylogenetic tree of 1,827 green plants and 52 outgroups using 78 plastid protein-coding genes. Li et al. (2019) reconstructed the angiosperm phylogeny based on 80 genes from 2,881 plastid genomes, representing 85% of extant families and all orders. In terms of phylogenetic analysis using the mitochondrial genome, Qiu et al. (2010) performed a phylogenetic analysis of 380 species of seed plants based on four mitochondrial gene sequences. Dong et al. (2020) conducted a phylogenetic analysis based on 38 mitochondrial genes from 91 representative angiosperm species. In addition, Soltis et al. (2011) conducted a two-group analysis of 640 plant species from 330 families. The first group included 17 genes representing all three plant genomes (i.e., nucleus, plastid, and mitochondrion); the second group contained 13 genes (representing only the nucleus and plastid). Jin et al. (2020) constructed 20 phylogenetic trees based on nucleotide and amino acid sequences of five gene sets from 89 plants using concatenation and coalescent approaches. The results of all the above analyses support that magnoliids form a sister clade to monocots and eudicots.

Magnoliids are sister to eudicots

Moore et al. (2011) analyzed the plastid inversion repeat sequences of 244 plants; Zeng et al. (2014) and Puttick et al. (2018) conducted phylogenetic analysis using transcriptome amino acid sequences from 61 and 103 plants, respectively; Wickett et al. (2014) systematically analyzed 852 protein-coding nuclear genes from 103 plants (92 transcriptomes and 11 nuclear genomes). The results of the above analysis support that magnoliids are sister plants to eudicots. In addition, Zhang et al. (2020) constructed a coalescent tree based on five different low-copy gene sets (comprising 1,167, 834, 683, 602, and 445 genes respectively) from 115 plants (44 nuclear genomes and 71 transcriptomes), most of which support that magnoliids are sister plants to eudicots. Guo et al. (2021) constructed phylogenetic trees based on nuclear genome sequencing of *Chloranthus spicatus* using four gene sets (257 single-copy genes, 937 single-copy genes, and 2,329 low-copy genes from 18 plants, and 612 single-copy genes from 218 plants, respectively), supporting the idea that magnoliids are sister to eudicots, while the results of chloroplast gene construction support that magnoliids form a sister clade to monocots and eudicots. The analysis suggested that ancient gene flow between monocots and eudicots might have occurred during the early evolution of angiosperms, resulting in inconsistent phylogenetic branches. In addition, Ma et al. (2021) sequenced the nuclear genome of *C. sessilifolius* and analyzed 1,689 single-copy genes concatenated nucleotide sequences based on nuclear genome data from 14 plants, supporting that magnoliids are sister plants to eudicots. At the same time, the analysis suggests that, in addition to hybridization, ILS may largely explain the observed phylogenetic inconsistencies among gene trees.

Magnoliids are sister to monocots or the evolutionary relationships remain unresolved

Zhao et al. (2021) conducted a phylogenetic analysis based on genome-wide data from 123 plants (covering 31 orders and 52 families); Zhang et al. (2012) constructed a concatenation tree using nucleotide and amino acids based on five low-copy nuclear genes obtained in 94 species; Endress and Doyle (2009) analyzed plastid and morphological data. The results of all these analyses suggest that magnoliids are sister plants to monocots. However, based on 1594, 756, and 296 gene sets from 151 angiosperms (including the five

TABLE 1 Phylogeny based on the sequencing of magnoliids own nuclear genome.

Phylogeny of magnoliids	Author	Species	Order	Family	Journal	Year of publication
Magnoliids form a sister clade to monocots and eudicots (8 references)	Rendón-Anaya et al.	*Persea americana*	Laurales	Lauraceae	*PNAS*	2019
	Chen et al.	*Phoebe bournei*	Laurales	Lauraceae	*Hortic. Res.*	2020
	Yin et al.	*Magnolia officinalis*	Magnoliales	Magnoliaceae	*iScience*	2021
	Dong et al.	*Magnolia biondii*	Magnoliales	Magnoliaceae	*Hortic. Res.*	2021
	Chen et al.	*Liriodendron chinense*	Magnoliales	Magnoliaceae	*Nat. Plants*	2019
	Hu et al.	*Piper nigrum*	Piperales	Piperaceae	*Nat. Commun.*	2019
	Jiang et al	*Cinnamomum camphora*	Laurales	Lauraceae	*Front. plant sci.*	2022
	Zhang et al	*Litsea coreana*	Laurales	Lauraceae	*Genomics*	2022
Magnoliids are sister plants to eudicots (11 references)	Chaw et al.	*Cinnamomum kanehira*	Laurales	Lauraceae	*Nat. Plants*	2019
	Sun et al.	*Cinnamomum camphora*	Laurales	Lauraceae	*J. Genet. Genomics*	2022
	Li et al.	*Cinnamomum burmannii*	Laurales	Lauraceae	*Ind. Crop. Prod.*	2022
	Lv et al.	*C. salicifolius*	Laurales	Calycanthaceae	*Plant J.*	2020
	Shang et al.	*Chimonanthus praecox*	Laurales	Calycanthaceae	*Genome Biol.*	2020
	Shen et al.	*C. praecox* 'Hongyun'	Laurales	Calycanthaceae	*Plant J*	2021
	Cui et al.	*Aristolochia contorta*	Piperales	Aristolochiaceae	*Hortic. Res.*	2022
	He et al	*Annona glabra*	Magnoliales	Annonaceae	*Nat. Eco. & Evol.*	2022
	Han et al.,	*Phoebe bournei*	Laurales	Lauraceae	*Plant Commun.*	2022
	Wang et al	*Cinnamomum camphora*	Laurales	Lauraceae	*Hortic. Res.*	2022
	Shen et al	*Cinnamomum camphora*	Laurales	Lauraceae	*Plant Biotechnol. J.*	2022
Magnoliids are sister plants to monocots (1 reference)	Qin et al.	*Aristolochia fimbriata*	Piperales	Aristolochiaceae	*Nat. Plants*	2021
The evolutionary relationships remain unresolved (1 reference)	Chen et al.	*Litsea cubeba*	Laurales	Lauraceae	*Nat. Commun.*	2020
No phylogenetic analysis of magnoliids (2 references)	Strijk et al.	*Annona muricata*	Magnoliales	Annonaceae	*Mol. Ecol. Resour.*	2021
	Nath et al.	*Persea americana*	Laurales	Lauraceae	*Hortic. Res.*	2022

major branches of the core angiosperms), Yang L. et al. (2020) employed both coalescent and concatenation approaches to infer phylogenetic trees of angiosperms. The authors believe that a fully bifurcated species tree may not be the best way to represent the early differentiation of angiosperms.

To sum up, among the 21 references on the phylogenetic analysis of magnoliids, 10 support that magnoliids form a sister clade to monocots and eudicots, 7 support that magnoliids form a sister clade to eudicots, 3 support that magnoliids form a sister clade to monocots, and one believes that the evolutionary relationships between magnoliids, monocots, and eudicots remain unresolved (Table 2).

Summary and perspectives

Magnoliids have important economic, ornamental and ecological values (Massoni et al., 2014; Massoni et al., 2015; Shen et al., 2021; Dong et al., 2021). They are also valuable materials for studying the origin, development and evolution of angiosperms (Massoni et al., 2015; Chen et al., 2019; Liu et al., 2020; Shang et al., 2020; Wu et al., 2021). Despite the large number of studies reporting the phylogenetic status of magnoliids, the evolutionary relationships between eudicots, monocots, and magnoliids remain inconsistent (Tables 1, 2). Long-branch attraction is a major obstacle to phylogenetic reconstruction, which may lead to the wrong inference of distantly

TABLE 2 Phylogeny of magnoliids based on other molecular data.

Phylogeny of magnoliids	Author	Journal	Year of publication
Magnoliids form a sister clade to monocots and eudicots (10 references)	Cai et al.	*BMC Evol. Biol.*	2006
	Moore et al.	*PNAS*	2007
	Moore et al.	*PNAS*	2010
	Soltis et al.	*Am. J. Bot.*	2011
	Qiu et al.	*J. Syst. & Evol.*	2010
	Ruhfel et al.	*BMC Evol. Biol.*	2014
	Li et al.	*Nat. Plants*	2019
	Gitzendanner et al.	*Am. J. Bot.*	2018
	Dong et al.	*PloS one*	2020
	Jin et al.	*Guihaia*	2020
Magnoliids are sister plants to eudicots (7 references)	Moore et al.	*Inter. J. Plant Sci.*	2011
	Zeng et al.	*Nat. Commun.*	2014
	Wickett et al.	*PNAS*	2014
	Puttick et al.	*Curr. Biol.*	2018
	Zhang et al.	*Nature*	2020
	Sun et al.	*J. Genet. genomics*	2022
	Ma et al.	*Nat. Commun.*	2021
Magnoliids are sister plants to monocots (3 references)	Zhao et al.	*Nat. Commun.*	2021
	Zhang et al.	*New Phytol.*	2012
	Endress et al.	*Am. J. Bot.*	2009
The evolutionary relationships remain unresolved (1 reference)	Yang et al.	*Plant Commun.*	2020

related lineages as close relatives (Qu et al., 2017; Shen et al., 2021). Meanwhile, ILS is the result of allele polymorphism in ancestral populations (Chen Y. C. et al., 2020). Many plant species have a century-long growth period, large population sizes, and limited interspecific differences. These factors have generated an important evolutionary network, which is deeply affected by the ILS. In addition, more attention should be paid to methodological choices in phylogenomic analysis, where the same data set may yield conflicting results (Guo et al., 2022). Different tree-building methods may be important factors contributing to the different evolutionary positions of magnoliids (Rendón-Anaya et al., 2019; Chen S.P. et al., 2020; Chen Y. C. et al., 2020; Shen et al., 2021). It is precisely because of the different tree-building methods, the existence of ILS effects, the number of orthologous genes, the limitation of numerical selection in different groups (Bergsten, 2005; Wiens, 2005), and the rapid differentiation of magnoliids in the early evolutionary stage that the results of research on the evolutionary status of magnoliids are different.

A fully resolved and well-supported phylogeny is of great significance for understanding the evolutionary history of magnoliids. Based on the comprehensive analysis of existing research results, how to adopt a more scientific strategy to analyze the phylogeny of magnoliids is a key consideration for future research on the evolution of magnoliids. For a long time, a large number of valuable plant species have not been sequenced due to the cost of sequencing and the complexity of the species' own genomes. In particular, there are still few genome sequencing samples of magnoliids, which also hinders the in-depth study of these issues to a certain extent. With the rapid development of sequencing technology and the reduction of sequencing cost, an increasing number of plant genome sequencing data will be published, especially more genomic data of magnoliids will be deciphered, and with the more mature means of phylogenetic research, it is believed that in the near future, there will be an industry-recognized result on the phylogenetic status of magnoliids.

Author contributions

ZS was the designer and principal of the project, and drafted the manuscript. XD, JC, and FW were responsible for collecting, sorting, and analyzing some relevant references. HY and MW contributed to the manuscript revision and read the submitted version. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by Henan Key R&D Project, China (Grant No. 221111110700), the Science and Technology Program on Developing Forestry of Henan, China (Grant No. YCHZ[2022]28), and the Fundamental Research Funds of Henan Academy of Forestry, China (Grant Nos. 2021JB01001 and 210612240).

Conflict of interest

Authors JC and FW were employed by Henan Colorful Horticulture Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics Int. J. Willi Hennig Soc.* 21 (2), 163–193. doi: 10.1111/j.1096-0031.2005.00059.x

Bremer, B., Bremer, K., Chase, M. W., Reveal, J. L., Soltis, D. E., and Soltis, P. S. (2003). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II [Review]. *Bot. J. Linn. Soc* 141 (4), 399–436. doi: 10.1111/j.1095-8339.2009.00996.x

Bremer, K., Chase, M. W., Stevens, P. F., Anderberg, A. A., Backlund, A., and Bremer, B. (1998). An ordinal classification for the families of flowering plants. *Ann. Mo. Bot. Gard.* 85 (4), 531–553. doi: 10.2307/2992015

Buggs, R. J. A. (2017). The deepening of darwin's abominable mystery. *Nat. Eco. Evol.* 1 (6), 169. doi: 10.1038/s41559-017-0169

Cai, Z., Penaflor, C., Kuehl, J. V., Leebens-Mack, J., Carlson, J. E., dePamphilis, C. W., et al. (2006). Complete plastid genome sequences of drimys, liriodendron, and piper: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* 6, 77. doi: 10.1186/1471-2148-6-77

Cantino, P. D., Doyle, J. A., Graham, S. W., Judd, W. S., Olmstead, R. G., Soltis, D. E., et al. (2007). Towards a phylogenetic nomenclature of tracheophyta. *Taxon* 56 (3), E1–E44. doi: 10.1002/tax.563001

Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc* 181 (1), 1–20. doi: 10.1111/boj.12385

Chaw, S. M., Liu, Y. C., Wu, Y. W., Wang, H. Y., Lin, C. I., Wu, C. S., et al. (2019). Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat. Plants* 5 (1), 63–73. doi: 10.1038/s41477-018-0337-0

Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., et al. (2019). Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nat. Plants* 5 (1), 18–25. doi: 10.1038/s41477-018-0323-6

Chen, F., Liu, X., Yu, C., Chen, Y., Tang, H., and Zhang, L. (2017). Water lilies as emerging models for darwin's abominable mystery. *Hortic. Res.* 4, 17051. doi: 10.1038/hortRes.2017.51

Chen, Y. C., Li, Z., Zhao, Y. X., Gao, M., Wang, J. Y., Liu, K. W., et al. (2020). The *Litsea* genome and the evolution of the laurel family. *Nat. Commun.* 11 (1), 1675. doi: 10.1038/s41467-020-15493-5

Chen, S. P., Sun, W. H., Xiong, Y. F., Jiang, Y. T., Liu, X. D., Liao, X. Y., et al. (2020). The phoebe genome sheds light on the evolution of magnoliids. *Hortic. Res.* 7, 146. doi: 10.1038/s41438-020-00368-z

Crepet, W. L., and Niklas, K. J. (2009). Darwin's second 'abominable mystery': Why are there so many angiosperm species? *Am. J. Bot.* 96 (1), 366–381. doi: 10.3732/ajb.0800126

Cronquist, A. (1988). *The evolution and classification of flowering plants. 2nd ed* (New York: New York Botanical Garden).

Cui, X., Meng, F., Pan, X., Qiu, X., Zhang, S., Li, C., et al. (2022). Chromosome-level genome assembly of *Aristolochia contorta* provides insights into the biosynthesis of benzylisoquinoline alkaloids and aristolochic acids. *Hortic. Res.* 9, uhac005. doi: 10.1093/hr/uhac005

Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E., and Savolainen, V. (2004). Darwin's abominable mystery: Insights from a supertree of the angiosperms. *PNAS* 101 (7), 1904–1909. doi: 10.1073/pnas.0308127100

Dong, S., Chen, L., Liu, Y., Wang, Y., Zhang, S., Yang, L., et al. (2020). The draft mitochondrial genome of magnolia biondii and mitochondrial phylogenomics of angiosperms. *PloS One* 15 (4), e0231020. doi: 10.1371/journal.pone.0231020

Dong, S., Liu, M., Liu, Y., Chen, F., Yang, T., Chen, L., et al. (2021). The genome of magnolia *biondii pamp.* provides insights into the evolution of magnoliales and biosynthesis of terpenoids. *Hortic. Res.* 8 (1), 38. doi: 10.1038/s41438-021-00471-9

Endress, P. K., and Doyle, J. A. (2009). Reconstructing the ancestral angiosperm flower and its initial specializations. *Am. J. Bot.* 96 (1), 22–66. doi: 10.3732/ajb.0800047

Engler, A. (1964). *Syllabus of plant families, volume 2: Angiosperms*. (Berlin-Nikolassee: Gebr. Borntraeger).

Friedman, W. E. (2009). The meaning of darwin's 'abominable mystery'. *Am. J. Bot.* 96 (1), 5–21. doi: 10.3732/ajb.0800150

Gitzendanner, M. A., Soltis, P. S., Wong, G. K., Ruhfel, B. R., and Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Am. J. Bot.* 105 (3), 291–301. doi: 10.1002/ajb2.1048

Giulietti, A. M., Harley, R. M., Queiroz, L., Graas, M. D., and Berg, C. (2005). Biodiversity and conservation of plants in brazil. *Conserv. Biol.* 19 (3), 632–639. doi: 10.1111/j.1523-1739.2005.00704.x

Guo, X., Fang, D., Sahu, S. K., Yang, S., Guang, X., Folk, R., et al. (2021). Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* 12 (1), 6930. doi: 10.1038/s41467-021-26922-4

Guo, C., Luo, Y., Gao, L. M., Yi, T. S., Li, H. T., Yang, J. B., et al. (2022). Phylogenomics and the flowering plant tree of life. *J. Integr. Plant Biol.* doi: 10.1111/jipb.13415

Han, X., Zhang, J., Han, S., Chong, S. L., Meng, G., Song, M., et al. (2022). The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Commun.* 3 (6), 100410. doi: 10.1016/j.xplc.2022.100410

He, Z., Feng, X., Chen, Q., Li, L., Li, S., Han, K., et al. (2022). Evolution of coastal forests based on a full set of mangrove genomes. *Nat. Eco. Evol.* 6 (6), 738–749. doi: 10.1038/s41559-022-01744-9

Hsu, B. S. (1984). Present day aspects and prospects of the study of angiosperms phylogeny. *Acta Bot. Yunnanica* 1), 1–10. doi: CNKI:SUN:YOKE.0.1984-01-000

Hutchinson, J. (1973). *The families of flowering plants* (Oxford: Oxford University Press).

Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat. Commun.* 10 (1), 4702. doi: 10.1038/s41467-019-12607-6

Jiang, R., Chen, X., Liao, X., Peng, D., Han, X., Zhu, C., et al. (2022). A chromosome-level genome of the camphor tree and the underlying genetic and climatic factors for its top-geoherbalism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.827890

Jin, X., Cheng, S., Yang, T., Yu, K., Duan, X., Ni, X., et al. (2020). Reconstruction of angiosperm phylogeny based on 5993 nuclear genes. *Guihaia* 40 (1), 44–59. doi: 10.11931/guihaia.gxzw201905048

Judd, W., Campbell, C., Kellogg, E., and Stevens, P. (1999). Plant systematics: A phylogenetic approach. *Taxon* 49 (1), 137–138. doi: 10.2307/1223950

Li, F., Huang, S., Mei, Y., Wu, B., Hou, Z., Zhan, P., et al. (2022). Genome assembly provided new insights into the *Cinnamomum burmannii* evolution and d-borneol biosynthesis differences between chemotypes. *Ind. Crop Prod.* 186, 115181. doi: 10.1016/j.indcrop.2022.115181

Liu, X., Liao, X. Y., Zheng, Y., Zhu, M. J., Yu, X., Jiang, Y. T., et al. (2020). Genome-wide identification of the YABBY gene family in seven species of magnoliids and expression analysis in *Litsea*. *Plants* 10 (1), 21. doi: 10.3390/plants10010021

Li, H., Yi, T., Gao, L., Ma, P., Zhang, T., Yang, J., et al. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* 5 (5), 461–470. doi: 10.1038/s41477-019-0421-0

Lv, Q., Qiu, J., Liu, J., Li, Z., Zhang, W., Wang, Q., et al. (2020). The *Chimonanthus salicifolius* genome provides insight into magnoliid evolution and flavonoid biosynthesis. *Plant J.* 103 (5), 1910–1923. doi: 10.1111/tpj.14874

Massoni, J., Couvreur, T. L., and Sauquet, H. (2015). Five major shifts of diversification through the long evolutionary history of magnoliidae (angiosperms). *BMC Evol. Biol.* 15, 49. doi: 10.1186/s12862-015-0320-6

Massoni, J., Forest, F., and Sauquet, H. (2014). Increased sampling of both genes and taxa improves resolution of phylogenetic relationships within magnoliidae, a large and early-diverging clade of angiosperms. *Mol. Phylogenet. Evol.* 70, 84–93. doi: 10.1016/j.ympev.2013.09.010

Ma, J., Sun, P., Wang, D., Wang, Z., Yang, J., Li, Y., et al. (2021). The *Chloranthus sessilifolius* genome provides insight into early diversification of angiosperms. *Nat. Commun.* 12 (1), 6929. doi: 10.1038/s41467-021-26931-3

Moore, M. J., Bell, C. D., Soltis, P. S., and Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *PNAS* 104 (49), 19363–19368. doi: 10.1073/pnas.0708072104

Moore, M. J., Hassan, N., Gitzendanner, M. A., Bruenn, R. A., Croley, M., Vandeventer, A., et al. (2011). Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *Inter. J. Plant Sci.* 172 (4), 541–558. doi: 10.1086/658923

Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *PNAS* 107 (10), 4623–4628. doi: 10.1073/pnas.0907801107

Nath, O., Fletcher, S. J., Hayward, A., Shaw, L. M., Masouleh, A. K., Furtado, A., et al. (2022). A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. *Hortic. Res.* 9, uhac157. doi: 10.1093/hr/uhac157

Puttick, M. N., Morris, J. L., Williams, T. A., Cox, C. J., Edwards, D., Kenrick, P., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* 28 (5), 733–745.e2. doi: 10.1016/j.cub.2018.01.063

Qin, L., Hu, Y., Wang, J., Wang, X., Zhao, R., Shan, H., et al. (2021). Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nat. Plants* 7 (9), 1239–1253. doi: 10.1038/s41477-021-00990-2

Qiu, Y. L., Chen, Z. D., Libo, L. I., Wang, B., Xue, J. Y., Hendry, T. A., et al. (2010). Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *J. Syst. Evol.* 48 (6), 391–425. doi: 10.1111/j.1759-6831.2010.00097.x

Qiu, Y. L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., et al. (1999). The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402 (6760), 404–407. doi: 10.1038/46536

Qu, X. J., Jin, J. J., Chaw, S. M., Li, D. Z., and Yi, T. S. (2017). Multiple measures could alleviate long-branch attraction in phylogenomic reconstruction of cupressoideae (Cupressaceae). *Sci. Rep.* 7, 41005. doi: 10.1038/srep41005

Rendón-Anaya, M., Ibarra-Laclette, E., Méndez-Bravo, A., Lan, T., Zheng, C., Carretero-Paulet, L., et al. (2019). The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *PNAS* 116 (34), 17081–17089. doi: 10.1073/pnas.1822129116

Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14, 23. doi: 10.1186/1471-2148-14-23

Shang, J., Tian, J., Cheng, H., Yan, Q., Li, L., Jamal, A., et al. (2020). The chromosome-level wintersweet (*Chimonanthus praecox*) genome provides insights into floral scent biosynthesis and flowering in winter. *Genome Biol.* 21 (1), 200. doi: 10.1186/s13059-020-02088-y

Shen, Z., Li, W., Li, Y., Liu, M., Cao, H., Provart, N., et al. (2021). The red flower wintersweet genome provides insights into the evolution of magnoliids and the molecular mechanism for tepal color development. *Plant J.* 108 (6), 1662–1678. doi: 10.1111/tpj.15533

Shen, T., Qi, H., Luan, X., Xu, W., Yu, F., Zhong, Y., et al. (2022). The chromosome-level genome sequence of the camphor tree provides insights into lauraceae evolution and terpene biosynthesis. *Plant Biotechnol. J.* 20 (2), 244–246. doi: 10.1111/pbi.13749

Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* 98 (4), 704–730. doi: 10.3732/ajb.1000404

Soltis, D. E., and Soltis, P. S. (2019). Nuclear genomes of two magnoliids. *Nat. Plants* 5 (1), 6–7. doi: 10.1038/s41477-018-0344-1

Soltis, P. S., Soltis, D. E., and Chase, M. W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402 (6760), 402–404. doi: 10.1038/46528

Soltis, D. E., Gitzendanner, M. A., and Soltis, P. S. (2007). A 567-taxon data set for angiosperms: the challenges posed by Bayesian analyses of large data sets. *Int. J. Plant Sci* 168, 137157. doi: 10.1086/509788

Strijk, J. S., Hinsinger, D. D., Roeder, M. M., Chatrou, L. W., Couvreur, T., Erkens, R., et al. (2021). Chromosome-level reference genome of the soursop (*Annona muricata*): A new resource for magnoliid research and tropical pomology. *Mol. Ecol. Resour.* 21 (5), 1608–1619. doi: 10.1111/1755-0998.13353

Sun, W. H., Xiang, S., Zhang, Q. G., Xiao, L., Zhang, D., Zhang, P., et al. (2022). The camphor tree genome enhances the understanding of magnoliid evolution. *J. Genet. Genomics* 49 (3), 249–253. doi: 10.1016/j.jgg.2021.11.001

Takhtajan, A. (1997). *Diversity and classification of flowering plants* (New York: NY: Columbia Univ. Press).

Tang, H., Lyons, E., and Schnable, J. C. (2014). Early history of the angiosperms. *Adv. Bot. Res.* 69, 195–222. doi: 10.1016/B978-0-12-417163-3.00008-1

The APG (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc* 161 (2), 105–121. doi: 10.1111/boj.12385

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature* 418 (6898), 671–677. doi: 10.1038/nature01014

Wang, X. D., Xu, C. Y., Zheng, Y. J., Wu, Y. F., Zhang, Y. T., Zhang, T., et al. (2022). Chromosome-level genome assembly and resequencing of camphor tree (*Cinnamomum camphora*) provides insight into phylogeny and diversification of terpenoid and triglyceride biosynthesis of *Cinnamomum. Hortic. Res.* 9, uhac216. doi: 10.1093/hr/uhac216

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *PNAS* 111 (45), E4859–E4868. doi: 10.1073/pnas.1323926111

Wiens, J. J. (2005). Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst. Biol.* 54 (5), 731–742. doi: 10.1080/10635150500234583

Wu, J. Y., Xue, J. Y., and Van de Peer, Y. (2021). Evolution of NLR resistance genes in magnoliids: Dramatic expansions of CNLs and multiple losses of TNLs. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.777157

Yang, L., Su, D., Chang, X., Foster, C., Sun, L., Huang, C. H., et al. (2020). Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant Commun.* 1 (2), 100027. doi: 10.1016/j.xplc.2020.100027

Yang, Y., Sun, P., Lv, L., Wang, D., Ru, D., Li, Y., et al. (2020). Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants* 6 (3), 215–222. doi: 10.1038/s41477-020-0594-6

Yin, Y., Peng, F., Zhou, L., Yin, X., Chen, J., Zhong, H., et al. (2021). The chromosome-scale genome of *Magnolia officinalis* provides insight into the evolutionary position of magnoliids. *iScience* 24 (9), 102997. doi: 10.1016/j.isci.2021.102997

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* 54, 956. doi: 10.1038/ncomms5956

Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., et al. (2020). The water lily genome and the early evolution of flowering plants. *Nature* 577 (7788), 79–84. doi: 10.1038/s41586-019-1852-5

Zhang, B., Yao, X., Chen, H., and Lu, L. (2022). High-quality chromosome-level genome assembly of *Litsea coreana* l. provides insights into magnoliids evolution and flavonoid biosynthesis. *Genomics* 114 (4), 110394. doi: 10.1016/j.ygeno.2022.110394

Zhang, N., Zeng, L., Shan, H., and Ma, H. (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* 195 (4), 923–937. doi: 10.1111/j.1469-8137.2012.04212.x

Zhao, T., Zwaenepoel, A., Xue, J. Y., Kao, S. M., Li, Z., Schranz, M. E., et al. (2021). Whole-genome microsynteny-based phylogeny of angiosperms. *Nat. Commun.* 12 (1), 3498. doi: 10.1038/s41467-021-23665-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fpls.2023.1137764

Check for updates

OPEN ACCESS

EDITED BY
Hui Song,
Qingdao Agricultural University, China

REVIEWED BY
Ruochen Wang,
Sichuan University, China
Qingtao Wang,
Hebei University of Engineering, China

*CORRESPONDENCE
Rui Dong
rdong@gzu.edu.cn
Chao Chen
cc690076748@163.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 04 January 2023
ACCEPTED 09 January 2023
PUBLISHED 27 January 2023

CITATION
Lu Z, Yang Z, Tian Z, Gui Q, Dong R and Chen C (2023) Genome-wide analysis and identification of microRNAs in *Medicago truncatula* under aluminum stress. *Front. Plant Sci.* 14:1137764. doi: 10.3389/fpls.2023.1137764

COPYRIGHT
© 2023 Lu, Yang, Tian, Gui, Dong and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide analysis and identification of microRNAs in *Medicago truncatula* under aluminum stress

Zhongjie Lu[1], Zhengyu Yang[2], Zheng Tian[1], Qihui Gui[1], Rui Dong[1]* and Chao Chen[1]*

[1]Department of Grassland Science, College of Animal Science, Guizhou University, Guiyang, China, [2]Department of Vehicle Engineering, Guizhou Technological College of Machinery and Electricity, Duyun, China

Numerous studies have shown that plant microRNAs (miRNAs) play key roles in plant growth and development, as well as in response to biotic and abiotic stresses; however, the role of miRNA in legumes under aluminum (Al) stress have rarely been reported. Therefore, here, we aimed to investigate the role of miRNAs in and their mechanism of Al tolerance in legumes. To this end, we sequenced a 12-strand-specific library of *Medicago truncatula* under Al stress. A total of 195.80 M clean reads were obtained, and 876 miRNAs were identified, of which, 673 were known miRNAs and 203 were unknown. A total of 55 miRNAs and their corresponding 2,502 target genes were differentially expressed at various time points during Al stress. Further analysis revealed that mtr-miR156g-3p was the only miRNA that was significantly upregulated at all time points under Al stress and could directly regulate the expression of genes associated with root cell growth. Three miRNAs, novel_miR_135, novel_miR_182, and novel_miR_36, simultaneously regulated the expression of four Al-tolerant transcription factors, GRAS, MYB, WRKY, and bHLH, at an early stage of Al stress, indicating a response to Al stress. In addition, legume-specific miR2119 and miR5213 were involved in the tolerance mechanism to Al stress by regulating F-box proteins that have protective effects against stress. Our results contribute to an improved understanding of the role of miRNAs in Al stress in legumes and provide a basis for studying the molecular mechanisms of Al stress regulation.

KEYWORDS

plant miRNA, *Medicago truncatula*, aluminum stress, transcript analysis, plant growth

1 Introduction

Aluminum (Al) toxicity is one of the most important factors limiting crop growth and production in acidic soils. Approximately 40% of global arable land is acidic, which is one of the main factors contributing to food shortages in regions such as those in Africa (Daspute et al., 2017). Al^{3+} is present in naturally acidic soils and is also the most harmful form of Al to plants, with micromolar concentrations of Al^{3+} inhibiting root elongation in a short period of time (Gui

et al., 2022). Chandran et al. (2008) stained the root tips of *Medicago truncatula* seedlings grown at 10-μm Al concentration with hematoxylin and showed that the root tips treated with Al for the longest time stained the darkest, indicating that they were the most severely damaged. Li et al. (2020) found that Al stress significantly enhanced the expression of *MsPG1* in the plasma membrane of *Medicago sativa* root apical epidermal cells, reduced the accumulation of Al in the cell wall, and improved the Al tolerance of *M. sativa*. However, some plants grown for a long time in acidic soils can chelate Al^{3+} in the vicinity of the root zone with organic acids such as malic, citric, and oxalic acids secreted by the root system, forming Al-organic acid complexes, and thus reducing the transport of Al to the interior of the root cell and enhancing Al tolerance (Chauhan et al., 2021).

MicroRNAs are a subset of the major non-coding RNAs in higher plants. They are typically 19–24 nucleotides long, and they are thought to post-transcriptionally regulate the cleavage of target mRNAs or repress their translation (Bartel, 2004; Kumar et al., 2017; Xu et al., 2019). miRNAs play key roles in gene expression, stress responses, growth and development, and other regulatory mechanisms in plants and animals (Zhang et al., 2015; Pegler et al., 2019). Liu et al. (2014) studied 22 conserved miRNA families in *Zea mays* and found that 72 genes targeted by 62 differentially expressed miRNAs may regulate maize ear development. Wang et al. (2011) identified two miRNAs, cbr-mir-241 and ath-miR854a, using microarray technology, that can directly regulate *Glycine max* resistance to blast rot through their targets (enzymes). Hoang et al. (2020) found that miRNAs play an important role in various processes of symbiotic nitrogen fixation with rhizobia in four legumes: *Lotus japonicus*, *M. truncatula*, *G. max*, and *Phaseolus vulgaris*.

Medicago truncatula is an annual legume that has been used as a model plant for legumes as it is a close relative of *M. sativa* and has the advantage of having a small genome, high similarity, and being a diploid plant (Chandran et al., 2008). In recent years, although researchers have conducted numerous studies on Al stress in plants, miRNA studies related to Al stress in legumes have rarely been reported. Therefore, we aimed to investigate the role of miRNAs in and their mechanism of Al tolerance in *M. truncatula*. High-throughput sequencing of root tip tissues under Al stress was performed to identify miRNAs associated with the Al stress response. The findings of this study provide new insights into the potential functions of miRNAs in the Al response mechanism.

2 Materials and methods

2.1 Plant materials and processing

Medicago truncatula A17 seeds were surface-disinfected with 1% NaClO solution for 10 min and rinsed five times in distilled water to remove residual disinfectant solution, then incubated in an artificial climate chamber at 25°C for 3 days protected from light, and then transferred to Hoagland's medium (pH 5.8) for 7 days (25°C, light/dark cycle of 16/8 h). The composition of Hoagland's culture solution is consistent with the experiments of Stephan and Procházka (1989).The nutrient solution was changed every 2 days during incubation. Seedlings with similar growth levels were divided into four groups, three of which were incubated in 10 μM $AlCl_3$ and 0.5 mM $CaCl_2$ (pH 4.5) solutions for 4, 24, and 48 h, respectively, and were recorded as T4, T24, and T48, respectively. The control group (T0) was incubated in a 0.5 mM $CaCl_2$ (pH 4.5) solution for 48 h. To treat the four groups of seedlings, two groups of seedlings, T0 and T48, were incubated simultaneously. After 24 h and 44 h of incubation of the above two groups of seedlings, respectively, the seedlings of T24 and T4 groups were incubated, and the four groups of seedlings were harvested simultaneously at 48 h to reduce the circadian effect. Three replicates were set up for each group, with 60 seedlings per replicate. After treatment, root tips of ~ 1.5 cm from each treated plant were collected, immediately frozen in liquid nitrogen, and stored at –80 °C.

2.2 Physiological indicators and fluorescein diacetate staining test

The activities of three enzymes, superoxide dismutase (SOD, SOD-BC0170), catalase (CAT, CAT-BC0200), and peroxidase (POD, POD-BC0090), were measured in four sets of samples using kits provided by Beijing Solarbio Science and Technology Co., Ltd. (Beijing, China). The specific operational steps of the test are presented in the instruction manual.

An image scanner (Perfection V800 photo, Epson, Suwa, Japan) with a WinRhizo root analysis system (WinRhizo Tron Pro 2009, Regent Instruments Inc., Quebec, Canada) was used to measure the root length and root surface area of four sets of samples (Pang et al., 2018). When treating root length and root surface area, we included a control group of plant material grown under Al-free conditions in a 0.5 mM $CaCl_2$ solution (pH 4.5) for the same duration as that of the corresponding Al treatment. The root vigor assay was performed using the naphthylamine method (BC5295, Beijing Solarbio Science and Technology Co., Ltd., Beijing, China). Sample root tips were treated in 2 g/mL FDA solution protected from light for 10 min according to the method of Ishikawa and Wagatsum (1998). The treated root tips were repeatedly rinsed at least five times with deionized water, followed by observation and photography using a fluorescence microscope (Axiolab5, ZEISS, Germany). Each sample was repeated three times. The aforementioned experimental operations were performed under dark conditions.

2.3 RNA extraction, library preparation of sRNA, and sequencing

RNA samples were extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) and tested for concentration, purity, and integrity (Liu et al., 2016). The resulting 3′ sRNA and 5′ sRNA were ligated for splicing. The first strand was synthesized by reverse transcription. Finally, polymerase chain reaction (PCR) amplification and size selection were performed. The target fragments were screened by PAGE, and the cut gels were recovered as fragments to obtain sRNA libraries. Finally, the PCR products were purified (AMPure XP system) and the library quality was assessed.

The resulting libraries were sequenced on the Illumina NovaSeq 6,000 platform from Biomarker Co., Ltd. (BMKcloud, Beijing, China). Clean data were obtained by removing reads containing adapters, poly-N, and low-quality reads from the raw data (Zhao et al., 2020). Using Bowtie tools, clean reads were aligned against Silva and other databases to screen

for other sRNAs, such as ribosomal RNA (rRNA), which were then eliminated. Finally, the unannotated reads were sequenced against the reference genome Medicago_truncatula.Mt4.0v2 using Bowtie2 (v1.0.0) software to obtain information on the position of the reads in the reference genome (Pertea et al., 2015).

2.4 Identification of miRNAs and prediction of new miRNAs

The reads that were aligned to the reference genome were then compared to the mature sequences of known miRNAs in the miRBase (v22) database and their upstream 2nt and downstream 5nt ranges, and the identified reads were considered known miRNAs. As miRNA precursors have a signature hairpin structure, the formation of the mature body is achieved by shearing of the Dicer/DCL enzyme. Based on these features, the miRDeep2 software package (Friedlander et al., 2012) was used, and the final prediction of new miRNAs was achieved by scoring using a Bayesian model (Zhang, 2015).

2.5 Analysis of differentially expressed miRNAs

Differential expression analysis of the two conditions/groups was performed using the DESeq2 R package (1.10.1) (Love et al., 2014). The resulting *P-values* were adjusted using Benjamini and Hochberg's approach to control for the false discovery rate. miRNAs with $|\log2(FC)| \geq 0.58$; *P-value* ≤ 0.05, found by DESeq2, were assigned as differentially expressed. Fold Change (FC) indicates the ratio of expression between two samples (groups).

2.6 Prediction, annotation, and functional analysis of DE miRNAs target genes

Target gene prediction was performed using TargetFinder (v1.6) software based on the gene sequence information of known miRNAs and newly predicted miRNAs in the corresponding species (Allen et al., 2005). The predicted target gene sequences were compared with the NCBI non-redundant protein sequence (Nr), SwissProt Protein databases, Gene Ontology database (GO), Clusters of Orthologous Groups of proteins (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Protein homology (KOG), Evolutionary genealogy of genes (eggNOG), and protein family (Pfam) (*E-value* $< 10^{-5}$) (Yu et al., 2022) databases using BLAST (v2.2.26) software to obtain the annotation information of the target genes. GO enrichment and KEGG enrichment of miRNA target genes were analyzed using Cluster Profiler (v3.10.1) software and KOBAS software (Mao et al., 2005), respectively.

2.7 Quantitative real-time polymerase chain reaction analysis

The extracted total RNA, combined with reverse random primers, was used for reverse transcription of miRNA. The qRT-PCR was performed using the MyiQ Single Color Real-time PCR system (Bio-Rad, Hercules, CA, USA) (95°C for 3 min and 45 cycles of 95°C for 5 s and 60°C for 30 s). Primers were designed using Primer Premier (v6.0) software and synthesized by Sangon Biotech Co., Ltd. (Shanghai, China) (Supplementary Table S1). The $2^{-\Delta\Delta Ct}$ method was used to the calculate relative gene expression levels.

3 Results

3.1 Physiological characteristics of roots under Al stress

The root length, root surface area, root activity, and plant growth of *M. truncatula* plants after Al treatment are shown in Figures 1A–C. With increasing duration of Al stress, the root length of plants in the Al-treated group differed significantly from that of the control at 48 h (Figure 1A), and their root length decreased by 5.0% compared with the control. The plant heel surface area of the Al-treated group was significantly different from that of the control group at both 24 and 48 h (Figure 1B), with reductions of 3.9% and 4.5%, respectively. Root activity showed an increasing trend and then decreasing, and was significantly lower at 48 h than at 24 h (Figure 1C). As shown in Figure 1D (I–IV), the root morphology of the Al-treated and control plants also showed significant differences as the stress time increased.

The FDA fluorescent staining method only stains live cells. Therefore, the FDA staining of the root system at each treatment time point revealed that damage to the plant root tip tissues had already begun after 4 h of Al treatment, and the damage to the root-tip tissues gradually increased with the Al treatment time (Figure 1D V–VIII).

To investigate whether the redox system was activated in *M. truncatula* subjected to Al stress, the enzymatic activities of SOD, CAT, and POD were examined in this study. The SOD activity reached its maximum at 4 h, after which, it started to decline, and decreased significantly at 48 h (Figure 2A). CAT and POD showed the same trend, with both significantly increasing at 4 h versus 24 h and significantly decreasing at 48 h (Figures 2B, C). These results indicated that the redox system in *M. truncatula* was rapidly activated under Al stress.

3.2 MicroRNA sequence analysis and mapping

In this study, 12 *M. truncatula* samples were sequenced for small RNA and a total of 296.01 M raw reads were obtained. After removing contaminants and reads of length < 18 and > 30 nucleotides, 195.80 M clean reads were obtained. There were not less than 11.96 M clean reads for each sample, the average GC content and base number were 50.99% and 1258.05 Mb, respectively, and the average Q30 was 86.82% (Supplementary Table S2). The average mapped reads, uniquely mapped reads, and multiple-mapped reads constituted 61.46%, 40.42%, and 21.05% of all libraries, respectively, and the comparison efficiency of the reads in each sample against the reference genome ranged from 51.76–68.47% (Supplementary Table S3). The length distribution of sRNAs was similar among the 12 libraries, with the highest abundance of 21-nucleotide sRNAs, followed by 24-nucleotide sRNAs (Supplementary Table S4), which is consistent with the previous findings of Bao et al. (2019).

10.3389/fpls.2023.1137764

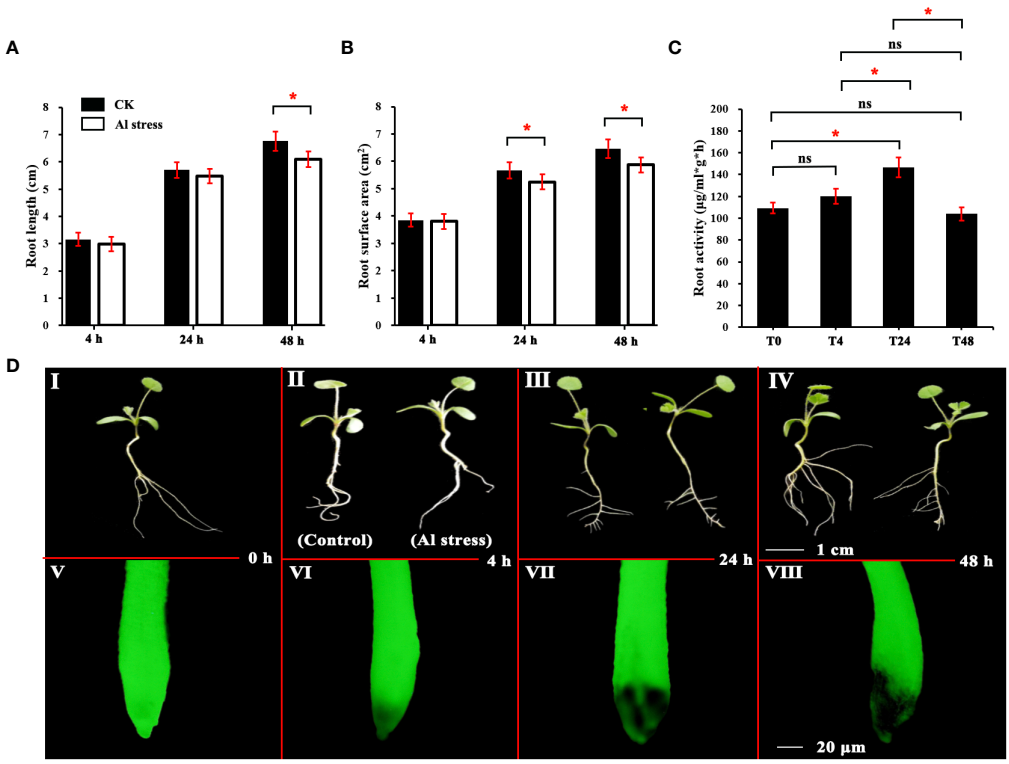


FIGURE 1
Analysis of root physiological characteristics of *M. truncatula* under aluminum stress. **(A)** Analysis of total root length of *M. truncatula*. **(B)** Analysis of root surface area of *M. Truncatula*. **(C)** Analysis of root activity of *M. Truncatula*. **(D)** I–IV is the growth comparison of plants under Al stress and the blank control at the same time point. Plant material grown under normal conditions and treated for the same amount of time as the corresponding aluminum treatment group were used as a control, scale = 1 cm; V–VIII is the comparison of root fluorescence of plants under Al stress and the blank control after FDA staining, scale = 20 μm. * indicates that there is a significant difference at the level of 0.05 according to Duncan's multiple range test, while ns indicates that the difference is not significant.

3.3 Identification of known and novel miRNAs, predictive analysis of target genes, and functional annotation

A total of 876 miRNAs were identified in this study, of which, 673 were known and 203 were unknown. There were 556 known miRNAs with family affiliation and 95 unknown miRNAs with family affiliation (Supplementary Table S5). Subsequently, miRNA target genes were predicted using TargetFinder software; 664 of the known miRNAs predicted 10,845 target genes, and 194 of the unknown miRNAs predicted 3,610 target genes (Supplementary Table S6). According to our functional annotation results, a total of 13,057 out of 13,148 target genes were annotated, with the number ranging from 3,852 (29.50%, KEGG) to 13,055 (99.98%, Nr), among which, the

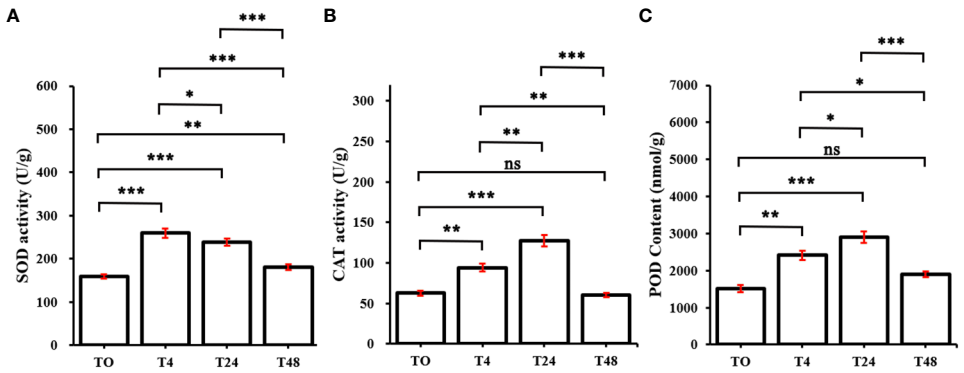


FIGURE 2
Changes of oxidoreductase in *M. truncatula* under continuous aluminum stress. **(A)** superoxide dismutase (SOD) activity. **(B)** Catalase (CAT) activity. **(C)** peroxidase activity (POD). The plotting data is the average and standard deviation of three repetitions. T0~T48 represent *M. truncatula* plant samples treated with aluminum for 0h, 4h, 24h, and 48h, respectively. According to Duncan's multiple range test, the asterisk represents the significant difference between aluminum treated samples and the control (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$), ns indicates that the difference is not significant.

most genes were annotated in Nr and eggNOG with 13,055 and 10,935, respectively (Supplementary Table S7).

3.4 MicroRNA differential expression analysis, predictive analysis, and functional annotation of target genes

In total, 55 differentially expressed miRNAs (Supplementary Table S8) were identified in this study. Comparison of the T0 *vs.* T4 treatment groups showed the highest number of differentially expressed miRNAs, 48, of which, 31 were upregulated and 17 were downregulated. Five miRNAs were differentially expressed in the comparison between the T0 *vs.* T24 treatment groups, with all of them exhibiting upregulation. Ten miRNAs were differentially expressed in the comparison between the T0 *vs.* T48 treatment groups, of which, five were upregulated and five were downregulated. A total of 2,502 differential target genes were predicted for the 55 differentially expressed miRNAs; the maximum number of target genes of differentially expressed miRNAs was 2,088 in the T0 *vs.* T4 treatment group and the minimum was 101 in the T0 *vs.* T24 treatment group. Subsequently, the target genes of the differentially expressed miRNAs were annotated using eight databases, including Nr, KOG, COG, Pfam, Swiss-Prot, eggNOG, GO, and KEGG, and all 2,502 target genes were annotated. The target genes of all three treatment groups were the most abundant annotated genes in the NR database, with 2088, 101, and 307, respectively (Supplementary Table S9).

3.5 Gene ontology function and KEGG pathway enrichment analysis

To determine whether miRNAs are functionally involved in the Al stress response and defense processes, we performed GO functional analysis of differential target genes of miRNAs. A total of 2,714 DEGs were enriched in the 145 GO treatments (Supplementary Table S10). Based on the functions of the enriched genes in each GO tree, a cluster analysis of 145 GO terms could be clustered into 30 major classes (Figure 3A). The three GO terms "Cellular metabolic process," "Organic substance metabolic process," and "Cellular component organization" had the highest number of DEGs, at 1,382, 378, and 333, respectively (Figure 3A).

KEGG pathway enrichment was performed to explore the function of genes differentially expressed under Al stress as well as metabolic pathways involved in the response. A total of 912 DEGs were enriched in 111 pathways (Supplementary Table S9). The 111 enriched pathways were clustered into 30 broad categories based on the function of the enriched genes in each pathway (Figure 3B). The highest number of DEGs was found in three enrichment pathways: "starch and sucrose metabolism," "plant hormone signal transduction," and "biosynthesis of amino acids," at 230, 213, and 71, respectively.

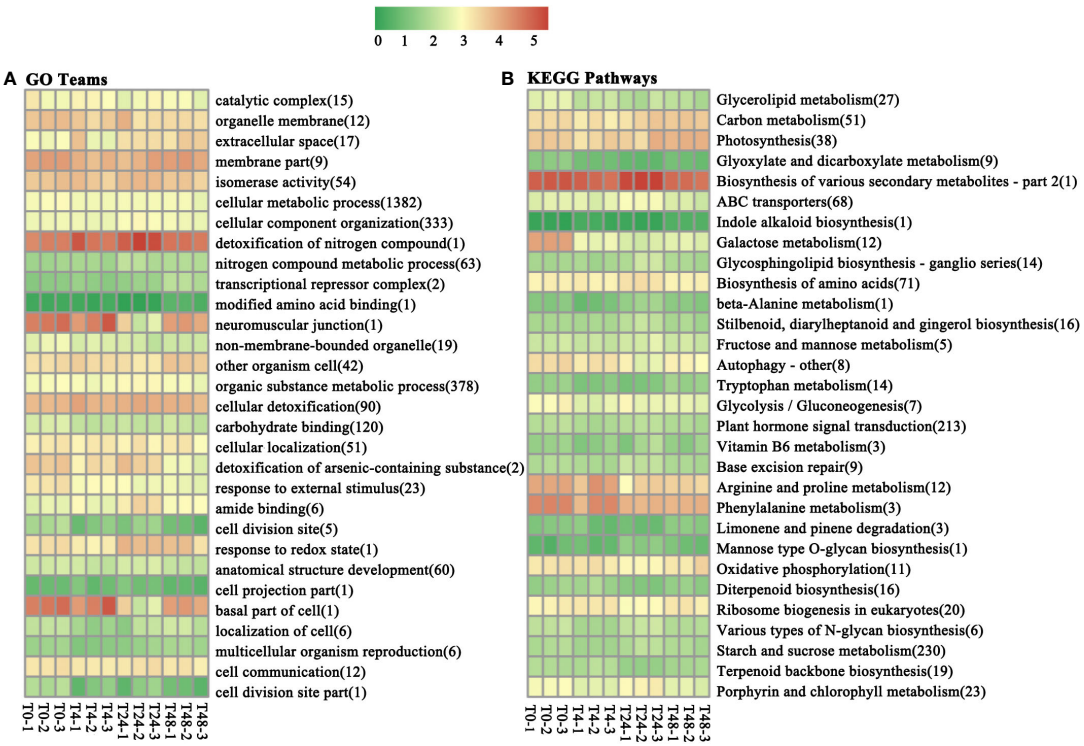


FIGURE 3
Under aluminum stress, 30 GO terms and KEGG pathways were enriched for the most DEGs. **(A)** GO teams. **(B)** KEGG pathways. The numbers in the abscissa represent the number of DEGs enriched for the GO term or KEGG pathway. T0 to T48 represent *M. truncatula* plant samples treated with aluminum for 0h, 4h, 24h, and 48h, respectively. The comparison is described as A *vs* B, which means the DEG was found in the A treatment relative to the B treatment.

3.6 qRT-PCR verification of DE miRNA and differential genes

Ten DE miRNAs were randomly selected for qRT-PCR validation to verify the reliability of the transcriptome data. The results showed that these miRNAs had similar expression levels and trends as observed in the miRNA-seq results (Figure 4) and the qRT-PCR results also demonstrated the reliability of the transcriptome data.

4 Discussion

Al stress contributes majorly to altered metabolic activity, root damage, cell wall damage, and cytoplasmic lysis, resulting in reduced photosynthetic efficiency, impaired water and nutrient uptake, and increased respiratory consumption and toxin accumulation; these effects limit crop growth (Xu et al., 2021), ultimately leading to plant death and yield loss (Osman et al., 2021; Abdel Latef et al., 2021). MicroRNAs widely regulate various physiological processes, such as development, signal transduction, and stress response in plants (Chen, 2012; Sunkar et al., 2012), and their mechanisms in plant stress tolerance have been demonstrated in *G.max* (Ning et al., 2019), *Z. mays* (Fu et al., 2017; Shan et al., 2020), and *Arabidopsis thaliana* (Shukla et al., 2018).

In this study, a total of 12 sRNA samples of *M. truncatula* under Al stress for 4 h, 24 h, 48 h and a control group were subjected to whole transcriptome sequencing, and a total of 876 miRNAs were obtained, among which, 203 were unknown miRNAs. Cao et al. (2018) similarly identified 876 miRNAs under salt/alkali stress in *M. truncatula*, supporting the reliability of the data in this study.

4.1 Effect of mtr-miR156g-3p on the root system of *M. truncatula* under Al stress

miR156 is one of the most abundantly expressed and highly evolutionarily conserved miRNAs in plants (Niu et al., 2015), and the physiological processes involved in the regulation of this family of miRNAs under abiotic stress have been demonstrated in *A. thaliana* (Guan et al., 2014; Niu et al., 2015), *Panax notoginseng* (Zheng et al., 2017), *Phalaenopsis* (Zhao et al., 2019), and other plants. miR156 is highly accumulated mainly in the seedling stage of plants, and its expression level decreases during the plant growth period, which is important for plant seedling development and for improving crop productivity and stress resistance (Wang et al., 2009; Wu et al., 2009; Jerome Jeyakumar et al., 2020). In the present study, mtr-miR156g-3p was the only DE miRNA involved in the response at 4, 24, and 48 h of Al stress, and it was upregulated compared to the control. Further annotation analysis of the target genes of mtr-miR156g-3p showed that six target genes, Medtr6g004260, Medtr4g056140, Medtr7g007440, Medtr5g077840, Medtr3g078260, and Medtr2g090610, regulate the formation of the root tip cell membrane. In addition, the expression levels of the above six target genes were significantly upregulated at 4 and 24 h of Al stress ($P < 0.05$) (Figure 5; Supplementary Table S10). Therefore, these findings suggest that mtr-miR156g-3p is activated in *M. truncatula* under Al stress and further overexpresses target genes with the ability to promote cell membrane formation in root tip cells to counteract the damage caused by stress. Niu et al. (2015) found that miR156 could enhance *A. thaliana* lateral root development by regulating related binding protein genes, which is consistent with the findings of the present study. Of note, in previous studies, mtr-miR156g-3p

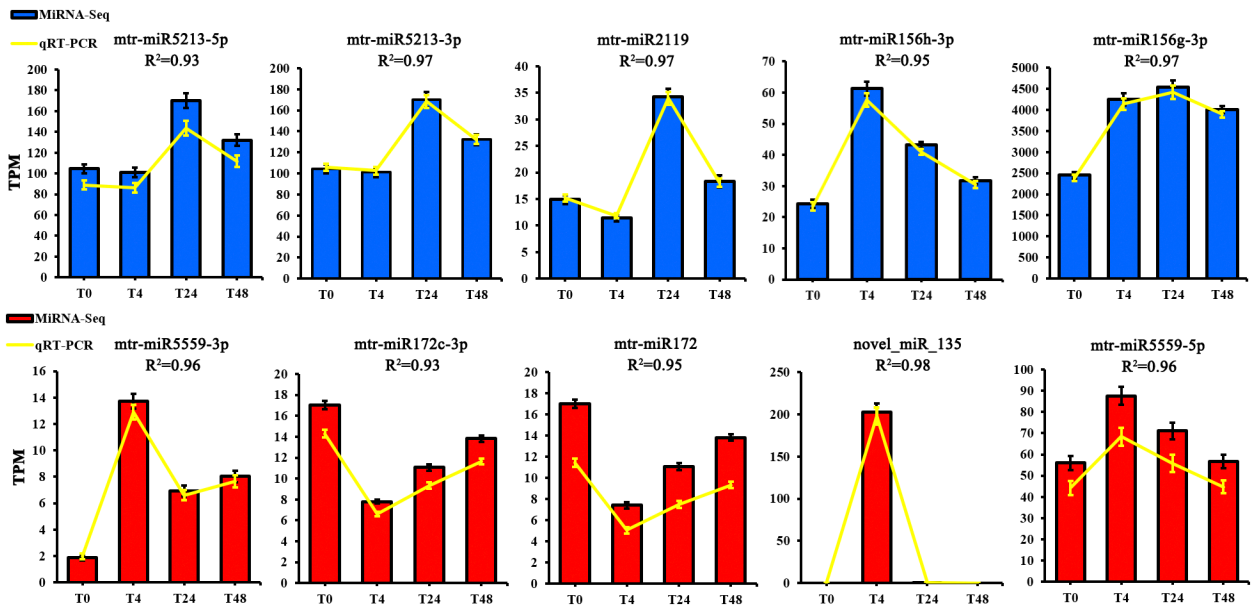


FIGURE 4
The expression patterns of ten selected miRNAs identified by RNA-Seq were verified by qRT-PCR. The Y-axis to the left of each histogram represents the expression level of RNA-seq (TPM). T0 to T48 represents *M. truncatula* plant samples treated with aluminum for 0h, 4h, 24h, and 48h, respectively.

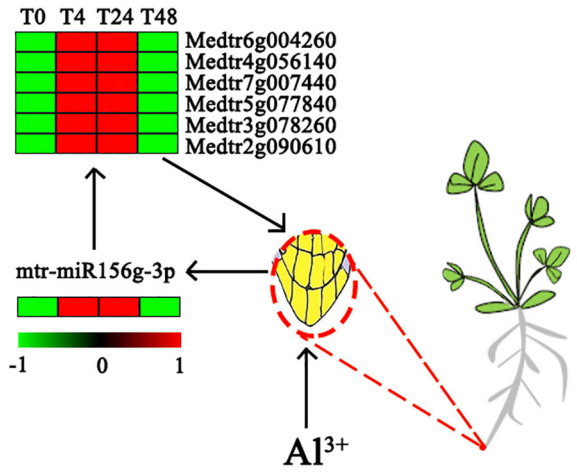


FIGURE 5
Analysis of expression patterns of miRNA and target genes affecting the root system of *M. truncatula* under aluminum stress.

under abiotic stress mediated the negative expression of the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene before regulating the expression of genes controlling root cell growth (Zheng et al., 2019; Rao et al., 2021). In contrast, there was no differential expression of genes associated with *SPL* in the present study, which may indicate that mtr-miR156g-3p in *M. truncatula* root tip tissue can directly regulate root cell growth genes. Zhao et al. (2019) found that mtr-miR156g-3p directly regulates anthocyanin formation in *Phalaenopsis* plants, and no expression of *SPL*-related genes was detected.

4.2 Regulation of transcription factors by miRNAs and target genes in the mechanism of Al resistance

Transcription factors play an important role in Al stress-tolerance mechanisms (Lin et al., 2021; Shu et al., 2022). The GRAS TF family is involved in a variety of biological processes such as biotic and abiotic plant stress, rootstock development, and meristem tissue formation (Wang et al., 2018). In the present study, seven target genes of four DE miRNAs were involved in the regulated expression of GRAS TFs, and all seven target genes were significantly upregulated at 4 h of Al stress, but not at 24 h and 48 h (Table 1A; Supplementary Table S10). The GRAS TF family consists of three members: GAI (GIBBERELLIC ACID INSENSITIVE), RGA (REPRESSOR OF GA1-3 MUTANT), and SCR (SCARECROW) (Fan et al., 2021), which are mainly expressed in roots and vascular cells with SHR (SHORT ROOT) and can enhance plant stress resistance (Cui et al., 2014). The MYB TF family members play an important role in plant cell wall formation, growth, and development, and are one of the major components of stress response mechanisms in plants (Sun et al., 2014). In this study, 14 target genes of 10 DE miRNAs regulated the expression of MYB TFs at 4 h of Al stress treatment, and all the 14 target genes were significantly upregulated. In contrast, no DE miRNAs regulating MYB TFs or their corresponding target genes were activated at 24 h of Al stress treatment. At 48 h, a total of two target genes of one DE miRNA were involved in the regulation of MYB TF expression, both of which were significantly downregulated (Table 1A; Supplementary Table S10). The MYB TFs are one of the largest protein families in plants (Ambawat et al., 2013), and most MYB proteins belong to the R2R3-MYB subfamily, which positively regulates salt tolerance in plants by mediating the expression of abscisic acid (ABA) and regulating cuticle formation (Wang et al., 2021). WRKY is a plant-specific zinc finger TF that is involved in various physiological processes in plants (Wu et al., 2014; Wang et al., 2022). In this study, a total of 13 target genes of 7 DE miRNAs regulated the expression of WRKY TFs at 4 h of Al treatment, and all 13 target genes were significantly upregulated. At 24 h, no DE miRNAs that regulated WRKY TFs or target genes were activated. At 48 h, one target gene of one DE miRNA was involved in the expression regulation of WRKY TFs, and this target gene was significantly downregulated (Table 1A; Supplementary Table S10). The WRKY TFs are not only involved in the regulation of various biological functions by themselves but also interact with other TFs to form a signaling network that regulates different biological processes and are an important component of the plant stress tolerance system (Wani et al., 2021). The bHLH class of TFs is the second largest class of TFs in plants and is involved in biological processes, such as light signaling, hormone and other signal transduction, root hair development, and stress tolerance mechanisms (Nakata et al., 2013). In the present study, a total of 10 target genes of 4 DE miRNAs were involved in the regulated expression of bHLH TFs, and all 10 target genes were significantly upregulated at 4 h of Al stress, whereas they were not expressed at 24 h and 48 h of Al stress (Table 1B; Supplementary Table S10). In the present study, the expression pattern of bHLH TFs was consistent with that of GRAS TFs, and both TFs play an important role in the growth and development of plant roots under stress (Nakata et al., 2013; Wang et al., 2018). Therefore, the present study suggests that related miRNAs promote the development of plant roots by regulating the expression of bHLH TFs and GRAS TFs at the early stage of Al stress in *M. truncatula*, thus enhancing the Al resistance of *M. truncatula*.

MicroRNAs and target genes mediate TFs, such as GRAS, MYB, WRKY, and bHLH, which play an important role in the Al response mechanism of *M. truncatula*. However, all target genes regulating the above TFs were significantly upregulated at 4 h during the initial stage of Al stress. In contrast, no genes regulating TFs were activated at 24 h. Although some target genes of individual TFs were activated at 48 h, they

10.3389/fpls.2023.1137764

TABLE 1A Information About Special MiRNA.

TFs family	miRNA	T0_TPM	T4_TPM	T24_TPM	T48_TPM	Regulated	Target genes
GRAS	mtr-miR156h-3p	24.35	61.27	–		up	Medtr7g027190
	novel_miR_135	119.16	202.35	–	–	up	Medtr4g133660
							Medtr4g077760
	novel_miR_182	–	202.12	–	–	up	Medtr4g077760
							Medtr4g133660
	novel_miR_36	–	202.35	–	–	up	Medtr4g133660
							Medtr4g077760
MYB	mtr-miR156h-3p	24.35	61.27	–	–	up	Medtr5g038910
							Medtr3g065440
	mtr-miR5559-5p	56.00	87.61	–	–	up	Medtr1g083630
	novel_miR_135	–	202.35	–	–	up	Medtr0193s0090
	novel_miR_143	26.28	49.27	–	–	up	Medtr2g096380
	novel_miR_180	74.14	113.42	–	–	up	Medtr6g055910
	novel_miR_182	–	202.12	–	–	up	Medtr0193s0090
	novel_miR_187	27.84	65.83	–	–	up	Medtr8g042410
	novel_miR_36	–	202.35	–	–	up	Medtr0193s0090
	novel_miR_45	70.44	112.34	–	–	up	Medtr2g100930
							Medtr3g077650
							Medtr4g057635
							Medtr3g074520
	novel_miR_9	78.60	123.77	–	–	up	Medtr7g011170
	mtr-miR397-5p	73.44	–	–	33.94	down	Medtr7g461410
							Medtr7g061330

-, indicates that the MiRNA is not expressed in the corresponding aluminum treatment group.

TABLE 1B Information About Special MiRNA.

TFs family	miRNA	T0_TPM	T4_TPM	T24_TPM	T48_TPM	Regulated	Target genes
WRKY	mtr-miR156h-3p	24.35	61.27	–	–	up	Medtr3g085710
	novel_miR_135	–	202.35	–	–	up	Medtr8g067650
							Medtr6g053200
							Medtr1g099600
	novel_miR_180	74.14	113.42	–	–	up	Medtr6g015675
	novel_miR_182	–	202.12	–	–	up	Medtr6g053200
							Medtr8g067650
							Medtr1g099600
	novel_miR_36	–	202.35	–	–	up	Medtr1g099600
							Medtr8g067650
							Medtr6g053200
	novel_miR_45	70.44	112.34	–	–	up	Medtr7g062220

(Continued)

10.3389/fpls.2023.1137764

TABLE 1B Continued

TFs family	miRNA	T0_TPM	T4_TPM	T24_TPM	T48_TPM	Regulated	Target genes
	novel_miR_9	78.60	123.77	–	–	up	Medtr3g056100
	novel_miR_193	63.10	–	–	23.93	down	Medtr4g132430
bHLH	mtr-miR172b	17.01	7.41	–	–	up	Medtr3g116770
							Medtr2g038040
	mtr-miR172c-3p	17.01	7.80	–	–	up	Medtr3g116770
							Medtr2g038040
	mtr-miR172d-3p	29.16	13.71	–	–	up	Medtr2g038040
							Medtr4g079760
	novel_miR_135	–	202.35	–	–	up	Medtr5g005110
							Medtr4g079760
							Medtr2g039620
	novel_miR_182	–	202.12	–	–	up	Medtr5g005110
							Medtr4g079760
							Medtr2g039620
	novel_miR_36	–	202.35	–	–	up	Medtr2g039620
							Medtr5g005110
							Medtr4g079760
F-box	miR2119	14.90	–	34.263	–	up	Medtr4g134390
							Medtr8g467670
							Medtr2g020070
							Medtr3g024110
	miR5213	104.43	–	170.14	–	up	Medtr5g069120
							Medtr7g079370

-, indicates that the MiRNA is not expressed in the corresponding aluminum treatment group.

were all downregulated. This study concluded that the above results could be attributed to the severe damage to the root tip tissue of *M. truncatula* with the increase in Al stress time, resulting in a large amount of cell inactivation. This reduces the ability of root tip cells to express Al tolerance genes and secrete secondary substances involved in the Al stress response. The above expression pattern of TFs was also consistent with the trend of the root physiological indicators in Figure 1. Notably, three miRNA target genes, novel_miR_135, novel_miR_182, and novel_miR_36, were involved in regulating the expression of the above four TFs, and they were significantly upregulated at the early stage of stress. Therefore, this study suggests that three miRNAs, novel_miR_135, novel_miR_182, and novel_miR_36, and their target genes may have potential roles in regulating Al responsive TFs.

4.3 Role of legume-specific miRNAs in the Al response mechanism of *M. truncatula*

miR2111, miR2119, and miR5213 are considered legume-specific miRNAs that regulate the expression of relevant defense genes when legumes are attacked by pathogens, thereby enhancing their own defense mechanisms (Kohli et al., 2014). This has been verified in plants such as *Prunella vulgaris* (Rosa et al., 2019) and *Cicer arietinum* (Kohli et al., 2014). In the present study, three target genes of miR2119 and one target gene of miR5213 were significantly upregulated at 24 h of Al stress, while their expression levels at other time points of Al stress were consistent with the treatment group Table 1A; Supplementary Table S10). Functional analysis of the four target genes revealed that they were involved in the regulation of an F-box protein containing a TIR structural domain (Supplementary Table S8). F-box proteins are receptors for plant growth hormones and are involved in plant defense responses (Dharmasiri et al., 2005; Navarro et al., 2006). Therefore, our findings suggest that legume-specific miR2119 and miR5213 may enhance Al tolerance in *M. truncatula* by regulating F-box protein overexpression *via* their target genes.

5 Conclusions

In this study, a total of 195.80 M clean reads were obtained from 12 sRNA libraries, and 876 miRNAs were identified, along with their corresponding 14,455 target genes. A total of 55 miRNAs were

differentially expressed during Al stress, with different miRNAs enhancing Al tolerance in *M. truncatula* through mechanisms such as the promotion of root tip cell formation, expression of Al-resistant TFs, and regulation of defense gene expression. These findings could provide useful genetic resources for subsequent biotechnological studies, such as miRNA silencing or gene overexpression, and may inform genetic improvement programs for the development of Al stress-tolerant genotypes in legume crops.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA908067.

Author contributions

RD and CC conceived the experiment. ZL, ZY, ZT, and QG carried it out. ZL and RD analyzed the data. RD and ZL wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (32060392) and Support by Guizhou Province Science and Technology Projects (Qian Ke He Zhi Cheng [2020] 1Y074) and GZMARS- Forage Industry Technology System.

Conflict of interest

The authors declare that the research has been conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Additional materials have been uploaded separately at the time of submission.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1137764/full#supplementary-material

References

Abdel Latef, A. A., Omer, A. M., Badawy, A. A., Osman, M. S., and Ragaey, M. M. (2021). Strategy of salt tolerance and interactive impact of azotobacter chroococcum and/or alcaligenes faecalis inoculation on canola (*Brassica napus* l.). plants grown in saline soil. *Plants* 10, 110. doi: 10.3390/plants10010110

Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121, 207–221. doi: 10.1016/j.cell.2005.04.004

Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi: 10.1007/s12298-013-0179-1

Bao, H., Chen, H., Chen, M., Xu, H., Huo, X., Xu, Q., et al. (2019). Transcriptome-wide identification and characterization of microRNAs responsive to phosphate starvation in populus tomentosa. *Funct. Integr. Genomics* 19, 953–972. doi: 10.1007/s10142-019-00692-1

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Cao, C., Long, R., Zhang, T., Kang, J., Wang, Z., Wang, P., et al. (2018). Genome-wide identification of microRNAs in response to salt/alkali stress in *Medicago truncatula* through high-throughput sequencing. *Int. J. Mol. Sci.* 19, 4076. doi: 10.3390/ijms19124076

Chandran, D., Sharopova, N., VandenBosch, K. A., Garvin, D. F., and Samac, D. A. (2008). Physiological and molecular characterization of aluminum resistance in *Medicago truncatula*. *BMC Plant Biol.* 8, 89. doi: 10.1186/1471-2229-8-89

Chauhan, D. K., Yadav, V., Vaculík, M., Gassmann, W., Pike, S., and Arif, N. (2021). Aluminum toxicity and aluminum stress-induced physiological tolerance responses in higher plants. *Crit. Rev. Biotechnol.* 41, 715–730. doi: 10.1080/07388551.2021.1874282

Chen, X. (2012). Small RNAs in development-insights from plants. *Curr. Opin. Genet. Dev.* 22, 361–367. doi: 10.1016/j.gde.2012.04.004

Cui, H., Kong, D., Liu, X., and Hao, Y. (2014). SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant J.* 78, 319–327. doi: 10.1111/tpj.12470

Daspute, A. A., Sadhukhan, A., Tokizawa, M., Kobayashi, Y., Panda, S. K., and Koyama, H. (2017). Transcriptional regulation of aluminum-tolerance genes in higher plants: Clarifying the underlying molecular mechanisms. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01358

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The f-box protein TIR1 is an auxin receptor. *Nature* 435, 441–445. doi: 10.1038/nature03543

Fan, Y., Yan, J., Lai, D., Yang, H., Xue, G., He, A., et al. (2021). Genome-wide identification, expression analysis, and functional study of the GRAS transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) moench]. *BMC Genomics* 22 (1), 1–21. doi: 10.1186/s12864-021-07848-z

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Fu, R., Zhang, M., Zhao, Y., He, X., Ding, C., Wang, S., et al. (2017). Identification of salt tolerance-related microRNAs and their targets in maize (*Zea mays* l.) using high-throughput sequencing and degradome analysis. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00864

Guan, X., Pang, M., Nah, G., Shi, X., Ye, W., Stelly, D. M., et al. (2014). miR828 and miR858 regulate homoeologous MYB2 gene functions in arabidopsis trichome and cotton fibre development. *Nat. Commun.* 5, 3050. doi: 10.1038/ncomms4050

Gui, Q., Yang, Z., Chen, C., Yang, F., Wang, S., and Dong, R. (2022). Identification and characterization of long noncoding RNAs involved in the aluminum stress response in *Medicago truncatula via* genome-wide analysis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1017869

Hoang, N. T., Tóth, K., and Stacey, G. (2020). The role of microRNAs in the legume–*Rhizobium* Nitrogen-fixing symbiosis. *Journal of Experimental Botany* 71, 1668–1680. doi: 10.1093/jxb/eraa018

Ishikawa, S., and Wagatsum, T. (1998). Plasma membrane permeability of root-tip cells following temporary exposure to Al ions is a rapid measure of Al tolerance among plant species. *Plant Cell Physiol.* 39, 516–525. doi: 10.1093/oxfordjournals.pcp.a029399

Jerome Jeyakumar, J. M., Ali, A., Wang, W. M., and Thiruvengadam, M. (2020). Characterizing the role of the miR156-SPL network in plant development and stress response. *Plants* 9, 1206. doi: 10.3390/plants9091206

Kohli, D., Joshi, G., Deokar, A. A., Bhardwaj, A. R., Agarwal, M., Katiyar-Agarwal, S., et al. (2014). Identification and characterization of wilt and salt stress-responsive MicroRNAs in chickpea through high-throughput sequencing. *PloS One* 9, e108851. doi: 10.1371/journal.pone.0108851

Kumar, V., Khare, T., Shriram, V., and Wani, S. H. (2017). Plant small RNAs: the essential epigenetic regulators of gene expression for salt-stress responses and tolerance. *Plant Cell Rep.* 26, 1–5. doi: 10.1007/s00299-017-2210-4

Lin, Y., Liu, G., Xue, Y., Guo, X., Luo, J., Pan, Y., et al. (2021). Functional characterization of aluminum (Al)-responsive membrane-bound NAC transcription factors in soybean roots. *Int. J. Mol. Sci.* 22, 12854. doi: 10.3390/ijms222312854

Li, J., Su, L., Lv, A., Li, Y., Zhou, P., and An, Y. (2020). MsPG1 alleviated aluminum-induced inhibition of root growth by decreasing aluminum accumulation and increasing porosity and extensibility of cell walls in alfalfa (*Medicago sativa*). *Environ. Exp. Bot.* 175, 104045. doi: 10.1016/j.envexpbot.2020.104045

Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., et al. (2014). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC Genomics* 15, 25. doi: 10.1186/1471-2164-15-25

Liu, W., Zhang, Z., Chen, S., Ma, L., Wang, H., Dong, R., et al. (2016). Global transcriptome profiling analysis reveals insight into saliva-responsive genes in alfalfa. *Plant Cell Rep.* 35, 561–571. doi: 10.1007/s00299-015-1903-9

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Nakata, M., Mitsuda, N., Herde, M., Koo, A., Moreno, J. E., Suzuki, K., et al. (2013). A bhlh-type transcription factor, aba-inducible bhlh-type transcription factor/ja-associated myc2-like1, acts as a repressor to negatively regulate jasmonate signaling in arabidopsis. *Plant Cell* 25, 1641–1656. doi: 10.1105/tpc.113.111112

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 21, 436–439. doi: 10.1073/pnas.0510928103

Ning, L. H., Du, W. K., Song, H. N., Shao, H. B., Qi, W. C., Sheteiwy, M. S. A., et al. (2019). Identification of responsive miRNAs involved in combination stresses of phosphate starvation and salt stress in soybean root. *Environ. Exp. Bot.* 167, 103823. doi: 10.1016/j.envexpbot.2019.103823

Niu, Y., Niu, Q. W., Ng, K. H., and Chua, N. H. (2015). The role of miR156/SPLs modules in arabidopsis lateral root development. *Plant J.* 83, 673–685. doi: 10.1111/tpj.12919

Osman, M. S., Badawy, A. A., Osman, A. I., and Abdel Latef, A. A. (2021). Ameliorative impact of an extract of the halophyte *Arthrocnemum macrostachyum* on growth and biochemical parameters of soybean under salinity stress. *J. Plant Growth Regul.* 40, 1245–1256. doi: 10.1007/s00344-020-10185-2

Pang, J. Y., Bansal, R., Zhao, H. X., Bohuon, E., Lambers, H., Ryan, M. H., et al. (2018). The carboxylate-releasing phosphorus-mobilizing strategy can be proxied by foliar manganese concentration in a large set of chickpea germplasm under low phosphorus supply. *New Phytologist* 219(2), 518–529. doi: 10.1111/nph.15200.

Pegler, J. L., Oultram, J. M. J., Grof, C. P. L., and Eamens, A. L. (2019). Profiling the abiotic stress responsive microRNA landscape of *Arabidopsis thaliana*. *Plants* 8, 58. doi: 10.3390/plants8030058

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T., Mendell, J. T., Salzberg, S. L., et al. (2015). StringTie enables improved reconstruction of a transcriptome from RNAseq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.312

Rao, S., Li, Y., and Chen, J. (2021). Combined analysis of MicroRNAs and target genes revealed miR156-SPLs and miR172-AP2 are involved in a delayed flowering phenomenon after chromosome doubling in black goji (*Lycium ruthencium*). *Front. Genet.* 12. doi: 10.3389/fgene.2021.706930

Rosa, C. D. L., Covarrubias, A. A., and Reyes, J. L. (2019). A dicistronic precursor encoding miR398 and the legume-specific miR2119 coregulates CSD1 and ADH1 mRNAs in response to water deficit. *Plant Cell Environ.* 42, 133–144. doi: 10.1111/pce.13209

Shan, T., Fu, R., Xie, Y., Chen, Q., Wang, Y., Li, Z., et al. (2020). Regulatory mechanism of maize (*Zea mays* l.) miR164 in salt stress response. *Russ J. Genet.* 56, 835–842. doi: 10.1134/S1022795420070133

Shukla, P. S., Borza, T., Critchley, A. T., Hiltz, D., Norrie, J., and Prithiviraj, B. (2018). Extract mitigates salinity stress in *Arabidopsis thaliana* By modulating the expression of miRNA involved in stress tolerance and nutrient acquisition. *PloS One* 13, e0206221. doi: 10.1371/journal.pone.020622

Shu, W., Zhou, Q., Xian, P., Cheng, Y., Lian, T., Ma, Q., et al. (2022). *GmWRKY81* encoding a WRKY transcription factor enhances aluminum tolerance in soybean. *Int. J. Mol. Sci.* 23, 6518. doi: 10.3390/ijms23126518

Stephan, U. W., and Procházka, Ż. (1989). Physiological disorders of the nicotianamine-auxotroph tomato mutant chloronerva at different levels of iron nutrition. i. growth characteristics and physiological abnormalities related to iron and nicotianamine supply. *Acta Botanica Neerlandica* 38, 147–153. doi: 10.1111/j.1438-8677.1989.tb02037.x

Sunkar, R., Li, Y. F., and Jagadeeswaran, G. (2012). Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17, 196–203. doi: 10.1016/j.tplants.2012.01.010

Sun, P., Zhu, X., Huang, X., and Liu, J. (2014). Overexpression of a stress-responsive MYB transcription factor of poncirus trifoliata confers enhanced dehydration tolerance and increases polyamine biosynthesis. *Plant Physiol. Biochem.* 78, 71–79. doi: 10.1016/j.plaphy.2014.02.022

Wang, J. W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in. *Arabidopsis thaliana. Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, C., Hao, X., Wang, Y., Maoz, I., Zhou, W., Zhou, Z., et al. (2022). Identification of WRKY transcription factors involved in regulating the biosynthesis of the anti-cancer drug camptothecin in ophiorrhiza pumila. *Hortic. Res.* 9, uhac099. doi: 10.1093/hr/uhac099

Wang, Y. X., Liu, Z. W., Wu, Z. J., Li, H., Wang, W. L., Cui, X., et al. (2018). Genome-wide identification and expression analysis of gras family transcription factors in tea plant (*Camellia sinensis*). *Sci. Rep.* 8, 3949. doi: 10.1038/s41598-018-22275-z

Wang, J., Liu, C., Zhang, L., Wang, J., Hu, G., Ding, J., et al. (2011). MicroRNAs involved in the pathogenesis of phytophthora root rot of soybean (*Glycine max*). *Agricultural Sciences in China* 10, 1159–1167. doi: 10.1016/S1671-2927(11)60106-5

Wang, X., Niu, Y., and Zheng, Y. (2021). Multiple functions of MYB transcription factors in abiotic stress responses. *Int. J. Mol. Sci.* 122, 6125. doi: 10.3390/ijms22116125

Wani, S. H., Anand, S., Singh, B., Bohra, A., Joshi, R., and (2021). And plant defense responses: latest discoveries and future prospects. *Plant Cell Rep.* 40, 1071–1085. doi: 10.1007/s00299-021-02691-8

Wu, G., Park, M. Y., Conway, S. R., Wang, J. W., Weigel, D., and Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in arabidopsis. *Cell* 138, 750–759. doi: 10.1016/j.cell.2009.06.031

Wu, Y., Wei, W., Pang, X., Wang, X., Zhang, H., and Dong, B. (2014). Comparative transcriptome profiling of a desert evergreen shrub, ammopiptanthus mongolicus, in response to drought and cold stresses. *BMC Genomics* 15, 671. doi: 10.1186/1471-2164-15-671

Xu, J., Hou, Q. M., Khare, T., Verma, S. K., and Kumar, V. (2019). Exploring miRNAs for developing climate-resilient crops: a perspective review. *Sci. Total Environ.* 653, 91–104. doi: 10.1016/j.scitotenv.2018.10.340

Xu, T., Zhang, L., Yang, Z., Wei, Y., and Dong, T. (2021). Identification and functional characterization of plant MiRNA under salt stress shed light on salinity resistance improvement through MiRNA manipulation in crops. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.665439

Yu, L., Huang, T., Qi, X., Yu, J., Wu, T., Luo, Z., et al. (2022). Genome-wide analysis of long non-coding RNAs involved in nodule senescence in *Medicago truncatula*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.917840

Zhang, B. H. (2015). MicroRNA: a new target for improving plant tolerance to abiotic stress. *Journal of Experimental Botany* 66, 1749–1761. doi: 10.1093/jxb/erv013

Zhang, Z., Jiang, L., Wang, J., Gu, P., and Chen, M. (2015). MTide: an integrated tool for the identification of miRNA-target interaction in plants. *Bioinformatics* 31, 290–291. doi: 10.1093/bioinformatics/btu633

Zhao, A., Cui, Z., Li, T., Pei, H., Sheng, Y., Li, X., et al. (2019). mRNA and miRNA expression analysis reveal the regulation for flower spot patterning in *Phalaenopsis* 'Panda'. *International Journal of Molecular Sciences*. 20, 4250. doi: 10.3390/ijms20174250

Zhao, M., Wang, T., Sun, T., Yu, X., Tian, R., Zhang, W. H., et al. (2020). Identification of tissue-specific and cold-responsive lncRNAs in *Medicago truncatula* by high-throughput RNA sequencing. *BMC Plant Biol.* 20, 99. doi: 10.1186/s12870-020-2301-1

Zheng, Y., Chen, K., Xu, Z., Liao, P., Zhang, X., Liu, L., et al. (2017). Small RNA profiles from *Panax notoginseng* roots differing in sizes reveal correlation between miR156 abundances and root biomass levels. *Sci. Rep.* 7, 9418. doi: 10.1038/s41598-017-09670-8

Zheng, C., Ye, M., Sang, M., and Wu, R. (2019). A regulatory network for miR156-SPL module in *Arabidopsis thaliana*. *International Journal of Molecular Sciences*. 20, 6166. doi: 10.3390/ijms20246166

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fpls.2023.1114988

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Zhiyuan Li,
Chinese Academy of Agricultural Sciences (CAAS), China
Rahat Sharif,
Yangzhou University, China
Qi Wang,
Tobacco Research Institute (CAAS), China

*CORRESPONDENCE
Zhaotang Ding
dzttea@163.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 03 December 2022
ACCEPTED 02 January 2023
PUBLISHED 02 February 2023

CITATION
Zaman S, Shen J, Wang S, Song D, Wang H, Ding S, Pang X, Wang M, Sabir IA, Wang Y and Ding Z (2023) Effect of shading on physiological attributes and comparative transcriptome analysis of *Camellia sinensis* cultivar reveals tolerance mechanisms to low temperatures. *Front. Plant Sci.* 14:1114988. doi: 10.3389/fpls.2023.1114988

COPYRIGHT
© 2023 Zaman, Shen, Wang, Song, Wang, Ding, Pang, Wang, Sabir, Wang and Ding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Effect of shading on physiological attributes and comparative transcriptome analysis of *Camellia sinensis* cultivar reveals tolerance mechanisms to low temperatures

Shah Zaman[1], Jiazhi Shen[1], Shuangshuang Wang[1], Dapeng Song[2], Hui Wang[2], Shibo Ding[2], Xu Pang[2], Mengqi Wang[2], Irfan Ali Sabir[3], Yu Wang[4] and Zhaotang Ding[1,4]*

[1]Tea Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China, [2]Tea Research Institute, Rizhao Academy of Agricultural Sciences, Rizhao, China, [3]School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [4]Tea Research Institute, Qingdao Agricultural University, Qingdao, China

Tea is a vital beverage crop all over the world, including in China. Low temperatures restrict its growth, development, and terrestrial distribution, and cold event variability worsens cold damage. However, the physiological and molecular mechanisms of *Camellia sinensis* under shade in winter remain unclear. In our study, tea leaves were utilized for physiological attributes and transcriptome analysis in November and December in three shading groups and no-shade control plants. When compared to the no-shade control plants, the shading group protected tea leaves from cold damage, increased photochemical efficiency (Fv/Fm) and soil plant analysis development (SPAD), and sustained chlorophyll *a*, chlorophyll *b*, chlorophyll, and carotenoid contents by physiological mean. Then, transcriptome analysis revealed 20,807 differentially expressed genes (DEGs) and transcription factors (TFs) in November and December. A comparative study of transcriptome resulted in 3,523 DEGs and many TFs under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD75% of shading in November and December. Statistically, 114 DEGs were downregulated and 72 were upregulated under SD0% *vs.* SD30%. SD0% *vs.* SD60% resulted in 154 DEGs, with 60 downregulated and 94 upregulated. Similarly, there were 505 DEGs of which 244 were downregulated and 263 were upregulated under SD0% *vs.* SD75% of shading throughout November. However, 279 DEGs were downregulated and 105 were upregulated under SD0% *vs.* SD30%. SD0% *vs.* SD60% resulted in 296 DEGs, with 172 downregulated and 124 upregulated. Finally, 2,173 DEGs were regulated in December, with 1,428 downregulated and 745 upregulated under SD0% *vs.* SD75%. These indicate that the number of downregulated DEGs in December was higher than the number of upregulated DEGs in November during low temperatures. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of differentially expressed genes were highly regulated in the photosynthesis, plant hormone signal transduction, and mitogen-activated protein kinase (MAPK) signaling pathways. However, qRT-PCR and RNA-seq relative expression of photosynthetic (DEGs)

Lhcb2 in both November and December, plant hormone (DEGs) *BRI1* and *JAZ* in November and *IAA* and *ERF1* in December, and key DEGs of MAPK signal transduction *FLS2*, *CHIB*, and *MPK4* in November and *RBOH*, *MKK4_5*, and *MEKK1* in December in three shading groups and no-shade control plants responded to tea cold tolerance. The enhanced expression of light-harvesting photosystem I gene *Lhca5*, light-harvesting photosystem II gene *Lhcb2*, and mitogen-activated protein kinases *MEKK1* and *MPK4/6* enhance the cold-tolerance mechanism of *C. sinensis*. These comprehensive transcriptomic findings are significant for furthering our understanding of the genes and underlying regulatory mechanisms of shade-mediated low-temperature stress tolerance in horticultural crops.

KEYWORDS

shading, tea, physiological attributes, transcriptome analysis, low temperature, signal transductions

1 Introduction

Tea *Camellia sinensis* is a beverage crop and a species of evergreen shrub or small tree with leaves, first discovered in China in 2737 B.C. The tea tree thrives in 10°C–30°C warm, humid climates (zone 8 climate or warmer). Due to climatic changes, tea plants undergo environmental and abiotic stresses such as cold, chilling, and hot and low temperatures, damaging plant development, quality, and yield (Zaman et al., 2022). Tea plants are susceptible to cold injury, especially during harsh winters (Li et al., 2019). Tea leaves are substantially harmed by cold stress, resulting in cellular disruption, decreased photosynthetic capability and chlorophyll formation, and other physiological changes that contribute to irregular growth and limit the tea plants' healthy development to survive in harsh environments (Chen et al., 2018). To offset the loss of tea yield and production during winters, innovative climatic and resilient agricultural practices are needed to alleviate abiotic stresses and harmful effects of low temperatures and reduce cold damage in tea plants (Zaman et al., 2022).

The temperature of the environment has a vital role in plant growth and development. It regulates photosynthesis (food production) and metabolism (food utilization). Each plant species has a temperature range in which it grows best based on its origin (e.g., tropical or temperate). The temperature has an enormous impact on plant resilience, fall color, and senescence (leaf fall). The capability of a plant to tolerate the average minimum temperature of a region without damage or death is determined by its hardiness. Even cold-hardy plants are vulnerable to the devastating effects of a rapid temperature drop to freezing (Li et al., 2019). Plants experience chilling and freezing as low-temperature stresses. Chilling temperatures (0°C–15°C) vary by plant kind and tolerance. Air temperature and wind speed affect chilling temperatures. At temperatures below (0°C), most of the plants initiate ice formation in cellular compartments and begin to freeze (Thomashow, 2010), which disrupts membrane permeability and the growth and development of horticultural plants. The gradual shifts in photoperiod and temperature throughout the year are detected by perennial plants such as tea trees to predict the upcoming season. These plants can shift from active growth to dormant phases and from frost sensitivity to cold tolerance in response to seasonal cues, including day length and temperature fluctuation. However, the seasonal cold adaptation of woody perennials is caused mainly by short autumn days, further enhanced by low temperatures (Aslam et al., 2022). Overwintering plants' morphological and physiological attributes coordinate with the climate's cycles, ultimately leading to their survival in low, moderate, and icy environments (Sakai and Larcher, 2012). During low temperatures, plants develop a mechanism known as cold acclimation (Zhang et al., 2019). This cold acclimation is an induced adaptive process that increases cold tolerance and is accomplished by exposing the plant to numerous low temperatures depending on the plant's adaptation to specific environments. Cold acclimation triggers a general response to cold stress. It can help avoid this cold damage, and the mechanisms of cold acclimation have been investigated incredibly well in forage legumes, cereal crops, and model plants like *Arabidopsis thaliana* (Baier et al., 2019). Tea plants may reduce the cold damage and sustain the photosystem systemic reaction of cold acclimation during cold stress (Li et al., 2019). Many significant crops, including rice, corn, soybean, potato, cotton, and tomato, are chilling sensitive and incapable of cold acclimation.

In contrast, other crops, such as oats, are chilling-tolerant but freezing-susceptible. However, barley, wheat, and rye are well suited to freezing conditions (Fu et al., 2011). Low temperatures alter the expression of thousands of differentially expressed genes (DEGs) and transcription factors, leading to metabolic rearrangement, physiological adaptation, and biochemical changes. Plant hormone changes are associated with multiple biological processes and complex regulatory networks (Thomashow, 1999). Changes in lipid composition, membrane fluidity, and structural organization occur in plants exposed to cold weather. Photosynthesis converts light energy into ATP and NADPH by chloroplast thylakoid membrane components. Low temperatures reduce the metabolic sink capacity for photosynthates; hence, photochemical rates must adjust. Low temperatures affect photosynthesis and carbon gain, determining

plant productivity, growth, and distribution. Cold-tolerant species may adapt to low temperatures. Changes in energy absorption and photochemical transformation *via* energy partitioning, as well as contemporaneous changes in chloroplast carbon metabolism, allocation, and partitioning, are mechanisms for overcoming the restrictions of short-term and long-term exposure to low temperatures. Thus, the mechanisms involved in photosynthetic acclimation to low temperatures arise from thylakoid membrane system alterations that affect photosynthetic electron transport. The genes and transcription factors that trigger the signals in the photosynthesis pathway regulate these photosynthetic processes during low temperatures, as well as post-transcriptional activation, increase the expression of sucrose synthesis enzymes, alter the expression of Calvin cycle enzymes, change the leaf protein content, and affect plant photosynthesis process. Plant hormones and their derivatives play a role in how plants react to cold temperatures (Fowler and Thomashow, 2002). Several hormones such as abscisic acid (ABA) and cytokinin (CA), brassinosteroids (BRs), gibberellins (GBs), salicylic acid (SA), and jasmonic acid (JA) contribute significantly under low-temperature stress (Ashraf and Rahman, 2018). Low-temperature stress causes plants to release hormones. Many studies suggest that plant hormones affect low-temperature responsiveness. Under low-temperature stress, plant response genes were ABA-dependent and ABA-independent, depending on ABA/AREB and DREB/CBF. The intracellular auxin response regulates cold-stressed plant growth. Salicylic acid reduces stress damage by boosting proline, antioxidants, heat shock protein, secondary metabolism, and sugar, improving plant stress tolerance. Low-temperature stress increases the endogenous salicylic acid level in cucumbers and the expression of COR genes in grafted cucumbers. Hormones such as ethylene boost cold resistance by enhancing the antioxidant enzyme system's activity. Under low-temperature stress, CRF2 and CRF3 regulate lateral root growth in *A. thaliana*. In fact, under low temperatures, hormone response is coordinated or antagonistic in horticultural plants. Thus, the plant hormone system's dynamic balance network is developed, which helps plants maintain normal growth and development under stress. The mitogen-activated protein kinase cascade is a signaling transduction module that converts extracellular inputs into intracellular responses. Many mitogen-activated protein kinases (MAPKs) have been found in horticultural plants, including tomatoes and apples, based on whole-genome sequencing and transcriptome analysis. Recent research has shown that the MAPK cascade is also important in the biotic and abiotic stress responses of horticulture plants, including low temperatures. The genes and transcription factors are triggered in plant hormone signal transduction and mitogen-activated protein kinase signaling pathways (Wu, 2022) to enhance plant immunity (Li et al., 2019) and increase tolerance. Mitogen-activated protein kinases are required to integrate diverse intracellular signals transmitted by stress-induced secondary messengers. Recent research indicates that mitogen-activated protein kinase signaling regulates the cold stress response in *A. thaliana* (Zhao et al., 2017).

Plants need to be shielded from the cold's destructive effects. Damage from chilling and freezing temperatures mimic that seen in natural settings. Greenhouses and tunnels provide significant protection from the cold for plants that are sensitive to the cold, but only if they are in good working order. Temperature fluctuations are less likely to cause damage in polyethylene-covered hoop homes. This is helpful for horticultural plants in preventing rapid temperature fluctuation within the greenhouse or tunnels. Perhaps a fast and exclusive strategy is needed to protect plants from cold temperatures in winter, and shading is one of the most fundamental and well-established techniques for crops implemented by growers all over the world (Mohotti, 2004). Shading is an effective strategy for limiting the quantity of sunlight that reaches tea plants. It helps to adjust important quality-related metabolites in tea leaves, improving tea growth, development, and flavor (Ku et al., 2010). Tea is a shade plant. In the past, we had successfully used different shading nets to protect tea plants from cold weather on physiological and physiochemical levels (Zaman et al., 2022), and shade is also one of the creative ways to protect plants from environmental factors (e.g., cold, chilling, frost, hot, light, and precipitation) and other harmful stresses or threats (Jiang et al., 2020) and cold stress (Zaman et al., 2022). The utilization of shade nets is a typical practice across various climatic settings to protect different crops on physiological, molecular, and cellular levels (Ji et al., 2018). Several researchers published studies on different tea cultivars in response to protective shading mechanisms under different environmental conditions (Ji et al., 2018; Zhang et al., 2022). Zaman et al. (2022) stated that different shade nets increased the physiological and biochemical attributes of *C. sinensis* (Zaman et al., 2022). Recently, a few studies were published on plant response to the shading on gene and transcriptional expression levels (Wu et al., 2020; Fang et al., 2022). The study of Wu et al. (2022) revealed that under three shading conditions compared to no-shade control plants, the phenotypic, physiological, and photosynthetic DEGs were enhanced in *Sambucus canadensis* (Wu et al., 2020). The authors also observed that the DEGs linked with photosynthesis, secondary metabolites, and other hormones were affected in shading groups compared to no-shade control plants. Thus, the result indicated an adaptive approach to retaining fitness under harsh conditions.

There are few studies on physiological responses, including DEG expression to cold stress in tea plants (Hao et al., 2018a; Hao et al., 2018b). Early low temperatures (4°C–16°C) may improve the tea plant's cold acclimation, and enzyme activities increase after 4°C–5°C of acclimatization (Li et al., 2019). Low temperatures affect several leaf metabolites, including tea plants, in which small RNAs changed after 1–48 h at 4°C (Zhang et al., 2014). These investigations are bound to one or a few varieties within specific environments and cannot find global responses across natural environments and germplasm variability. Consequently, our understanding of the dynamic changes of DEG expression and the transcription factors and regulatory mechanisms of *C. sinensis* under different shading nets is still limited. Smart, short-term, and resilient agrarian solutions are essential to alleviate the harmful effects of critical environmental stresses and enhance the growth and productivity of tea plants under low temperatures. Therefore, in the present study, we used high-throughput RNA-seq of tea cultivar *C. sinensis* cv. Zhong Cha 102 to identify the important DEG expression and transcription factors responsible for enhancing cold-tolerance mechanism with response to shading during low temperatures. Several DEGs and transcription factors enriched in the present study play an essential role in photosynthesis, plant hormone signaling, and mitogen-activated protein kinase signaling transduction in three shading groups

compared with no-shade control plants in November and December during low temperatures.

The aim of this study is to understand the physiological and transcriptome responses of important DEGs and transcription factors regulated under shading and no-shade responsible for the cold-tolerance mechanism of *C. sinensis* during low temperatures.

2 Materials and methods

2.1 Plant material and experimental setup

Experiments were carried out in the experimental field throughout during low temperatures at Rizhao, Shandong, China (35°514′N, 119° 662′E). In this experiment, we tested the 5-year-old Zhong Cha 102 C. *sinensis* cultivar in three shading groups compared to the no-shade group during low temperatures. Above the tea leaves, a single layer of black polytene shades was set up in four rows, each 50 m long and supported by four metal fence posts. The main difference between black clothing shades was their intensity, such as 0% no shade or control (SD1), 30% of shading (SD2), 60% of shading (SD3), and 75% shading (SD4) were used (Table 1). The trial was carried out in November and December 2021, and the first harvesting was performed at 11:00 a.m. (24/11) and the second at 11:00 a.m. (23/12) during day time. The first, second, and third upper mature leaves were collected from each replication. A total of 30 sub-samples were collected in three shading groups compared to the no-shade control. All samples with six biological replicates were obtained from each row, carefully stored with high moisture in light barrier laminate packaging, and sent to the laboratory for further physiological and gene expression analyses.

2.2 Physiological attributes

2.2.1 Cold damage analysis

Cold plant injuries in three shading groups and no-shade control plants were recorded. Browning symptoms (on leaf tissue) were confirmed to be a cold-damage indication (Alisoltani et al., 2015), and the percentage of leaf incurring cold injury on each plant was noted according to Zhang et al. (2020) and Zaman et al. (2022).

2.2.2 Chlorophyll fluorescence measurement

A portable photosynthesis system (Li-6400XT, LI-COR, Inc., Lincoln, NE, USA) was used to measure chlorophyll fluorescence. The photochemical efficiency of photosystem II (Fv/Fm) was also measured. In brief, the fourth developed leaf from the shoot tip was acclimated in the dark for 30 min in three shading groups compared with no-shade control plants during low temperatures. As mentioned in our previous work, soil plant analysis development (SPAD) values in three shading groups and no-shade control plants were also recorded (Zaman et al., 2022).

2.2.3 Determination of chlorophyll and carotenoid contents

For measuring the chlorophyll and carotenoid concentrations in three shading groups in comparison with no-shade control plants during low temperatures, 50 g of fresh tea leaves was cut into small pieces and kept in the dark at 25°C for 48 h in acetone–anhydrous ethanol solution (1:1) with minor modifications, and data were recorded according to our previous work (Zaman et al., 2022).

2.3 Transcriptome analysis

Tea leaf total RNA was extracted using RNAprep Pure, and the library was constructed following the RNA detection method. After quality control, 150-bp paired-end reads were generated using Illumina HiSeq. HISAT2 software utilized the aligned reference genome of *C. sinensis* 'Shuchazao'. Feature Counts program tallied the genes (Kim et al., 2015). FPKM quantified gene expression mapped reads will be released on 14 February 2023 in the database of the National Center for Biotechnology Information (NCBI) with accession number PRJNA905739. DESeq2 software examined DEGs between comparison groups (Liao et al., 2014). The Benjamini–Hochberg method was used to correct *p*-values and calculate the false discovery rate (FDR). |log2 (fold change)|2 or FDR = 0.05 was used to screen DEGs. After adjustment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with p = 0.05 were considered enriched. Transcriptome analysis was completed at Wuhan Metware Biotechnology Co., Ltd.

TABLE 1 Shade specification, measurement, and PAR rate under three shading treatments compared to non-shade control plants in November and December during low temperature.

			November	December
Treatments	Shade cloth specification	Shade level (%)	PAR (μmol m−2 s−1)	PAR (μmol m−2 s−1)
SD 0	None	0%	793 a	530 a
SD 1	Black polyethylene net curtains	30%	353 b	262 b
SD 2	Black polyethylene net curtains	60%	261 c	202 c
SD 3	Black polyethylenenet curtains	75%	77 d	48 d

Specification as per the manufacturer: Treatments; SD0 as control, SD1 30% shading, SD2 60% shading, and SD3 75% shading. Black polyethylene net curtains, single layer, purchased from Shouguang lvyuan plastic products factory (Weifang, China). The photosynthetically active radiation (PAR). Different letters represent a significant difference between the treatments indicated by LSD test at $p < 0.05$.

2.4 qRT-PCR analysis

To verify the accuracy of the transcriptome data, 11 DEGs were selected for expression-level validation. Primers were designed using Beacon Designer 8, and the primer sequences are shown in Supplementary Table 6. Quantitative real-time PCR (qRT-PCR) was performed on an analytikjena-qTOWER2.2 fluorescence quantitative PCR instrument (Germany) using 2× SYBR® Green master mix (DF, China). Three biological replicates were analyzed, with glyceraldehyde 3-phosphate dehydrogenase (CsGAPDH) gene used as the internal reference gene.

2.5 Statistical analyses

Statistix 8.1 software was used to determine significant differences between groups, and one-way analysis of variance (ANOVA) and Duncan's multiple-range tests (p-values < 0.05) were used to analyze physiological and gene expression data under shading group and no-shade control plants during low temperatures. GraphPad Prism software was used for making figures. Graphics were formatted using Adobe Photoshop CC 2019.

3 Results

3.1 Weather situation under shading during low temperatures

Monthly air temperature and relative humidity were measured at Rizhao, Shandong, China, from November to December 2021 in three shading groups compared to no-shade control plants during low temperatures (Figure 1). In November, the maximum air temperature was 18°C, the minimum temperature was −2.6°C, the maximum RH was 100%, and the minimum RH was 15.1%. Under SD0% control, the highest temperature was 16.4°C, the minimum temperature was −10.5°C, the maximum RH was 100%, and the minimum RH was 13.4%, as recorded in December 2021. Similarly, under SD30%, the maximum temperature was 17.9°C, the minimum temperature was −2.4°C, the maximum RH was 100%, and the minimum RH was 18.9%, as recorded in November; the maximum temperature was 16.9°C, the minimum temperature was −10.6°C, the maximum RH was 100%, and the minimum RH was 17.4% in December. In November, the maximum temperature was 18.1°C, the minimum temperature was −1.9°C, the maximum RH was 100, and the minimum RH was 18.9%. In December, the maximum temperature was 16.6°C, the minimum temperature was −11.1°C, the maximum RH was 100%, and the minimum RH was 17.3%. Similarly, in November, the maximum temperature was 19.1°C, the minimum temperature was −1.9°C, the maximum RH was 100%, and the minimum RH was 18.9%; in December, the maximum temperature was 16.3°C, the minimum temperature was −11.2, the maximum RH was 99.2%, and the minimum RH was 15.7%, recorded under SD75% of shading during low temperatures. Throughout this investigation, the highest temperature was 19.1°C, the lowest temperature was −11.2°C, the highest relative humidity was 100%, and the lowest was 13.4% (Supplementary Table 1).

3.2 Physiological attributes of *Camellia sinensis* under shading during low temperatures

To investigate the effect of shading on *C. sinensis*, the maximum cold injury was found under SD0% no-shade conditions throughout the experiment. The three levels of shading treatment significantly reduced plant cold damage compared to no-shade control plants under low temperatures. Interestingly, SD75% of shading-treated plants had the lowest cold injury in November and December. However, leaf injury was more severe in December than in November for all the treatments due to low temperatures (Figures 2A, B). Light quality has a significant influence on photosynthesis in plants. Photochemical efficiency (Fv/Fm) is the light quality that significantly impacts plants' photosynthesis. Fv/Fm illustrates the efficiency of light energy conversion inside the PSII reaction center or plants' possible maximum photosynthetic capability. As demonstrated in Figures 2C, D, the Fv/Fm efficiency

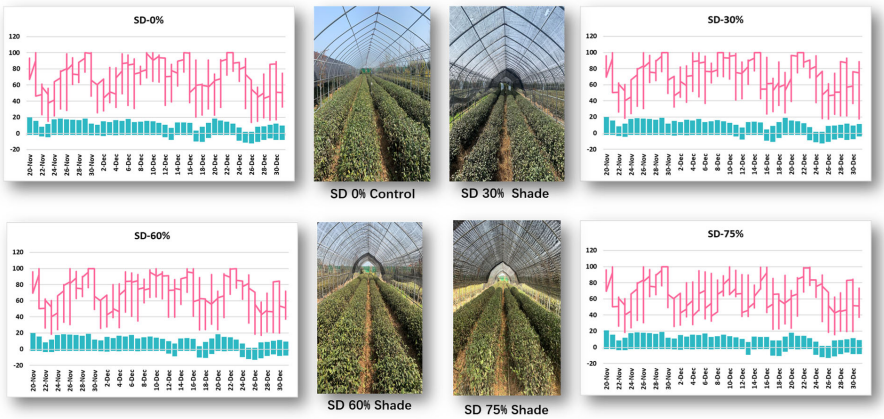


FIGURE 1
Weather situation under shading and no-shade conditions in November and December during low temperatures.

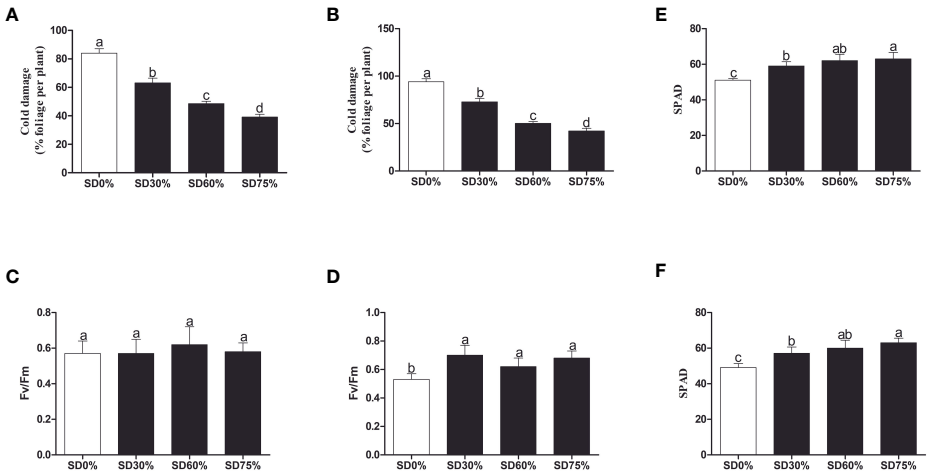


FIGURE 2
Tea plant performance under shading and no-shade in November and December during low temperatures. **(A)** Cold damage % in November. **(B)** Cold damage % in December. **(C)** Photochemical efficiency (Fv/Fm) in November. **(D)** Photochemical efficiency (Fv/Fm) in December. **(E)** Soil plant analysis development (SPAD) in November. **(F)** Soil plant analysis development (SPAD) in December. Different letters on the top of the columns within each figure represent significant differences between the treatments indicated by LSD test at $p < 0.05$. LSD, least significant difference.

was not significantly different in the three shading groups compared to no-shade control plants during low temperatures. However, a substantial decrease in Fv/Fm was seen in the SD0% no-shade control plants compared to the three shading groups in December during low temperatures in winter. The SPAD value is a parameter that evaluates the plant's relative photosynthetic pigments or indicates the plant's degree of greenness. As shown in Figure 2E, the SPAD value under unshaded control plants SD0% (50.700) was significantly lower than that under SD30% (58.700) and SD60% of shading (62.120 and 63.160) and significantly increased in SD75% of shading in November. Similarly, the minimum value (49.260) under SD0% (56.620), SD30% (60.080), and SD60% of shading (63.200) was significantly recorded under SD75% of shading in December during low temperatures (Figure 2F).

As shown in Figure 3A, the chlorophyll *a* and *b*, chlorophyll, and carotenoid levels were not significantly different in the three shading groups compared to no-shade control plants during low temperatures in November. In the shading group, chlorophyll *a* and *b* and total chlorophyll (Chll) slightly increased as compared to no-shade control plants. The carotenoid (Cx) content increased in the no-shade control plants compared to the three shading groups. Similarly, chlorophyll *a* and *b* and chlorophyll (Chll) increased in the shading group, but there were no significant differences between SD30%, SD60%, and no-shade control plants during November. However, under SD75% of shading, the levels of chlorophyll *a* and *b* and chlorophyll (Chll) and carotenoid (Cx) were significantly reduced in comparison with no-shade control plants in December during winters (Figure 3B). These indicate that the cold temperature in December affects tea leaves' green pigmentation during low temperatures.

3.3 Transcriptome profiles of *Camellia sinensis* under shading during low temperatures

Three shading groups were compared with no-shade control plants during low temperatures to understand the effects of shading on the molecular mechanisms of *C. sinensis*. The tea leaf samples in

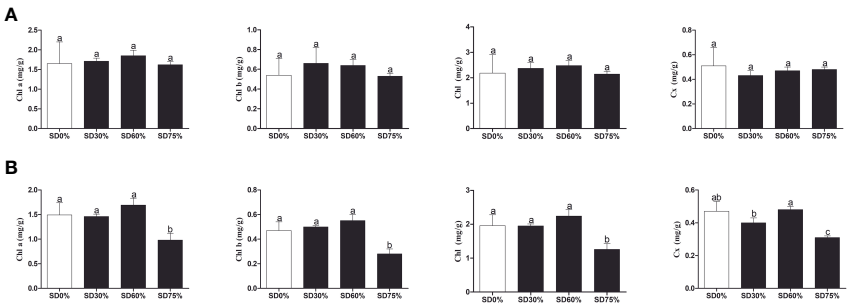


FIGURE 3
Physiological attributes of tea plants under shading and no shade in November and December during low temperatures. **(A)** Chlorophyll (Chl *a*), chlorophyll (Chl *b*), chlorophyll (Chl), and carotenoid (Cx) contents in November. **(B)** Chlorophyll (Chl *a*), chlorophyll (Chl *b*), chlorophyll (Chl), and carotenoid (Cx) contents in December. Different letters on the top of the columns within each figure represent significant differences between the treatments indicated by LSD test at $p < 0.05$. LSD, least significant difference.

three shading groups were compared with no-shade control plants used for RNA-seq analysis with three biological replicates. A total of 165.3 Gb of clean data was obtained from 24 libraries. Each library was corrected with a Q20 percentage greater than 96.80%, and Q30 is 91.38%. The quality of each sample in transcriptome data was high and suggested that the analysis of transcriptome data was accurate and reliable (Supplementary Table 2).

3.4 Comparison of DEGs under shading during low temperatures

In the present study, we conducted an inclusive transcriptome analysis using RNA-seq of tea leaves in three shading groups compared to no-shade control plants during low temperatures. Our goal was to understand better the processes by which tea leaves were affected by low temperatures. We were able to obtain 20,807 DEGs (Figure 4E). The patterns of these DEGs are combined into 10 primary cluster categories (Supplementary Table 3). In comparison with no-shade control plants, sub-class 1 (99 genes), sub-class 2 (125 genes), sub-class 3 (268 genes), sub-class 4 (128 genes), sub-class 5 (257 genes), sub-class 6 (239 genes), sub-class 7 (586 genes), sub-class 8 (366 genes), sub-class 9 (197 genes), and sub-class 10 (542 genes) showed a different pattern of response in response to three shading groups in November and December under low temperatures (Figure 4F). We compared the DEGs of each shading group with no-shade control plants during low temperatures, for instance, SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD75% of shading in November and December during low temperatures. In all group comparisons, a total of 3,523 DEGs were found. Figures 4A, B show that the number of downregulated DEGs was higher than that of upregulated DEGs. This indicates that cold temperature affected the numbers of DEGs and transcription factors in three shading groups compared with no-shade control plants during low temperatures. In the Venn diagram, 13 DEGs were shared in three shading groups in comparison with no-shade control plants during low temperatures in November, and 105 DEGs were shared in December in three shading groups in comparison with no-shade control plants during low temperatures (Figures 4C, D).

3.5 GO functional enrichment analysis under shading during low temperatures

The DEGs were estimated using Gene Ontology (GO) and the KEGG pathway to identify the genes associated with tea leaves in three shading groups compared with no-shade control plants in November and December during low temperatures. These genes are mainly involved in cellular processes, metabolic processes, signaling, response to stimulus, growth, and regulation of biological processes in the biological process (BP) category. In the cellular component (CC) process category, the enriched genes were involved in protein-containing complex cellular, anatomical entities. Similarly, genes related to catalytic activity, binding, transcriptional regulator activity, molecular transducer activity, and molecular function regulator enriched in the molecular function (MF) category under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD75% in three shading groups were compared with no-shade control plants in November and December during low temperatures (Figures 5A, B; Supplementary Table 4).

3.6 Pathway enrichment analysis under shading during low temperatures

Different DEGs were expressed into various KEGG pathways under shading and no-shade control plants in November and

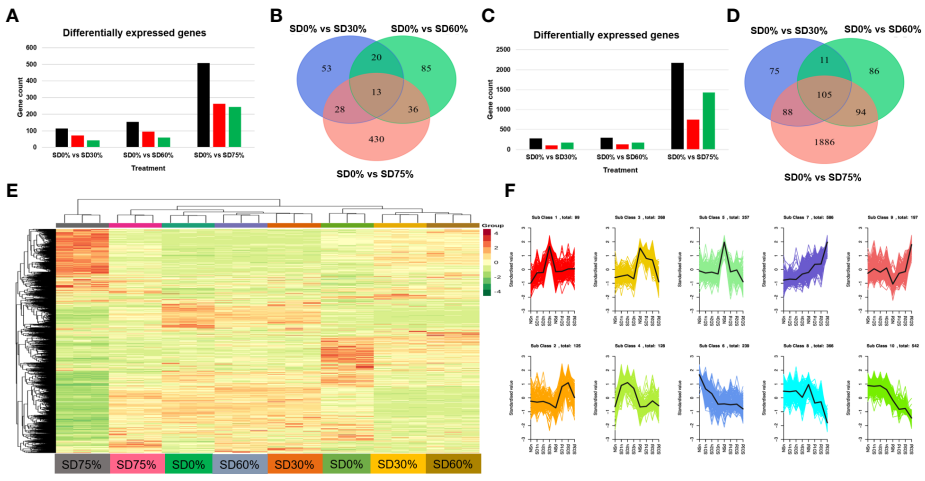


FIGURE 4
Transcriptome analysis of tea plant under shading and no shade in November and December during low temperatures. **(A)** Number of differentially expressed genes under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD 75% during November **(B)** Venn diagram of differentially expressed genes (DEGs) and shared DEGs under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD 75% during November. **(C)** Number of differentially expressed genes under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD 75% during December. **(D)** Venn diagram of differentially expressed genes and shared DEGs under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD 75% during December. **(E)** Heatmap of DEG regulation under different shading and no-shade control plants in November and December. **(F)** Cluster analysis of differentially expressed genes under different shading and no-shade control plants in November and December.

10.3389/fpls.2023.1114988

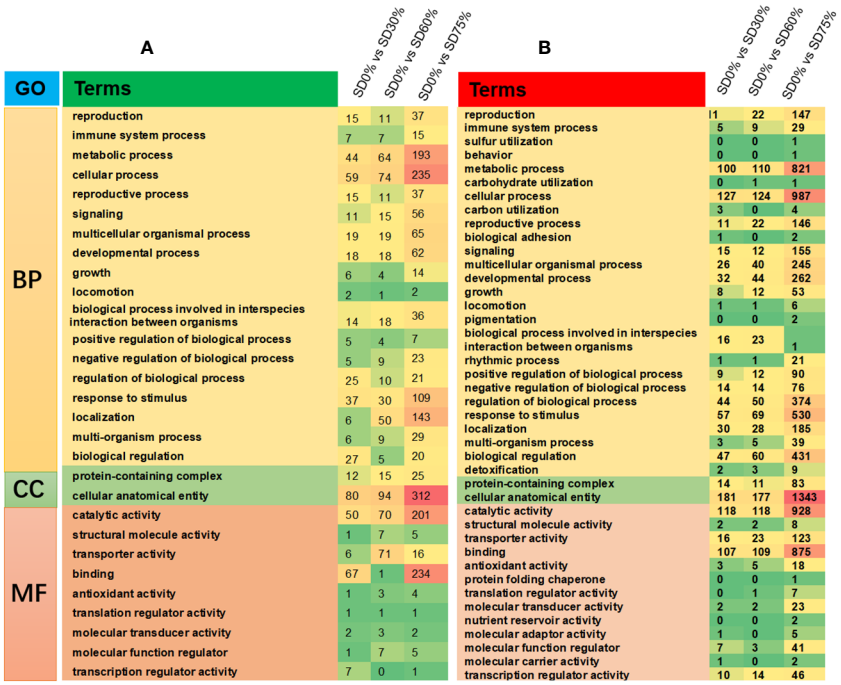


FIGURE 5
The classification of Gene Ontology (GO) terms of tea plants under shading and no shade in November and December during low temperatures. **(A)** Representation of DEGs enriched in biological process (BP), cellular component (CC), and molecular function (MF) under SD0% *vs.* 30%, SD0% *vs.* 60%, and SD0% *vs.* SD75% in November. **(B)** Representation of DEGs enriched in BP, CC, and MF under SD0% *vs.* 30%, SD0% *vs.* 60%, and SD0% *vs.* SD75% in December. DEGs, differentially expressed genes.

December during low temperatures (Supplementary Table 5). The key pathways focused on in this study were photosynthesis, plant hormone signal transduction, and mitogen-activated protein kinase signaling transduction plant (Figure 6D). Interestingly, plant hormone signal transduction and mitogen-activated protein kinase signal transduction were enriched under SD0% *vs.* SD30% (Figure 6A). Photosynthesis and mitogen-activated protein kinase signal transduction were enriched under SD0% *vs.* SD60% (Figure 6B), and photosynthesis-antenna, plant hormone signal transduction, and mitogen-activated protein kinase signal transduction were enriched under SD0% *vs.* SD75% during low temperatures (Figure 6C). The KEGG enrichment analysis showed that significant enrichment pathways increased with low temperatures and that the DEGs significantly enriched in these pathways in three shading groups compared with no-shade control plants are likely to play essential roles in coping with low-temperature stress in tea leaves (Figure 6).

3.7 Key DEGs in photosynthesis pathway under shading during low temperatures

The expression of crucial DEGs in the photosynthesis pathway led us further to evaluate the functional annotation of these expressed DEGs. We find that the main DEGs-encoded proteins were found in chloroplasts, where they act as antenna proteins, components of photosystems I and II, cytochromes, and electron transporters (Figure 7). A comparison of these DEGs' expression in three shading groups in comparison with no-shade control plants in November and December during low temperatures revealed seven genes that had similar patterns of expression, and some of these DEGs were regulated more than one time. Interestingly, we find that light harvesting in photosystem I gene *Lhca5* and light harvesting in photosystem II gene *Lhcb2* had quite different expression patterns under SD0% *vs.* SD75% in November and December under low temperatures (Figure 7C). *Lhca5* encrypts a protein in the membrane involved in photosynthesis system I, and Lhcb2 encodes light-harvesting chlorophyll *a*/*b*-binding (LHC) proteins that constitute the antenna system of the photosynthetic apparatus (Figure 7A). We validated the key DEG *Lhcb2* under SD0% *vs.* SD75% in both November and December during low temperatures. The expression level using qRT-PCR is shown in Figures 7E, F. Hence, high expressions of *Lhcb2* were positively correlated in November and December during low temperatures. The other two essential genes *PsbQ* photosystem II subunit q and *PsbR* subunit of photosystem II were highly expressed under SD0% *vs.* SD75% in November under low temperatures (Figures 7C, D). At the same time, the subunit K of photosystem I reaction center *PsaQ* and subunit of photosystem I *PsaO* were highly expressed under SD0% *vs.* SD75% in December under low temperatures (Figures 7C, D). As mentioned earlier, all DEGs were highly expressed during low temperatures in November and December (Supplementary Table 7). Other essential DEGs of PSII K protein *psbK*, photosystem II subunit s *psbS*, subunit K of photosystem I reaction center *psaK*, two photosynthetic electron transporter, and multi-subunit enzyme energy-transducing membranes of chloroplasts were downregulated under SD0% *vs.* SD60% and SD0% *vs.* SD75% in three shading groups compared to no-shade control plants during low temperatures (Figures 7B, D).

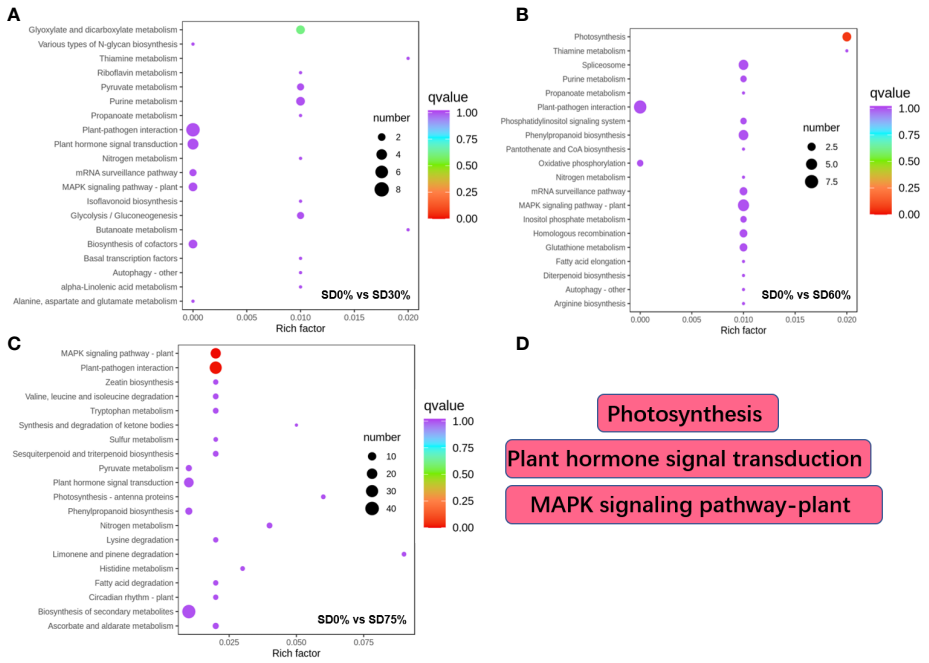


FIGURE 6
The numbers of DEGs expressed in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in tea plants under shading and no shade in November and December during low temperatures. **(A)** Numbers of DEGs enriched under SD0% *vs.* SD30%. **(B)** Numbers of DEGs enriched under SD0% *vs.* SD60%. **(C)** Numbers of DEGs enriched under SD0% *vs.* SD75%. **(D)** The key pathways in present study were photosynthesis, plant hormone signal transduction, and MAPK signaling pathway. DEGs, differentially expressed genes.

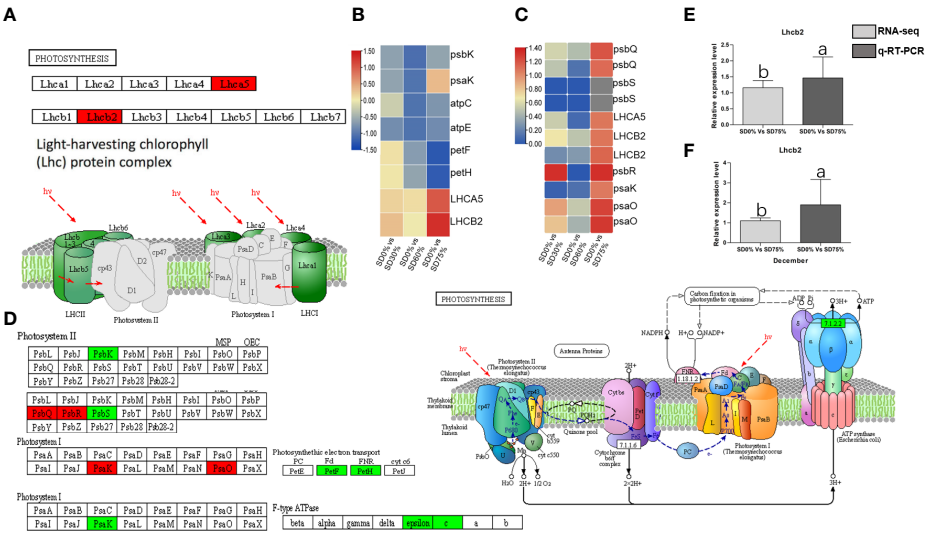


FIGURE 7
The numbers of DEGs were expressed in photosynthesis pathway in tea plants under shading and no shade in November and December during low temperatures. **(A)** Numbers of DEGs represented in red photosynthesis. **(B)** Heatmap representation of key DEGs' regulation in photosynthesis pathway under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD75% in November during low temperatures. **(C)** Heatmap representation of key DEGs' regulation in photosynthesis pathway under SD0% *vs.* SD30%, SD0% *vs.* SD60%, and SD0% *vs.* SD75% in December during low temperatures. **(D)** The numbers of DEGs were expressed in photosystem II and photosystem I; DEGs with green color indicate downregulation, and DEGs in red color indicate upregulations. **(E)** The relative expression level of key DEG Lhcb2 was validated by qRT-PCR under SD0% *vs.* SD75% in November. **(F)** The relative expression level of key DEG *Lhcb2* was validated by qRT-PCR under SD0% *vs.* SD75% in December. DEGs, differentially expressed genes.

3.8 Key DEGs and TFs activated in signal transduction pathways under shading during low temperatures

As most enrichment results are associated with signal transduction pathways, key DEGs and transcription factors generally play an essential part in these pathways, and we performed additional analysis on differentially expressed genes and transcription factors in three shading groups in comparison with no-shade control plants in November and December during low temperatures. The two signaling transduction pathways map (ko04075) "Plant hormone signal transduction" and map (ko04016) "mitogen-activated protein kinase signaling pathway" activated and enriched the key DEGs and transcription factors in three shading groups in comparison with no-shade control plants during low temperatures (Supplementary Table 8). Interestingly, 35 DEGs and transcription factors are associated with "Plant hormone signal transduction". Four key DEGs and transcription factors were initially activated under SD0% *vs.* SD30%. SD0% *vs.* SD60% accounted for six key DEGs and transcription factors, and the late response of signaling triggered 25 key DEGs and transcription factors under SD0% *vs.* SD75%. However, 21 DEGs and transcription factors were shared in three shading groups compared with no-shade control plants in November and December during low temperatures, as summarized in Table 2. This indicates that key DEGs and transcription factors play an essential role during low temperatures in *C. sinensis*. These DEGs and transcription factors were mapped into hormone signal transduction of auxin (ABA), CA, abscisic acid (AB), GBs, ethylene (ET), BRs, JA, and SA, as shown in Figure 8. KEGG pathway enrichment analysis of DEGs and transcription factors revealed that the mitogen-activated protein kinase signaling pathway was the most prominent in the three shading groups compared with no-shade control plants in November and December during low temperatures. Furthermore, 39 DEGs and transcription factors are associated with the mitogen-activated protein kinase signaling pathway. Among them, three DEGs and transcription factors were enriched under SD0% *vs.* SD30%, nine DEGs and transcription factors were activated under SD0% *vs.* SD60%, and 27 DEGs and transcription factors were triggered under SD0% *vs.* SD75% of shading (Table 3). Twenty-two DEGs and transcription factors were found to be common in three shading groups compared to no-shade control plants in November and December during low temperatures. Most MAPK cascades were activated and transmitted as primary, intermediate, and late responses during cold temperatures. The expression of utmost key DEGs *MEKK1* and *MPK4/6* functioning in cold stress was significantly higher in November and December during low temperatures (Figure 9).

4 Discussion

4.1 Physiological attributes of *Camellia sinensis* under shading during low temperatures

Temperature changes are among the most common abiotic stressors encountered by plants. Low-temperature stress affects plants' various morphological, physiological, biochemical, and molecular characteristics, ultimately inhibiting plant growth and development (Wang et al., 2018). Therefore, short-term strategies are needed to alleviate important environmental stresses and increase plant growth and development. Shade is one of the finest strategies to protect plants from various temperature stressors (Zaman et al., 2022) to increase growth and productivity and reduce cold damage or leaf injury in plants (Zhang et al., 2022). The present study showed cold damage to *C. sinensis* leaves in three shading groups compared with no-shade control plants in November and December during low temperatures. The results demonstrate that shaded tea leaves were shielded from leaf injury in November and December during low temperatures compared to no-shade control plants. It means that shading protected tea leaves from cold damage due to stopping ice accumulation in extracellular tissues (Zuther et al., 2018). Therefore, it might be helpful for alleviating the harmful effects of cold temperatures. However, different shading conditions modify the intensity and affect the photosynthetic capacity of plant growth and development under different environmental conditions, thereby regulating plant tolerance to high photosynthetically active radiation (PAR) or low PAR (Rezai et al., 2018). Our study revealed that the PAR (μmol m^{-2} s^{-1}) rate was significantly lower in three shading groups compared to no-shade control plants in November and December during low temperatures. Our results are like those of the previous study, which reported a higher PAR rate in the controlled group than in the shading groups (Thakur et al., 2019). We also measured the soil plant and development (SPAD) in three shading groups compared with no-shade control plants during low temperatures to quantify the relative amount of chlorophyll in the leaves. We found that SPAD values were significantly higher in three shading groups in comparison with no-shade control plants in November and December during low temperatures. Our results are consistent with those of previous studies (Kumari et al., 2014; Zaman et al., 2022). Increased SPAD values suggest higher chlorophyll concentrations in the leaf because plants grown under shade exhibited higher photosynthetic activity due to their higher photosynthetic pigments. Variations in photosynthetic pigmentation in response to low temperatures have received much attention in recent years. Photochemical efficiency is a reliable predictor for plant adoption under abiotic stress (Zuther et al., 2018), including low temperatures. However, we did not find many variations in the three shading groups compared to no-shade control plants in November during low temperatures. Nevertheless, when the temperature was colder in December, the cold weather affects the photochemical efficiency under no-shade control plants. These results indicate that cold reduces the photochemical efficiency, but different shades alleviated the harmful effects of low temperatures and sustained photochemical machinery, which might be the reason that shading is one of the effective systems in preventing leaf cold injury of tea plants during low temperatures. In fact, plants regularly interact with shade by strengthening light use efficiency through a shade tolerance strategy (Gong et al., 2015). In this work, we found no significant differences in chlorophyll *a* and *b*, chlorophyll, and carotenoid contents in November. A lower amount of chlorophyll *a* and *b* and a higher chlorophyll concentration were beneficial for photosynthesis under different shades, consistent with an earlier study (Li et al., 2020). The degradation in chlorophyll concentration under SD0% no-shade plants in November indicates that cold affects tea

10.3389/fpls.2023.1114988

TABLE 2 DEGs involved in plant hormone signal transduction pathway.

	SD0% Vs. SD30%	
Gene Symbol	Manually annotated function	Shared DEG
SAUR	SAUR family protein	
BRI1	protein brassinosteroid insensitive 1	YES
JAZ	jasmonate ZIM domain-containing protein	YES
MYC2	transcription factor MYC2	YES
	SD0% Vs. SD60%	
IAA	auxin-responsive protein IAA	
GID1	gibberellin receptor GID1	
MKK4_5	mitogen-activated protein kinase kinase 4/5	
BAK1	brassinosteroid insensitive 1-associated receptor kinase 1	
BRI1	protein brassinosteroid insensitive 1	YES
MYC2	MYC2	YES
	SD0% Vs. SD75%	
IAA	auxin-responsive protein IAA	
ARF	auxin response factor	
ARR-B	two-component response regulator ARR-B family	YES
ARR-A	two-component response regulator ARR-A family	YES
GID1	gibberellin receptor GID1	YES
DELLA	DELLA protein	YES
PIF4	phytochrome-interacting factor 4	YES
PP2C	protein phosphatase 2C	YES
CTR1	serine/threonine-protein kinase CTR1	YES
MKK4_5	mitogen-activated protein kinase 4/5	
EBF1_2	EIN3-binding F-box protein	
ERF1	ethylene-responsive transcription factor 1	YES
BAK1	brassinosteroid insensitive 1-associated receptor kinase 1	YES
BRI1	protein brassinosteroid insensitive 1	YES
BSK	BR-signaling kinase	
BIN2	protein brassinosteroid insensitive 2	YES
BZR1_2	brassinosteroid resistant 1/2	YES
TCH4	Xyloglucan xyloglucans transferase TCH4	YES
CYCD3	cyclin D3, plant	
JAR1_4_6	jasmonic acid-amino synthetase	
JAZ	jasmonate ZIM domain-containing protein	YES
MYC2	transcription factor MYC2	YES
TGA	transcription factor TGA	
PR1	pathogenesis-related protein 1	YES
MPK6	mitogen-activated protein kinase 6	

10.3389/fpls.2023.1114988

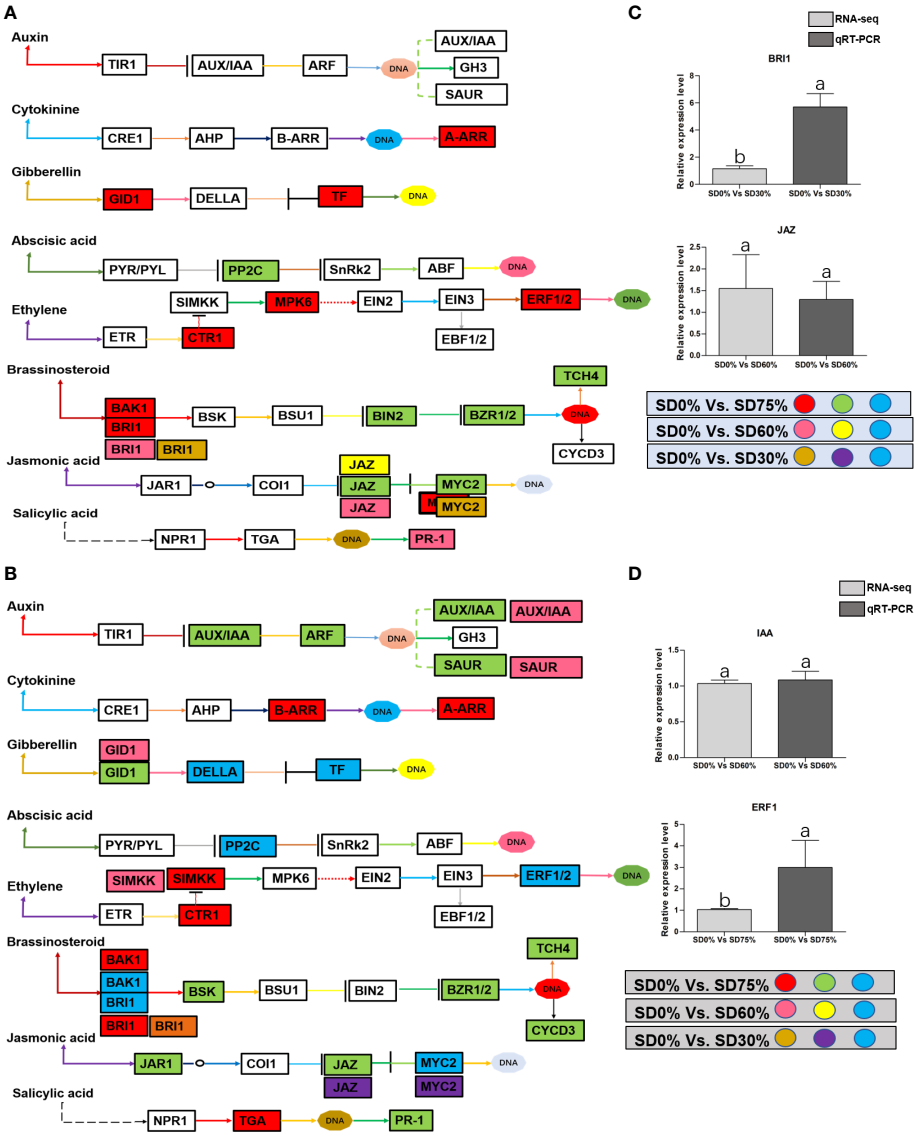


FIGURE 8
(A) The numbers of DEGs were expressed in Plant hormone signal transduction pathway in tea plant under shading and no-shade in November during low temperature. First, the up-regulation of these DEGs in red, down-regulation in green and both up and down regulation in sky blue color under SD0% Vs. SD75%. Second, the up-regulation of DEGs in pink, down-regulation in yellow and both up and down regulation shown in sky blue colors under SD0% Vs. SD60%. Third, the up-regulation of DEGs in brown, down-regulation in purple and both up-and down regulation in sky blue colors under SD0% Vs. SD75% **(C)** The relative expression level of key DEGs *BRI1* and *JAZ* were validated by q-RT-PCR under SD0% Vs. SD30, SD0% Vs. SD60% and % SD0% Vs. SD75% in November during low temperature. **(B)** The numbers of DEGs were expressed in Plant hormone signal transduction pathway in tea plant under shading and no-shade in December during low temperature. First, the up-regulation of these DEGs in red, down-regulation in green and both up and down regulation in sky blue color under SD0% Vs. SD75%. Second, the up-regulation of DEGs in pink, down-regulation in yellow and both up and down regulation shown in sky blue colors under SD0% Vs. SD60%. Third, the up-regulation of DEGs in brown, down-regulation in purple and both up-and down regulation in sky blue colors under SD0% Vs. SD75% **(D)** The relative expression level of key DEGs *IAA* and *ERF1* were validated by q-RT-PCR under SD0% Vs. SD30, SD0% Vs. SD60% and %SD0% Vs. SD75% in November during low temperature.

leaves and disrupts membrane fluidity in cell compartments during low temperatures. However, when the temperature decreases, we found that the concentrations of chlorophyll *a* and *b*, chlorophyll, and carotenoid were degraded under SD75% of shading in December during low temperatures. The reasons are unknown. Moving forward, under SD30% and SD60% of shade conditions, the concentrations of chlorophyll *a* and *b* and chlorophyll were higher as compared to no-shade control plants in December during low temperatures. This indicates that under different shade groups, the higher chlorophyll concentration and more pigment-binding proteins are associated with their photons collected by photosystem II per unit. The light-harvesting complex of photosystem II is mainly composed of chlorophyll *b* (Wu, 2022). In contrast, the low PAR, photochemical efficiency, and lower chlorophyll concentration in no-shade control plants indicate that cold temperatures might disrupt the membrane permeability in the cell compartment of tea leaves exposed to low temperatures. Thus, the above findings revealed that different shade groups protected against cold damage, improved the physiological attributes of tea leaves, and alleviated the harmful effects of cold during low temperatures.

TABLE 3 DEGs involved in MAPK signaling pathway.

	SD0% Vs. SD30%	
Gene Symbol	Manually annotated function	Shared DEG
ANP1	mitogen-activated protein kinase kinase ANP1	YES
MYC2	transcription factor MYC2	YES
ER/ERLs	LRR receptor-like serine/threonine-protein kinase ERECTA	
	SD0% VS SD60%	
WRKY33	WRKY transcription factor 33	
ANP1	mitogen-activated protein kinase ANP1	YES
PR1	pathogenesis-related protein 1	YES
CHIB	basic endochitinase B	YES
FLS2	LRR receptor-like serine/threonine-protein kinase FLS2	YES
BAK1	brassinosteroid insensitive 1-associated receptor kinase 1	YES
MKK4_5	mitogen-activated protein kinase 4/5	
CALM	Calmodulin	YES
ER/ERLs	LRR receptor-like serine/threonine-protein kinase ERECTA	YES
	SD0% VS SD75%	
WRKY33	WRKY transcription factor 33	
MPK3	mitogen-activated protein kinase 3	
WRKY22	WRKY transcription factor 22	
ANP1	mitogen-activated protein kinase ANP1	YES
CTR1	serine/threonine-protein kinase CTR1	YES
ERF1	ethylene-responsive transcription factor 1	YES
MPK4/6	mitogen-activated protein kinase 4/6	YES
MYC2	transcription factor MYC2	YES
PP2C	protein phosphatase 2C	YES
MAPK17_18	mitogen-activated protein kinase 17/18	YES
CALM	Calmodulin	YES
ER	LRR receptor-like serine/threonine-protein kinase ERECTA	YES
FLS2	LRR receptor-like serine/threonine-protein kinase FLS2	YES
BAK1	brassinosteroid insensitive 1-associated receptor kinase 1	YES
PR1	pathogenesis-related protein 1	YES
FRK1	senescence-induced receptor-like serine/threonine-protein kinase	
OXI1	serine/threonine-protein kinase OXI1	
MEKK1	mitogen-activated protein kinase 1	
PR1	pathogenesis-related protein 1	
MKK9	mitogen-activated protein kinase 9	
EBF1_2	EIN3-binding F-box protein	
MKK9	mitogen-activated protein kinase 9	
CHIB	basic endochitinase B	YES
CAT1	catalase	
RBOH	respiratory burst oxidase	
MKK4_5	mitogen-activated protein kinase 4/5	
SPCH	transcription factor SPEECHLESS	

4.2 Key pathways enriched in *Camellia sinensis* under shading during low temperatures

According to various transcriptome data, several genes and transcription factors play an essential role in photosynthesis, plant hormone signal transduction, and mitogen-activated protein kinase signal cascades under different abiotic stresses (Li et al., 2019; Ermilova, 2020; Ritonga and Chen, 2020; Aslam et al., 2022). In this study, the annotation of Gene Ontology and KEGG enrichment analysis demonstrated that low temperatures regulated the expression of several DEGs in the photosynthetic pathway of photosynthesis (Wu et al., 2020) and activated numerous DEGs and transcription factors in signal transduction pathways (Wang et al., 2018; Wu, 2022) under shade and no-shade control plants in November and December, respectively.

4.3 Photosynthesis-related DEGs revealed in *Camellia sinensis* under shading during low temperatures

Photosynthesis is the essential process that drives the growth and development of a plant. In the photosynthesis process, photosystem I and II pigment–protein complexes are related *via* the cytochrome complex and the electron transport chain (Mamedov et al., 2015) from plastocyanin to Fd (Busch and Hippler, 2011). In this state, the reaction centers of the thylakoid membrane receive light energy that pigments have collected in the LHC proteins in photosystem I. However, a unique complex called photosystem II can absorb light and break down water (Shi et al., 2012a). Oxygen-evolving enhancer proteins, including *PsbB*, *PsbC*, *PsbR*, *PsbO*, *PsbP*, and *PsbQ*, are essential components of the oxygen-evolving complex (Silveira and Carvalho, 2016). Photosystem I contains the reaction center of subunits *PsaA* and *PsaB*; binding subunits *PsaG*, *PsaK*, *PsaH*, *PsaL*, *PsaO*, and *PsaP*; ferredoxin docking subunits *PsaD* and *PsaE*; plastocyanin docking subunits *PsaF* and *PsaN*; and *FeS* apoprotein (Ichino et al., 2014). In the current study, photosynthesis-related GO terms and KEGG pathways differed significantly between the three groups compared to no-shade control plants. Photosynthesis-related DEGs are enriched in photosynthesis, photosynthesis-antenna proteins, and carbon fixation in photosynthetic organisms' pathways. Photosynthesis involves photosystem II, cytochrome b6/f

10.3389/fpls.2023.1114988

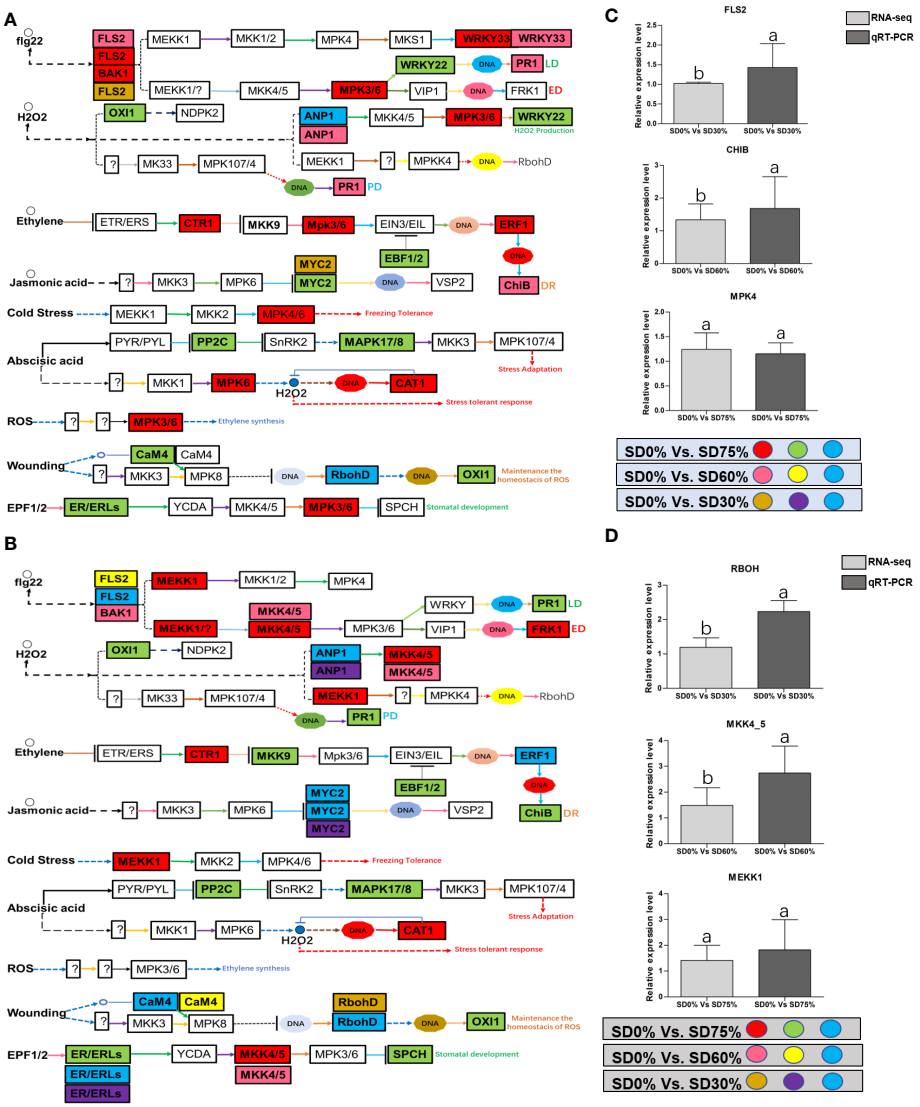


FIGURE 9
(A) The numbers of DEGs were expressed in MAPK signaling transduction pathway in tea plant under shading and no-shade in November during low temperature. First, the up-regulation of these DEGs in red, down-regulation in green and both up and down regulation in sky blue color under SD0% Vs. SD75%. Second, the up-regulation of DEGs in pink, down-regulation in yellow and both up and down regulation shown in sky blue colors under SD0% Vs. SD60%. Third, the up-regulation of DEGs in brown, down-regulation in purple and both up-and down regulation in sky blue colors under SD0% Vs. SD75% **(C)** The relative expression level of key DEGs *FLS2*, *CHIB* and *MPK4* were validated by q-RT-PCR under SD0% Vs. SD30, SD0% Vs. SD60% and % SD0% Vs. SD75% in November during low temperature. **(B)** The numbers of DEGs were expressed in *MAPK* signaling transduction pathway in tea plant under shading and no-shade in December during low temperature. First, the up-regulation of these DEGs in red, down-regulation in green and both up and down regulation in sky blue color under SD0% Vs. SD75%. Second, the up-regulation of DEGs in pink, down-regulation in yellow and both up and down regulation shown in sky blue colors under SD0% Vs. SD60%. Third, the up-regulation of DEGs in brown, down-regulation in purple and both up- and down regulation in sky blue colors under SD0% Vs. SD75% **(D)** The relative expression level of key DEGs *RBOH*, *MKK4_5* and *MEKK1* were validated by q-RT-PCR under SD0% Vs. SD30, SD0% Vs. SD60% and %SD0% Vs. SD75% in November during low temperature.

complex, electron transport, and ATP synthase. Most DEGs were increased and enriched in three shading groups compared with no-shade control plants in November and December during low temperatures. The enrichment and high expression of light-harvesting chlorophyll (LHCI), *Lhca5* in photosystem under SD0% *vs.* SD75%, may be a key DEG for coping with cold stress in November and December during low temperatures. LHCII-related DEG *Lhcb2* was increased under low temperatures. This suggests that shade promoted the regulation of LHC-related genes under low temperatures and improved the photosystem machinery of tea leaves during low temperatures. Our findings are congruent with those of a previous study in which authors discovered that the high expression of light-harvesting chlorophyll Lhca2 in the photosystem is an important contributor to cold tolerance in tea leaves (Li et al., 2019). In addition, the core complex subunits of photosystem II *PsbR* and *PsbQ* were highly expressed under SD0% *vs.* SD75% in December during low temperatures, and the photosystem I core complex contains reaction center subunits *PsaK* and *PsaO*, which were also highly expressed under SD0% *vs.* SD75% during low temperatures. At the same time, *psaK* was found to be downregulated under SD0% *vs.* SD75% in November during low temperatures. We validated the highly regulated key DEG *Lhcb2* expressed in November and

December under shading and no-shade control plants, respectively. The relative expression level was satisfactory between RNA-seq and qRT-PCR analysis of *Lhcb2*.

4.4 Plant hormone signal transduction DEGs and TFs activated in *Camellia sinensis* under shading during low temperatures

In order to survive in low-temperature environments, plants have evolved a set of complex signaling pathways (Wang et al., 2018). Among them, plant hormone signal transduction is a crucial pathway for activating DEGs and transcription factors under different shading conditions (Lorrain et al., 2008; Gommers et al., 2017). These DEGs and transcription factors are active during low temperatures (Wu, 2022). In the present study, several critical DEGs and transcription factors were regulated in three shading groups compared with no-shade control plants in November and December during low temperatures. *BRI1* and *MYC2* were upregulated and *JAZ* was downregulated under SD0% *vs.* SD30%, and under SD0% *vs.* SD60%, *BRI1*, *JAZ*, and *PR-1* were expressed highly. *AA-R*, *GIDI1*, *CTRI1*, *MPK6*, *ERF1/2*, *BAK1*, and *BRI1* were also expressed highly and *PP2C*, *BIN2*, *BZR1/2*, *TCH4*, *JAZ*, and *MYC2* were downregulated in November during low temperatures. In contrast, *BRI1* was upregulated and *JAZ* and *MYC2* were downregulated under SD0% *vs.* SD30%. Under SD0% *vs.* SD60%, *AUX/AA*, *GIDI*, *SIMKK*, *BAK1*, *BRI1*, and *MYC2* were found to be expressed highly. However, *AUX/AA*, *ARF*, *SAUR*, *B-AAR*, *A-AAR*, *GIDI*, *DELLA*, *PPC2*, *CTRI*, *SIMKK*, *EBF1/2*, *ERF1/2*, *BAK1*, *BRI1*, *BSK*, *BZR1/2*, *TCH4*, *CYCD3*, *JAR1*, *JAZ*, *MYC2*, *TGA*, and *PR-1* were differentially regulated in December during low temperatures. This indicates that several important DEGs and transcription factors are associated with hormone signal transduction playing an essential role in three shading groups in comparison with no-shade control plants in November and December during low temperatures in *C. sinensis*, and the molecular mechanism of these DEGs is illustrated in Figure 8. Abiotic stress responses in plants are primarily regulated by the ABA signaling pathway, which causes profound variations in gene expression and adaptive physiological responses (Danquah et al., 2014). DEGs involved in the production of plant hormones like auxin (IAA), ABA and CA, BRs, GBs, SA, and JA are differentially expressed due to variations in the expression of genes triggered in the signal cascade mechanism (Wang et al., 2018). The signaling route is triggered by auxin in response to low temperatures. Auxin is the first hormone released and initiates the hormone signal transduction process in plants. Auxin/indole-3-acetic acid and *AUX/IAA* family, response factor *ARF* family, and the short auxin RNA-SAUR family (Luo et al., 2018) are all auxin-responsive factors that help plants swiftly detect and respond to changes in auxin levels under low-temperature stress. The results of this study demonstrated that DEGs of auxin play a significant part in mitigating the negative effects of cold weather by regulating the expression of *AUX1*, *ARF*, and *SAUR*. Moreover, most known reactions of plant cytokinin metabolism or signaling systems to low temperatures are inhibitory (Argueso et al., 2009), but the responses are complex (Pavlů et al., 2018). Interestingly, cold temporarily activates type A *ARR* expression in a cytokinin- and ethylene-dependent way (Shi et al., 2012b). In our work, cytokinin *B-ARR* and type A *ARR* were expressed highly in three shading groups compared with no-shade control plants in November and December during low temperatures. Our results demonstrate that these two cytokinins interact to manage cold stress signals and alleviate the harmful effects of low temperatures in tea leaves. These findings align with earlier research, in which the author stated that mutations in cytokinin-type genes increase cold tolerance in *A. thaliana* during cold temperatures (Jeon et al., 2010). The classic phytohormones that control plant growth and abiotic stressors are gibberellic acid. The soluble receptor of *GID1* controls plant growth and development by interacting with *DELLA* protein in the gibberellic signaling pathway. The involvement of these DEGs in signaling pathways improved cold tolerance, and the overexpression of *GID1* and *DELLA* was regulated during low temperatures in three shading groups compared with no-shade control plants during low temperatures. The enzymes involved in ABA synthesis are zeaxanthin epoxidase, 9-*cis*-epoxycarotenoid dioxygenase, and abscisic aldehyde oxidase (Finkelstein, 2013). In a previous study, two ZEP genes, *NCDE 1* gene and *SnRK*s, were identified and expressed differentially under cold stress (Li et al., 2019), which functioned as part of a double-negative regulatory network that enhanced ABA signal transduction with *PYR/PYL*, *PP2C*, and *SnRK2*. In our study, we found that *PPC2* regulated differentially in three shading groups in contrast with no-shade control plants in November and December during low temperatures might play an important role during cold stress. BRs are a class of steroid hormones that play an important role in plant growth and development (Sharif et al., 2022a). Both BRs and ET are well known for their roles in plant growth and development during biotic and abiotic stressors (Jiroutova et al., 2018), including low temperatures. Our research identifies DEGs and transcription factors involved in brassinosteroids and ethylene, which might induce cold tolerance in tea leaves in three shading groups in comparison with no-shade control plants in November and December during low temperatures, respectively; these DEGs and transcription factors may be cooperatively controlled to facilitate tea plant adaptation to cold stress. More studies are needed to fully understand how they interact to regulate when plants are subjected to cold conditions. Furthermore, plants use jasmonic acid and its derivative as signaling molecules in response to biotic or abiotic environmental stress. This activates a variety of jasmonic-related genes and transcription factors that control plants' protective responses to various stresses (Gális et al., 2009). Among them, MYC family transcription factors are the important transcription factors involved in the hormone signaling transduction pathway in plants' response to low temperatures. To put it simply, MYC2, one of the newly identified plants MYC transcription factors, which is found in the model plant *A. thaliana* and mostly used in in-depth investigations under different abiotic stress, plays a regulatory role by establishing the *COI1*/*JAZs*/*MYC2* complex (An et al., 2022). In this study, we also discovered DEGs of JAZ and transcription factor MYC2 expressed differentially in three shading groups in comparison with no-shade control plants in November and December during low temperatures. Our result is consistent with a previous study in which *MYC2*, *JAZ1*, *JAZ2*, *JAZ3*, and *JAZ12* are involved in the jasmonic signal transduction pathway and were significantly upregulated in *Phoebe bournei* under shade (An et al., 2022). We assume that the various shading strategies in our work are reliable for coping with cold tolerance in tea and played a

significant part in the differential expression of JAZ and MYC2 during low temperatures. Moreover, Figures 8A, B show that there was a certain alteration in the expression of DEGs and transcription factors involved in the plant hormone signaling pathway, indicating a dramatic shift in hormone signaling. Undoubtedly, different stresses cause a cross-talk between hormone channels in plants (Murphy, 2015; Li et al., 2019). In fact, the supreme hormone auxin is still less understood in response to environmental stresses, which requires further exploration (Sharif et al., 2022b). Therefore, the reported findings of DEGs and transcription factors in our study are consistent with the hormonal cross-talk way that might raise cold tolerance and thereby alleviate cold damage in *C. sinensis* under low temperatures. However, we validated the high expression of key DEGs *BRI1* and *JAZ* under shading and no-shade control plants in November and *IAA* and *ERF1* in December. The relative expression level was reasonable between RNA-seq and qRT-PCR analysis of confirmed DEGs enriched in plant hormone signal transduction pathway (Figures 8C, D).

4.5 Mitogen-activated protein kinase cascades activated in *Camellia sinensis* under shading during low temperatures

Mitogen-activated protein kinase cascades are essential in regulating many distinct biological processes, including cell function, proliferation, and response to various abiotic stresses (Hao et al., 2015), including low temperatures. In the current study, several WRKY, FLS, MYC2, ERLs, and ERFs and MAPKKK, MAPKK, and MAPK families were regulated with key DEGs that transmitted signaling in tea leaves in three shading groups in comparison with no-shade control plants in November and December during low temperatures. Though, the regulations of these genes and transcription factors transmitted signaling in three steps. In the first step, the initial cold sensor was triggered in cell compartments when tea leaves were subjected to less cold temperatures in three shading groups compared to no-shade control plants during low temperatures. However, when cold temperature increases, the high expression of DEGs triggered signal transduction into the nucleus in three shading groups compared with no-shade control plants during low temperatures. Moreover, the expression of upregulation and downregulation of DEGs and transcription factors in three shading groups in comparison with no-shade control plants was responsible for the transmission of signaling, and transduction events occurred inside the cell during low temperatures. This indicates that cold weather triggers the cascade, and weather fluctuations owing to different shades regulate the expression of DEGs and transcription factors in tea leaf cell compartments. The activation of these cascades may help maintain membrane permeability, preventing ice formation in cell compartments of tissues at low temperatures. Low temperatures restrict cell membrane fluidity and create ice formation in the cell compartment of plants. However, the cascade activation plays an essential role in signaling events' genes and transcription factors, enhancing plant tolerance responses to low temperatures (Chen and Thelen, 2013). Regarding shading conditions, the GO categorization results of their DEGs are involved in cellular processes, membranes, signaling, and binding. According to the processes mentioned earlier, low-temperature signaling events occur in the cellular compartment and the transmission of signals cascades into the membrane and binds the membrane fluidity under low temperatures. The ethylene-responsive DEGs were highly expressed in three shading groups compared with no-shade control plants in November and December during low temperatures (Figures 9A, B). To cope with low temperatures, ET, ABA, and JA hormone derivatives might be transmitted signals under different shades. Our result is consistent with that of Müller and Munné-Bosch (2015), who reported that ET, ABA, and JA could activate ERF genes (Müller and Munné-Bosch, 2015). ERFs bind to the GCC box and DRE elements at low temperatures, providing cold tolerance to certain plants (Ritonga and Chen, 2020). Similarly, we identified the essential transcription factor families MAPKKK, MAPKK, and MAPK were activated in three shading groups compared with no-shade control plants in November and December during low temperatures. It indicates that mitogen-activated protein kinases play a crucial role in integrating diverse intracellular signals provided by shading-induced secondary messengers to low temperatures. In most cases, signal transduction will activate dormant MAPKKKs, the first kinase in the mitogen-activated protein kinases cascade. After being activated, MAPKKKs phosphorylate MAPKKs on a conserved serine/threonine and then trigger both (Ritonga and Chen, 2020). The signaling pathway of MAPK by activated MAPKKs is a further step in MAPK engagement. Mitogen-activated protein kinase signaling plays a role in *A. thaliana* cold stress response *via* regulating the ICE1 pathway (Zhao et al., 2017). Moreover, we also validated the highly expressed key DEGs regulated in the mitogen-activated protein kinase signaling transduction pathway. The key genes were *FLS2*, *CHIB*, and *MPK4* expressed in November and *RBOH*, *MKK4_5*, and *MEKK1* in December under shading and no-shade control plants, and the relative expression level was similar between RNA-seq and qRT-PCR analysis (Figures 9C, D). Moreover, we identified two significant DEGs, *MEKK1* and *MPK4/6*, enriched in the mitogen-activated protein kinase signaling pathway, a novel discovery for our tea cultivar Zong Cha 102. Both DEGs were highly regulated under shading groups compared with no-shade control plants in November and December during low temperatures. These DEGs are positive indicators for cold tolerance in *C. sinensis* (Figure 10). Our novel findings are consistent with Teige et al. (2004), in which the authors reported that *MEKK1* directly mediates *MPK4* and *MPK6* and that the regulation of these two key genes in the mitogen-activated protein kinase signaling pathway in model plant *A. thaliana* is mainly responsible for cold stress tolerance (Teige et al., 2004).

5 Conclusion

In summary, we revealed some novel physiological and transcriptomic information from the tea plant at low temperatures in three shading groups and no-shade control plants. We discovered that the adaptive regulation of tea leaves under shading conditions is primarily related to physiological characteristics such as photochemical machinery, photochemical efficiency of soil plant development, and chlorophyll content and gene expression analysis. In comparison to no-shade control plants, the shading group performed a main role in the protection mechanism of tea leaves against cold damage. Furthermore, DEGs and transcription factors associated with photosynthesis, plant

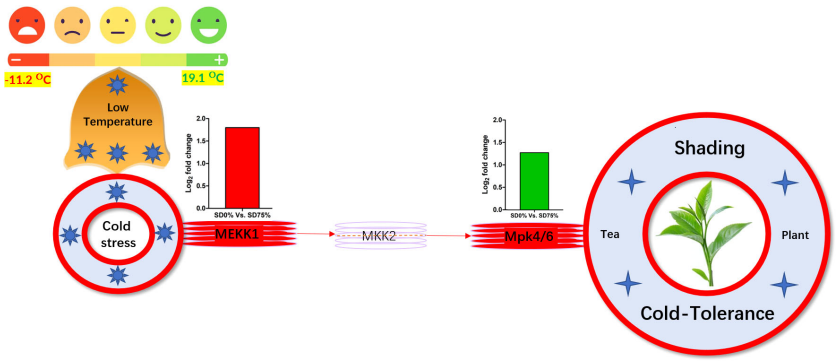


FIGURE 10
Systematic representation of key DEGs *MEKK1* and *MPK4/6* activated under shading in mitogen-activated protein kinase signaling pathway during low temperatures might be responsible for cold-tolerance mechanism of tea. DEGs, differentially expressed genes.

hormone signaling, and mitogen-activated protein kinase signaling pathways played an important role in shading and no-shade control plants. The light-harvesting photosystem I gene *Lhca5* and light-harvesting photosystem II gene *Lhcb2* synthesized the CO_2 fixation and maintained photosynthetic machinery, sustaining chlorophyll concentration. Finally, DEGs involved in plant hormone signaling and mitogen-activated protein kinase signaling cascades were activated, and shading might protect the tea plants from cold injury due by retaining the ice accumulation and maintaining membrane fluidity in cellular compartments of leaf tissues under low temperatures. Mitogen-activated protein kinase signal transduction DEGs *MEKK1* and *MPK4/6* are key players in this study for strengthening the cold-tolerance mechanisms in *C. sinensis*. Taking it all together, in this study, we confirm that different shading groups compared to no-shade conditions played a substantial role as a cover for sheltering tea cultivar Zhong Cha 102 on physiological and molecular levels during low temperatures. Also, shading is a fast and reliable technique for enhancing the cold-tolerance mechanism in *C. sinensis*. This study paves the way for researchers to better understand the protective approach for shielding horticultural crops under low-temperature scenarios.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization of work and evaluation by SZ and ZD. Methodology and investigation by SW and DS. Physiological analysis performed by HW, XP, and MW. Physiological data analysis performed by JS, DS, and SD. Manipulation of transcriptome data and manuscript written by SZ. The article was revised and organized by IS. Project direction, resources, and funding achieved by YW and ZD. All authors contributed to the article and approved the submitted version.

Funding

The research was funded by the Technology System of Modern Agricultural Industry in Shandong Province (SDAIT-19-01), the Special Foundation for Distinguished Taishan Scholar of Shandong Province (No. ts201712057), the Scientific and Technological Innovation Project of SAAS (CXGC2022B03), and the Special Talent Program of SAAS (CXGC2022E18).

Acknowledgments

We would like to thank Wuhan Metware Biotechnology Company for the transcriptome analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1114988/full#supplementary-material

SUPPLEMENTARY TABLE 1
Climatic data.

SUPPLEMENTARY TABLE 2
Summary of raw reads.

SUPPLEMENTARY TABLE 3
Cluster analysis.

SUPPLEMENTARY TABLE 4
GO terms.

SUPPLEMENTARY TABLE 5
Enrichment of KEGG pathways.

SUPPLEMENTARY TABLE 6
Primer's information used in qRT-PCR detection.

SUPPLEMENTARY TABLE 7
DEGs regulations in photosynthesis pathway.

SUPPLEMENTARY TABLE 8
DEGs regulations in hormone signaling pathway and MPAK signaling transduction pathway.

References

Alisoltani, A., Shiran, B., Fallahi, H., and Ebrahimie, E. (2015). Gene regulatory network in almond (Prunus dulcis mill.) in response to frost stress. *Tree Genet. Genomes* 11, 100.

An, J., Wei, X., and Huo, H. (2022). Transcriptome analysis reveals the accelerated expression of genes related to photosynthesis and chlorophyll biosynthesis contribution to shade-tolerant in phoebe bournei. *BMC Plant Biol.* 22, 268. doi: 10.1186/s12870-022-03657-y

Argueso, C. T., Ferreira, F. J., and Kieber, J. J. (2009). Environmental perception avenues: The interaction of cytokinin and environmental response pathways. *Plant Cell Environ.* 32, 1147–1160. doi: 10.1111/j.1365-3040.2009.01940.x

Ashraf, M. A., and Rahman, A. (2018). "Hormonal regulation of cold stress response," in *Cold tolerance in plants: Physiological, molecular and genetic perspectives*. Eds. S. H. Wani and V. Herath (Cham: Springer International Publishing), 65–88. doi: 10.1007/978-3-030-01415-5_4

Aslam, M., Fakher, B., Ashraf, M. A., Cheng, Y., Wang, B., and Qin, Y. (2022). Plant low-temperature stress: Signaling and response. *Agronomy* 12, 702. doi: 10.3390/agronomy12030702

Baier, M., Bittner, A., Prescher, A., and van Buer, J. (2019). Preparing plants for improved cold tolerance by priming. *Plant Cell Environ.* 42, 782–800. doi: 10.1111/pce.13394

Busch, A., and Hippler, M. (2011). The structure and function of eukaryotic photosystem I. *Biochim. Biophys. Acta (BBA) - Bioenerg.* 1807, 864–877. doi: 10.1016/j.bbabio.2010.09.009

Chen, M., and Thelen, J. J. (2013). ACYL-LIPID DESATURASE2 is required for chilling and freezing tolerance in arabidopsis. *Plant Cell* 25, 1430–1444. doi: 10.1105/tpc.113.111179

Chen, L., Zhao, Y., Xu, S., Zhang, Z., Xu, Y., Zhang, J., et al. (2018). OsMADS57 together with OsTB1 coordinates transcription of its target OsWRKY94 and D14 to switch its organogenesis to defense for cold adaptation in rice. *New Phytol.* 218, 219–231. doi: 10.1111/nph.14977

Danquah, A., de Zelicourt, A., Colcombet, J., and Hirt, H. (2014). The role of ABA and MAPK signaling pathways in plant abiotic stress responses. *Biotechnol. Adv.* 32, 40–52. doi: 10.1016/j.biotechadv.2013.09.006

Ermilova, E. (2020) *Cold stress response: An overview in chlamydomonas. frontiers in plant science* (Accessed November 7, 2022).

Fang, Z.-T., Jin, J., Ye, Y., He, W.-Z., Shu, Z.-F., Shao, J.-N., et al. (2022). Effects of different shading treatments on the biomass and transcriptome profiles of tea leaves (Camellia sinensis l.) and the regulatory effect on phytohormone biosynthesis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.909765

Finkelstein, R. (2013). Abscisic acid synthesis and response. *Arabidopsis Book* 11, e0166. doi: 10.1199/tab.0166

Fowler, S., and Thomashow, M. F. (2002). Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* 14, 1675–1690. doi: 10.1105/tpc.003483

Fu, Z., Yue, J., LiPing, B., Liang, Z., LiJing, C., HaoGe, L., et al. (2011). The ICE-CBF-COR pathway in cold acclimation and AFPs in plants. *Middle East J. Sci. Res.* 8, 493–498.

Gális, I., Gaquerel, E., Pandey, S. P., and Baldwin, I. T. (2009). Molecular mechanisms underlying plant memory in JA-mediated defence responses. *Plant Cell Environ.* 32, 617–627. doi: 10.1111/j.1365-3040.2008.01862.x

Gommers, C. M. M., Keuskamp, D. H., Buti, S., van Veen, H., Koevoets, I. T., Reinen, E., et al. (2017). Molecular profiles of contrasting shade response strategies in wild plants: Differential control of immunity and shoot elongation. *Plant Cell* 29, 331–344. doi: 10.1105/tpc.16.00790

Gong, W. Z., Jiang, C. D., Wu, Y. S., Chen, H. H., Liu, W. Y., and Yang, W. Y. (2015). Tolerance vs. avoidance: Two strategies of soybean (Glycine max) seedlings in response to shade in intercropping. *Photosynthetica* 53, 259–268. doi: 10.1007/s11099-015-0103-8

Hao, X., Tang, H., Wang, B., Yue, C., Wang, L., Zeng, J., et al. (2018a). Integrative transcriptional and metabolic analyses provide insights into cold spell response mechanisms in young shoots of the tea plant. *Tree Physiol.* 38, 1655–1671. doi: 10.1093/treephys/tpy038

Hao, X., Wang, B., Wang, L., Zeng, J., Yang, Y., and Wang, X. (2018b). Comprehensive transcriptome analysis reveals common and specific genes and pathways involved in cold acclimation and cold stress in tea plant leaves. *Sci. Hortic.* 240, 354–368. doi: 10.1016/j.scienta.2018.06.008

Hao, L., Wen, Y., Zhao, Y., Lu, W., and Xiao, K. (2015). Wheat mitogen-activated protein kinase gene TaMPK4 improves plant tolerance to multiple stresses through modifying root growth, ROS metabolism, and nutrient acquisitions. *Plant Cell Rep.* 34, 2081–2097. doi: 10.1007/s00299-015-1853-2

Ichino, T., Fuji, K., Ueda, H., Takahashi, H., Koumoto, Y., Takagi, J., et al. (2014). GFS9/TT9 contributes to intracellular membrane trafficking and flavonoid accumulation in arabidopsis thaliana. *Plant J.* 80, 410–423. doi: 10.1111/tpj.12637

Jeon, J., Kim, N. Y., Kim, S., Kang, N. Y., Novák, O., Ku, S.-J., et al. (2010). A subset of cytokinin two-component signaling system plays a role in cold temperature stress response in arabidopsis*. *J. Biol. Chem.* 285, 23371–23386. doi: 10.1074/jbc.M109.096644

Jiang, X., Zhao, H., Guo, F., Shi, X., Ye, C., Yang, P., et al. (2020). Transcriptomic analysis reveals mechanism of light-sensitive albinism in tea plant camellia sinensis 'Huangjinju.'. *BMC Plant Biol.* 20, 216. doi: 10.1186/s12870-020-02425-0

Ji, H.-G., Lee, Y.-R., Lee, M.-S., Hwang, K. H., Park, C. Y., Kim, E.-H., et al. (2018). Diverse metabolite variations in tea (Camellia sinensis l.) leaves grown under various shade conditions revisited: A metabolomics study. *J. Agric. Food Chem.* 66, 1889–1897. doi: 10.1021/acs.jafc.7b04768

Jiroutova, P., Oklestkova, J., and Strnad, M. (2018). Crosstalk between brassinosteroids and ethylene during plant growth and under abiotic stress conditions. *Int. J. Mol. Sci.* 19, 3283. doi: 10.3390/ijms19103283

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Ku, K. M., Choi, J. N., Kim, J., Kim, J. K., Yoo, L. G., Lee, S. J., et al. (2010). Metabolomics analysis reveals the compositional differences of shade grown tea (Camellia sinensis l.). *J. Agric. Food Chem.* 58, 418–426. doi: 10.1021/jf902929h

Kumari, P., Panwar, S., Singh, K., and Raju, D. V. S. (2014). Artificial lighting systems for ornamental plants - a review.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: An efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Li, Y., Jeyaraj, A., Yu, H., Wang, Y., Ma, Q., Chen, X., et al. (2020). Metabolic regulation profiling of carbon and nitrogen in tea plants [Camellia sinensis (L.) o. kuntze] in response to shading. *J. Agric. Food Chem.* 68, 961–974. doi: 10.1021/acs.jafc.9b05858

Li, Y., Wang, X., Ban, Q., Zhu, X., Jiang, C., Wei, C., et al. (2019). Comparative transcriptomic analysis reveals gene expression associated with cold adaptation in the tea plant camellia sinensis. *BMC Genomics* 20, 1–17. doi: 10.1186/s12864-019-5988-3

Lorrain, S., Allen, T., Duek, P. D., Whitelam, G. C., and Fankhauser, C. (2008). Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors. *Plant J.* 53, 312–323. doi: 10.1111/j.1365-313X.2007.03341.x

Luo, J., Zhou, J.-J., and Zhang, J.-Z. (2018). Aux/IAA gene family in plants: Molecular structure, regulation, and function. *Int. J. Mol. Sci.* 19, 259. doi: 10.3390/ijms19010259

Mamedov, M., Govindjee,, Nadtochenko, V., and Semenov, A. (2015). Primary electron transfer processes in photosynthetic reaction centers from oxygenic organisms. *Photosynth. Res.* 125, 51–63. doi: 10.1007/s11120-015-0088-y

Mohotti, A. J. (2004). Shade in tea. is it beneficial?

Müller, M., and Munné-Bosch, S. (2015). Ethylene response factors: A key regulatory hub in hormone and stress signaling. *Plant Physiol.* 169, 32–41. doi: 10.1104/pp.15.00677

Murphy, A. (2015). Hormone crosstalk in plants. *J. Exp. Bot.* 66, 4853–4854. doi: 10.1093/jxb/erv339

Pavlů, J., Novák, J., Koukalová, V., Luklová, M., Brzobohatý, B., and Černý, M. (2018). Cytokinin at the crossroads of abiotic stress signalling pathways. *Int. J. Mol. Sci.* 19, 2450. doi: 10.3390/ijms19082450

Rezai, S., Etemadi, N., Nikbakht, A., Yousefi, M., and Majidi, M. M. (2018). Effect of light intensity on leaf morphology, photosynthetic capacity, and chlorophyll content inSage (Salvia officinalis l.).

Ritonga, F. N., and Chen, S. (2020). Physiological and molecular mechanism involved in cold stress tolerance in plants. *Plants* 9, 560. doi: 10.3390/plants9050560

Sakai, A., and Larcher, W. (2012). *Frost survival of plants: Responses and adaptation to freezing stress* (Springer Science & Business Media).

Sharif, R., Su, L., Chen, X., and Qi, X. (2022a). Involvement of auxin in growth and stress response of cucumber. *Vegetable Res.* 2, 9. doi: 10.48130/VR-2022-0013

Sharif, R., Su, L., Chen, X., and Qi, X. (2022b). Hormonal interactions underlying parthenocarpicfruit formation in horticultural crops. *Hortic. Res.* 9, 17. doi: 10.1093/hr/uhab024

Shi, L.-X., Hall, M., Funk, C., and Schröder, W. P. (2012a). Photosystem II, a growing complex: Updates on newly discovered components and low molecular mass proteins. *Biochim. Biophys. Acta (BBA) - Bioenerg.* 1817, 13–25. doi: 10.1016/j.bbabio.2011.08.008

Shi, Y., Tian, S., Hou, L., Huang, X., Zhang, X., Guo, H., et al. (2012b). Ethylene signaling negatively regulates freezing tolerance by repressing expression of CBF and type-a ARR genes in arabidopsis. *Plant Cell* 24, 2578–2595. doi: 10.1105/tpc.112.098640

Silveira, J. A. G., and Carvalho, F. E. L. (2016). Proteomics, photosynthesis, and salt resistance in crops: An integrative view. *J. Proteomics* 143, 24–35. doi: 10.1016/j.jprot.2016.03.013

Teige, M., Scheikl, E., Eulgem, T., Dóczi, R., Ichimura, K., Shinozaki, K., et al. (2004). The MKK2 pathway mediates cold and salt stress signaling in arabidopsis. *Mol. Cell* 15, 141–152. doi: 10.1016/j.molcel.2004.06.023

Thakur, M., Bhatt, V., and Kumar, R. (2019). Effect of shade level and mulch type on growth, yield and essential oil composition of damask rose (Rosa damascena mill.) under mid hill conditions of Western Himalayas. *PloS One* 14, e0214672. doi: 10.1371/journal.pone.0214672

Thomashow, M. F. (1999). PLANT COLD ACCLIMATION: Freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 571–599. doi: 10.1146/annurev.arplant.50.1.571

Thomashow, M. F. (2010). Molecular basis of plant cold acclimation: Insights gained from studying the CBF cold response pathway. *Plant Physiol.* 154, 571–577. doi: 10.1104/pp.110.161794

Wang, K., Bai, Z., Liang, Q., Liu, Q., Zhang, L., Pan, Y., et al. (2018). Transcriptome analysis of chrysanthemum (Dendranthema grandiflorum) in response to low temperature stress. *BMC Genomics* 19, 319. doi: 10.1186/s12864-018-4706-x

Wu, W. (2022). Comparative transcriptome analysis revealed the freezing tolerance signaling events in winter rapeseed (Brassica rapa l.). doi: 10.37247/PAGEN2ED.2.22.16

Wu, M., Li, Z., and Wang, J. (2020). Transcriptional analyses reveal the molecular mechanism governing shade tolerance in the invasive plant solidago canadensis. *Ecol. Evol.* 10, 4391–4406. doi: 10.1002/ece3.6206

Zaman, S., Shen, J., Wang, S., Wang, Y., Ding, Z., Song, D., et al. (2022). Effects of shading nets on reactive oxygen species accumulation, photosynthetic changes, and associated physiochemical attributes in promoting cold-induced damage in camellia sinensis (L.) kuntze. *Horticulturae* 8, 637. doi: 10.3390/horticulturae8070637

Zhang, Q., Bi, G., Li, T., Wang, Q., Xing, Z., LeCompte, J., et al. (2022). Color shade nets affect plant growth and seasonal leaf quality of camellia sinensis grown in Mississippi, the united states. *Front. Nutr.* 4. doi: 10.3389/fnut.2022.786421

Zhang, Q., Li, T., Wang, Q., LeCompte, J., Harkess, R. L., and Bi, G. (2020) *Screening tea cultivars for novel climates: Plant growth and leaf quality of camellia sinensis cultivars grown in Mississippi, united states. frontiers in plant science* (Accessed November 7, 2022).

Zhang, P., Li, S., Zhao, P., Guo, Z., and Lu, S. (2019). Comparative physiological analysis reveals the role of NR-derived nitric oxide in the cold tolerance of forage legumes. *Int. J. Mol. Sci.* 20, 1368. doi: 10.3390/ijms20061368

Zhang, Y., Zhu, X., Chen, X., Song, C., Zou, Z., Wang, Y., et al. (2014). Identification and characterization of cold-responsive microRNAs in tea plant (Camellia sinensis) and their targets using high-throughput sequencing and degradome analysis. *BMC Plant Biol.* 14, 271. doi: 10.1186/s12870-014-0271-x

Zhao, C., Wang, P., Si, T., Hsu, C.-C., Wang, L., Zayed, O., et al. (2017). MAP kinase cascades regulate the cold response by modulating ICE1 protein stability. *Dev. Cell* 43, 618–629. doi: 10.1016/j.devcel.2017.09.024

Zuther, E., Lee, Y. P., Erban, A., Kopka, J., and Hincha, D. K. (2018). "Natural variation in freezing tolerance and cold acclimation response in arabidopsis thaliana and related species," in *Survival strategies in extreme cold and desiccation: Adaptation mechanisms and their applications advances in experimental medicine and biology*, 81–98. doi: 10.1007/978-981-13-1244-1_5

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 February 2023
DOI 10.3389/fpls.2023.1141295

Check for updates

OPEN ACCESS

EDITED BY
Hui Song,
Qingdao Agricultural University, China

REVIEWED BY
Jihui Chen,
Guizhou University, China
Kai Luo,
Hainan University, China

*CORRESPONDENCE
Yating Li
yatingli@ldu.edu.cn
Yanling Yin
894337103@qq.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 10 January 2023
ACCEPTED 31 January 2023
PUBLISHED 15 February 2023

CITATION
Fan S, Amombo E, Avoga S, Li Y and Yin Y (2023) Salt-responsive bermudagrass microRNAs and insights into light reaction photosynthetic performance. *Front. Plant Sci.* 14:1141295. doi: 10.3389/fpls.2023.1141295

COPYRIGHT
© 2023 Fan, Amombo, Avoga, Li and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Salt-responsive bermudagrass microRNAs and insights into light reaction photosynthetic performance

Shugao Fan[1†], Erick Amombo[2†], Sheila Avoga[3], Yating Li[1*] and Yanling Yin[1*]

[1]School of Resources and Environmental Engineering, Ludong University, Yantai, China, [2]African Sustainable Agriculture Institute, Mohammed VI Polytechnic University, Laayoune, Morocco, [3]Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Science, Wuhan, China

Introduction: Bermudagrass (*Cynodon dactylon* L.) is a warm-season grass with high drought and salt tolerance. However, its cultivation as a silage crop is limited by its lower forage value when compared to other C4 crops. Because of its high genetic variability in abiotic stress tolerance, bermudagrass-mediated genetic breeding offers significant promise for introducing alternative fodder crops in saline and drought-affected regions, and improved photosynthetic capacity is one way for increasing forage yield.

Methods: Here, we used RNA sequencing to profile miRNAs in two bermudagrass genotypes with contrasting salt tolerance growing under saline conditions.

Results: Putatively, 536 miRNA variants were salt-inducible, with the majority being downregulated in salt-tolerant vs sensitive varieties. Also, seven miRNAs putatively targeted 6 genes which were significantly annotated to light reaction photosynthesis. Among the microRNAs, highly abundant miRNA171f in the salt tolerant regime targeted Pentatricopeptide repeat-containing protein and dehydrogenase family 3 member F1 both annotated to electron transport and Light harvesting protein complex 1 genes annotated to light photosynthetic reaction in salt tolerant regime vs salt sensitive counterparts. To facilitate genetic breeding for photosynthetic capacity, we overexpressed miR171f in *Medicago tracantula* which resulted in a substantial increase in the chlorophyll transient curve, electron transport rate, quantum yield of photosystem II non photochemical quenching, NADPH and biomass accumulation under saline conditions while its targets were downregulated. At ambient light level the electron transport was negatively correlated with all parameters while the NADPH was positively associated higher dry matter in mutants.

Discussion: These results demonstrate that miR171f improves photosynthetic performance and dry matter accumulation via transcriptional repression of genes in the electron transport pathway under saline conditions and thus a target for breeding.

KEYWORDS

electron transport, microRNA, salt stress, bermudagrass, target gene

1 Introduction

Soil and water salinity are serious abiotic factors globally (Cheeseman, 1988; Allakhverdiev et al., 2000). Other than natural causes, soil salinity has been aggravated by excessive use of saline water for irrigation (Shrivastava and Kumar, 2015). As a result, soil structure and water-holding capacity have deteriorated dramatically over the years (Yadav et al., 2011). Soil salinity is mostly constituted by NaCl due to its remarkable solubility in water (Tavakkoli et al., 2010). Excessive buildup of Na^+ and Cl^- ions are phytotoxic and imposes physiological drought (Hasegawa et al., 2000; Munns, 2002). Therefore, plants growing in these soils must adapt to survive. This poses a serious threat to plant growth and crop productivity, especially in arid and semi-arid regions where drought is already a major challenge (Alori et al., 2020). Understanding soil salinity *in toto* is a complex process. This is attributed to the soil's highly heterogeneous nature of salt occurrence making it hard to develop a regional level policy (Hassani et al., 2021). Concentrating on the various bioprocesses that plants demonstrate to survive salt stress at genotypic and phenotypic levels is therefore more viable.

Genetic breeding and cultivating high-yielding forages in saline areas are ideal for addressing food insecurity through livestock production (Fita et al., 2015). The photosynthetic pathway is one of the most reliable indicators of plant performance and productivity and thus a target for breeding (Ashraf and Harris, 2013). For example, photosynthesis is the primary driver of biomass accumulation, and the efficiency with which a crop transforms CO_2 into metabolites is a crucial determinant of eventual crop yield (Simkin et al., 2019). Also, other than being reliable abiotic stress tolerance level indicators (Guidi et al., 2019), the slow induction of chlorophyll *a* fluorescence kinetics variable such as the maximum quantum yield of photosystem II, the electron transport rate, nonphotochemical quenching, and the photosynthetic performance index have been used as targets for forward and reverse genetic breeding for efficient yield gain (Sherstneva et al., 2022). Thus, increased salt tolerance and forage yields can be achieved by maintaining a balance of light use efficiency and photosynthetic carbon uptake and assimilation efficiency (Busch et al., 2018).

Bermudagrass (*Cynodon dactylon* L.) is a warm-season grass with excellent saline-land potential (Chen et al., 2019). Despite the incredible potential of photosynthetic efficiency of light conversion to harvestable biomass in bermudagrass, it has yet to be extensively examined due to a scarcity of genetic information for candidate gene discovery. As a result, bermudagrass remains to be a potentially rich genetic resource for breeding alternative forages for salt tolerance.

Small RNAs, which influence the transcription of their targets *via* RNA-RNA interactions, are one of the most remarkable regulatory molecules (Helge and Witold, 2008), with microRNAs being one of the best-understood groups. MicroRNAs influence gene expression *via* transcriptional repression (Voinnet, 2009). Llave et al. (2002); Park et al. (2002) and Reinhart et al. (2002) rigorously described miRNA and their targets in Arabidopsis which opened doors for subsequent studies. Many miRNAs have since been implicated with salt tolerance in various grass family members, including switchgrass (*Panicum virgatum*) (Xie et al., 2013) and wild emmer wheat (*Triticum turgidum*) (Feng et al., 2017).

While miRNAs' potential roles in photosynthesis have been elucidated in Arabidopsis, tobacco, and rice (Zhang et al., 2017; Pan et al., 2018), studies that comprehensively focus on photosynthetic regulation by miRNAs under salt stress are scarce. Closing this gap will be critical in availing a genomic platform on the photosynthetic response of bermudagrass to salt stress and identifying potential targets for breeding for photosynthetic performance under saline conditions. Here, we did comparative research in two bermudagrass cultivars with differing salt tolerance levels to investigate the regulatory network of miRNA, their targets in photosynthesis, and their potential function in plant breeding.

2 Materials and methods

Here, plant resources comprised two bermudagrass varieties, cultivar 43(C43, salt tolerant) and cultivar 198 (C198, salt sensitive), collected from the germplasm center of the Wuhan Botanical Garden, Chinese Academy of Science. We screened for salt tolerance in our previous study (Hu et al., 2015). Freshly harvested stolons with two internodes were planted in plastic pots containing soil mixed with sand (ratio of 2:1, v/v) in the greenhouse. After 14 days of growth and maturation, the plants were transplanted into plastic pots (7 cm in diameter and 9 cm in depth) filled with coarse silica aggregate as the plant anchor medium. The pots were hung in rectangular containers filled with half-strength Hoagland's solution, replenished three times a week at one-day intervals, and refreshed weekly. For 14 hours, plants were cultivated in a greenhouse with a temperature regime of 25°C at night and 28°C during the day, with photosynthetically active radiation levels of 800 mol m^{-2} s^{-1}. After 21 days, the half-strength Hoagland's solution was added with 200 mM NaCl, and after 7 days of salt treatment, 0.5 g of fresh leaves were excised and instantly frozen in liquid nitrogen. Samples were collected from multiple pots from each treatment for RNA sequencing (three repetitions), quantitative real-time PCR (real-time qPCR) validation (three repeats), and physiological measurements (four biological replicates for each physiological parameter) all at the same time.

2.1 RNA extraction, sequencing, and small RNA library construction and data processing

Total RNA was extracted using Trizol reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. The amount and purity of RNA were determined using the Bioanalyzer 2100 and the RNA 6000 Nano LabChip Kit (Agilent, CA, USA). The experiment was carried out in line with Illumina Incorporation's standard methods, which included library preparation and sequencing studies. The TruSeq Small RNA Sample Prep Kits were used to construct small RNA sequencing libraries (Illumina, San Diego, USA). The characteristics of the miRNA were applied after the total RNA was extracted from the sample. RNA ligase enzyme was used to attach single-stranded DNA 3' and 5' sticky ends to short RNAs in sequence. The short RNA sequence with the connected ends was reverse-transcribed using primers

complementary to the three ends, and the resulting cDNA sequence was PCR-amplified. Finally, the PCR product of 140-160 base pairs (bp) in length was recovered using a 6% polyacrylamide Tris-borate-EDTA gel to conclude the library preparation. Illumina Hiseq2000/2500 was used to sequence the library that was created. The data quality of the sequence was assessed using Illumina FastQC software (Ichikawa Biosystems ACGT101-miR LC Sciences, Houston, Texas, USA), which was created separately by the firm. Sequencing junctions, unusual miRNA sequences, and incomplete sequences were removed from the analysis. Following that, base lengths of 18 and >25 nucleotides (nt) were screened, as well as miRNA-free sequences such as mRNA, RFam (containing rRNA, tRNA, snRNA, snoRNA, and so on), and Repbase (www.girinst.org).

2.2 Detection of known and novel miRNAs

Following the cleanup and removal of the remaining sequences, Bowtie2 (Langmead and Salzberg, 2012) was used to align the precursors of specific species in the miRBase database. Variations in length at the 5' and 3' ends and a mismatch within the sequence were permitted in the alignment analysis. The miRNAs that were matched to the mature sequence component of the given species were categorized as known miRNAs. If the identified sequence could not be linked to the corresponding site of a known miRNA, it was designated a novel 5p or 3p miRNA candidate sequence. The new mismatched sequences were again processed through a Bowtie alignment to precursors of other chosen species in miRBase, and the aligned miRNA precursor sequences were recognized as unique. The RNA secondary structures were predicted using RNAfold soft (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

2.3 Analysis of differentially expressed miRNAs and prediction of their targets

The expression levels of differently expressed miRNAs in the four regimes were examined using the Fisher exact test. The fold changes of miRNA readings were calculated as the ratio of salt treatment over corresponding control for each cultivar, as well as salt-tolerant C43 salt vs salt-sensitive C198 salt. The fold change and P-value were used to determine the significance of the expression (significant *absolute fold change 1 and P-value 0.05; highly significant **, absolute fold change 1 and P-value 0.01). To predict DEM target genes, we utilized the website service psRNATarget (http://plantgrn.noble.org/psRNATarget/?function=3) in conjunction with our prior bermudagrass EST data (Hu et al., 2015). The gene ontology (GO) keywords and KEGG Pathway of the miRNA targets were also annotated to diverse functions.

2.4 Expression of photosynthesis-related miRNAs and their targets by real-time qPCR

To identify and validate the differentially expressed miRNAs in the photosynthesis pathway, six miRNAs with high expression levels (at least 1000) in all libraries were selected from conserved families (miR171f, miR319, MiR156a, and miR159), and non-conserved (MIR305714, MIR46018, and MIR2836) and amplified using real-time qPCR to examine their expression. The forward miRNAs primers were designed based on the full miRNAs sequence, while the reverse primer was the universal reverse primer (5'GTGCAGGGTCCGAGGT3'). The stem-loop primer, used for miRNA cDNA synthesis, was designed according to Chen et al. (2005). The 18S rRNA was taken as the reference gene. Similarly, the expression profiles of the respective miRNA target genes were assayed by real-time qPCR. In summary, a 10-μg aliquot of RNA was used for the RT reaction followed by the addition of oligo(dT) primer. The RT reaction was performed by MMLV reverse transcriptase (Toyobo, Osaka, Japan) according to the supplier's manual. For real-time qPCR, *ACTIN* for bermudagrass was used as a housekeeping gene as the internal control. To calculate the relative expression, the $2^{-\Delta\Delta Ct}$ method was used and Student's *t*-test was performed to compare differences in expression profile. The means were considered significantly different when $P \leq 0.05$.

2.5 Heterologous overexpression of miR171f

After confirming that miR171f responds to salt stress, we generated transgenic *Medicago* constitutively expressing miR171f. The constitutive expression construct of miR171f was introduced into wild-type (WT) plants *via Agrobacterium tumefaciens*-mediated transformation. The selectable marker gene, Hyg conferring hygromycin resistance was amplified from the genomic DNA of regenerated plants and controls. The expression levels of pre-miR171f and mature miR171f were then compared between the wildtype and transgenic plants to determine whether the bermudagrass pre-miR171f was successfully integrated into the genome of *Medicago*, transcribed, and properly processed.

2.6 Chlorophyll a fluorescence and NADPH activity

After dark adaptation, the maximum chlorophyll fluorescence (F_M) was determined every 30s by saturation pulse (800 ms, 2,700 μmol quanta $m^{-2}{\cdot}s^{-1}$). Light intensity-dependent parameters of $_{\Phi}$PSII, ETR, and NPQ, were measured with a range of intensities from 0, 200, 400, 600, 800, to 1000 $\mu mol{\cdot}m^{-2}{\cdot}s^{-1}$ and calculated. Measurement of photosynthetic light response curves was performed on a single fully expanded leaf exposed to a light source. All the measurements were conducted using the Li-6400 portable photosynthesis system. The NADPH activity was determined by the reduction of the tetrazolium salt XTT by O^{2-} following Kaundal et al. (2012) protocol. In the presence of O^{2-}, XTT generated a soluble yellow formazan that can be quantified spectrophotometrically.

3 Results

3.1 Identification of known and novel miRNAs

In the C43 CK, C43 salt, C198 CK, and C198 salt libraries, there were 10,945,629, 9,755,632, 12,520,523, and 19,469,714 raw reads as

well as 2,775,967, 3,047,166, 2,804,775, and 3,647,171 unique reads. After excluding non-coding RNA groups such as rRNAs, tRNAs, snoRNAs, snRNAs, and other contaminants, as well as those with a length of 18nt and >25 nt, a total of 4,897,423, 4,464,136, 4,670,166, and 4,852,132 valid reads, as well as 1,833,027, 1,855,458, 1,320,178, and 1,556,539 unique reads were obtained (Table 1).

After length filtering, most of the length data lay between 20 and 24 nt. C198 ck was dominated by 21 nt (19.92%) in salt-sensitive libraries, whereas C198 salt was dominated by 22 nt (22.29%). C43 ck was dominated by 24 nt (27.66%), whereas C43 salt was dominated by 22 nt (20.30%). The 18 nt had the least miRNAs in all the libraries (Figure 1; Table S1).

After matching the identified miRNAs with miRbase, we classified them into three groups: gp1, gp2, and gp3. Gp1 was made up of 20 mature miRNAs that were mapped to specific species' miRNAs/pre-miRNAs, and the pre-miRNAs were then mapped to our bermudagrass EST. Gp2 consisted of 61 mature reads that were mapped to miRNAs/pre-miRNAs of select species in miRbase but were not further mapped to bermudagrass EST. As a result, the known miRNAs were established by these two groups (gp1 and gp2). Gp3 had 59 mature miRNAs that were not linked to specific species miRNAs in miRbase, however, the readings were mapped to our bermudagrass EST (Figure 2; Table S2).

As a result, we considered these to be novel sequences. The sequences were thus labeled as p3/p5 to distinguish them from the previously published sequences. The sequences were further probed to match Hofacker's prediction requirements for novel miRNAs (Hofacker, 2003). As a result, 59 had very low abundance with dissociation energy ranging from -32.3 to -142.30 kcal mol^{-1} which is consistent with the degradation principle of miRNA during the early generation process. Their GC content ranged from 30-81%, indicating that the novel miRNAs were stable. The minimal free energy index (MFE) ranged from 0.8 to 1.5 which is higher than that of other small noncoding RNAs reported by Zhang et al. (2018). In addition, uracil (U) and adenine (A) were the dominant nucleotides in the first positions of the miRNAs (Figures 3A–C; Table S3) which is not only a characteristic of mature miRNA but also plays a vital role in the miRNA-target interaction.

TABLE 1 Small RNA reads from sequencing of four bermudagrass libraries.

	C198_ck		C198_salt		C43_ck		C43_salt	
Type*	Total	unique	Total	unique	Total	unique	Total	unique
Raw reads	582,201	118,587	349311	80,185	208,395	45,208	186,778	55,831
3ADT&length filter	637,930	138,105	427479	102,454	286,206	68,519	209,685	70,560
Junk reads	681,016	155,326	532807	136,668	471,632	108,808	285,641	105,371
Rfam	1,120,506	262,989	1116018	321,331	1,091,790	338,433	973,513	360,002
Repeats	656,616	230,351	935309	356,690	1,031,775	420,110	1,329,076	514,502
valid reads	421,789	144,721	561401	187,692	609,651	235,709	541,705	267,458
rRNA	407,452	217,025	589133	269,087	860,842	506,969	624,339	376,593
tRNA	162,656	53,074	340674	102,432	337,132	109,271	313,399	105,141
miRNA	4,670,166	1,320,178	4852132	1,556,539	4,897,423	1,833,027	4,464,136	1,855,458

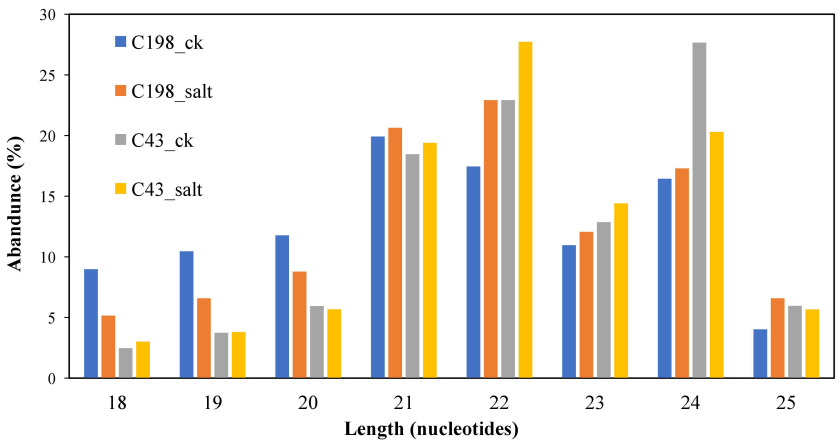


FIGURE 1
Length distribution of unique sequences after cleaning.

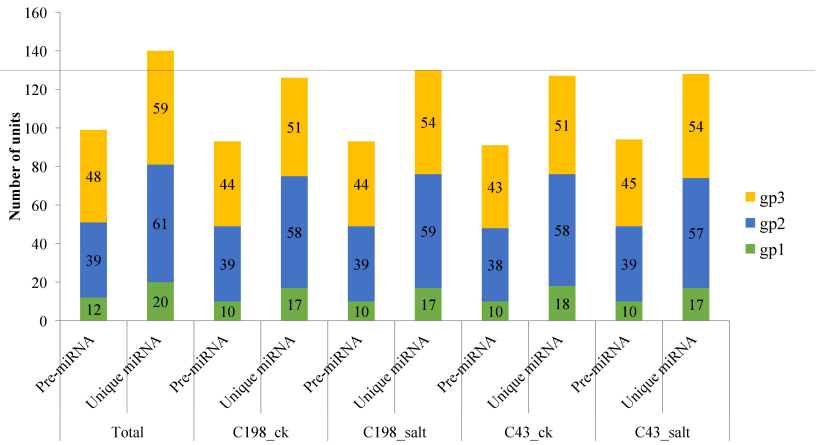


FIGURE 2
Plot showing the characterization of predicted vs mature microRNAs (a) group 1 (b) group 2 and (c) group 3. Group 1 represents sequences that were mapped to miRNAs/pre-miRNAs of specific species in miRbase as well as bermudagrass EST; group 2 represents sequences that were mapped to miRNAs/pre-miRNAs of selected species in miRbase, and the mapped pre-miRNAs were not further mapped to genome/EST. group 3 represented the reads that were not mapped to pre-miRNAs of selected species in miRbase but were rather mapped to genome/EST.

3.2 Conservation profile of detected miRNA

All the known miRNAs were shown as variants, which were defined as 'isomiRs. For example, R-n means that the detected miRNA sequence was n base/s less than the known representative miRNA sequence in miRbase (rep_miRSeq) in the right side; L+n means the detected miRNA sequence is n base more than known rep_miRSeq in the left side; R+n means the detected miRNA sequence is n base more than known rep_miRSeq on the right side; 2ss5TC13TA means 2 substitutions (ss), which are T≥C at position 5 and T≥A at position 13. If there was no matching annotation, the detected miRNA sequence is the same as known rep_miRSeq. Among the known group, miRNAs belonged to 16 conserved families while the rest belonged to 13 non-conserved families (Figure 4; Table S4).

3.3 Identification of differentially expressed miRNAs

To identify differentially expressed miRNAs (DEM) under salinity stress, the analysis of differential expression patterns was

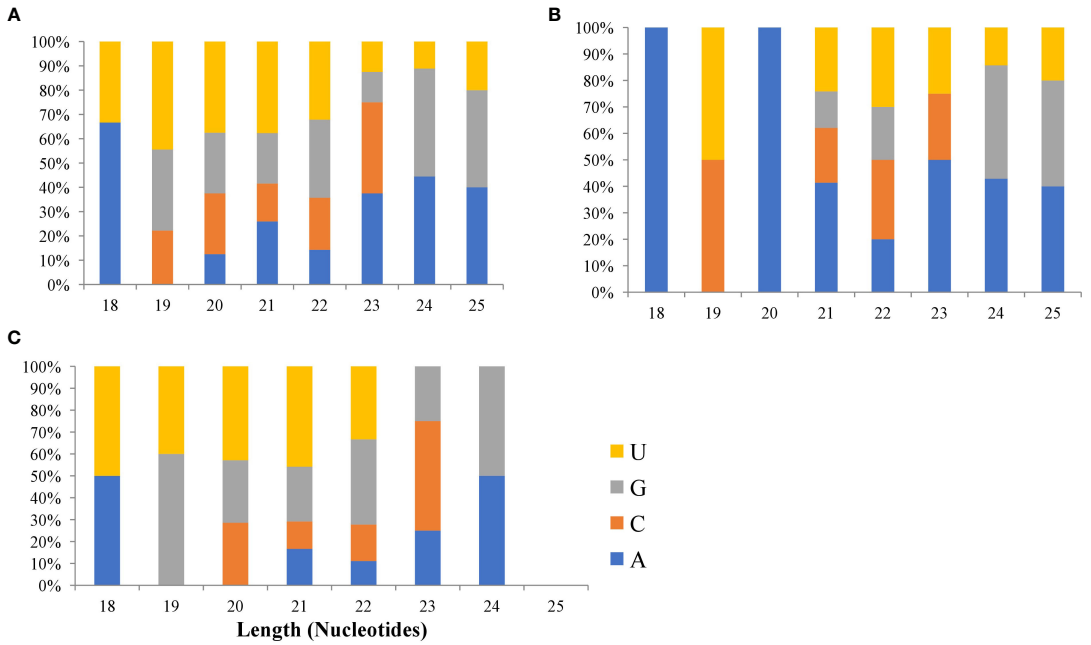


FIGURE 3
Plots of first nucleotide bias of **(A)** group 3 **(B)** group 2 and **(C)** group 1 microRNAs. The upper plot represents the first nucleotide. Gp1 represents sequences that were mapped to miRNAs/pre-miRNAs of specific species in miRbase as well as bermudagrass EST; group 2 represents sequences that were mapped to miRNAs/pre-miRNAs of selected species in miRbase, and the mapped pre-miRNAs were not further mapped to genome/EST. group 3 represented the reads that were not mapped to pre-miRNAs of selected species in miRbase but were rather mapped to genome/EST.

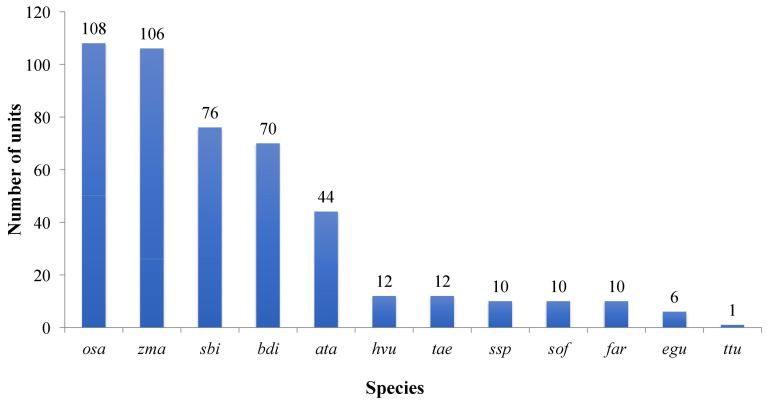


FIGURE 4
Plot showing the conservation profile of identified miRNAs. Ata- *Aegilops tauschii*, bdi- *Brachypodium distachyon*, egu- *Elaeis guineensis*, far- *Fragaria ananassa*, Hvu- *Hordeum vulgare*, osa-*Oryza sativa*, sbi- *Sorghum bicolor*, sof- *Saccharum officinarum*, Tae-*Triticum aestivum*- ssp- *Saccharum sp*, zma-*Zea mays*, ttu-*Triticum turgidum*.

performed by statistical comparison between four libraries. In each library, the expression level of mature miRNAs was normalized to Transcripts Per Kilobase Million (TPM) and compared between salt treatment vs control as well as salt-tolerant vs salt-sensitive cultivars. Totally, 536 miRNAs were expressed in the four libraries. Among them, 111 miRNAs were shared among all four libraries. In addition, in the C43 vs C198 salt regime, the total DEM belonged to 12 conserved families i.e., miR160, miR164, miR166, miR167, miR156, osa-miR171, osa-miR172, miR390, miR396, miR162, miR444 and miR393, 6 non-conserved families i.e., miR812, miR6255, miR818, miR2118, miR9482, and miR5384, while 57 were novel and most of the DEM were downregulated (72 downregulated and 63 upregulated) (Figure 5A; Table S5). Among the novel miRNAs, the expression abundance was very low, for instance, only 11 miRNAs had an expression abundance of more than 10 reads. Out of these relatively abundant novel miRNAs, 6 were downregulated while five were upregulated under salt stress. Cluster analysis revealed a total of 7 miRNA clusters in the bermudagrass. Cluster 2 constituted the largest number of miRNA variants i.e., MIR444c, miR444a, MIR444d, miR444c, and MIR444c. Some pre-miRNAs in the same cluster had similar expression patterns with their mature counterparts, for example in cluster 6 MIR159d was not only located in the same genomic locus with miR159c but were all downregulated in both the control and salt regimes of C43 vs C198 (Table S6). A total of 111 microRNAs were shared among all the four libraries. Exclusive microRNAs were 1, 2, 2, and 0 in ck198, ck43, salt 198 and salt 43 respectively (Figure 5B).

3.4 miRNA targeting salt-tolerant and photosynthesis-related genes

To understand the potential transcriptional roles of the DEM, it was crucial to focus on their targets in salt tolerant vs sensitive regimes. Putative 1891 genes were targeted by 139 mature miRNAs. The number of targets per miRNA ranged from 1 to 287. Notably, microRNAs targeting most genes are novel with very low abundance. Also, many miRNAs targeted a single gene (Table S7). To understand the potential function of the miRNA targets, we integrated their gene functions in the GO database with previously published literature. A total of 584 target genes were annotated to various GO terms which were classified into three major categories i.e., 'molecular function', 'cellular component', and 'biological process' (Figure 6A; Table S8).

The KEGG analysis revealed the top 20 enriched pathways which were identified with 55 genes "plant hormone signal transduction" was the most significantly enriched term with respect to the richness factor and gene number (Figure 6B; Table S9). This suggested that most of the targets in this category might play an important role in

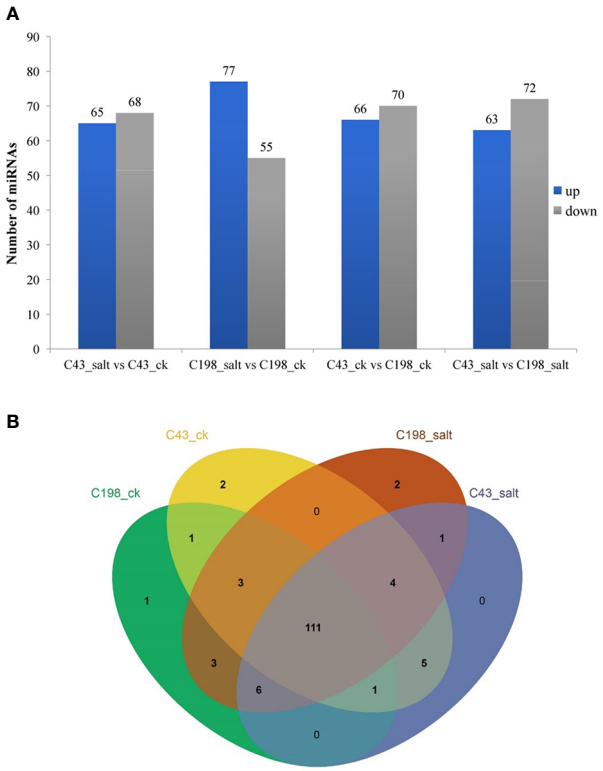


FIGURE 5
(A) Bar plot of differentially expressed miRNAs and their expression level. **(B)** Venn diagram showing miRNA sharing in four bermudagrass libraries with contrasting salt tolerance levels.

hormonal salt stress signal perception and transduction by roots of bermudagrass.

Among them key salt tolerance pathways are significantly enriched such as the transcription factor complex. Under this category transcription factors from major salt-tolerant families such as NAC, WRKY, ERF, bHLH, MADs, and SOS, are enriched. Component 2 which accounted for 31% of the variation composed of members salt tolerant genes including WRKY22 (Li et al., 2018), NAC021 (Yang et al., 2020), CAT7 (Karam et al., 2016) and two photosynthesis genes *light harvesting complex 1* (Yang et al., 2019) and *aldehyde dehydrogenase 1* (Figure 7).

To study the role of miRNA in photosynthesis, effectively, we filtered highly abundant miRNAs whose targets were assigned to relevant gene ontologies and three conserved miRNAs miR171f, miR319a and miR159c were highly abundant. The most abundant is miR171f targeting *Pentatricopeptide repeat-containing protein At3g62470* which is significantly enriched to a biological function 'electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity. miR319a and miR159c each target *NAD-dependent protein deacetylase SRT1* which are significantly enriched to the biological function 'NAD$^+$ binding'. The rest are non-conserved miRNAs MIR305714, MIR46018, MIR2836, and MIR305714, three targeting genes annotated to electron carrier activity (MIR305714, MIR46018, MIR2836) and MIR305714 targeting protein *SCARECROW SCR* significantly annotated to the stomatal complex morphogenesis (Table 2).

To proceed with these miRNAs, we validated the expression *via* real-time PCR using primers (Table S10). All the miRNAs are downregulated in sequencing and PCR analysis except miR171f whose target is downregulated. Therefore, we considered miR171f a target for forward genetic functional analysis (Figures 8A, B).

3.5 Heterologous MiR171f overexpression

MiR171f precursor sequence and its promoter were cloned from the C43 bermudagrass precursor sequence and the salt tolerance of transgenic *M. truncatula* was analyzed. The results showed that

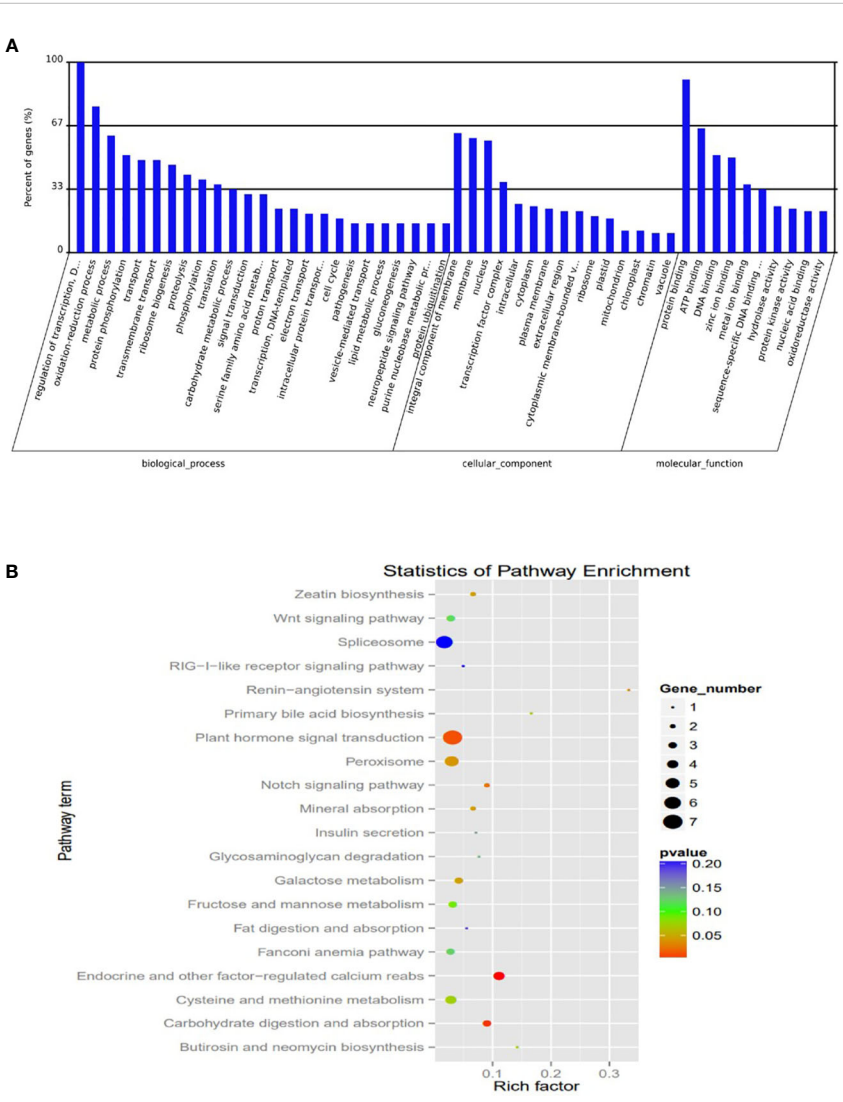


FIGURE 6
Plots of Characterization was based on **(A)** Gene ontology (GO) categories in the root apex of four bermudagrass libraries **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment.

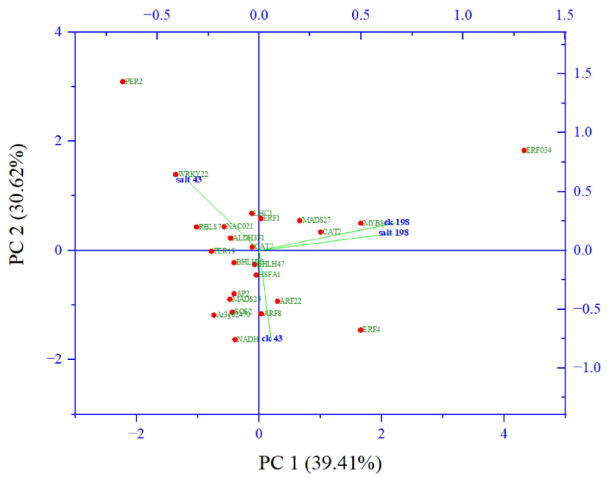


FIGURE 7
Principal component analysis of abundant salt responsive and photosynthesis target genes from gene ontologies.

MiR171f contained an intact stem-loop structure (Figure S1A). Overexpression of the MiR171f precursor in *M. truncatula* induced morphological changes both in the shoots and roots. Notable changes included greener leaves and denser canopy compared to the wild as well as the roots as depicted by Figure S1B. The mutant under salt stress had the highest accumulation of NADPH followed by the mutant control. While salinity caused a decline in NADPH in the wild compared to the control. However, the decline was not significant (Figure 9A). Also, there is a notable increase in dry matter yield in the mutant compared to the wild (Figure 9B).

The results revealed that different treatments resulted in considerably varied chlorophyll fluorescence behavior. For instance, the wild type of salt treatment had the lowest F_0 of 441, whereas the control wild had a much higher F_0 of 820. The F_0 was medium in the two mutants. Generally, after the first F_0, all treatments experienced an O-J increase that occurred between 0.00001-0.0001 s. To achieve the greatest F_M, the J-I-P phase of the fluorescence induction curve rise time spanned from 0.0001-0.001 s. The salt mutant had the highest F_M value of 2300 a.u, followed by the wild control, and the wild type of treatment had the lowest of 730.4 au. Notably, the J-I-P rise was much delayed in the wild type of salt regime (Figure 10A). The salt regime of the mutant exhibited the sharpest decline in the electron transport rate while the mutant salt treatment had the second highest (Figure 10B). The mutant also exhibited the highest initial PhiSII (Figure 10C) which despite the drop with the increase in light was the highest among the treatments. A faster rise in NPQ is also observed in the mutant (Figure 10D).

Multivariate analysis of the mutant reveals high contribution of NADPH to the dry matter content followed by the F_V/F_M. The least correlation is observed in ETR. Other high positive correlation is observed in F_V/F_M and phiPSII, and F_V/F_M vs NPQ (Figure 11).

4 Discussion

Since photosynthesis is one of the most stress-responsive functional phenotypes, the impact of salinity on the photosynthetic machinery is critical (Hameed et al., 2021). Typically, salinity imposes physiological drought disrupting the balance between foliar transpiration and root water intake (dos Santos et al., 2022). For most crops, the water-stressed leaves harbor 90% of total plant chlorophyll which constitutes the primary light-harvesting complex responsible for triggering the initial photochemical events (Lokstein et al., 2021). Effectively, salinity rapidly disrupts the reduction-oxidation properties of photosystem II acceptors and reduces the photosynthetic electron transport efficiency in both PSI and PSII (Mathur et al., 2014). Thus, attributed to this sensitivity, the genetic responses to stress such as salinity can be very dynamic and complex in nature (Nongpiur et al., 2016). Here, to gain more insight into the mechanism of salt perception and tolerance in bermudagrass, we comparatively analyze how the grass modifies its gene expression post-transcriptionally in photosynthetic leaves.

In addition to their roles in the growth and development and maintenance of genome integrity, miRNAs are important

TABLE 2 Target gene ontologies for highly downregulated microRNAs related to photosynthesis.

microRNA	ID	Gene target annotation	Abundance	Gene ID	GO term
miR156a	comp89548_c0	Premnaspirodiene oxygenase OS=Hyoscyamus muticus GN=*CYP71D55*	115	GO:0009055	electron carrier activity
MIR46018	comp105422_c0	Uncharacterized protein At5g39865 OS=*Arabidopsis thaliana* GN=*At5g39865*	763	GO:0009055	electron carrier activity
miR171f	comp116117_c0	Aldehyde dehydrogenase family 3 member F1 OS=*Arabidopsis thaliana* GN=*ALDH3F1*	1104	GO:0006118	electron transport
miR171f	comp124273_c0	"Pentatricopeptide repeat-containing protein At3g62470, mitochondrial OS=*Arabidopsis thaliana* 3g62470	3172	GO:0045156	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity
miR171f	comp121036_c0	Light harvesting protein complex1 OS=*Oryza sativa* subsp. indica GN= *LHC1*	1753	GO:0070403	Light harvesting protein complex
miR159c	comp121036_c0	NAD-dependent protein deacetylase *SRT1* OS=*Oryza sativa* subsp. indica GN=*SRT1*	1755	GO:0070403	NAD+ binding
MIR305714	comp122900_c0	Protein SCARECROW OS=*Zea mays* GN=*SCR*	505	GO:0010103	stomatal complex morphogenesis

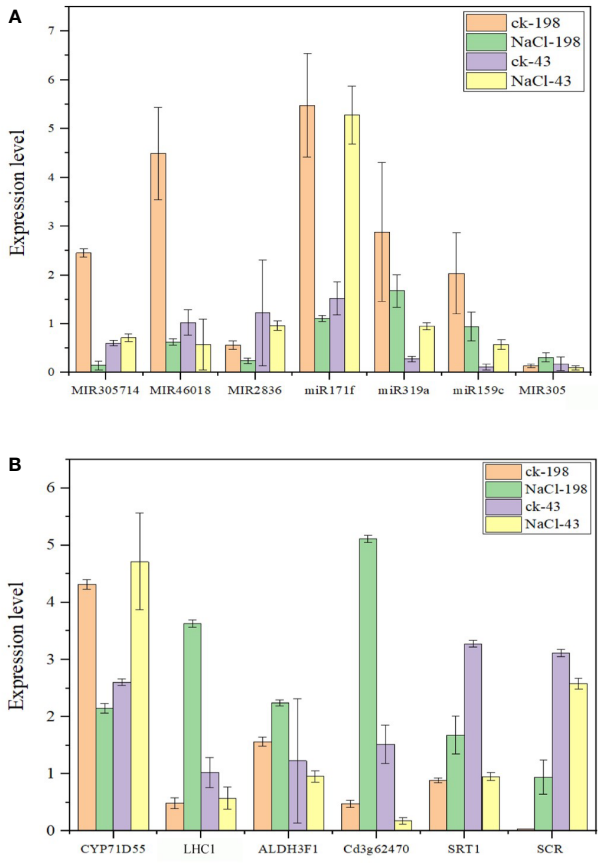


FIGURE 8
Bar plots showing real-time qPCR validation of miRNAs with high expression levels from sequencing data (at least 1000 in salt tolerant vs sensitive regime) **(A)** and respective target genes for the selected miRNAs **(B)**. The data are the means of four replicates. Fold change below 1 is downregulation and above 1 is upregulation.

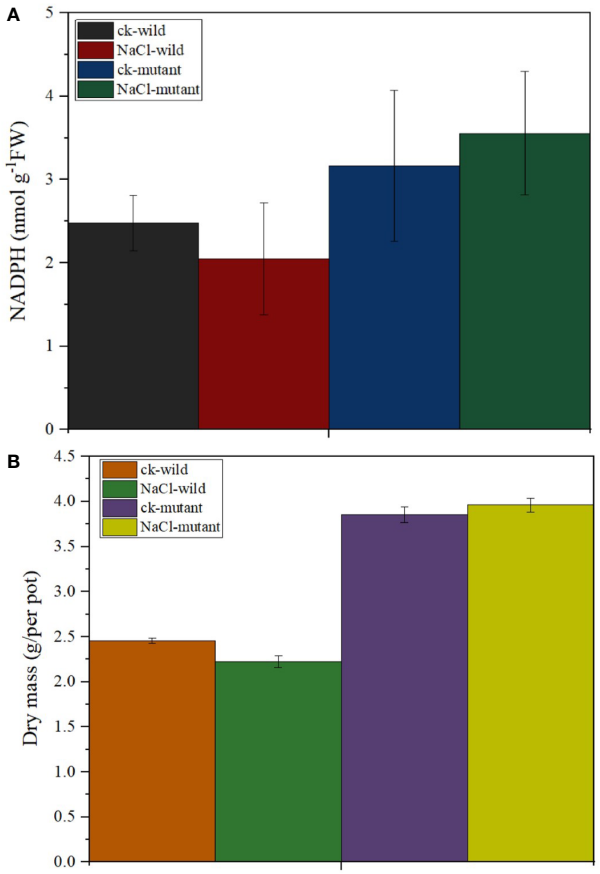


FIGURE 9
(A) Stem-loop secondary structures of MIR171f of bermudagrass. Segments corresponding to the mature miR171f are shown in orange, **(B)** Morphological variation in the control, wild type, and mutant barrelclover plants. NADPH and dry matter accumulation levels.

components in plant stress responses (Zhang, 2015). In this study, most of the novel miRNAs detected have a remarkably low abundance (less than 10 reads) indicating that Illumina sequencing was an effective tool in offering a rich source of unreported small RNA data in bermudagrass leaves. The dominance of 22 nt under salt treatment in both regimes suggests that the miRNAs with this length might play more dominant roles in bermudagrass leaf response to salt stress in both cultivars. This is consistent with previous studies in other genera members such as maize (*Z. mays*) (Barber et al., 2012), and barley (*Hordeum vulgare*) (Deng et al., 2015), as well as radish (*Raphanus sativus*) (Sun et al., 2015) and Chinese populus (*Populus tomentosa)* (Ren, 2013) growing under salt stress. Also, most members from conserved families such as miR166, miR156, miR167, miR172, and miR168 had a remarkably high abundance (more than 1000 in the four libraries). This indicates that the overexpression of these miRNAs might be involved in maintaining normal biological functions in the four bermudagrass libraries. Despite a majority having a low abundance, some of the novel miRNAs identified from four libraries had a relatively high abundance (≥10), and their expression levels changed significantly in salt-tolerant regimes. For example, six of the 11 most abundant novel miRNAs are downregulated while only five are upregulated. This similar trend is observed in total miRNAs expressed in C43 vs C198 salt regime whereby most miRNAs are downregulated. These results suggested that the downregulation of not only known but also novel miRNAs play a role in salt tolerance in bermudagrass leaves.

Noteworthy, some of the abundant miRNAs whose targets were not related to photosynthesis from our gene ontology study have been recently confirmed to play vital roles in photosynthesis in other species. For example, suppression of miR166 through gene editing influenced CO_2 assimilation in rice (Iwamoto, 2022). MiR396 was not only found to be salt responsive in Arabidopsis (Song et al., 2019) and maize (Ding et al., 2009) but also to regulate the expression of chlorophyll biosynthetic genes (Wang et al., 2020). Pan et al. (2016) observed that miR172 improves photosynthetic performance in soybean. Furthermore, alfalfa genotypes overexpressing miR156 exhibited altered photosynthesis activity under salinity conditions (Arshad et al., 2017). Also, it has been shown that rice mutants with suppressed miR390 exhibited an increase in chlorophyll accumulation (Ding et al., 2016). For miRNAs whose targets are in the photosynthesis pathway in this study, 4 are conserved (miR171f,

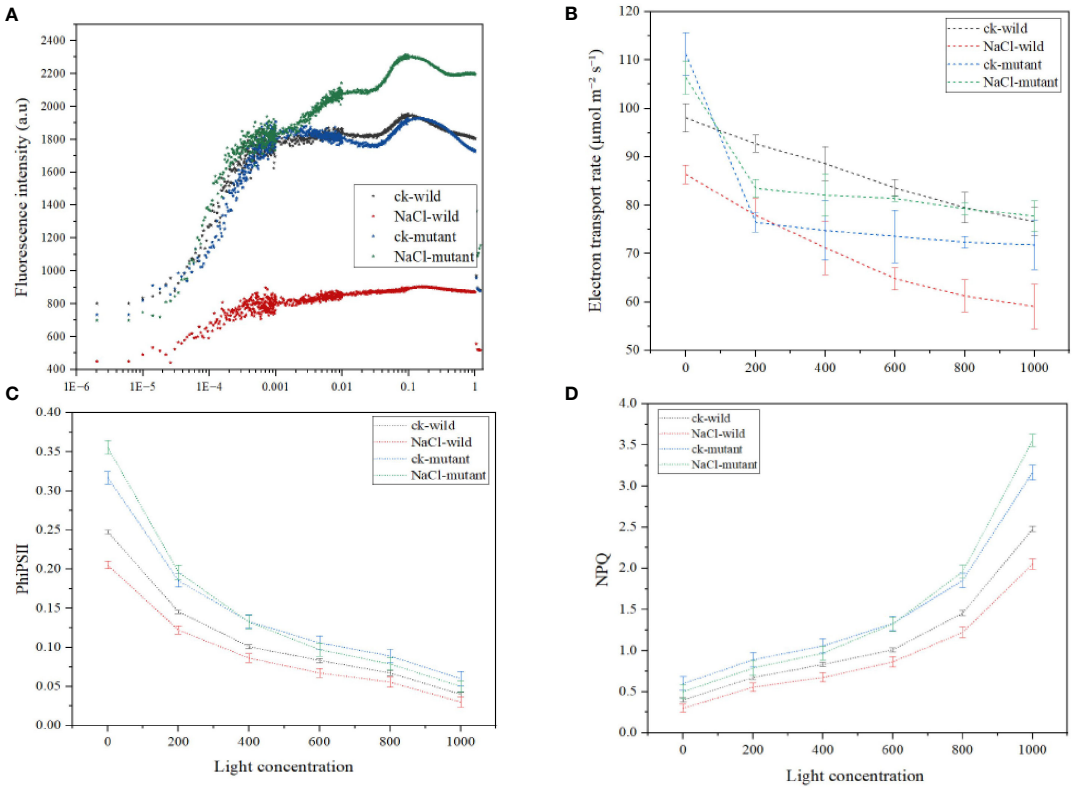


FIGURE 10
Chlorophyll a fluorescence induction kinetics parameters under salt treatments **(A)** OJIP curve **(B)** electron transport rate **(C)** quantum yield of photosystem II **(D)** nonphotochemical quenching.

miR319, miR156, and miR159) while 3 are nonconserved. Interestingly, the targets fall in the electron transport photosynthesis and NAD^+ binding. In a closely related species to bermudagrass, it was reported that the overexpression of miR319 in creeping bentgrass significantly enhanced tolerance to salt stress by enhancing the photosynthetic performance. Generally, these observations indicate that the downregulation of these miRNAs in the C43 vs C198 salt regime demonstrates a possible regulatory role in photosynthetic regulation through transcriptional repression under salt stress.

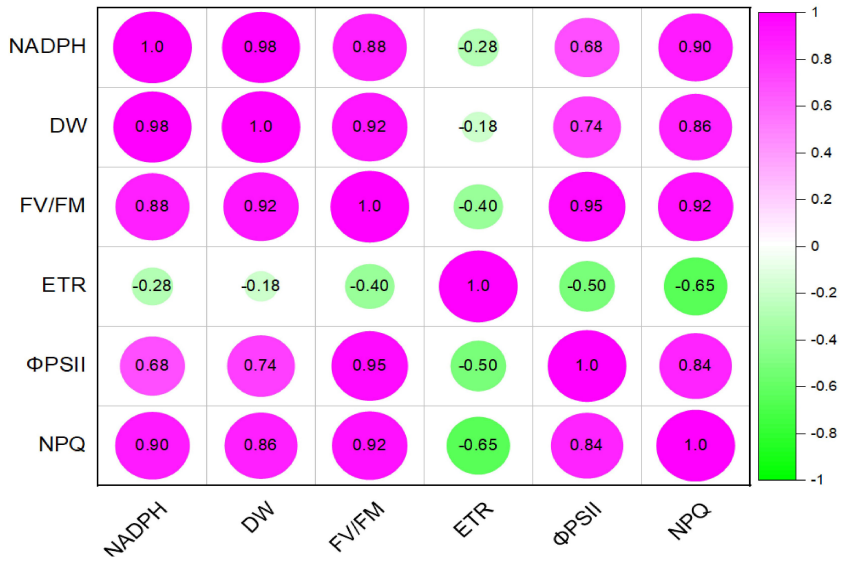


FIGURE 11
pearssons correlation plot of various light response chlorophyll a fluorence parameters, NADPH and dry matter of mutant *Medicago truncatula* overexpressing bermudagrass miR171f under saline conditions.

Generally, miRNA activity leads to the repression of genes and transcription factors they target. Therefore, the upregulation of miR171f under saline conditions in salt tolerant variety demonstrates that the suppression of gene targets is a prerequisite for desirable phenotypes. miR171f targets the *Pentatricopeptide repeat-containing protein At3g62470* gene located in the chloroplasts which are associated with electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity. This is consistent with the fact that light-dependent reactions of photosynthesis take place in the thylakoid membrane, inside chloroplasts (Merchant and Sawaya, 2005). It is also in the chloroplasts where the light-harvesting complex is located. Here, chlorophyll fluorescence occurs transforming light energy into chemical energy (Papageorgiou, 2004). In this study, following the initial F_0, all treatments saw an O-J rise suggesting that salinity modifications had no effect on this phase. This phase represents the photochemical step of Chl fluorescence induction. Thus, higher F_0 values in the mutant relative to the wild type showed a larger physical separation of the PSII reaction center from their corresponding pigment antennae, which has been shown to contribute to better photosynthetic performance by restricting energy input into the electron transport chain (Strasser et al., 2000). The J-I-P phase of the fluorescence induction curve rising time was set at 0.0001-0.001 s to produce the highest F_M. Notably, the wild-type J-I-P increase was substantially delayed, demonstrating plastoquinone accumulation, whereas it increased in the mutant. The large increase in I translates to slower electron transit to the PSI acceptors (Joliot and Johnson, 2011). In mutants, there is also a higher plateau, indicating a bigger number of PSI end acceptors, which are typically linked with alternative electron transfer routes that function as electron sinks (Popova et al., 2022). During the photochemical reactions, plants use water to make NADPH which is then metabolized to release electrons (Wasilewska-Dębowska et al., 2022). the NPQ influences overall plant photosynthesis, biomass, and yield by preventing photoinhibition and temporarily limiting photosynthetic quantum output. (Murchie and Ruban, 2020). In this study, a positive correlation between NPQ and F_V/F_M is observed from the boot stage indicating that most light dissipation occurred at this stage while at flowering some varieties had begun to exhibit photoinhibition. Further, the NPQ operation is related to the state of reduction of the thylakoid electron transport (Alboresi et al., 2019). This is an important point because the level of protection required depends on the balance between the amount of absorbed irradiance and the ability of the system to use this energy within the various 'sinks' for electrons (Joliot and Johnson, 2011). Thus, it is important to determine the rate of flow of electrons. The NADPH is a product of the first level of photosynthesis (Johnson, 2017). It helps to fuel the reactions that occur in the second stage of the process of photosynthesis (Lim et al., 2020). Higher production of NADPH in mutants indicates that miR171f promotes higher electron donation under salt stress. The released electrons must travel through the photosystem II protein and down the electron transport chain then they pass through photosystem I (Cooper, 2000). Thus, higher accumulation of PhiPSII therefore indicated high efficiency in the transfer of electrons in the mutants that the wild type.

Chlorophyll constitutes the primary light-harvesting complex responsible for triggering the initial photochemical events (Lokstein et al., 2021). Here, we proposed that miR171f mediated transcriptional regulation of LHC1 modulates the reduction-oxidation properties of photosystem II acceptors and enhances the photosynthetic electron transport efficiency in both PSI and PSII (Mathur et al., 2014) which influences the NADPH synthesis from NADP (Figure 12).

5 Conclusions

In conclusion, we used RNA sequencing to profile miRNAs in two bermudagrass cultivars with variable degrees of salt tolerance when subjected to salt stress. The bulk of the 536 salt-inducible miRNA variations were downregulated in salt-tolerant cultivars. Seven miRNAs targeted six genes associated with electron transport, dark reaction photosynthesis, and NAD+ binding. MiR171f targeting a Pentatricopeptide repeat-containing protein *Cd3g62470* and Aldehyde dehydrogenase family 3 member F1, both of which are considerably enriched in the electron transport route of light response photosynthesis, were dramatically increased in salt tolerant regimes. To enable genetic breeding for photosynthetic capacity, we

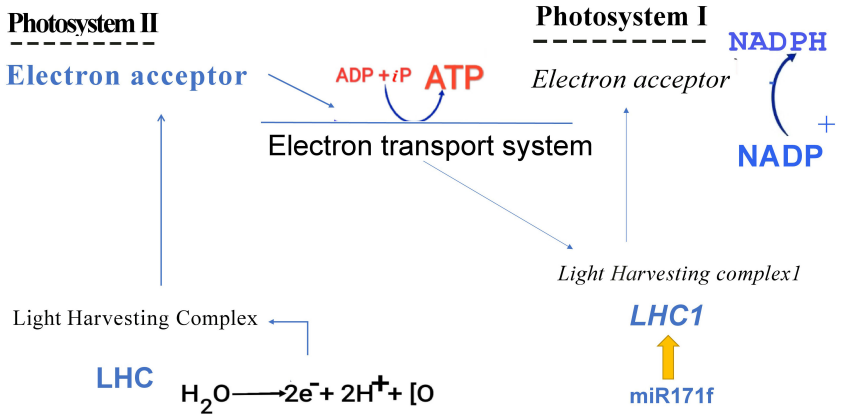


FIGURE 12
Graphical representation of proposed mechanism.

overexpressed miR171f in *Medicago truncatula*, which led in a considerable improvement in photosynthetic performance and biomass accumulation in the mutants under saline circumstances compared to the wild type. While not very qualitative, these findings provide significant insights into plant breeding.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number PRJNA287735.

Author contributions

SF performed the statistical analysis and wrote the first draft of the manuscript, EA conducted the research, SA wrote sections of the manuscript, YL and YY contributed to conception and design of the study. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Natural Science Foundation of China, grant number 31901396 and 42107128, Shandong Provincial Natural Science Foundation (ZR2021QD049).

Acknowledgments

We would like to thank Mr. Yinkun Zhang for helping us take care of experimental materials in the greenhouse.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1141295/full#supplementary-material

References

Alboresi, A., Storti, M., and Morosinotto, T. (2019). Balancing protection and efficiency in the regulation of photosynthetic electron transport across plant evolution. *New Phyto.* 221.1, 105–109. doi: 10.1111/nph.15372

Allakhverdiev, S. I., Sakamoto, A., Nishiyama, Y., Inaba, M., and Murata, N. (2000). Ionic and osmotic effects of NaCl-induced inactivation of photosystems I and II in synechococcus sp. *Plant Physiol.* 123, 1047–1056. doi: 10.1104/pp.123.3.1047

Alori, E. T., Emmanuel, O. C., Glick, B. R., and Babalola, O. O. (2020). Plant-archaea relationships: a potential means to improve crop production in arid and semi-arid regions. *World J. Microbiol. Biotechnol.* 36 (9), 133. doi: 10.1007/s11274-020-02910-6

Arshad, M., Gruber, M. Y., Wall, K., and Hannoufa, A. (2017). An insight into microRNA156 role in salinity stress responses of alfalfa. *Front. Plant Sci.* 8, 356. doi: 10.3389/fpls.2017.00356

Ashraf, M., and Harris, P. J. C. (2013). Photosynthesis under stressful environments: An overview. *Photosynthetica* 51, 163–190. doi: 10.1007/s11099-013-0021-6

Barber, T. W., Wei, Z., Kranthi, H., Dorweiler, J., Hudson, M. E., and Moose, S. P. (2012). Repeat-associated small RNAs vary among parents and following hybridization in maize. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10444–10449. doi: 10.1073/pnas.1202073109

Busch, F. A., Sage, R. F., and Farquhar, G. ,. D. (2018). Plants increase CO_2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nat. Plants.* 4 (1), 46–54. doi: 10.1038/s41477-017-0065-x

Cheeseman, J. M. (1988). Mechanisms of salinity tolerance in plants. *Plant Physiol.* 87, 547–550. doi: 10.1104/pp.87.3.547

Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., et al. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33, 179. doi: 10.1093/nar/gni178

Chen, J. B., Zong, J. Q., Li, D. D., Chen, Y., Wang, Y., Guo, H. L., et al. (2019). Growth response and ion homeostasis in two bermudagrass (*Cynodon dactylon*) cultivars differing in salinity tolerance under salinity stress. *Soil Sci. Plant Nutr.* 65 (4), 419–429. doi: 10.1080/00380768.2019.1631125

Cooper, G. M. (2000). *The cell: A molecular approach. 2nd edition* (Sunderland (MA: Sinauer Associates).

Deng, P., Bian, J., Yue, H., Wang, M., Du, X., Song, W., et al. (2015). Characterization of microRNAs and their targets in wild barley (*Hordeum vulgare* subsp. spontaneum) using deep sequencing. *Genome* 59, 339–348. doi: 10.1139/gen-2015-0224

Ding, D., Zhang, L., Wang, H., Liu, Z., Zhang, Z., and Zheng, Y. (2009). Differential expression of miRNAs in response to salt stress in maize roots. *Ann. Bot.* 103, 29–38. doi: 10.1093/aob/mcn205

Ding, Y., Ye, Y., Jiang, Z., Wang, Y., and Zhu, C. (2016). MicroRNA390 is involved in cadmium tolerance and accumulation in rice. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00235

dos Santos, T. B., Ribas, A. F., de Souza, S. G., Budzinski, I. G., and Domingues, D. S. (2022). Physiological responses to drought, salinity, and heat stress in plants: a review. *Stresses.* 2(1), 113–35.

Feng, K., Nie, X., Cui, L., Deng, P., Wang, M., and Song, W. (2017). Genome-wide identification and characterization of salinity stress-responsive miRNAs in wild emmer wheat (*Triticum turgidum* ssp. *Dicoccoides). Genes (Basel)* 8, 6. doi: 10.3390/genes8060156

Fita, A., Rodríguez-Burruezo, A., Boscaiu, M., Prohens, J., and Vicente, O. (2015). Breeding and domesticating crops adapted to drought and salinity: A new paradigm for increasing food production. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00978

Guidi, L., Lo-Piccolo, E., and Landi, M. (2019). Chlorophyll fluorescence, photoinhibition and abiotic stress: Does it make any difference the fact to be a C3 or C4 species? *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00174

Hameed, A., Ahmed, M. Z., Hussain, T., Aziz, I., Ahmad, N., Gul, B., et al. (2021). Effects of salinity stress on chloroplast structure and function. *Cells* 10 (8), 2023. doi: 10.3390/cells10082023

Hasegawa, P. M., Bressan, R. A., Zhu, J. K., and Bohnert, H. J. (2000). Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Mol. Biol.* 51, 463–499. doi: 10.1146/annurev.arplant.51.1.463

Hassani, A., Azapagic, A., and Shokri, N. (2021). Global predictions of primary soil salinization under changing climate in the 21st century. *Nat. Commun.* 12, 6663. doi: 10.1038/s41467-021-26907-3

Helge, G., and Witold, F. (2008). The expanding world of small RNAs. *Nature* 451, 414–416. doi: 10.1038/451414a

Hofacker, I. L. (2003). Vienna RNA Secondary structure server. *Nucleic Acids Res.* 31, 3429–3431. doi: 10.1093/nar/gkg599

Hu, L., Li, H., Chen, L., Lou, Y., Amombo, E., and Fu, J. (2015). RNA-Seq for gene identification and transcript profiling in relation to root growth of bermudagrass (*Cynodon dactylon*) under salinity stress. *BMC Genomics* 16, 575. doi: 10.1186/s12864-015-1799-3

Iwamoto, M. (2022). In-frame editing of transcription factor gene RDD1 to suppress miR166 recognition influences nutrient uptake, photosynthesis, and grain quality in rice. *Sci. Rep.* 12 (1), 10795. doi: 10.1038/s41598-022-14768-9

Johnson, M. P. (2017). Photosynthesis [published correction appears in essays biochem. *61(4):429]. Essays Biochem.* 60 (3), 255–273. doi: 10.1042/EBC20160016

Joliot, P., and Johnson, G. N. (2011). Regulation of cyclic and linear electron flow in higher plants. *Proc. Natl. Acad. Sci. U S A.* 108 (32), 13317–13322. doi: 10.1073/pnas.1110189108

Karam, M. A., Abd-Elgawad, M. E., and Ali, R. M. (2016). Differential gene expression of salt-stressed *Peganum harmala* l. *J. Genet. Eng. Biotechnol.* 14 (2), 319–326. doi: 10.1016/j.jgeb.2016.10.005

Kaundal, A., Rojas, C. M., and Mysore, K. S. (2012). Measurement of NADPH oxidase activity in plants. *Bio-protocol* 2 (20), e278. doi: 10.21769/BioProtoc.278

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9 (4), 357–359. doi: 10.1038/nmeth.1923

Li, G. Z., Wang, Z. Q., Yokosho, K., Ding, B., Fan, W., Gong, Q. Q., et al. (2018). Transcription factor WRKY22 promotes aluminum tolerance *via* activation of OsFRDL4 expression and enhancement of citrate secretion in rice (*Oryza sativa*). *New Phytol.* 219 (1), 149–162. doi: 10.1111/nph.15143

Lim, S. L., Voon, C. P., Guan, X., Yang, Y., Gardeström, P., and Lim, B. L.. (2020). *In planta* study of photosynthesis and photorespiration using NADPH and NADH/NAD^+ fluorescent protein sensors. *Nat. Commun.* 11, 3238. doi: 10.1038/s41467-020-17056-0

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of scarecrow-like mRNA targets directed by a class of arabidopsis miRNA. *Science* 297, 2053–2056. doi: 10.1126/science.1076311

Lokstein, H., Renger, G., and Götze, J. P. (2021). Photosynthetic light-harvesting (Antenna) complexes-structures and functions. *Molecules* 26 (11), 3378. doi: 10.3390/molecules26113378

Mathur, S., Agrawal, D., and Jajoo, A. (2014). Photosynthesis: response to high temperature stress. *J. Photochem. Photobiol. B.* 137, 116–126. doi: 10.1016/j.jphotobiol.2014.01.010

Merchant, S., and Sawaya, M. R. (2005). The light reactions: a guide to recent acquisitions for the picture gallery. *Plant Cell* 17 (3), 648–663. doi: 10.1105/tpc.105.030676

Munns, R. (2002). Comparative physiology of salt and water stress. *Plant Cell Environ.* 25, 239–250. doi: 10.1046/j.0016-8025.2001.00808.x

Murchie, E. H., and Ruban, A. V. (2020). Dynamic non-photochemical quenching in plants: from molecular mechanism to productivity. *Plant J.* 101 (4), 885–896. doi: 10.1111/tpj.14601

Nongpiur, R. C., Singla-Pareek, S. L., and Pareek, A. (2016). Genomics approaches for improving salinity stress tolerance in crop plants. *Curr. Genomics* 17 (4), 343–357. doi: 10.2174/1389202917666160331202517

Pan, J., Huang, D., Guo, Z., Kuang, Z., Zhang, H., Xie, X., et al. (2018). Overexpression of microRNA408 enhances photosynthesis, growth, and seed yield in diverse plants. *J. Integr. Plant Biol.* 60 (4), 323–340. doi: 10.1111/jipb.12634

Pan, W. J., Tao, J. J., Cheng, T., Bian, X. H., Wei, W., Zhang, W. K., et al. (2016). Soybean miR172a improves salt tolerance and can function as a long-distance signal. *Mol. Plant* 69, 1337–1340. doi: 10.1016/j.molp.2016.05.010

Papageorgiou, G. C. (2004). *Chlorophyll a fluorescence: A signature of photosynthesis* (Dordrecht: Springer Netherlands).

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in arabidopsis thaliana. *Curr. Biol.* 12, 1484–1495. doi: 10.1016/S0960-9822(02)01017-5

Popova, A. V., Stefanov, M., Ivanov, A. G., and Velitchkova, M. (2022). The role of alternative electron pathways for effectiveness of photosynthetic performance of *Arabidopsis thaliana*, wt and *Lut2*, under low temperature and high light intensity. *Plants (Basel).* 11 (17), 2318. doi: 10.3390/plants11172318

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev.* 16, 1616–1626. doi: 10.1101/gad.1004402

Ren, Y. (2013). Identification and characterization of salt-responsive microRNAs in populus tomentosa by high-throughput sequencing. *Biochimie* 95, 743–750. doi: 10.1016/j.biochi.2012.10.025

Sherstneva, O., Khlopkov, A., Gromova, E., Yudina, L., Vetrova, Y., Pecherina, A., et al. (2022). Analysis of chlorophyll fluorescence parameters as predictors of biomass accumulation and tolerance to heat and drought stress of wheat (*Triticum aestivum*) plants. *Funct. Plant Biol.* 49 (2), 155–169. doi: 10.1071/FP21209

Shrivastava, P., and Kumar, R. (2015). Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi J. Biol. Sci.* 22 (2), 123–131. doi: 10.1016/j.sjbs.2014.12.001

Simkin, A. J., López-Calcagno, P. E., and Raines, C. A. (2019). Feeding the world: improving photosynthetic efficiency for sustainable crop production. *J. Exp. Bot.* 70 (4), 1119–1140. doi: 10.1093/jxb/ery445

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant-environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334

Strasser, R. J., Srivastava, A., and Tsimilli-Michael, M. (2000). "The fluorescence transient as a tool to characterize and screen photosynthetic samples," in *Probing Photosynthesis: Mechanism, Regulation & Adaptation*, 443–480. Available online at: http://www.hansatech-instruments.com/docs/the%20fluorescence%20transient.pdf/

Sun, X., Xu, L., Wang, Y., Yu, R., Zhu, X., Luo, X., et al. (2015). Identification of novel and salt-responsive miRNAs to explore the miRNA-mediated regulatory network of salt stress response in radish (*Raphanus sativus* l.). *BMC Genomics* 16, 197. doi: 10.1186/s12864-015-1416-5

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046

Wang, L., Tian, Y., Shi, W., et al. (2020). The miR396-GRFs module mediates the prevention of photo-oxidative damage by brassinosteroids during seedling de-etiolation in arabidopsis. *Plant Cell.* 32 (8), 2525–2542. doi: 10.1105/tpc.20.00057

Wasilewska-Dębowska, W., Zienkiewicz, M., and Drozak, A. (2022). How light reactions of photosynthesis in C4 plants are optimized and protected under high light conditions. *Int. J. Mol. Sci.* 23 (7), 3626. doi: 10.3390/ijms23073626

Xie, F., Stewart, C. N., Taki, F. A., He, Q., Liu, H., and Zhang, B. (2013). High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. *Plant Biotechnol. J.* 12, 354–366. doi: 10.1111/pbi.12142

Yadav, S., Mohammad, I., Aqil, A., and Shamsul, H. (2011). Causes of salinity and plant manifestations to salt stress: A review. *J. Environ. Biol.* 32, 667–685.

Yang, Z., Li, W., Su, X., Ge, P., Zhou, Y., Hao, Y., et al. (2019). Early response of radish to heat stress by strand-specific transcriptome and miRNA analysis. *International journal of molecular sciences*. 20(13):3321.

Yang, Z., Li, J. L., Liu, L. N., Xie, Q., and Sui, N. (2020). Photosynthetic regulation under salt stress and salt-tolerance mechanism of sweet sorghum. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01722

Yang, Z., Zhu, P., Kang, H., et al. (2020). High-throughput deep sequencing reveals the important role that microRNAs play in the salt response in sweet potato (*Ipomoea batatas* l.). *BMC Genomics* 21, 164. doi: 10.1186/s12864-020-6567-3

Zhang, B. H. (2015). MicroRNA: a new target for improving plant tolerance to abiotic stress. *J. Expt. Bot.* 66, 1749–1761. doi: 10.1093/jxb/erv013

Zhang, J. P., Yu, Y., Feng, Y. Z., Zhou, Y. F., Zhang, F., Yang, Y. W., et al. (2017). MiR408 regulates grain yield and photosynthesis *via* a phytocyanin protein. *Plant Physiol.* 175 (3), 1175–1185. doi: 10.1104/pp.17.01169

Zhang, B., Liu, J., Wang, X., and Wei, Z. (2018). Full-length RNA sequencing reveals unique transcriptome composition in bermudagrass. *Plant physiology and biochemistry*. 132, 95–103.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpls.2023.1137077

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Aamir Manzoor,
Anhui Agricultural University, China

REVIEWED BY
Helal Mostofa,
Shanxi Academy of Agricultural Sciences, China
Muhammad Ahmad Hassan,
Anhui Agricultural University, China
Jiming Liu,
Nanyang Technological University, Singapore

*CORRESPONDENCE
Wirulda Pootakham
wirulda@alumni.stanford.edu

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 04 January 2023
ACCEPTED 06 February 2023
PUBLISHED 16 February 2023

CITATION
Nawae W, Naktang C, Charoensri S, U-thoomporn S, Narong N, Chusri O, Tangphatsornruang S and Pootakham W (2023) Resequencing of durian genomes reveals large genetic variations among different cultivars.
Front. Plant Sci. 14:1137077.
doi: 10.3389/fpls.2023.1137077

COPYRIGHT
© 2023 Nawae, Naktang, Charoensri, U-thoomporn, Narong, Chusri, Tangphatsornruang and Pootakham. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Resequencing of durian genomes reveals large genetic variations among different cultivars

Wanapinun Nawae[1], Chaiwat Naktang[1], Salisa Charoensri[1], Sonicha U-thoomporn[1], Nattapol Narong[1], Orwintinee Chusri[2], Sithichoke Tangphatsornruang[1] and Wirulda Pootakham[1]*

[1]National Omics Center (NOC), National Science and Technology Development Agency (NSTDA), Thailand Science Park, Pathum Thani, Thailand, [2]Chantaburi Horticulture Research Center, Horticulture Research Institute, Department of Agriculture, Chantaburi, Thailand

Durian (*Durio zibethinus*), which yields the fruit known as the "King of Fruits," is an important economic crop in Southeast Asia. Several durian cultivars have been developed in this region. In this study, we resequenced the genomes of three popular durian cultivars in Thailand, including Kradumthong (KD), Monthong (MT), and Puangmanee (PM) to investigate genetic diversities of cultivated durians. KD, MT, and PM genome assemblies were 832.7, 762.6, and 821.6 Mb, and their annotations covered 95.7, 92.4, and 92.7% of the embryophyta core proteins, respectively. We constructed the draft durian pangenome and analyzed comparative genomes with related species in Malvales. Long terminal repeat (LTR) sequences and protein families in durian genomes had slower evolution rates than that in cotton genomes. However, protein families with transcriptional regulation function and protein phosphorylation function involved in abiotic and biotic stress responses appeared to evolve faster in durians. The analyses of phylogenetic relationships, copy number variations (CNVs), and presence/absence variations (PAVs) suggested that the genome evolution of Thai durians was different from that of the Malaysian durian, Musang King (MK). Among the three newly sequenced genomes, the PAV and CNV profiles of disease resistance genes and the expressions of methylesterase inhibitor domain containing genes involved in flowering and fruit maturation in MT were different from those in KD and PM. These genome assemblies and their analyses provide valuable resources to gain a better understanding of the genetic diversity of cultivated durians, which may be useful for the future development of new durian cultivars.

KEYWORDS

durian, genome assembly, pangenome, comparative genomics, genomic structural variation

1 Introduction

Durio zibethinus (family Malvaceae) or durian is an endemic plant species of Southeast Asia (Teh et al., 2017). Cultivated durians produce thorny fruits, known as the "King of Fruit", with sweet, delicious, and richly aromatic arils inside (Ketsa et al., 2020). The expensive price of their fruits makes durians an important economic crop in Thailand, Malaysia, and Indonesia (Ketsa et al., 2020). Several durian cultivars have been developed, *via* outcrossing and selection, and cultivated in these countries (Ketsa et al., 2020). Some popular durian cultivars in Thailand are Monthong, Kanyao, Kradumthong, Puangmanee, and Chanee. These cultivars exhibit different tree sizes, fruiting periods each year, fruit sizes, fruit ripening, and the number, flavor, texture, and aroma intensity of arils (Ketsa et al., 2020). The resistance levels to *Phytophthora* and other diseases also vary among cultivars (O'Gara et al., 2004; Ketsa et al., 2020). To understand the genetic basis of durian, the genome of Musang King cultivar has been sequenced (Teh et al., 2017). The analysis of this genome data provided insights into the genome evolution of *D. zibethinus*, including the expansion of the methionine ɣ-lyase gene family, which played a role in the production of volatile sulfur compounds (Teh et al., 2017). The reanalysis of this genome showed a hexaploidization event in the durian genome about 19–21 million years ago (Mya) (Wang et al., 2019). Another analysis of the same genome sequence revealed a catalog of 2,586 resistance gene analogs in this durian cultivar (Cortaga et al., 2022). This genome sequence has proven itself to be an important resource for understanding the agronomic characteristics and evolution of durian.

Plant genomes are dynamic and a single reference genome sequence for each species might not reflect genomic diversity within species (Danilevicz et al., 2020). The resequencing of genomes from multiple individuals with different phenotypes within a species and the construction of pangenome enabled the listing of total core genes (found in all individuals) and dispensable or accessory genes (found in subsets of individuals) (Bayer et al., 2020; Danilevicz et al., 2020; Della Coletta et al., 2021). One of the approaches for constructing a pangenome was map-to-pan, which included the mapping of resequenced reads to a reference genome, *de novo* assembly of unmapped reads, and the incorporating of assembled contigs with the reference genome (Hu et al., 2020; Della Coletta et al., 2021). Pangenomes were constructed from hundreds to thousands of individuals in sunflower (Hübner et al., 2019), tomato (Gao et al., 2019), and Asian rice (Wang et al., 2018). A much smaller number of individuals were used to construct the pangenomes of walnut (Trouern-Trend et al., 2020), banana (Rijzaani et al., 2022), maize (Haberer et al., 2020), and pepper (Kim et al., 2021). The key knowledge obtained from the analyses of pangenomes was extensive structural variants (SVs), including presence-absence variation (PAV) and copy number variation (CNV) (Della Coletta et al., 2021). These variations were shown to be linked to agronomic traits and were useful for crop improvement (Danilevicz et al., 2020). For example, the analyses of the pangenome showed the presence of an allele in the *TomLoxC* promoter that contributed to desirable tomato flavor and the negative selection of disease-resistance genes during the domestication of tomatoes (Gao et al., 2019). The variations of several disease-resistance gene families were reported from the analyses of walnut (Trouern-Trend et al., 2020) and pepper (Kim et al., 2021) pangenomes. The effects of SVs on agronomically important traits highlighted the value of genome resequencing on multiple individuals with different phenotypes within a species (Della Coletta et al., 2021).

In this study, we resequenced the genomes of three popular durian cultivars, including Kradumthong (KD), Monthong (MT), and Puangmanee (PM). The resequencing allowed us to identify the genetic variations among durian cultivars and between durians and other species. We also analyzed transcriptome data to see groups of transcripts highly expressed in different durian cultivars. The results improved our understanding of the durian genome evolutions and provided other genetic information that could be used to guide the breeding of new durian cultivars.

2 Materials and methods

2.1 Plant materials and DNA/RNA isolation

The MT, KD, and PM trees whose materials were used in this study were maintained at Chanthaburi horticultural research center, Chanthaburi, Thailand. They grow in the non-flooded area and receive 150 liters/tree/day of water supply. Fertilizers and pesticides are supplied regularly. For this study, healthy leaves were collected, immediately frozen in liquid nitrogen, and stored at −80°C. DNA was extracted and purified using the QIAGEN Genomic-tip 100/G following the manufacturer's protocol (Qiagen, Germany). DNA quality was assessed using 0.75% pulsed-field gel electrophoresis and the concentration was tested with Qubit® dsDNA BR Assay Kits (Thermo Fisher Scientific) and NanoDrop (Thermo Fisher Scientific).

Total RNA was extracted from healthy leaves using CTAB buffer and 25:24:1 phenol:chloroform:isoamyl alcohol. Contaminated DNA was removed by using DNA-free™ DNA Removal Kit (Invitrogen™). The quality and quantity of RNA were evaluated with the fragment analyzer machine (Agilent). Poly(A) mRNAs were enriched from total RNA samples using the Dynabeads mRNA Purification Kit (Thermo Fisher Scientific, Waltham, MA, USA).

2.2 Library preparation and sequencing

For library preparation and sequencing, one nanogram of high quality, high molecular weight DNA was used for the 10x Genomics linked-read library preparation using the Chromium Genome Library Kit & Gel Bead Kit v2, the Chromium Genome Chip Kit v2, and the Chromium i7 Multiplex Kit according to the manufacturer's instructions (10x Genomics, Pleasanton, USA). The library quality was assessed using Bioanalyzer DNA High Sensitivity DNA Assay (Agilent) and the concentration was tested with Qubit® dsDNA BR Assay Kits (Thermo Fisher Scientific). The 10x Genomics library was sequenced on the Illumina HiSeq X Ten (150 bp paired-

end reads). For RNA, we constructed cDNA libraries according to the MGIEasy RNA Library Prep set protocol. The libraries were sequenced with the MGISEQ-2000RS machine.

2.3 Genome assembly

The linked-read data were assembled using the Supernova assembler version 2.1.1 with the default parameter setting (https://support.10xgenomics.com/de-novo-assembly/software/pipelines/latest/using/running; 10x Genomics, Pleasanton, USA). For the quality assessment, short-read DNA sequence data obtained from this study were mapped back to the final assembly sequences using minimap2 (Li, 2018), and the percentage of successful mapping was identified. We also employed the Benchmarking Universal Single-Copy Orthologs (BUSCO) version 4.0.5 (Simão et al., 2015) to evaluate the assembly by testing for the presence and completeness of the orthologs using the embryophyta OrthoDB release 10 database (Zdobnov et al., 2021).

2.4 Genome annotation

We used RepeatModeler (Flynn et al., 2020) with default parameters to construct libraries of the consensus sequences of TEs from the assemblies of four durian cultivars. The assembly of MK was retrieved from the Genomes – NCBI Datasets database. The consensus sequences of all cultivars provided by RepeatModeler were used with RepeatMasker version 4.1.2 (http://www.repeatmasker.org) with default parameters to identify repeats in all durian assemblies. The LTR compositions in each assembly were identified by LTRharvest (Ellinghaus et al., 2008) and LTR_FINDER_parallel (Ou and Jiang, 2019) with default parameters. The LTR annotations from both programs were used with LTR_retriever (Ou and Jiang, 2018) to calculate the insertion of times of LTR sequences based on the rate of nucleotide substitution per site per year of 3.5×10^{-9} (Wang et al., 2021).

We used MAKER2 (Holt and Yandell, 2011) to annotate gene regions in repeat masked sequences of MT, KD, and PM assemblies. The assembled transcript sequences of each durian cultivar and the protein sequences of Arabidopsis, grape, rice, and soybean were as described in the MK annotation (Teh et al., 2017). In the MAKER pipeline, this evidence was used to generate an initial set of gene predictions. Snap (Korf, 2004) and Augustus (Stanke et al., 2006) were used for ab initio gene predictions based on the first round of gene predictions. We employed the BUSCO version 4.0.5 (Simão et al., 2015) to evaluate the annotation results by testing for the presence and completeness of the orthologs using the embryophyta OrthoDB release 10 database (Zdobnov et al., 2021).

2.5 Comparative genomics and phylogenetic analysis

We used OrthoFinder (Emms and Kelly, 2019) to identify orthologous groups (protein families) from protein sequences of four durian cultivars (*Durio zibethinus*), two cottons (*Gossypium arboreum* and *Gossypium raimondii*) and cacao (*Theobroma cacao*), Arabidopsis (*Arabidopsis thaliana*) and papaya (*Carica papaya*). The protein sequences of Musang King durian, *G. arboreum*, *G. raimondii*, *T. cacao*, *A. thaliana*, and *C. papaya* were downloaded from the NCBI database with the accession numbers GCF_002303985.1, GCF_000612285.1, GCF_000327365.2, GCF_000208745.1, GCF_000001735.4, and GCF_000150535.2, respectively. The sequences from single-copy orthologous groups were aligned with MUSCLE software (Edgar, 2004; Edgar, 2021; https://github.com/rcedgar/muscle/). The alignments were further processed by trimming gap-rich regions with trimAl (Capella-Gutiérrez et al., 2009) using the automated1 heuristic method and concatenating with catsequences (https://github.com/ChrisCreevey/catsequences). The final concatenated alignment was subjected to ModelTest-NG (Darriba et al., 2020) for identifying the substitution model of each alignment block. The RAxML-ng (Kozlov et al., 2019; https://github.com/amkozlov/raxml-ng) with default MRE-based bootstrapping parameter was used to construct a maximum likelihood phylogenetic tree from the concatenated alignment and substitution models. Protein family expansions/contractions were analyzed with CAFE version 5 (Mendes et al., 2020) based on the numbers of proteins in each family and phylogenetic tree. Possible functions of protein families were annotated based on Gene Ontology (GO) and MapMan4 function classes. We performed sequence homology searches between the representative sequence (the longest protein sequence) of each family with the MapMan4 bins (Schwacke et al., 2019) using Mercator4 version 5.0 (https://plabipd.de/portal/mercator4) and the NCBI non-redundant protein sequences (nr) databases (https://ftp.ncbi.nlm.nih.gov/blast/db/) using Blast2GO (Conesa et al., 2005).

2.6 Presence/absence variation analysis

We used the EUPAN pipeline (Hu et al., 2017) to identify PAV profiles from annotated gene contents of the genomes. The following steps were adopted for each of the MT, KD, and PM *de novo* assemblies. The contigs were aligned to MK reference assemblies (Teh et al., 2017). The unaligned sequences were blasted against the NCBI nonredundant nucleotide (nr/nt) database (https://ftp.ncbi.nlm.nih.gov/blast/db/) to filter out contaminated sequences (we kept only sequences matched with plant sequences in the database) and redundant sequences were removed. The draft pangenome was built by combining the reference genome and a set of non-redundant sequences. We used Liftoff (Shumate and Salzberg, 2021) for the annotation by transferring the annotations of the original assemblies to the draft pangenome based on sequence identity of 90% and gene coverage of 80%. High-quality DNA reads, which were used for genome assemblies, were mapped to draft pangenome. Gene coverage and gene PAVs were calculated based on mapping results and the annotations.

For transcriptome analysis, quality RNA sequencing reads were aligned to the draft pangenome with HISAT2 and were assembled

with StringTie to get full-length transcripts (Pertea et al., 2016). We followed the get_homologues-est pipeline (Contreras-Moreira et al., 2017) to process transcripts and to get PAV profiles of durian cultivars. In brief, coding regions of transcripts were obtained using transcripts2cdsCPP.pl, and clusters of orthologous sequences were generated with get_homologues-est.pl, PAVs were calculated with compare_clusters.pl and parse_pangenome_matrix.pl, and domain enrichment was obtained by using pfam_enrich.pl script. The assembled transcripts were annotated with GO and MapMan4 function classes as mentioned above.

3 Results

3.1 Genome assembly and annotation

We obtained 128.92, 119.96, and 115.08 Gb of raw reads from the genome sequencing of KD, MT, and PM cultivars, respectively. The raw reads represented 140-157X coverage of the 738 Mb of Musang King (MK) reference assembly (Teh et al., 2017). Using 10x Genomics linked-reads library sequencing and ragtag for reference-guided scaffolding, we obtained the assemblies of 839.7 (N50 = 21.6 Mb), 762.6 (N50 = 18.3 Mb), and 821.6 (N50 = 19.0 Mb) Mb for KD, MT and PM (Table 1), respectively. The numbers of scaffolds longer than 10 Mb (30 scaffolds) were the same among the three cultivars and MK (Teh et al., 2017). The alignments of KD, MT, and PM assembly sequences to the MK reference sequence showed contiguous matches of these scaffolds (Supplementary Figure 1). A total of 93.5%, 92.6%, and 91.3% of the KD, MT, and PM assemblies, respectively, could be aligned to the MK assembly (Teh et al., 2017). The results showed that 79.8%, 79.4%, and 76.8% of the KD, MT, and PM assemblies were aligned to the MK assembly with an identity value of >50%. The low percent identity areas were found in the alignment with the repeat regions of the MK assembly. These results revealed the variations of genome sequences among these cultivars. The numbers of annotated protein-coding genes in KD, MT, and PM assemblies were 47,980, 45,705, and 44,814, respectively (Table 2), which were similar to that in MK assembly (45,335) (Teh et al., 2017). BUSCO analyses of the annotations of KD, MT, and PM assemblies showed 95.7, 92.4, and 92.7% completeness based on the odb10 embryophyta database.

TABLE 1 Genome statistics.

	Kradumthong	Puangmanee	Monthong
Raw read (Gb)	128.92	115.08	119.96
N50 (bp)	21,597,206	19,022,827	18,254,889
Total bases (Mb)	839.66	821.59	762.61
Number of Contigs	59,633	99,407	92,437
Number of Contigs >10 Mb	30	30	30
Longest contig (bases)	37,760,038	30,624,853	32,015,873
GC content (%)	33.76	33.21	33.06
Repeat (%)	62.27	60.27	61.27
Gypsy LTR (%)	33.30	30.56	30.29

The unaligned fragments of MT, KD, and PM assemblies were merged with the MK reference assembly to generate the draft pangenome of these four popular durian cultivars (Hu et al., 2017). Only sequences that were equal to or longer than 1 kb were kept for downstream analyses. The total length of this draft pangenome was 745.9 Mb (31 Mb sequences were added to the MK reference assembly). The annotations of MK, KD, MT, and PM assemblies were transferred to the pangenome based on sequence alignments (Shumate and Salzberg, 2021). KD, MT, and PM annotations mapped to the same regions as MK annotations were filtered out. A total of 77,401 annotated proteins (from 50,112 genes) were transferred to the draft pangenome. The transferred proteins included all 63,007 annotated proteins from the MK assembly and 4,994, 3,074, and 6,326 proteins from KD, MT, and PM assemblies, respectively.

3.2 The analyses of repeat regions

We obtained about 1,200 consensus sequences of repeats in MK and KD (with a mean length of 1,471 and 1,534 bases, respectively) and about 1,600 sequences in MT and PM (with a mean length of 1,031 and 1,049 bases, respectively). About 60-63% (451-518 Mb) of these assemblies were masked as repeat regions. Among all repeats, the Gypsy elements of the long terminal repeat retrotransposon class (LTR/Gypsy) occupied the largest proportion of the assemblies (30-33%). We identified LTR/Gypsy elements in all four durians, cacao (*Theobroma cacao*), and two cotton species (*Gossypium arboreum*, *Gossypium raimondii*) for the interspecies comparative analysis. The numbers of intact LTR/Gypsy in MK and KD (228 and 105 sequences) were higher than that in MT and PM (30 and 23 sequences) and cacao (88 sequences). LTR/Gypsy elements in all durians were present in a significantly lower number than those in cottons (591 sequences in *G. raimondii* and 2,402 sequences in *G. arboretum*). The insertion times of LTR/Gypsy elements (the first appearance time of these elements in genomes) were estimated to understand their evolution in genomes. The average insertion times of all LTR/Gypsy elements in MK, KD, MT, and PM were estimated to be 7.6 ± 4.7, 10.1 ± 4.7, 11.8 ± 3.8, and 11.0 ± 4.9 Mya, respectively. The alignment and phylogenetic analysis of the LTR/Gypsy elements showed that the insertion time of the elements that could be found only in durians was estimated to be 8.2 ± 2.8 Mya. We found one LTR/Gypsy group in the phylogenetic tree that contained the elements from cacao (1 sequence), durians (25 sequences), and cottons (22 sequences) assemblies, and their insertion times were 21.9, 13.4 ± 3.7, and 5.0 ± 3.8 Mya, respectively (Supplementary Figure 2). The results together suggested that the amplification of LTR/Gypsy elements was most active in cottons. Among durian cultivars, the amplification was most active in MK.

TABLE 2 Annotation statistics.

	Kradumthong	Puangmanee	Monthong
Number of predicted gene model	47,980	44,814	45,705
Total gene length (Mb)	136.68	132.40	130.25
Average gene size (nt)	2849	2954	2850
Average number of exons/genes	5.285	5.333	5.274
Average exon length (nt)	228.4	225.3	229.2
Average number of introns/genes	4.355	4.378	4.368
Average intron length (nt)	406.8	416.2	405.8
BUSCO (% of embryophyta core genes)	95.7	92.7	92.4

3.3 Comparative analysis

We identified protein families by comparing the protein sequences of all durians with those of cottons (*G. arboretum* and *G. raimondii*), cacao (*T. cacao*), Arabidopsis (*Arabidopsis thaliana*) and papaya (*Carica papaya*). Cottons and cacao were representatives of Malvales species, while Arabidopsis and papaya were Brassicales species and were considered outgroups. We built a maximum likelihood phylogenetic tree from the sequences in 314 single-copy protein families. The tree showed that durians and cottons formed a monophyletic clade that was split from cacao (Figure 1A). Within the durian group, MK was isolated from other durians and, for the group of durians of Thailand, MT was separated from PM and KD.

The comparative analysis showed that 31,184 protein families had proteins from at least one durian cultivar (Figure 1B). Comparison among durian cultivars showed that about 1-2% of the families with durian proteins were cultivar specific (Figure 1C). Although both durian and cotton were in the clade that split from cacao (Figure 1A), the number of durian-cacao specific families (the families commonly found only in durians and cacao, 622 families) was about five times larger than that of cotton-cacao specific families (128 families). These results suggested that the rates of protein evolution in durian and cotton were different. We identified expanded and contracted families for every internal and leaf node within the phylogenetic tree to see the evolution of protein families (Figure 1A and Supplementary Table 1). The analysis revealed 401 expanded and 143 contracted families in the durian common ancestor (Figure 1A) when compared to the families in the common ancestor of cotton and durian (Figure 1A). Functions of protein families were annotated and classified based on MapMan function classes (Schwacke et al., 2019). The results showed that the families of proteins involved in cell division were expanded in the durian common ancestor. We found high proportions of both expanded and contracted families involved in transcriptional regulation and protein phosphorylation (Figure 2). The proportion of expanded families was higher than the number of contracted families for the proteins involved in, for example, ribosome biogenesis and pathogen response. On the other hand, the proportion of contracted families was higher than that of the expanded families for the proteins involved in solute transport and pectin metabolism.

For four durian leaf nodes, the analysis revealed the expansion of 1,617 protein families (the families found in multiple durian cultivars were counted only once) and the contraction of 1,330 families (Figure 1A). We also classified these families based on MapMan function classes and found several classes that contained both contracted and expanded families from the same cultivar. We selected the top 30 function classes with the highest difference between the numbers of expanded and contracted families to see highly adaptive protein functions in durians (Figure 3). Like those of the durian common ancestor, most of the rapidly evolved protein families in each durian cultivar were involved in transcriptional regulation and protein phosphorylation (Figure 3). These proteins were also involved in the responses of plants to abiotic and biotic stresses, the sensing of light quality, and the regulation of plant growth and organ development. The number of protein members of these two and most of the other families were the highest in MK and declined in KD, MT, and PM, respectively. The adaptability of some families was lesser in some durian cultivars than that in other cultivars. For example, the families of proteins related to pathogen response were present in the list of the top 30 most adaptive functions of all cultivars but not in the list of MT, indicating that these families were less adaptive in MT. The families involved in pectin metabolism and the circadian clock system were less adaptive in KD and PM, respectively. Some other families were highly adaptive in only one cultivar. The examples were families involved in nucleus protein translocation in MK, light response in MT, s-glutathionylation protein modification in KD, and sucrose metabolism in PM.

3.4 Presence/absence variations

We analyzed the gene presence-absence variations (PAVs) among cultivars based on the annotation of the draft pangenome. The analysis showed that 49,631 genes or about 99% of the total genes were present in all analyzed cultivars (Supplementary Figure 3). We found 414 genes that were absent in one to three cultivars. These genes were referred to as PAV genes. To reduce redundancy, sequences with 40% identity and 30% coverage to their longer homologous sequences were filtered out. The putative functions of PAV genes were annotated based on the homology search against the NCBI database. The PAV sequences matched with the transposon-related functions and uncharacterized genes were removed. We searched for the homologs of the PAV genes in the protein family analysis results. The PAV genes that had

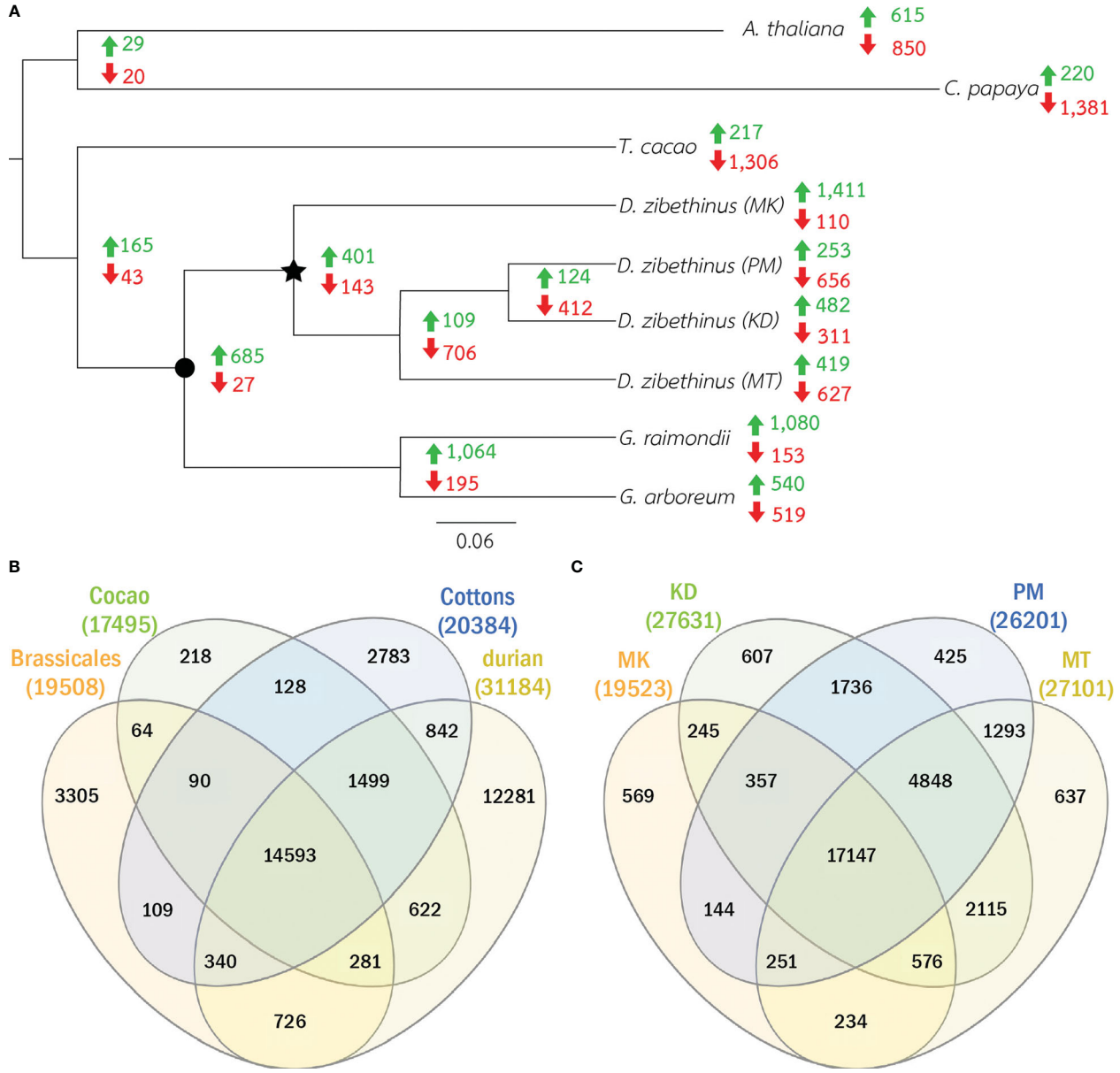


FIGURE 1
Phylogenetic tree and the numbers protein families. **(A)** The phylogenetic relations of four durian cultivars and other five related species are shown. Star node represents common ancestor of all durians and circle node represents the common ancestor of cottons and durians. Green and red numbers at nodes and leaves showed the numbers of expanded and contracted families. **(B)** The Venn diagram of the numbers of protein families in durians, cottons, cacao and Brassicales species (Arabidopsis and papaya) is shown. **(C)** The Venn diagram of the numbers of protein families in all four durian cultivars is shown.

homologs in the same genome were discarded. Some PAV genes were not assigned to any protein family due to the uniqueness of their sequences. This group of PAV genes was kept for further analysis.

We finally obtained 39 PAV genes (Table 3). Among them, seven PAV genes were simultaneously absent in more than one cultivar, while other 32 PAV genes were absent in a cultivar-specific manner. The numbers of PAV genes that were specifically absent in MK, MT, KD, and PM were nine, eleven, five, and seven, respectively (Figure 4). For each cultivar, the homologs of each of its PAV genes were searched in the genomes of the other three cultivars based on the protein family analysis results. These homologs were referred to as PAV homologs. We counted the copy number of PAV homologs and found that the PAV genes in MK had the highest number of homologs (Figure 4). For example, the cytochrome P450 78A7-like encoding gene was the PAV gene in MK and the copy numbers of the homolog of this gene in KD, PM, and MT were twenty-one, five, and two, respectively (Table 3). Other PAV genes in MK with similar PAV homolog profiles were UBN2 domain-containing protein and microtubule-associated

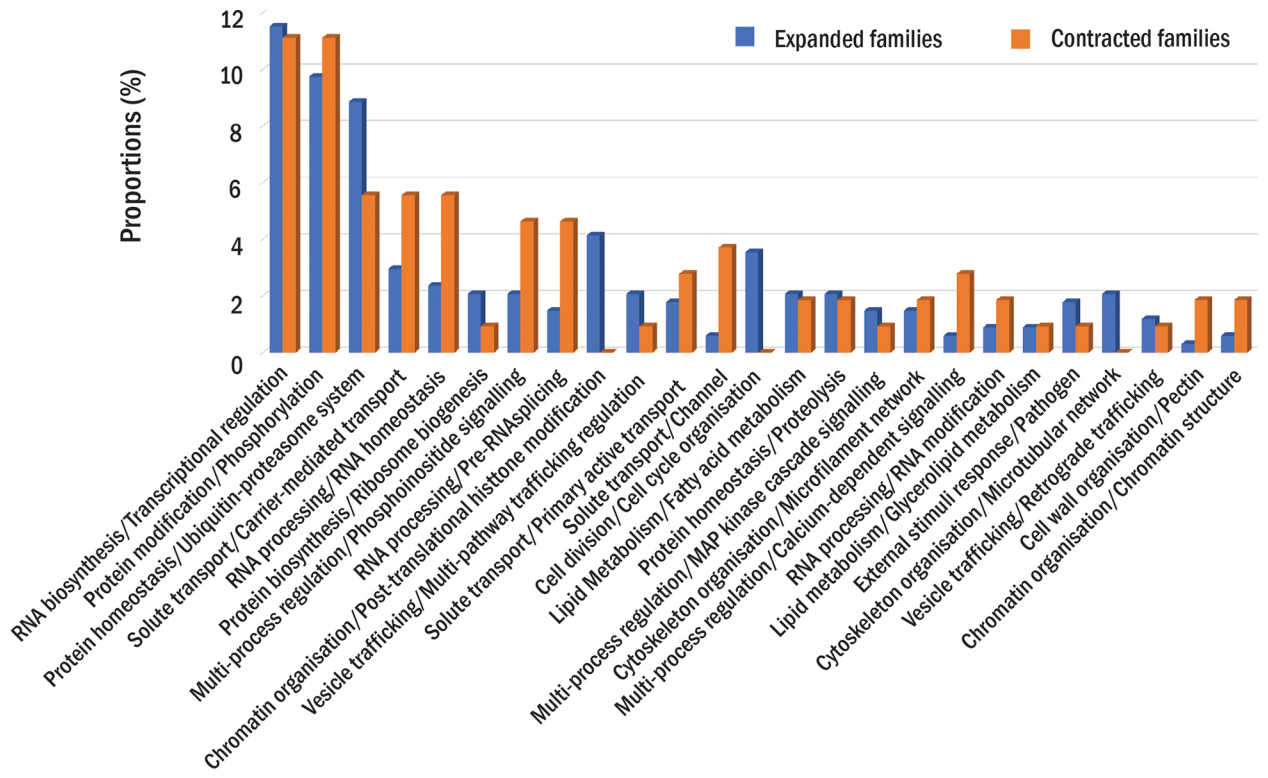


FIGURE 2
Protein families in the durian ancestor. The proportions of expanded (blue) or contracted (orange) families in each function class are shown. The proportions are calculated as the percentage of the number expanded or contracted families in each function class to the total number of expanded or contracted families.

protein TORTIFOLIA coding genes. The total copy number of the PAV homologs in MK (75 copies) was significantly higher than those in MT, KD, or PM (Figure 4). These results suggested that the MK genome might evolve differentially from the MT, KD, and PM genomes.

The number and function of the PAV genes also varied among Thai durian cultivars (Figure 4 and Table 3). The total number of PAV genes was highest in MT, followed by those in PM and KD, respectively. Despite their high number, the PAV genes in MT had a low copy number of PAV homologs (Figure 4). Additionally, several genes missing in MT were potentially involved in defense response. Examples of these PAV genes were probable disease resistance protein At1g61300, probable disease resistance protein At1g12280, NB-ARC domain-containing disease resistance protein, and lysine histidine transporter 1 encoding genes (Table 3). The results indicated that the PAV genes in MT were present as accessory genes in other cultivars. For PM and KD, the numbers of PAV genes in these two cultivars were similar and were lower than that in MT (Figure 4). The numbers of their PAV homologs appeared to be lower than that in MK. Additionally, most of the genes absent in KD and PM were involved in the metabolism of biomolecules, which differed from the functions of genes absent in MT (Table 3). These results together showed three different PAV profiles in MK, MT, and the group of KD and PM, which was related to the phylogenetic relationship and origins of these durian cultivars.

3.5 Pantranscriptome

We identified the PAVs at the gene expression level. In addition to KD, MT, PM, and MK durians, Salika (SK) cultivar was included in this analysis. The leaf samples of KD, MT, PM, and SK were obtained from plants grown under the same environment in a small cultivation plot (the distance between plants was about 10 meters) at the Chanthaburi horticultural research center. MK was considered as an outgroup in this analysis as this cultivar grew under a different environment. The analysis of homology-based pantranscriptome analysis showed a total of 48,779 transcript orthologous groups. We found that 41.3% of these orthologous groups contained transcripts of all five cultivars (core orthologous groups). The domain enrichment analysis (based on the protein sequences translated from transcript sequences) using all orthologous groups as background showed the enrichment of helicase conserved C-terminal domain, pentatricopeptide repeat domain, and F-box domain in the core group (Supplementary Table 2). A majority of sequences with helicase conserved C-terminal domain were members of the DEAD-box ATP-dependent RNA helicase family. We found that 32.3% of all orthologous groups contained transcripts from single cultivars, including 9.5%, 6.4%, 5.6%, 5.4%, and 5.4% from MK, SK, MT, PM, and KD, respectively (Figure 5). No enriched domains were found for each of these groups.

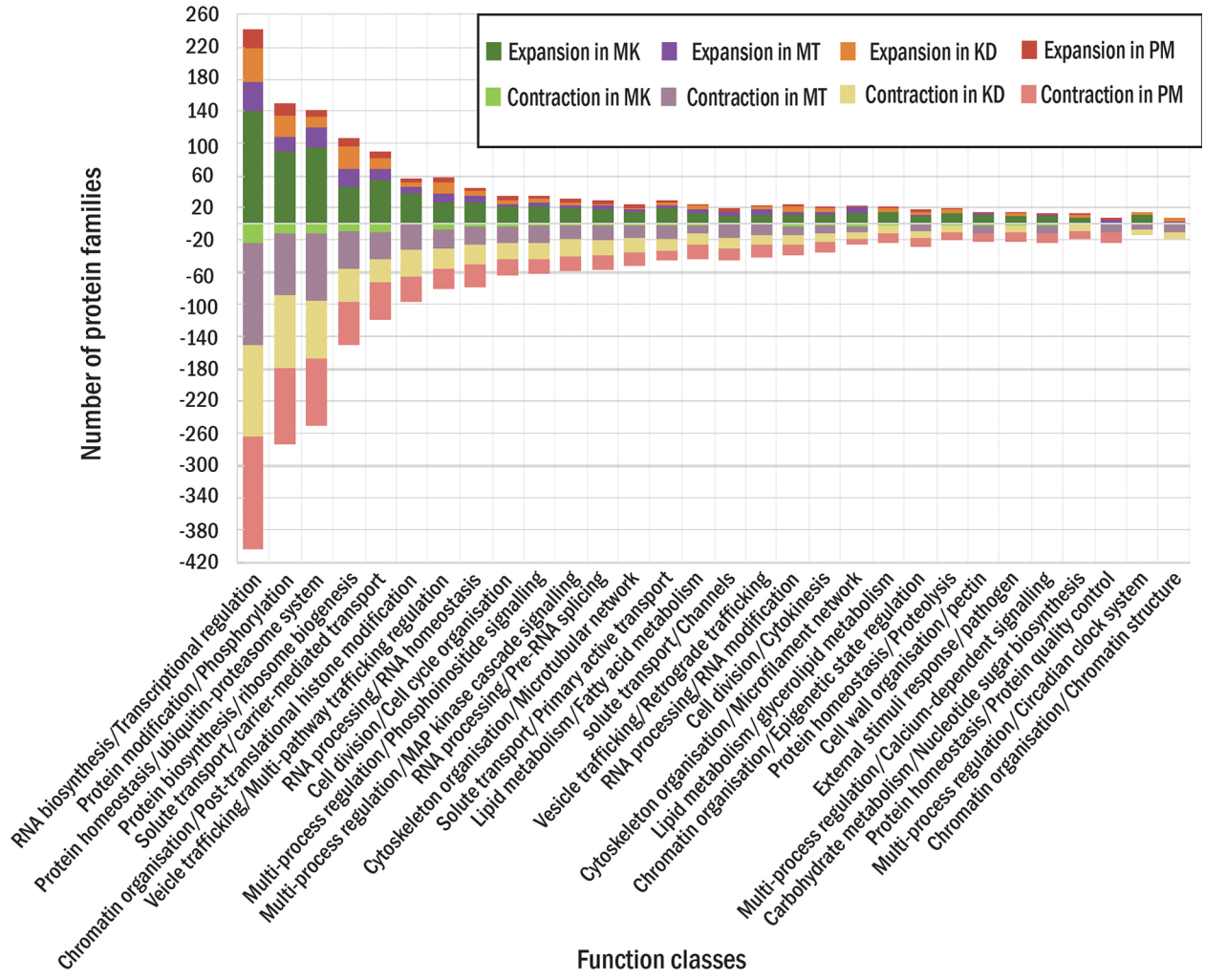


FIGURE 3
Protein families in each durian cultivars. The numbers of expanded (positive side) and contracted (negative side) families in each function class for MK, MT, KD and PM are shown in different colors.

TABLE 3 List of selected PAV genes and the numbers of proteins in the associated protein families.

	PAVs*				Protein Numbers			
Annotations	MK	MT	KD	PM	MK	MT	KD	PM
cytochrome P450 78A7-like	0	1	1	1	0	2	21	5
cytochrome P450 CYP736A12-like	0	1	1	1	0	0	1	1
MDIS1-interacting receptor like kinase 2-like	0	1	1	1	0	0	0	2
microtubule-associated protein TORTIFOLIA1 isoform X1	0	1	1	1	0	1	2	4
non-functional NADPH-dependent codeinone reductase 2-like	0	1	1	1	0	2	5	1
probable disease resistance protein At1g12280	0	1	1	1	–	–	–	–
putative sel repeat-containing protein L21	0	1	1	1	0	0	1	2
putative wall-associated receptor kinase-like 16	0	1	1	1	0	1	1	1
UBN2 domain-containing protein	0	1	1	1	0	6	9	5

(Continued)

10.3389/fpls.2023.1137077

TABLE 3 Continued

	PAVs*				Protein Numbers			
Annotations	MK	MT	KD	PM	MK	MT	KD	PM
lysine histidine transporter 1	1	0	1	1	–	–	–	–
NADH-plastoquinone oxidoreductase subunit 7	1	0	1	1	–	–	–	–
NB-ARC domain-containing disease resistance protein, putative	1	0	1	1	1	0	0	0
octapeptide-repeat protein T2-like	1	0	1	1	–	–	–	–
PLAC8 family protein	1	0	1	1	–	–	–	–
probable disease resistance protein At1g12280 isoform X2	1	0	1	1	3	0	0	0
putative F-box only protein 10	1	0	1	1	–	–	–	–
receptor-like protein 12	1	0	1	1	–	–	–	–
receptor-like protein 9DC3 isoform X2	1	0	1	1	–	–	–	–
serine/threonine-protein kinase PBL13 isoform X1	1	0	1	1	–	–	–	–
SKP1-like protein 1A	1	0	1	1	–	–	–	–
cycloartenol-C-24-methyltransferase	1	1	0	1	0	1	0	1
F3H9.20 protein	1	1	0	1	1	1	0	0
glutamate dehydrogenase 1	1	1	0	1	–	–	–	–
glycine-rich cell wall structural protein 1-like	1	1	0	1	4	4	0	4
G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	1	1	0	1	–	–	–	–
arginase 1, mitochondrial	1	1	1	0	–	–	–	–
cytochrome P450 81D1-like	1	1	1	0	1	1	0	0
cytosolic sulfotransferase 15-like	1	1	1	0	1	0	1	0
transcription factor MYB16-like	1	1	1	0	–	–	–	–
tryptophan synthase beta chain 1	1	1	1	0	–	–	–	–
tyrosine-protein phosphatase RLPH2	1	1	1	0	–	–	–	–
zinc finger BED domain-containing protein RICESLEEPER 2-like	1	1	1	0	2	0	1	0
beta-glucuronosyltransferase GlcAT14A-like isoform X1	0	0	1	1	–	–	–	–
serpin-ZX-like	0	0	1	1	0	0	1	1
putative DNA (cytosine-5)-methyltransferase CMT1	1	0	0	1	1	0	0	0
zinc finger CCCH domain-containing protein 1-like	1	0	0	1	1	0	0	1
glycoside hydrolase family 2 protein	1	0	1	0	–	–	–	–
Histone acetyltransferase MCC1	1	0	1	0	1	0	1	0
glutamate dehydrogenase 1	1	1	0	0	–	–	–	–

*0 = presence, 1 = absence

The transcripts in the other 58.7% orthologous groups were not expressed in at least one cultivar (Figure 5). These orthologous groups were referred to as accessory orthologous groups. The number of accessory groups in MT (25.9% of the total orthologous groups) was higher than that in KD (22.8%), PM (22.1%), SK (21.4%), and MK (20.2%), respectively. Some domains could be enriched in multiple cultivars because some accessory groups were shared by multiple cultivars (Table 4). The domains enriched in the accessory groups of four cultivars included multicopper oxidase domain (unenriched in MK and enriched in all other cultivars), no apical meristem (NAM) protein (unenriched in PM), and late embryogenesis abundant protein (unenriched in SK) (Supplementary Table 2). Examples of enriched domains in three cultivars were the thaumatin family (thaumatin-like protein) in MT, KD, and PM, probable lipid transfer (xylogen-like protein) in MT, PM and SK, and SRF-type transcription factor (agamous-like MADS-box protein) in MK, PM, and SK. Some domains enriched in two cultivars were AP2 (ethylene-responsive transcription factor), xylanase inhibitor N-terminal (aspartyl protease family protein), gibberellin regulated protein (GAST1 protein homolog

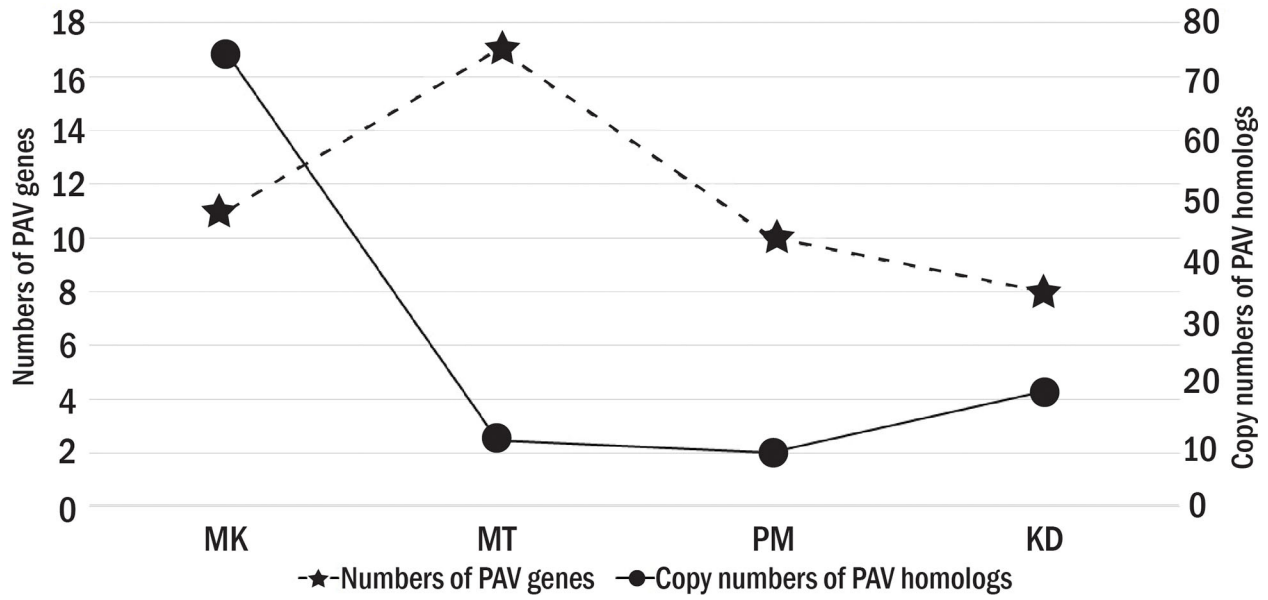


FIGURE 4
The number of PAV genes and PAV homologs. The primary y axis shows the numbers of PAV genes with star symbols that are connected by dash lines. The secondary y axis shows the copy numbers of PAV homologs with circle symbols that are connected by solid lines.

and Snakin-1), and pectinesterase (probable pectinesterase) domains in MT and PM, wall-associated receptor kinase galacturonan-binding (rust resistance kinase Lr10) domain in MT and KD, and NB-ARC (putative disease resistance families) domain in KD and PM. Finally, the domains enriched in single cultivar included, X8, peroxidase, plant invertase/pectin methylesterase inhibitor, C2H2-type zinc finger, and epidermal patterning factor protein domains in MT; d-mannose binding lectin, s-locus glycoprotein, and berberine like domains in MK; MYB-like DNA-binding domain in SK; and auxin-responsive protein, transcriptional repressor-ovate, cotton fiber expressed protein, and sulfotransferase domain in PM.

4 Discussion

Durian exhibited high genetic diversities because they are generally highly outcrossed during cultivation (Teh et al., 2017; Mursyidin et al., 2022). In this study, we sequenced the genomes of three popular durian cultivars of Thailand (Kradumthong, Monthong, and Puangmanee) and comparatively analyzed them with the genome of the Malaysian durian (Musang King) and other related species. The aim was to identify the similarities and differences among durian genomes to understand their evolution and diversity.

Assembly sequences of MT, KD, and PM were aligned with MK reference sequences to identify their conserved sequences. The alignments showed 30 conserved scaffolds among four durian genomes. The chromosome numbers of *Durio* species varied between 2n = 54 and 69, and the chromosome number of *D. zibethinus* was 2n = 56 (IBPGR, 1986). The number of conserved scaffolds was close to the estimated haploid chromosome number. A set of nonredundant DNA fragments was collected using the map-to-pan strategy (Hu et al., 2017) to generate the draft pangenome of all four popular durian cultivars. Size of the draft pangenome was increased by only 4% compared to size of the MK reference assembly because genome sequences were highly similar among four durians. Correspondingly, in sense of the genome annotations, the protein family analysis showed that only 1-2% of the families were unique to each durian cultivar.

The evolution of durian genomes among Malvales species was investigated by comparing protein families and LTR/Gypsy repeat elements in durians with those in cottons and cacao. In the phylogenetic tree, durians and cottons were in a monophyletic clade that was split from cacao (Figure 1A). The insertion timelines of LTR/Gypsy sequences in these genomes were consistent with this placement, *i.e.*, the repeats were the oldest in cacao, followed by those in durians and cottons, respectively. We also found that the number of protein families that cacao specifically shared with durians was higher than that it specifically shared with cotton (Figure 1B). These results indicated that changes in durian proteins after the split from the cacao-durian ancestor were lower than those in cotton proteins. The number of intact LTR/Gypsy sequences also suggested that the amplification of these elements was less active in durian genomes than that in cotton genomes. The results were consistence with the finding that the evolution rate of the durian genome was significantly slower than that of cotton genomes (Wang et al., 2019).

The expansion and contraction of protein families in the common ancestors of durians and cottons were compared to see the difference in their evolution paths. Several rapidly evolved families in durians were involved in transcription, protein phosphorylation, and protein ubiquitination processes. The rapid evolutions of these families have been linked to their functions,

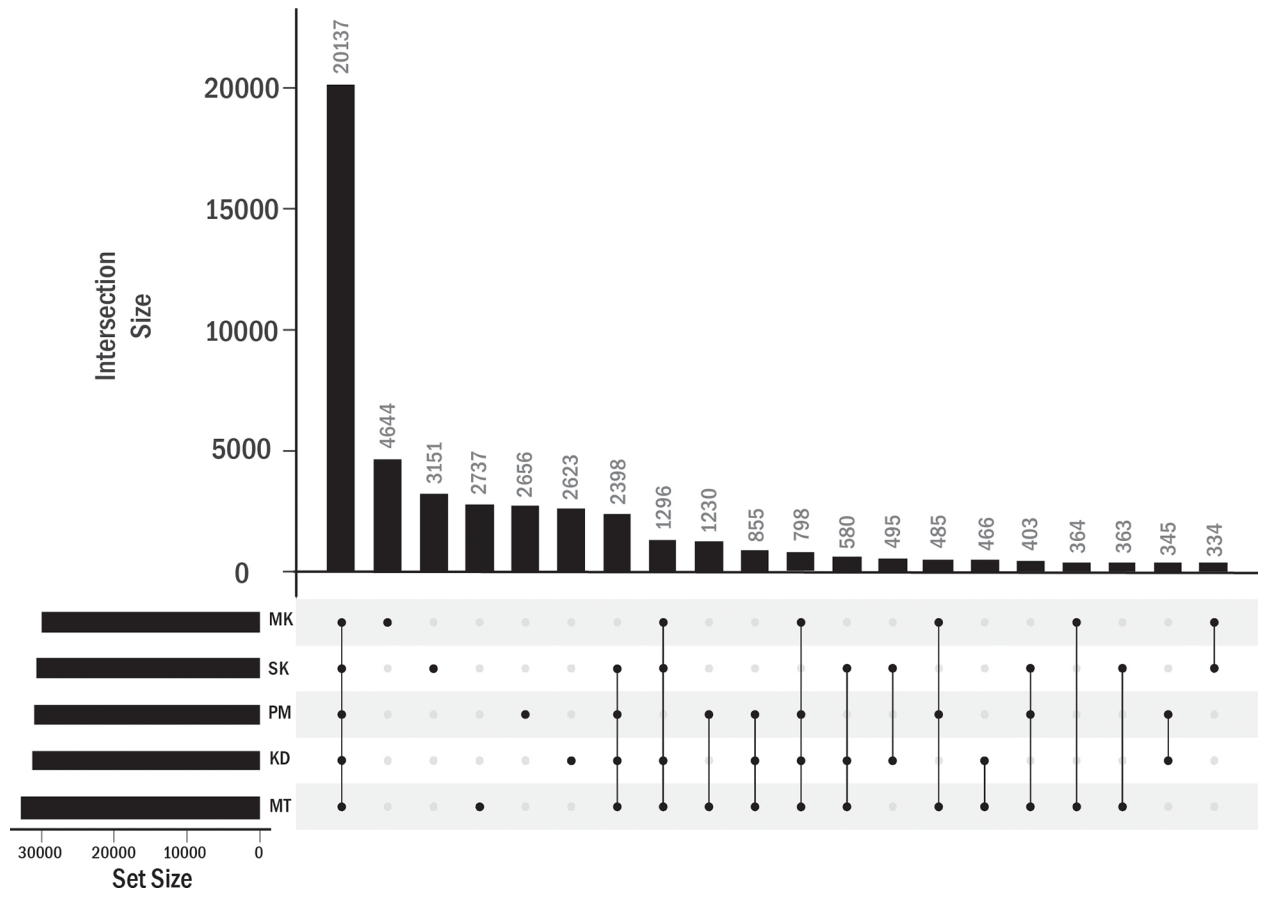


FIGURE 5
Transcript orthologous groups in each durian cultivars. The number of orthologous groups in each durian cultivars is shown by the horizontal bar. Vertical bars showed the numbers of orthologous groups found in single (single dot in column) cultivar or shared by multiple cultivars (connected dots).

which were involved in the interactions of an organism with environments (Demuth and Hahn, 2009). For example, we found several expanded families of transcription factors (TF) that were involved in the responses of plants to pathogens and other environmental stimuli. These TFs included transcription factor MYC2 (*MYC2*) family, which could play roles in abiotic and biotic stress responses and the circadian clock (Kazan and Manners, 2013), probable WRKY transcription factor 53 family (*WRKY53*), which played roles in JA signaling, leaf senescence (Miao and Zentgraf, 2007) and the basal resistance against *Pseudomonas syringae* (Hu et al., 2012), and probable WRKY transcription factor 33 (*WRKY33*), which could confer resistance to fungal pathogens *Botrytis cinerea* and *Alternaria brassicicola* (Zheng et al., 2006). These families were expanded (*MYC2*), contracted (*WRKY33*), or unchanged (*WRKY53*) in the common ancestor of cotton. Other protein families related to pathogen responses and cell cycle were also highly expanded in durians. The results suggested that the evolutions of the proteins implicated in the responses of plants to environmental stimuli, especially pathogen infections, were different between durian and cotton. The expansion of these families might be related to the demand for a higher dosage of defense-responsive genes in durians to survive in high rainfall regions compared to cottons that were cultivated in arid to semiarid regions of the tropics and subtropics (Demuth and Hahn, 2009).

In this study, we found several similarities and differences among durian genomes. The phylogenetic tree and the profiles of CNV and PAV showed that MK was most different from all other cultivars (Figures 1, 3 and Table 3). Among the durians of Thailand, MT was isolated from KD and PM in the phylogenetic tree (Figure 1A). PAV and CNV analysis results showed that pathogen-responsive genes were one of the gene groups that exhibited high variations among durian cultivars. For example, we found a higher copy number of putative disease resistance RPP13-like protein 1, which conferred resistance to downy mildew caused by *Peronospora parasitica* (Bittner-Eddy et al., 2000), in MK than that in other cultivars. In contrast, the cysteine-rich receptor-like protein kinase 8 (*CRK8*) family, which could confer resistance to *P. syringae* in cottons (Hussain et al., 2022), was significantly contracted in MK but expand in KD and MT. Correspondingly, the resistance levels against pathogen infections have been shown to vary among durian cultivars (Vawdrey et al., 2005). The diversification of disease-resistance genes could be occurred not only after speciation but also after

10.3389/fpls.2023.1137077

TABLE 4 List of selected enriched domains .

PfamID	Description	Enrichment*				
		KD	MT	MK	PM	SK
PF00394	Multicopper oxidase	1	1	0	1	1
PF02365	No apical meristem (NAM) protein	1	1	1	0	1
PF03168	Late embryogenesis abundant protein	1	1	1	1	0
PF00560	Leucine Rich Repeat	1	0	1	1	0
PF00314	Thaumatin family	1	1	0	1	0
PF02984	Cyclin, C-terminal domain	1	1	0	0	1
PF14368	Probable lipid transfer	0	1	0	1	1
PF00319	SRF-type transcription factor	0	0	1	1	1
PF13947	Wall-associated receptor kinase galacturonan-binding	1	1	0	0	0
PF00931	NB-ARC domain	1	0	0	1	0
PF00847	AP2 domain	0	1	0	1	0
PF14543	Xylanase inhibitor N-terminal	0	1	0	1	0
PF02704	Gibberellin regulated protein	0	1	0	1	0
PF00190	Cupin	0	1	0	1	0
PF03188	Eukaryotic cytochrome b561	0	1	0	1	0
PF05910	Plant protein of unknown function (DUF868)	0	1	0	1	0
PF01095	Pectinesterase	0	1	0	1	0
PF00786	P21-Rho-binding domain	0	1	0	1	0
PF00462	Glutaredoxin	0	1	0	1	0
PF07983	X8 domain	0	1	0	0	0
PF00141	Peroxidase	0	1	0	0	0
PF04043	Plant invertase/pectin methylesterase inhibitor	0	1	0	0	0
PF13912	C2H2-type zinc finger	0	1	0	0	0
PF17181	Epidermal patterning factor proteins	0	1	0	0	0
PF02519	Auxin responsive protein	0	0	0	1	0
PF04844	Transcriptional repressor, ovate	0	0	0	1	0
PF02469	Fasciclin domain	0	0	0	1	0
PF05553	Cotton fibre expressed protein	0	0	0	1	0
PF00332	Glycosyl hydrolases family 17	0	0	0	1	0
PF00685	Sulfotransferase domain	0	0	0	1	0
PF00249	Myb-like DNA-binding domain	0	0	0	0	1
PF01453	D-mannose binding lectin	0	0	1	0	0
PF00954	S-locus glycoprotein domain	0	0	1	0	0
PF08031	Berberine and berberine like	0	0	1	0	0

*0 = unenriched, 1 = enriched

the divergence within species (Kim et al., 2021). Some redundant resistance genes might be deleted after whole-genome duplication events, while some other resistance genes might be retained related to the presence of particular pathogen pressures (Golicz et al., 2016; Rijzaani et al., 2022). The variation of gene contents in cottons could also be associated with geographic disjunction (Grover et al., 2017). For durian, MK was popularly grown in Malaysia, while MT was originally grown in the southern region of Thailand and PM

and KD were originally grown in the central region of Thailand. The variation of gene contents among durian cultivars might be linked to the differences in their cultivation areas and breeding programs.

The expression of genes involved in flower formation and fruit ripening varied among durian cultivars. From the transcriptome analysis, we found the enrichment of plant invertase/pectin methylesterase inhibitor (PMEI) domain in MT. PMEI played a role in flower formation, fruit development, and biotic stress responses (Coculo and Lionetti, 2022). At the time that we collected samples, MT, KD, PM, and SK were in the flowering stage. Young fruit setting was also detected. The detection of transcripts with the PMEI domain might be related to flowering. The duration of fruit maturity varied among durian cultivars (Somsri, 2014). The growth and development period from anthesis to maturity of MT fruit was 120-127 days, which was longer than 95-100 days in KD (Sangwanangkul and Siriphanich, 2000; Somsri, 2014). Different maturity periods might be related to different levels of the transcript with the PMEI domain. Another gene whose expression was associated with fruit ripening in durian was the aminocyclopropane-1-carboxylic acid synthase encoding gene (*ACS*) (Teh et al., 2017). The expressions of *ACS* appeared to be similar among durian cultivars (Teh et al., 2017). In this study, the ACS protein family was shown to be contracted in papaya and cacao and expanded in cottons and durians, which was consistent with the study in MK (Teh et al., 2017). The copy numbers of this gene differed slightly among durian cultivars and cottons. In contrast to *ACS* that was a key ethylene-production enzyme in durian ripening (Liu et al., 2015; Teh et al., 2017), the expression of PMEI containing genes was not directly induced by ethylene (Srivastava et al., 2012). The role of the PMEI domain in durian maturity could be of interest to further research.

5 Conclusion

In this study, we resequenced the gnomes of three popular and agronomically different durian cultivars in Thailand (MT, KD, and PM) and comparatively analyzed them with the genomes of Malaysian durian (MK) and other related species to understand their genetic diversity. We found slower evolution rates of protein-coding genes and repeat elements in durian genomes compared to those in cotton genomes. Among durian cultivars, the highest expanded protein families in the Malaysian durian cultivar, which was different from Thai durian cultivars as shown in the phylogenetic tree. Among Thai durian cultivars, MT was most different from the other two Thai cultivars. The families of proteins involved in pathogen response in MT were less adaptive than those in KD and PM cultivars. We also observed that the number of homologs missing in MK but present in MT, KD, or PM genomes was higher than the number of homologs missing in MT, KD, or PM but present in the MK genome. The PAV analysis also showed that the missing genes in MT were involved in pathogen response, while the missing genes in KD and PM were involved in the metabolism of biomolecules. Additional analysis showed a higher abundance of transcripts with PMEI domain in MT than those in other cultivars, which was of interest to further test if the expression of the genes of these transcripts was involved in durian fruit ripening. Our results demonstrated genetic variations among the selected durian cultivars, which yielded arils with different flavors and textures and had different disease resistance levels. It was likely that these four durian cultivars were developed from different origins. In this study, we reported the draft version of the pangenome generated from the high-quality assemblies of four famous durian cultivars. For further works, commercial, local, and wild durian cultivars should be considered to construct a more complete version of the durian pangenome.

Data availability statement

The datasets presented in this study can be found in online repositories. The genome sequences and RNA reads presented in the study are deposited in the NCBI database (https://www.ncbi.nlm.nih.gov/), accession number PRJNA641506 and PRJNA909034.

Author contributions

WN performed bioinformatics analyses and wrote the manuscript. CN and SC performed bioinformatics analyses. SU-T and NN extracted DNA and RNA and performed sequencing. OC maintained durian plants and collected leaf samples. ST and WP initiated the project and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Science and Technology Development Agency (NSTDA), Thailand, grant number: P-19-51818.

Acknowledgments

We are grateful to the National Science and Technology Development Agency (NSTDA), Thailand for facilities. We thank the team at Chanthaburi Horticultural Research Center for fieldwork and sample collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1137077/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Whole genome sequence alignments. **(A-C)** Dot pots of the alignments of **(A)** MT, **(B)** KD and **(C)** PM assembly sequences (on vertical axis) to the MK reference sequence (on horizontal axis) are shown. The alignment regions with the identity of >80% are in green color, while those with the identity of <80% are in orange color.

SUPPLEMENTARY FIGURE 2
The phylogenetic tree of LTR/Gypsy elements in the studied Malvales genomes. The LTR/Gypsy elements of cacao (CO), cottons (GA = *G. arboreum* and GR = *G. raimondii*) and durians are shown in orange, pink and blue colors.

SUPPLEMENTARY FIGURE 3
The numbers of DNA mapped genes. The Venn diagram showed the numbers of genes on the draft pangenome that are mapped by DNA reads of each durian cultivar.

References

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., and Edwards, D. (2020). Plant pan-genomes are the new reference. *Nat. Plants* 6, 914–920. doi: 10.1038/s41477-020-0733-0

Bittner-Eddy, P. D., Crute, I. R., Holub, E. B., and Beynon, J. L. (2000). RPP13 is a simple locus in arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in peronospora parasitica. *Plant J.* 21, 177–188. doi: 10.1046/j.1365-313x.2000.00664.x

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Coculo, D., and Lionetti, V. (2022)The plant Invertase/Pectin methylesterase inhibitor superfamily (Accessed November 24, 2022).

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Contreras-Moreira, B., Cantalapiedra, C. P., García-Pereira, M. J., Gordon, S. P., Vogel, J. P., Igartua, E., et al. (2017)Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species (Accessed November 26, 2022).

Cortaga, C. Q., Latina, R. A., Habunal, R. R., and Lantican, D. V. (2022). Identification and characterization of genome-wide resistance gene analogs (RGAs) of durian (Durio zibethinus l.). *J. Genet. Eng. Biotechnol.* 20, 29. doi: 10.1186/s43141-022-00313-8

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E., and Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* 54, 18–25. doi: 10.1016/j.pbi.2019.12.005

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., and Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biol.* 22, 3. doi: 10.1186/s13059-020-02224-8

Demuth, J. P., and Hahn, M. W. (2009). The life and death of gene families. *BioEssays* 31, 29–39. doi: 10.1002/bies.080085

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* 5, 113. doi: 10.1186/1471-2105-5-113

Edgar, R. C. (2021). MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv*. 06, 20. doi: 10.1101/2021.06.20.449169

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinf.* 9, 18. doi: 10.1186/1471-2105-9-18

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117, 9451–9457. doi: 10.1073/pnas.1921046117

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016). The pangenome of an agronomically important crop plant brassica oleracea. *Nat. Commun.* 7, 13390. doi: 10.1038/ncomms13390

Grover, C. E., Arick, M. A., Conover, J. L., Thrash, A., Hu, G., Sanders, W. S., et al. (2017). Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (Gossypieae) yields insights into genome downsizing. *Genome Biol. Evol.* 9, 3328–3344. doi: 10.1093/gbe/evx248

Haberer, G., Kamal, N., Bauer, E., Gundlach, H., Fischer, I., Seidel, M. A., et al. (2020). European Maize genomes highlight intraspecies variation in repeat and gene content. *Nat. Genet.* 52, 950–957. doi: 10.1038/s41588-020-0671-9

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12, 491. doi: 10.1186/1471-2105-12-491

Hu, Y., Dong, Q., and Yu, D. (2012). Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen pseudomonas syringae. *Plant Sci.* 185–186, 288–297. doi: 10.1016/j.plantsci.2011.12.003

Hu, Z., Sun, C., Lu, K., Chu, X., Zhao, Y., Lu, J., et al. (2017). EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics* 33, 2408–2409. doi: 10.1093/bioinformatics/btx170

Hu, Z., Wei, C., and Li, Z. (2020). "Computational strategies for eukaryotic pangenome analyses," in *The pangenome: Diversity, dynamics and evolution of genomes*. Eds. H. Tettelin and D. Medini (Cham, Switzerland: Springer International Publishing), 293–307. doi: 10.1007/978-3-030-38281-0_13

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0

Hussain, A., Asif, N., Pirzada, A. R., Noureen, A., Shaukat, J., Burhan, A., et al. (2022). Genome wide study of cysteine rich receptor like proteins in gossypium sp. *Sci. Rep.* 12, 4885. doi: 10.1038/s41598-022-08943-1

IBPGR (1986). *Genetic resources of tropical & sub tropical fruits & nuts(excluding musa)* (Rome: IBPGR).

Kazan, K., and Manners, J. M. (2013). MYC2: The master in action. *Mol. Plant* 6, 686–703. doi: 10.1093/mp/sss128

Ketsa, S., Wisutiamonkul, A., Palapol, Y., and Paull, R. E. (2020). "The durian," in *Horticultural reviews* (Hoboken, United States: John Wiley & Sons, Ltd), 125–211. doi: 10.1002/9781119625407.ch4

Kim, M.-S., Chae, G. Y., Oh, S., Kim, J., Mang, H., Kim, S., et al. (2021). Comparative analysis of *de novo* genomes reveals dynamic intra-species divergence of NLRs in pepper. *BMC Plant Biol.* 21, 247. doi: 10.1186/s12870-021-03057-8

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5, 59. doi: 10.1186/1471-2105-5-59

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Liu, M., Pirrello, J., Chervin, C., Roustan, J.-P., and Bouzayen, M. (2015). Ethylene control of fruit ripening: Revisiting the complex network of transcriptional regulation. *Plant Physiol.* 169, 2380–2390. doi: 10.1104/pp.15.01361

Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022

Miao, Y., and Zentgraf, U. (2007). The antagonist function of arabidopsis WRKY53 and ESR/ESP in leaf senescence is modulated by the jasmonic and salicylic acid equilibrium. *Plant Cell* 19, 819–830. doi: 10.1105/tpc.106.042705

Mursyidin, D. H., Makruf, M. I., Badruzsaufari,, and Noor, A. (2022). Molecular diversity of exotic durian (Durio spp.) germplasm: A case study of kalimantan, Indonesia. *J. Genet. Eng. Biotechnol.* 20, 39. doi: 10.1186/s43141-022-00321-8

O'Gara, E., Vawdrey, L., Tania, M., Sangchote, S., van Thanh, H., Binh, L. N., et al. (2004). "Screening for resistance to phytophthora," in *Diversity and management of phytophthora in southeast Asia* (Canberra: Australian: Centre for International Agricultural Research, ACIAR Monograph), 194–199.

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Ou, S., and Jiang, N. (2019). LTR_FINDER_parallel: Parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA* 10, 48. doi: 10.1186/s13100-019-0193-0

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Rijzaani, H., Bayer, P. E., Rouard, M., Doležel, J., Batley, J., and Edwards, D. (2022). The pangenome of banana highlights differences between genera and genomes. *Plant Genome* 15, e20100. doi: 10.1002/tpg2.20100

Sangwanangkul, P., and Siriphanich, J. (2000)Growth and maturation of durian fruit cv. monthong. In: *Thai Journal of agricultural science* (Thailand) (Accessed November 25, 2022).

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12, 879–892. doi: 10.1016/j.molp.2019.01.003

Shumate, A., and Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics* 37, 1639–1643. doi: 10.1093/bioinformatics/btaa1016

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Somsri, S. (2014). Current status of durian breeding program in Thailand.. *Acta Hortic.* 1024, 51–59. doi: 10.17660/ActaHortic.2014.1024.3

Srivastava, S., Gupta, S. M., Sane, A. P., and Nath, P. (2012). Isolation and characterization of ripening related pectin methylesterase inhibitor gene from banana fruit. *Physiol. Mol. Biol. Plants* 18, 191–195. doi: 10.1007/s12298-012-0102-1

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinf.* 7, 62. doi: 10.1186/1471-2105-7-62

Teh, B. T., Lim, K., Yong, C. H., Ng, C. C. Y., Rao, S. R., Rajasegaran, V., et al. (2017). The draft genome of tropical fruit durian (Durio zibethinus). *Nat. Genet.* 49, 1633–1641. doi: 10.1038/ng.3972

Trouern-Trend, A. J., Falk, T., Zaman, S., Caballero, M., Neale, D. B., Langley, C. H., et al. (2020). Comparative genomics of six juglans species reveals disease-associated gene family contractions. *Plant J.* 102, 410–423. doi: 10.1111/tpj.14630

Vawdrey, L. L., Martin, T. M., and De Faveri, J. (2005). A detached leaf bioassay to screen durian cultivars for susceptibility to phytophthora palmivora. *Australas. Plant Pathol.* 34, 251–253. doi: 10.1071/AP05005

Wang, M., Li, J., Wang, P., Liu, F., Liu, Z., Zhao, G., et al. (2021). Comparative genome analyses highlight transposon-mediated genome expansion and the evolutionary architecture of 3D genomic folding in cotton. *Mol. Biol. Evol.* 38, 3621–3636. doi: 10.1093/molbev/msab128

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9

Wang, J., Yuan, J., Yu, J., Meng, F., Sun, P., Li, Y., et al. (2019). Recursive paleohexaploidization shaped the durian genome. *Plant Physiol.* 179, 209–219. doi: 10.1104/pp.18.00921

Zdobnov, E. M., Kuznetsov, D., Tegenfeldt, F., Manni, M., Berkeley, M., and Kriventseva, E. V. (2021). And OrthoDB in 2020: Evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 49, D389–D393. doi: 10.1093/nar/gkaa1009

Zheng, Z., Qamar, S. A., Chen, Z., and Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J.* 48, 592–605. doi: 10.1111/j.1365-313X.2006.02901.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 February 2023
DOI 10.3389/fpls.2023.1126660

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Aamir Manzoor,
Anhui Agricultural University, China

REVIEWED BY
Lin Zhang,
Hubei University of Chinese Medicine, China
Cheng Song,
West Anhui University, China

*CORRESPONDENCE
Ling-Li Wu
wulingli0307@163.com
Jian-An Li
lja0731@126.com

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 18 December 2022
ACCEPTED 31 January 2023
PUBLISHED 20 February 2023

CITATION
Wang Y-J, Wu L-L, Sun M-h, Li Z, Tan X-F and Li J-A (2023) Transcriptomic and metabolomic insights on the molecular mechanisms of flower buds in responses to cold stress in two *Camellia oleifera* cultivars. *Front. Plant Sci.* 14:1126660. doi: 10.3389/fpls.2023.1126660

COPYRIGHT
© 2023 Wang, Wu, Sun, Li, Tan and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Transcriptomic and metabolomic insights on the molecular mechanisms of flower buds in responses to cold stress in two *Camellia oleifera* cultivars

Ya-Jun Wang[1,2,3,4†], Ling-Li Wu[1,2,3,4*†], Min-hong Sun[1,2], Ze Li[1,2,3,4], Xiao-Feng Tan[1,2,3,4] and Jian-An Li[1,2,3,4*]

[1]Key Laboratory of Cultivation and Protection for Non-wood Forest Trees, Ministry of Education, and the Key Laboratory of Non-Wood Forest Products, Forestry Ministry, Central South University of Forestry and Technology, Changsha, China, [2]Engineering Technology Research Center of Southern Hilly and Mountainous Ecological Non-Wood Forest Industry of Hunan Province, Changsha, China, [3]Camellia Oil Tree Research Institute of Central South University of Forestry and Technology, Changsha, China, [4]The Belt and Road International Union Research Center for Tropical Arid Non-wood Forest in Hunan Province, Changsha, China

Introduction: The *Camellia oleifera* (*C. oleifera*) cultivars 'Huashuo' (HS) and 'Huaxin' (HX) are new high-yielding and economically valuable cultivars that frequently encounter prolonged cold weather during the flowering period, resulting in decreased yields and quality. The flower buds of HS sometimes fail to open or open incompletely under cold stress, whereas the flower buds of HX exhibit delayed opening but the flowers and fruits rarely drop.

Methods: In this study, flower buds at the same development stage of two *C. oleifera* cultivars were used as test materials for a combination of physiological, transcriptomic and metabolomic analyses, to unravel the different cold regulatory mechanisms between two cultivars of *C. oleifera*.

Results and discussion: Key differentially expressed genes (DEGs) and differentially expressed metabolites (DEMs) involved in sugar metabolism, phenylpropanoid biosynthesis, and hormone signal transduction were significantly higher in HX than in HS, which is consistent with phenotypic observations from a previous study. The results indicate that the flower buds of HX are less affected by long-term cold stress than those of HS, and that cold resistance in *C. oleifera* cultivars varies among tissues or organs.This study will provide a basis for molecular markers and molecular breeding of *C. oleifera*.

KEYWORDS

***Camellia oleifera*, cold stress, phenylpropanoid, plant hormones, starch and sucrose, metabolome, transcriptome**

1 Introduction

Camellia oleifera Abel. (*C. oleifera*) is a member of the genus *Camellia* in the family Theaceae. *C. oleifera* is one of four major woody oil tree species in China and has a wide range of applications. The main product from *C. oleifera* is tea oil, which is extracted from the seeds. Tea oil is a high-quality edible oil containing a significant proportion of unsaturated fatty acids (>90%), primarily linoleic acid and oleic acid (>80%), as well as squalene, tea polyphenols, tocopherol, and phytosterol (Liu et al., 2018; Zhang et al., 2022). Tea oil induces antioxidant enzymes *in vitro* and *in vivo*, and protects against oxidative damage to liver tissues and gastrointestinal mucosa (Teixeira and Sousa, 2021). It is frequently referred to as "Oriental olive oil" due to its high oleic acid content. Numerous byproducts are produced in the process of extracting tea oil from the seeds, including tea meal and tea shell. The latter is an important industrial raw material for extracting tea saponin, which is widely used in the production of laundry products, organic fertilizers, and insecticides (Quan et al., 2022). *C. oleifera* is currently cultivated in red soils in hilly areas throughout much of southern China. Hunan, Jiangxi, and Guangxi provinces are the primary production areas and account for 76% of the total area of production in the country (Qin et al., 2018). Hunan has a subtropical monsoon humid climate and four distinct seasons, with mean temperatures of 10–12°C in autumn and winter and 4–8°C in January. *C. oleifera* typically flowers from October to February, and thus frequently encounters cold temperatures during the flowering period (Gao et al., 2015).

Cold is among the most important abiotic stresses affecting plants, and can negatively impact overall growth and development. Consequently, the growth of nutritional (leaves) and reproductive organs (flowers) can be inhibited when plants are exposed to temperatures below their optimal growth temperatures, resulting in reduced yields (Tuteja et al., 2011; Soualiou et al., 2022). Studies focused only on physiological responses to stress cannot identify the mechanisms by which plants adapt to environmental stresses, and reproductive traits may be a better marker of plant adaptive responses (Kumar et al., 2012). The reproductive stages of flowering plants are typically highly sensitive to temperature, and cold frequently leads to delayed flowering, induces pollen sterility, and disrupts mitosis I and II (Zinn et al., 2010). Furthermore, studies have shown that while dormant buds are not sensitive to cold temperatures, expanding buds and flowers are vulnerable to cold. For example, cold can delay the development of apple (*Malus domestica*) buds, and the flowers are more cold-sensitive during the active period than during the dormant period (Proebsting and Mills, 1978; Salazar-Gutiérrez et al., 2016).

The *C. oleifera* cultivars 'Huashuo' (HS) and 'Huaxin' (HX) are new, high-yielding national cultivars bred from common *C. oleifera* seedling (Tan et al., 2011; Tan et al., 2012). HS forms a half open round crown, and has smooth, yellowish-brown bark (Figure 1A). The leaves are oval and dark green in color, with an average thickness of 0.48 mm (Figure 1B). This cultivar generally flowers from early November to early December and the fruit is a yellow-brown, five-sided ovate capsule (Figure 1C). Average fruit weight can reach 68.75 g at the height of the fruiting season. HX is tall tree with a naturally round crown and yellow-brown bark (Figure 1D). The leaves are leathery, broadly ovate, and dark green in color, with an average thickness of 0.44 mm (Figure 1E). The flowering period generally lasts from October to mid-December, and the fruit is ovate and green-yellow in color (Figure 1F). Average fruit weight can reach 48.83 g at the height of the fruiting season. These two cultivars are currently cultivated throughout Hunan, as 8–14°C is the optimum flowering temperature for *C. oleifera*, and damage due to cold stress can significantly affect its reproductive output.

Research into the cold resistance of *C. oleifera* cultivars has mainly focused on physiological and biochemical responses in vegetative growth organs (mainly leaves), whereas little research has been conducted on the molecular mechanisms involved in cold stress responses during the flowering period (Dong et al., 2020). Previous studies have shown that under cold stress, the flower buds of HS fail to open, open only partially, or wilt and drop. However, whereas the flowers HX exhibit delayed opening, the flowers and fruits of this cultivar exhibit little dropping due to the presence of a nectar-like sticky material at the base of the flowers, which may be a physiological response that protects the floral apparatus from cold damage (Wu

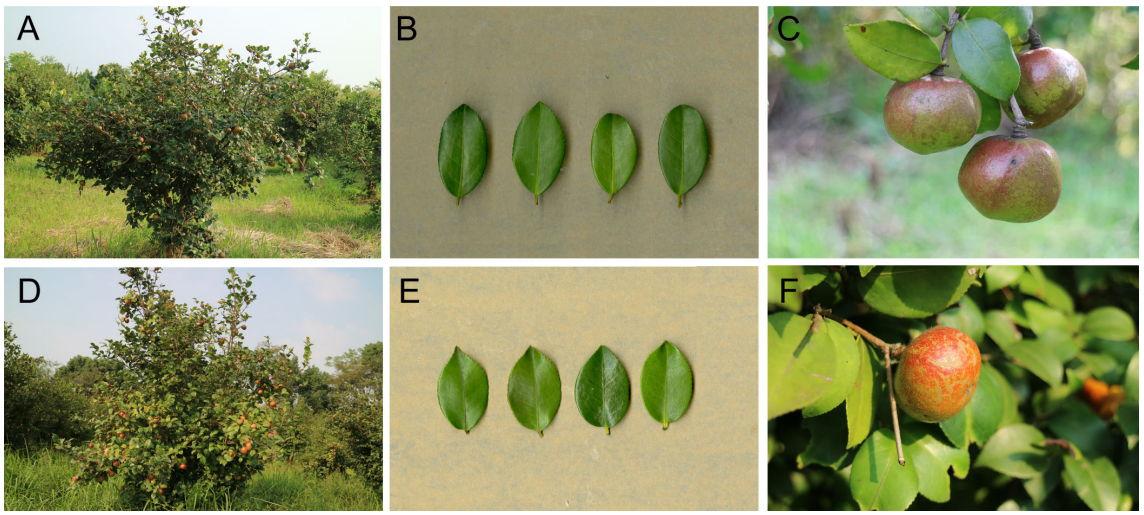


FIGURE 1
Trees, leaves, and fruits of HS **(A–C)** and HX **(D–F)**.

et al., 2020a). In addition, biochemical reactions to cold stress differed between the two cultivars (Wu et al., 2020b).

To explaining the differences of phenotypes from the molecular perspective, we investigate changes in metabolism and transcriptional levels in the flower buds of the two cultivars under cold stress, we used flower buds as test materials, and performed untargeted metabolomics and transcriptomics using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) and Illumina next generation sequencing (NGS) technology. Our analyses will provide a basis for molecular markers and molecular breeding of *C. oleifera*.

2 Materials and methods

2.1 Plant materials and cold treatments

Samples of HX and HS were obtained from the Seedling Center of Hunan. In March 2018 we selected 120 robust, 2-year-old oil seedlings and transplanted them into plastic pots (22 × 22 × 20 cm) filled with a mix of peat, loess, and perlite (2:1:1). Transplants were placed on the roof of the Life Science Building at the Central South University of Forestry and Technology (Changsha, Hunan, 28° 10′ N, 113° 23′ E) and were subject to similar water and fertilization regimes. The cold stress experiment was conducted between November 2019 and January 2020 in an artificial climate chamber. On November 3, 2019, we selected 40 dwarfed but robust plants (20 plants per cultivar) with numerous flower buds, and exposed them to cold stress treatments (6°C) in an artificial climate chamber. Other parameters in the chamber were as follows: relative humidity = 70–80%, photoperiod = 12 h (8:00 a.m.–8:00 p.m.), photon flux density = 200 $\mu mol \cdot m^{-2s}$, and average carbon dioxide concentration = 450 $\mu mol \cdot mol^{-1}$. No fertilizer was applied during the cold treatments. Plants were given 500 mL water every 3 days to maintain soil moisture.

Samples were consistently collected in the same order and processed at approximately 10:00 a.m. Six samples of each cultivar were collected and each sample contained three unopened buds during four batch of sampling: the first was collected prior to placing plants into the artificial climate chamber as CK (0 days) with 6 samples labelled CK_HX1 to CK_HX6 and CK_HS1 to CK_HS6, the second batch was collected after 1 day in the artificial climate chamber as ST (1 days; short-term stress) with 6 samples labelled ST_HX1–ST_HX6 and ST_HS1–ST_HS6, the third after 7 days as MT (7 days; medium-term stress) with 6 samples labelled MT_HX1–MT_HX6 and MT_HS1–MT_HS6, and the fourth after 25 days as LT (long-term stress; LT) with 6 samples labelled LT_HX1–LT_HX6 and LT_HS1–LT_HS6. Samples were immediately wrapped in tin foil and labeled, snap-frozen in liquid nitrogen for 30 min, and then stored at –80°C in an ultra-cold freezer. Each sample was divided into two parts, one for untargeted metabolomics and the other for transcriptomics.

2.2 Sugar and endogenous hormone measurements

The 24 samples obtained from this experiment were analyzed by gas chromatography (GC)-MS for the absolute quantification of 13 sugars and UPLC-MS/MS for the absolute quantification of three plant hormones (Supplementary Table S1). Each test was replicated three times and followed methods described in a previous study (Wu et al., 2022).

2.3 Metabolomic sample preparation and extraction

The samples were freeze-dried and crushed for 1.5 min at 30 Hz using a mixer mill (MM 400; Retsch GmbH, Haan, Germany) with a zirconia bead. We weighed 100 mg powder and extracted it overnight at 4°C in 0.6 mL 70% aqueous methanol. Following centrifugation at 10,000 g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 mL; ANPEL, Shanghai, China) and filtered (SCAA-104, pore size 0.22 μm; ANPEL) in preparation for the UPLC-MS/MS analysis.

The sample extracts were analyzed on a UPLC-ESI-MS/MS system (UPLC-MS/MS: Shim-pack UFLC SHIMADZU CBM30A system; Shimadzu, Kyoto, Japan; GC-MS: 4500 quadrupole-linear ion trap [Q TRAP]; Applied Biosystems, Waltham, MA, USA). The analytical conditions for UPLC were as follows. The column was an Acquity UPLC HSS T3 C18 column (1.8 μm, 2.1 mm × 100 mm; Waters Corporation, Milford, MA, USA). The mobile phase consisted of solvent A (pure water with 0.04% acetic acid) and solvent B (acetonitrile with 0.04% acetic acid). Acetic acid and acetonitrile were purchased from Merck (Darmstadt, Germany). Sample measurements were performed along a gradient, with the starting conditions 95% A and 5% B. Within 10 min, a linear gradient to 5% A + 95% B was applied, and the composition of 5% A + 95% B was maintained for 1 min. Then the composition was adjusted to 95% A + 5.0% B within 0.10 min and maintained for 2.9 min. The column oven was set to 40°C, and the injection volume was 4 μL. The effluent was alternatively connected to an ESI-triple Q TRAP-MS. The qualitative identification of metabolites is based on the database MWDB.

2.4 Library preparation and RNA sequencing

RNA samples were sent to Wuhan Metware Biotechnology Co., Ltd. (Wuhan, China), where the libraries were produced and sequenced. Total RNA was extracted from the flower buds using an RNA Prep Pure Plant Kit (Tiangen, Beijing, China) following the manufacturer's protocols. RNA purity and concentration were determined using a NanoPhotometer spectrophotometer (Implen, Munich, Germany). The input material consisted of 3 μg RNA per sample. Sequencing libraries were generated using a NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) following the manufacturer's protocols. Library construction and inspection were conducted as previously described (Li et al., 2018). Polymerase chain reaction (PCR) products were purified using an AMPure XP system (Beckman Coulter, Brea, CA, USA) and library quality was assessed using an Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Libraries

were sequenced on the Illumina HiSeq platform (Illumina Inc., San Diego, CA, USA) and 125 bp/150 bp paired-end reads were generated.

2.5 Transcriptome assembly and gene functional annotation

Clustering of the index-coded samples was performed using a cBot Cluster Generation System with a TruSeq PE Cluster Kit v3-cBot-HS (Illumina Inc.) following the manufacturer's protocols. Following cluster generation, libraries were sequenced on the Illumina HiSeq platform and 125 bp/150 bp paired-end reads were generated. Image data obtained from the high-throughput sequencer were transformed into raw data using Casava base recognition. After filtering, the original data were used to determine the sequencing error rate and GC content distribution to obtain clean reads for subsequent analyses. Clean reads were assembled using a Trinity assembler, and the obtained transcripts were used as reference sequences in subsequent analyses. The transcripts were clustered hierarchically using read numbers and the expression patterns of the aligned transcripts. The longest cluster sequence obtained from corset hierarchical clustering was used as a unigene in subsequent analyses. Gene function was annotated using the NR, Pfam, euKaryotic Orthologous Groups (KOG), COG, eggNOG, Swiss-Prot, KEGG, and GO databases.

2.6 Differential expression analyses

Fragments per kilobase of transcript per million fragments mapped was used as a measure of transcript or gene expression level. DEGs between the two biological conditions were obtained by inter-sample group analysis using DESeq2 (Love et al., 2014). Following differential analysis, the false discovery rate (FDR) was obtained by correcting p-values for multiple hypothesis testing using the Benjamini-Hochberg method. The screening conditions for DEGs were $|\log2\text{Fold Change}| \geq 1$ and $\text{FDR} < 0.05$.

A combination of fold change and variance influence on projection (VIP) values from the orthogonal projections to latent structures discriminant analysis (OPLS-DA) model was used to identify DEMs. Metabolites with fold changes ≥ 2 or ≤ 0.5 were selected. Metabolites with a difference ≥ 2 or ≤ 0.5 in the control (CK) and experimental groups (ST, MT, and LT) were considered significantly different. Metabolites with VIP ≥ 1 were also selected.

2.7 Statistical analyses

Unsupervised principal component analysis (PCA) was conducted using the *prcomp* function in R (R: The R Project for Statistical Computing (r-project.org)). Set the prcomp function parameter scale=True, which means unit variance scaling (UV) normalization of data. The data were unit variance scaled prior to conducting the PCA. DEGs and DEMs were visualized in heat map form using TBtools v1.098763 software, and Venn diagrams were drawn using the Metware Cloud website (https://cloud.metware.cn/#/home). Figures 1, 2 were drawn in Adobe Photoshop 2020. Figures 3–8 were drawn in Adobe Illustrator 2022.

3 Results

3.1 Phenotypic and physiological changes

The growth form of each cultivar prior to being placed in the artificial climate chambers is shown in Figure 2. The morphology and size of the flower buds were similar among the two cultivars. The calyx was wrapped around the buds and the calices and bracts were hairy. HS exhibited no obvious changes with respect the size of flower buds between Days 1 and 7 of the cold stress treatments, and the sepals remained closed and tightly wrapped around the internal flower structures (Figures 2a1, a2). However, a small number of sepals had unfolded by Day 25, exposing the petals (Figure 2a3). Some sepals gradually opened in HX, and the petals were visible in approximately one third of the buds by Day 7 (Figure 2b2). Most buds were completely open by Day 25 (Figure 2b3). The petals and stamens of HX dropped after 32 days of cold stress, and a honey-like mucilaginous substance appeared at the base of the flowers (Figure 2b4). This substance became stickier and more abundant as the duration of cold stress increased (observation period = 60 days). Observations indicated that the flower buds of HX opened one at a time under cold stress, whereas the flower buds of HS required an adaptive dormant period. Both cultivars bloomed under LT cold stress; however, the timing and duration of flowering differed.

To further assess changes in flower buds in response to cold stress, we measured the sugar and phytohormone contents of buds. Thirteen sugars and three phytohormones were detected. The three most abundant sugars were D-fructose, glucose, and sucrose, and most abundant phytohormones were salicylic acid (SA), abscisic acid (ABA), and indole-3-acetic acid (IAA). The contents of D-fructose, glucose, and sucrose in both cultivars increased after a single day of cold stress; however, while the contents of these sugars changed little in HS over the treatment period, they increased continuously in HX. Furthermore, contents of the three soluble sugars were consistently higher in HX than in HS, particularly after 25 days of treatment (Figures 2C–E). Cold stress promotes the accumulation of soluble sugars in HX, which may influence the cold tolerance of this cultivar. Phytohormones also differed between the two cultivars during cold stress: while the ABA content in HS did not vary significantly throughout the treatment period, it increased substantially in HX, peaking at 25 days and exceeding the levels observed in HS (Figure 2F). The trends in IAA content were similar among cultivars, first decreasing and then increasing; however, the IAA content of HS was higher than that of HX (Figure 2G). The SA content in HS first decreased and then increased as the duration of cold stress increased, but changes were not significant after 7 days, whereas the SA content in HX increased gradually with the lengthening of the duration of cold stress. SA content was higher in HS than in HX after 25 days of cold stress, but the difference was not significant (Figure 2H). These results indicate that differences in cold tolerance among the two cultivars might be related to sugar, ABA, and IAA.

10.3389/fpls.2023.1126660

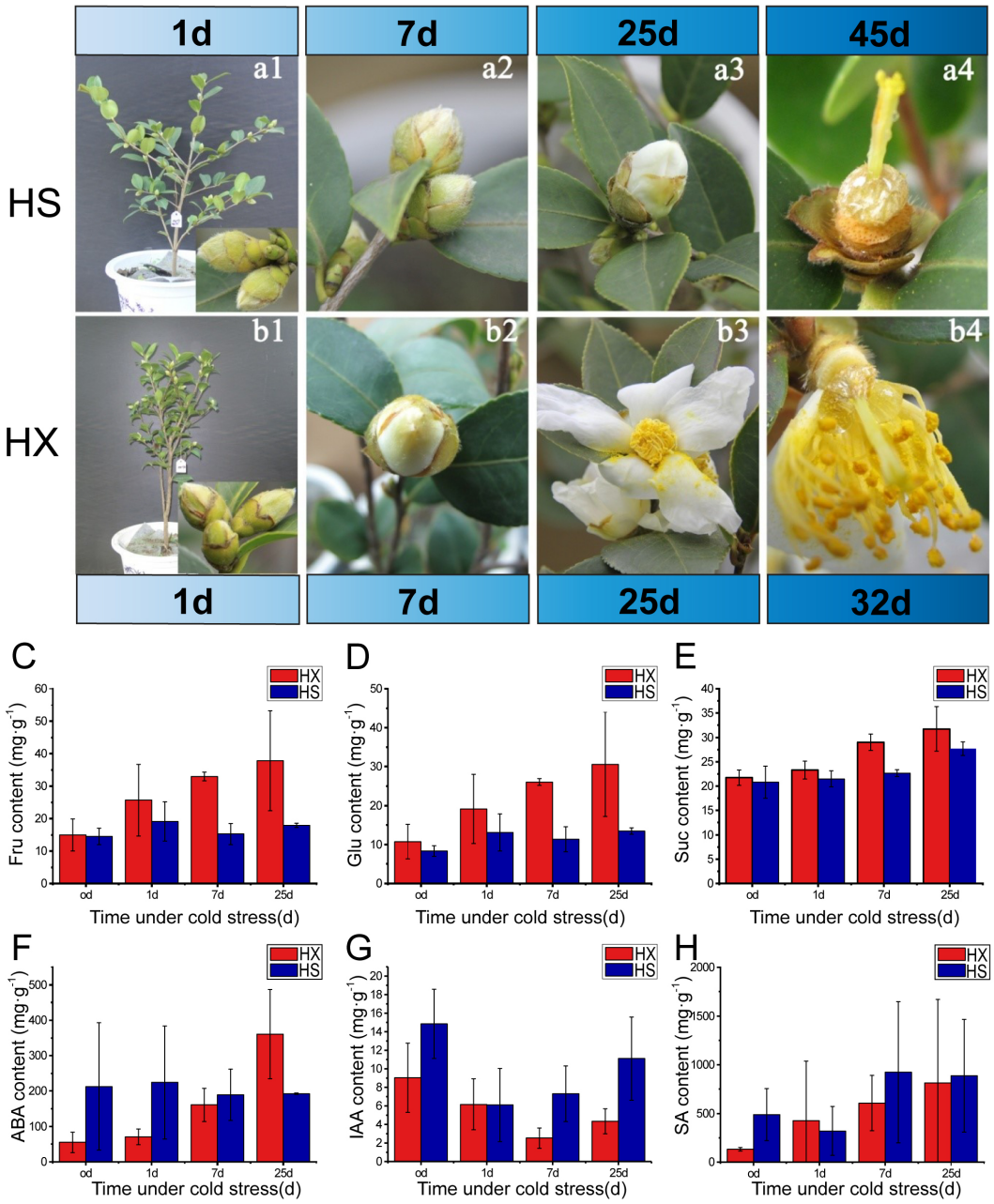


FIGURE 2
Effects of cold stress on phenotype, sugar content, and phytohormone content: bud morphology and flowering of HS **(a1-a4)**, and HX **(b1-b4)** during the treatment period; the contents of **(C)** D-fructose, **(D)** glucose, **(E)** sucrose, **(F)** ABA, **(G)** IAA, and **(H)** SA in flower buds of HS and HX under different cold treatments.

3.2 Metabolite profile analysis

3.2.1 Metabolite analysis

A total of 578 metabolites were detected. These metabolites were divided into 10 main groups, including alkaloids (31 metabolites), amino acids and derivatives (61), flavonoids (170), lignans and coumarins (12), lipids (63), nucleotides and derivatives (37), organic acids (31), phenolic acids (102), tannins (24), and terpenoids (4), plus an additional 43 compounds that did not fit into any of these groups (Figure 3A). The most abundant metabolites were flavonoids, phenolic acids, lipids, and amino acids and derivatives. PCA indicated that principal components 1–5 explained more than 62.5% of the cumulative variance in metabolites (Figure 3B). The PCA score plot indicated that the close pooling of the mixed group samples used for quality control demonstrated remarkable experimental repeatability; furthermore, principal components 1 and 2 accounted for 34.5% and 11.4% of variance, respectively. PC1 separated HS from HX, indicating that major differences in metabolite levels between the two cultivars, whereas PC2 separated samples exposed to cold stress for 0–1 day from those subjected to cold stress for 7–25 days (Figure 3C). In addition, while samples of both cultivars migrated upward along PC2 over the course of the treatment period, HS migrated farther than HX. Samples from the LT-HS group became increasingly clustered in the upper left as

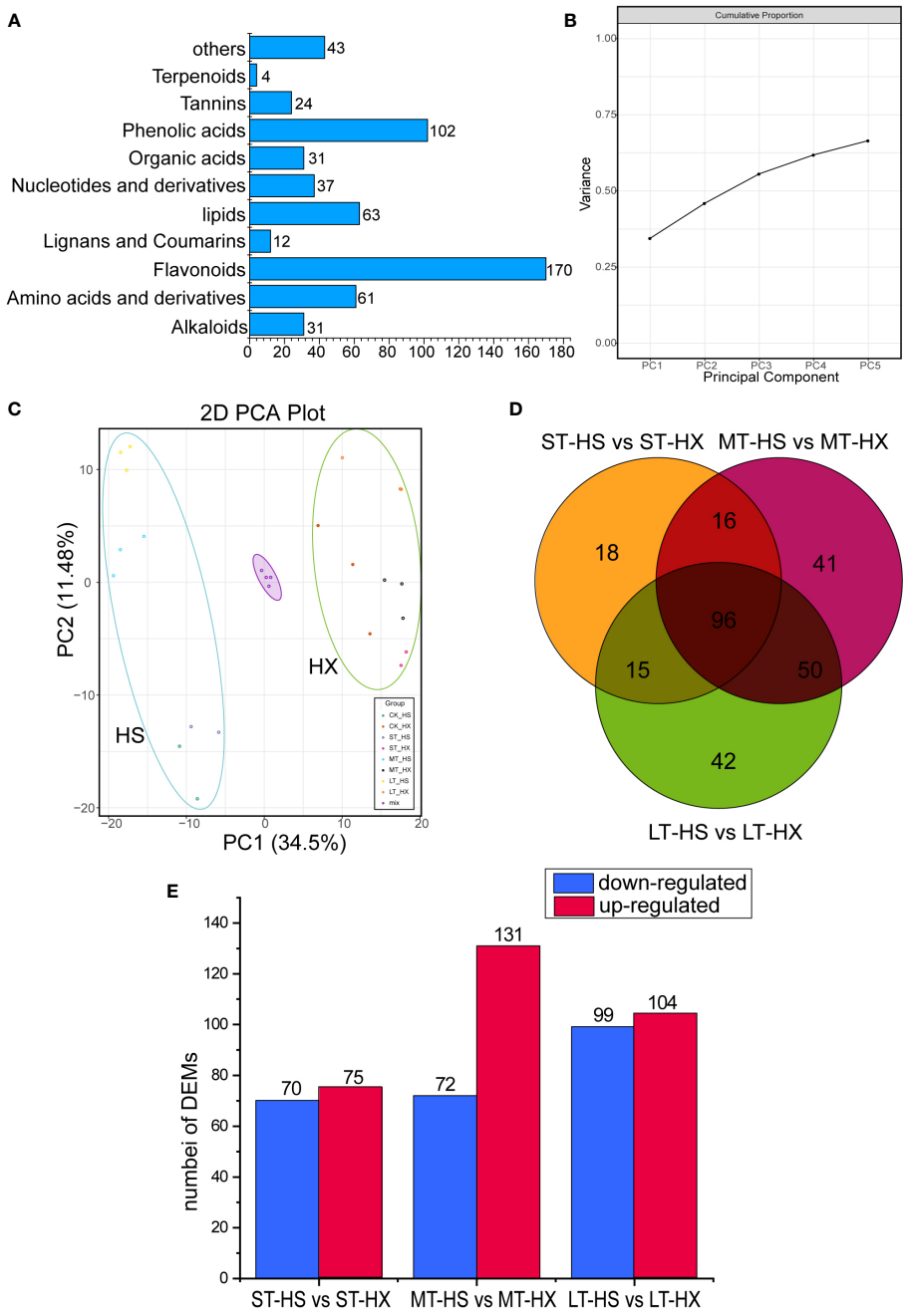


FIGURE 3
Overview and analysis of metabolites. **(A)** classification of the 578 metabolites. **(B)** variance explained by PCs 1–5. **(C)** PCA of metabolome data at the four sampling points (0, 1, 7, and 25 days; three biological replicates per point). **(D)** Venn diagrams of DEMs under different cold treatments. **(E)** up- and down regulated metabolites.

the duration of cold stress increased, suggesting that metabolic changes were more pronounced in HX than in HS.

We used a comparative analysis of changes in metabolites to identify differences in accumulation patterns between the cultivars in response to cold stress. We compared the DEMs of the two cultivars based on the criteria of fold change ≥ 2 or ≤ 0.5. We identified 145 DEMs after 1 day of cold treatment (75 upregulated and 70 downregulated), 203 DEMs after 7 days (131 upregulated and 72 downregulated), and 203 DEMs after 25 days (104 upregulated and 99 downregulated) (Figure 3E). Furthermore, 96 commonly enriched DEMs were identified in different comparisons, and 42 metabolites were identified specifically in the comparison of the LT-HS and LT-HX groups (Figure 3D). Most metabolites, including flavonoids, phenolic acids, and lipids, were more highly induced in HX than in HS, whereas the opposite was generally true for amino acids and derivatives. Moreover, the content of pyridoine, D-glucoronic acid, sodium ferulate, (R)-pantetheine, 5,7-dyhydroxy-1(3H)-isobenzofuran-one-O-glucoside, androsin, and D-(+)-melezitose were higher in HX than in HS, whereas nicotinic acid, riboflavin, and maltotetraose were more strongly induced in HS. Cold stress significantly induced certain metabolites in both cultivars, including L-Proline, succinic acid, and coniferin (Supplementary Figure S1).

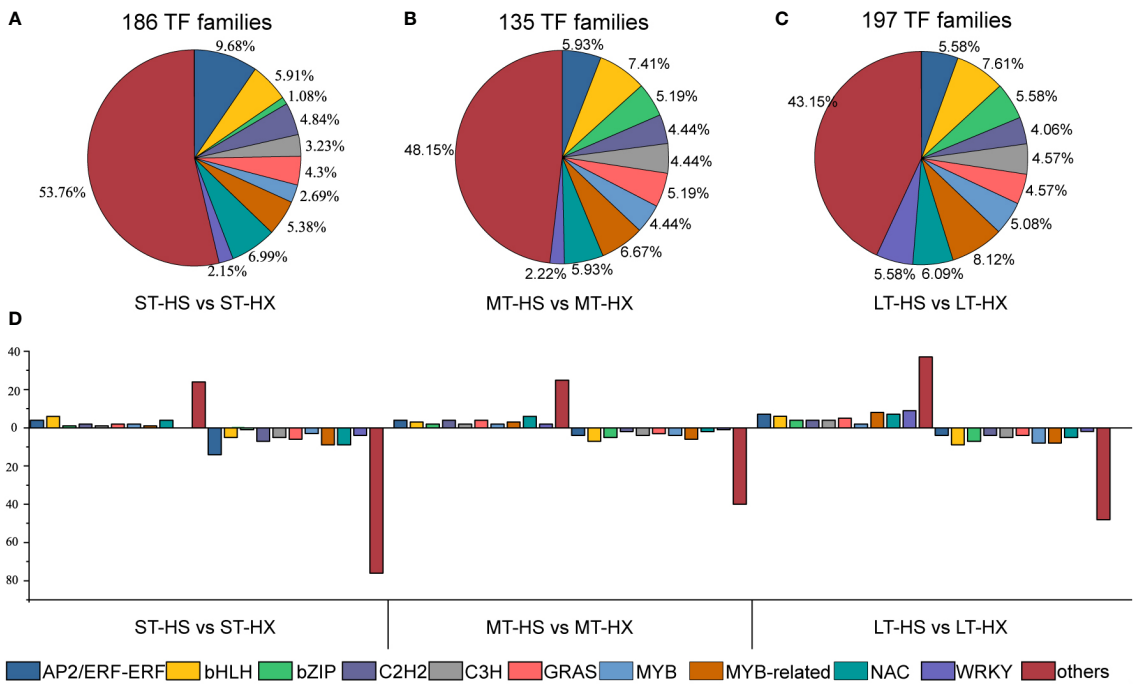


FIGURE 4
Cold stress-responsive TFs. **(A–C)** Distribution of TF families under ST, MT, and LT cold stress, and **(D)** up- and down regulated TF families.

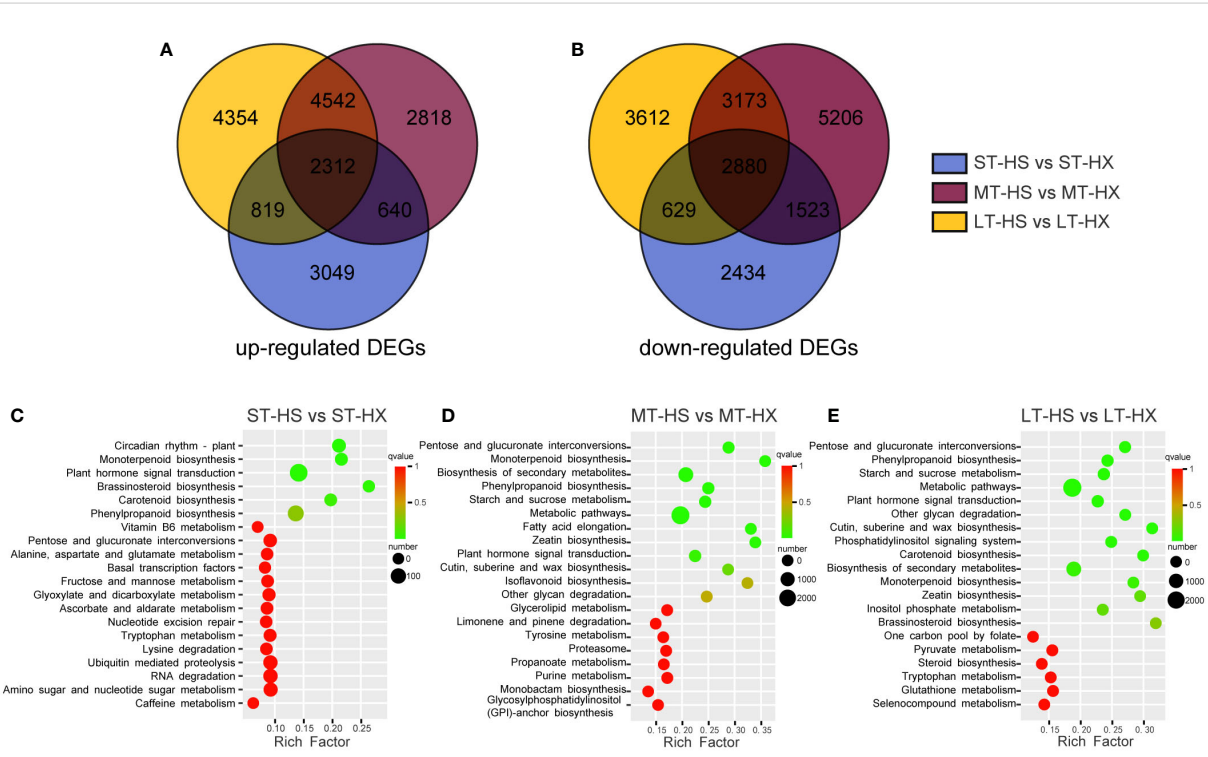


FIGURE 5
DEG analysis. The Venn diagram shows DEGs that were **(A)** up regulated or **(B)** down regulated; KEGG enrichment analysis of DEGs after **(C)** 1, **(D)** 7, and **(E)** 25 days of cold stress treatments.

3.2.2 KEGG pathway mapping of DEMs

The KEGG pathways in different pairwise comparisons (ST-HS *vs* ST-HX, MT-HS *vs* MT-HX, and LT-HS *vs* LT-HX) typically included the valine, leucine, and isoleucine degradation, tyrosine metabolism, phenylpropanoid biosynthesis, and carbapenem biosynthesis pathways. The key enrichment pathways between ST-HS and ST-HX included the flavone and flavonol biosynthesis, anthocyanin biosynthesis, and ubiquinone and another terpenoid–quinone

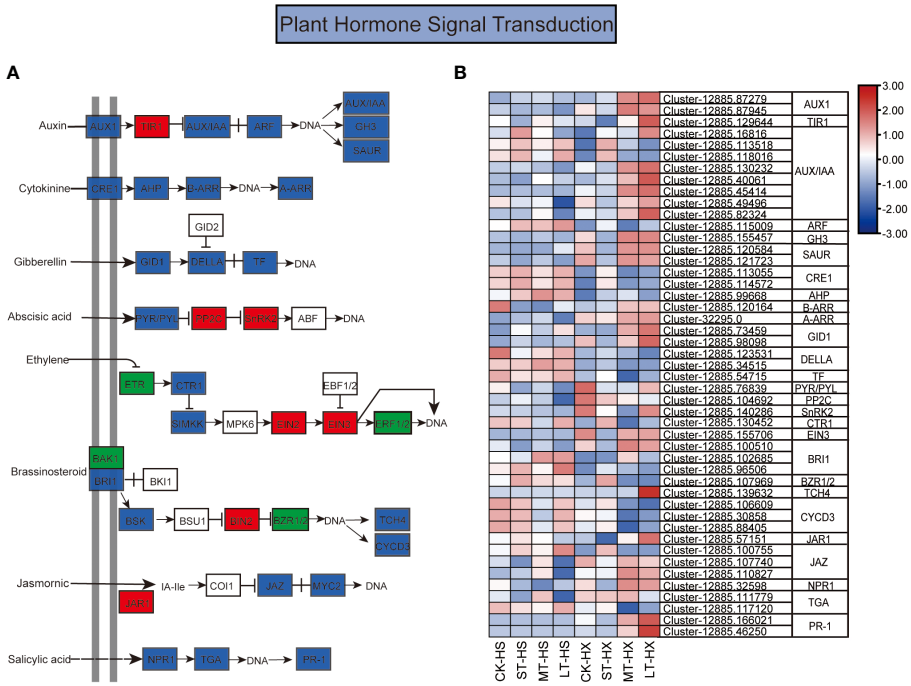


FIGURE 6
The KEGG pathway annotation diagram of the plant hormone signal transduction pathway of DEGs. **(A)** Schematic diagram of the KEGG pathway for the enrichment of DEGs in the plant hormone signal transduction pathway. Red boxes indicate upregulation, green boxes indicate downregulation, and blue boxes indicate both up- and downregulation. **(B)** Heatmap showing the expression of DEGs associated with different hormones. Box color indicates the expression level of each gene: blue = decreased expression and red = increased expression.

biosynthesis pathways, whereas the key enrichment pathways between MT-HS and MT-HX included the valine, leucine, and isoleucine degradation, phenylpropanoid biosynthesis, and valine, leucine and isoleucine biosynthesis pathways. Significantly enhanced pathways between LT-HS and LT-HX included the phenylpropanoid biosynthesis, propanoate metabolism, purine metabolism, and pyrimidine metabolism pathways (Supplementary Figure S2). The results indicate that metabolites in the flower buds of the two cultivars differ significantly under long-term cold stress.

3.3 Transcriptome analysis

3.3.1 Transcription factors in response to cold stress

Transcription factors (TFs) are essential for gene expression in plants under abiotic stress, and a better understanding of TFs and their downstream target genes will facilitate the development of stress-tolerant crops with improved quality and yields (Khan et al., 2018). We identified 74 TF families in our analysis, of which the AP2/ERF-ERF family was the largest. Moreover, 186 TFs were identified in comparisons between the ST-HS and ST-HX groups (47 upregulated and 139 downregulated), 135 between the MT-HS and MT-HX groups (57 upregulated and 78 downregulated), and 197 between the LT-HS and LT-HX groups (93 upregulated and 104 downregulated) (Figure 4D). TF families that were susceptible to cold stress included MYB, MYB-related, GRAS, bHLH, C2H2, C3H, NAC, bZIP, WRKY, and AP2/ERF-ERF. In the comparison of ST-HS and ST-HX, the AP2/ERF-ERF family was the largest group (9.68%), followed by NAC (6.99%) and bHLH (5.91%) (Figure 4A). The four most abundant families in the MT-HS *vs.* MT-HX comparison were bHLH (7.41%), MYB-related (6.67%), AP2/ERF-ERF (5.93%), and NAC (6.09%) (Figure 4B). The three most abundant families in the LT-HS *vs.* LT-HX comparison were MYB-related (8.12%), bHLH (7.61%), and NAC (6.09%) (Figure 4C). Most genes in families such as AP2/ERF-ERF, MYB-related, NAC, and WRKY were more strongly upregulated in HX than in HS after 25 days of cold treatment, indicating that TF gene expression in the flower buds was more active in HX than in HS under long-term cold stress.

3.3.2 DEGs analysis

The 24 samples were analyzed *via* transcriptome sequencing. After removing low-quality reads, a total of 1,393,597,534 clean data points were obtained, with the percentage of Q30 bases > 93% and the percentage of GC > 43.32%, indicating that the transcriptome sequencing was reliable and of high quality. Based on screening conditions |log2Fold Change| ≥ 1 and FDR < 0.05, 14,285 DEGs were identified between ST-HS and ST-HX (6,820 upregulated and 7,465 downregulated), 23,093 between MT-HS and MT-HX (10,312 upregulated and 12,781 downregulated), and 22,320 between LT-HS and LT-HX (22,320 upregulated and 22,320 downregulated) (Supplementary Table S2). In these three comparisons, 2,312 upregulated and 2,880 downregulated DEGs overlapped. Interestingly, 3,049, 2,818, and 4,354 DEGs were specifically upregulated after 1, 7, and 25 days of cold stress, respectively, whereas 2,434, 5,206 and 3,612 DEGs were specifically downregulated (Figures 5A, B). Upregulated genes became more numerous as the duration of stress increased, and may be among

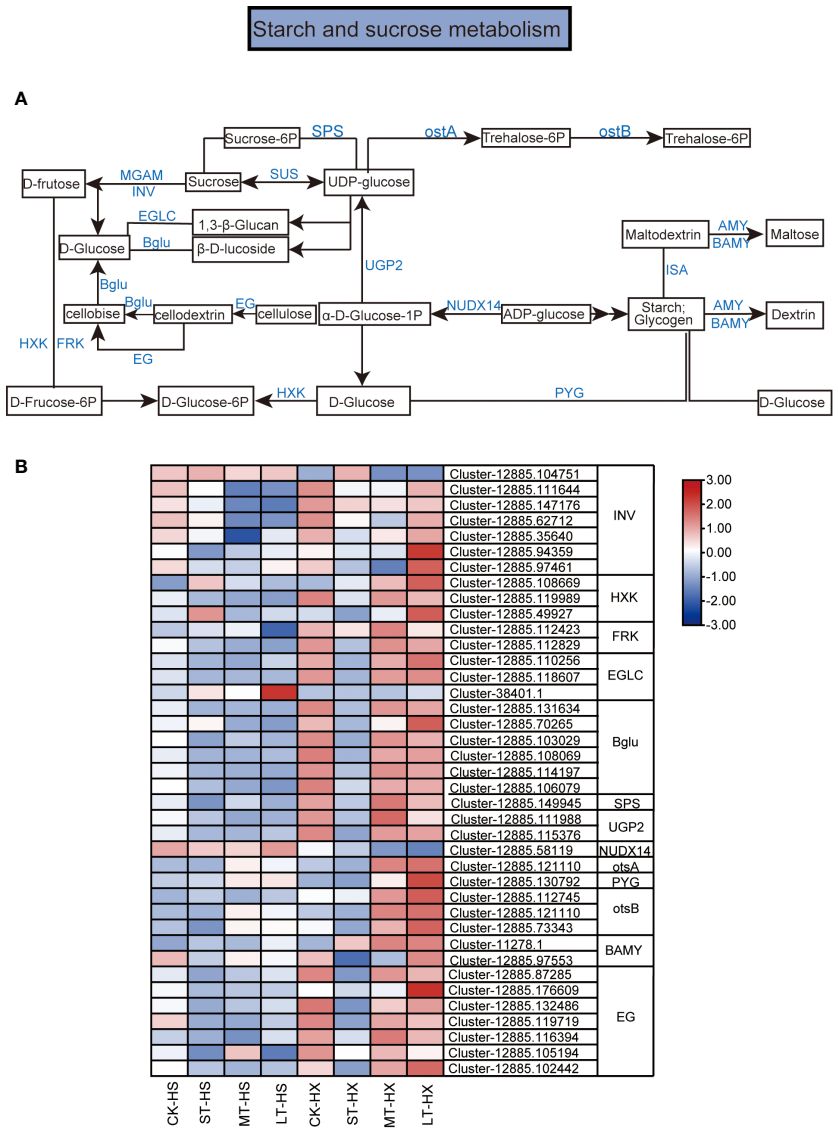


FIGURE 7
KEGG pathway annotation diagram for the starch and sucrose metabolism pathway of DEGs. **(A)** Schematic diagram of the KEGG pathway for the enrichment of DEGs in the starch and sucrose metabolism pathway. Blue text indicates enzymes and black text indicates metabolites. **(B)** Heatmap showing the expression of DEGs. Box color indicates the expression level of each gene: blue = decreased expression and red = increased expression.

the key cold resistance mechanisms in HX. These results imply that the two cultivars respond differently to cold stress. The numerous DEGs shared among groups independent of the time of cold stress might be attributable to genetic differences between two cultivars. As such, the focus of our discussion from here will be the analysis of specific genes identified in the comparison of the LT-HS and LT-HX groups.

3.3.3 Gene function annotation of DEGs

The unigene sequence was compared to the KEGG, NR, Swiss-Prot, GO, COG/KOG, and Trembl databases using BLAST software, and the amino acid sequence of the unigene was predicted and compared to the Pfam database using HMMER software to obtain annotation information for the unigene (Supplementary Figure S3A). The comparison of 10,951 transcripts with the NR library facilitated the identification of transcript sequence similarities between species and the obtention of functional information for homologous sequences. The results indicate that the species most similar to *C. oleifera* is *Vitis vinifera* (Supplementary Figure S3B). Following GO annotation of 234,040 unigenes, the annotated genes were classified based on the next level of three GO categories: biological processes, cellular components, and molecular functions. The terms with the highest number of transcripts in the biological processes category were "cellular and metabolic processes" and "biological regulation and stimulus-response" (Supplementary Figure S3C). For cellular components, the most abundant terms were "cell" and "cell portion," whereas the most abundant term for molecular function was "binding." A total of 52,336 transcripts were annotated and classified

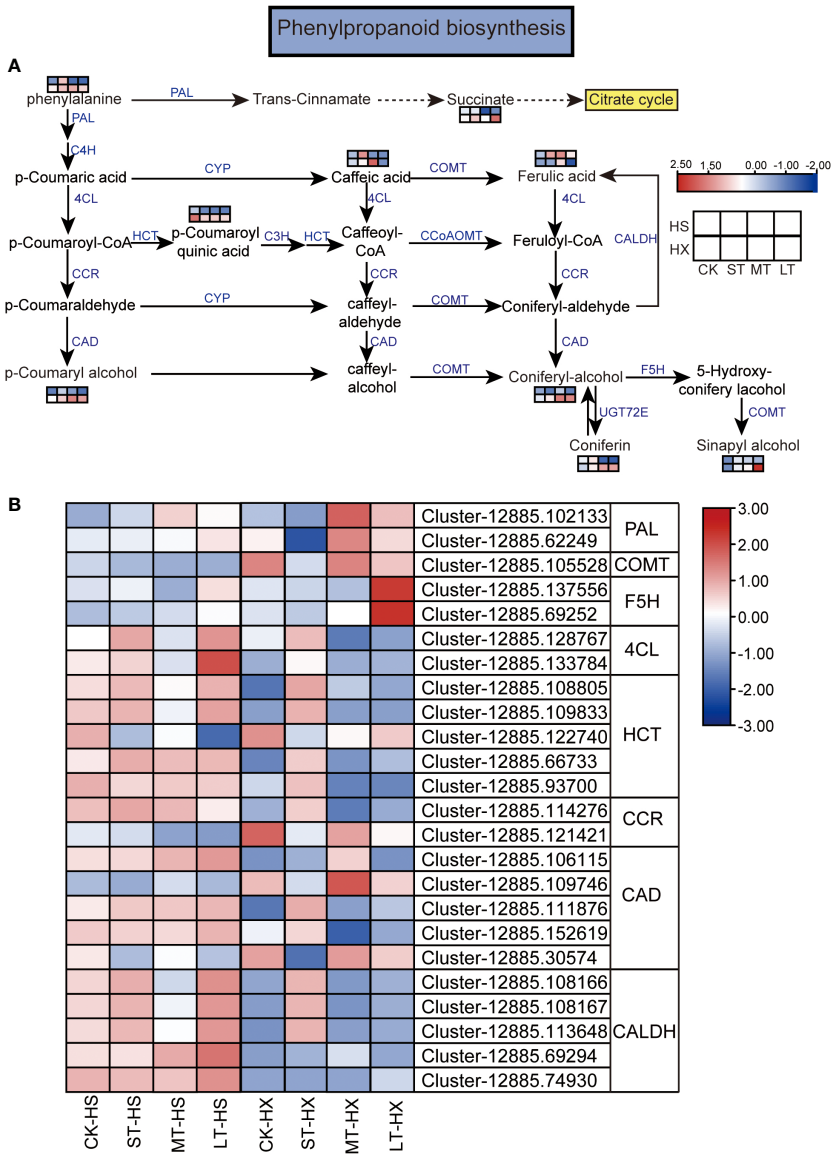


FIGURE 8
KEGG pathway annotation diagram of the phenylpropanoid biosynthesis pathway of DEGs and DEMs. **(A)** Schematic diagram of the KEGG pathway for the enrichment of DEGs and DEMs involved in phenylpropanoid biosynthesis. Blue text indicates enzymes, black text indicates metabolites, and the dotted lines indicate the omission of certain reactions. **(B)** Heatmap showing the expression of DEGs. Box color indicates the expression level of each gene: blue = decreased expression and red = increased expression.

into 25 categories in the KOG database. Among these categories, general function prediction had the largest number of transcripts (11,774), followed by signal transduction mechanisms (5,078), posttranslational modification, protein turnover, and chaperones (4,909), and carbohydrate transport and metabolism (2,925) (Supplementary Figure S3D). These results indicate that sugar metabolism plays an important role in the responses of the two cultivars to cold stress. KEGG annotation-based enrichment analysis of the two cultivars at different points during the treatments revealed that DEGs were significantly enriched in the metabolic, secondary metabolic, starch and sucrose metabolism, plant hormone signal transduction, and phenylpropanoid biosynthesis pathways (Figures 5C–E). These results suggest that metabolites involved in starch and sucrose metabolism and plant hormone signal transduction may be key metabolites to the cold stress responses of the two cultivars.

3.3.4 Plant hormone signal transduction pathways in response to cold stress

Hormones play a central role in regulating responses to cold stress in plants, which is essential to plant growth and development (Zhao et al., 2021). Our analysis identified a large number of DEGs between HS and HX under cold stress; these DEGs are involved in hormone signal transduction for a range of hormones, including IAA, ABA, cytokinin (CTK), gibberellin (GA), ethylene (ET),brassinosteroid (BR), jasmonic acid (JA), and SA (Figure 6A). The DEGs exhibited distinct expression patterns in the two cultivars, revealing complicated cold stress response mechanisms involving

phytohormones. The expression of most DEGs was relatively stable in HS over the course of the treatments but exhibited significant upward trends in HX, peaking at 25 days. Fifteen DEGs in the auxin signal transduction pathway were significantly differentially expressed under different cold treatments, including genes encoding auxin influx carrier (AUX1), transport inhibitor response 1 (TIR1), auxin-induced protein/auxin-responsive protein (AUX/IAA), auxin response factor (ARF), auxin-responsive GH3 family (GH3), and SAUR family proteins (SAUR). The expression levels of two *AUX1* genes, one *TIR1* gene, six *AUX/IAA* genes, one *GH3* gene, and two *SAUR* genes were significantly higher in LT-HX than in LT-HS. By contrast, only two *AUX/IAA* genes and one *ARF* gene were downregulated in LT-HS compared to LT-HX. These results suggest that the expression of auxin signal transduction related genes was generally repressed in HS under prolonged cold stress. In the ABA pathway, DEGs including pyrabaction resistant/PYRlike (*PYR/PRL*), protein phosphatase 2C gene (*PP2C*), and sucrose non-fermenting 1-related protein kinase 2 gene (*SnRK2*) exhibited decreased expression in both HS and HX after 1 day of cold stress. Expression of these DEGs increased in HX throughout the treatment period, whereas in HS their expressions briefly increased and then decreased. Meanwhile, there was a significant increase in expression after 25 days in LT-HX compared to LT-HS. Other hormone signalling genes, including *B-ARR*, *A-ARR*, *GIDI*, *EIN3*, *TCH4*, *JAZ*, and *NPR1*, exhibited more pronounced upregulation in HX compared to HS when cold stress persisted for more than 7 days (Figure 6B).

3.3.5 Starch and sucrose metabolism pathways analyses

The KEGG enrichment analysis indicates that many of the DEGs between the two cultivars under different cold treatments are associated with starch and sucrose metabolism. Figure 7A shows a schematic diagram of the metabolic pathways for starch and sucrose. In all, 39 DEGs were selected for this study. These DEGs encode enzymes such as beta-fructofuranosidase (INV), hexokinase (HXK), fructokinases (FRK), glucan endo-1,7-beta-D-glucosidase (EGLC), beta-glucosidase (Bglu), sucrose-phosphate synthase (SPS), UTP-glucose-1-phosphate uridylyltranferase, (UGP2), ADP-sugar diphosphatase (NUDX14), trehalose 1–4-phosphate synthase (otsA), glycogen phosporylase (PYG), trehalose 11-phosphate phosphatase (otsB) beta-amylase (BAMY), and endoglucanase (EG). Among these genes, one *INV* gene, one *EGLC* gene, and one *NUDX14* gene exhibited significantly lower expression levels in LT-HX compared to LT-HS. However, the expression of most DEGs was significantly higher in LT-HX than in LT-HS, particularly in the 25-day treatment (Figure 7B). The higher expression levels may lead to better cold tolerance in HX.

3.4 Combined transcriptomic and metabolomic analyses

We combined our metabolomic and transcriptomic data to further clarify the responses of HS and HX to cold stress. Common pathways established based on the KEGG pathways of DEGs and DEMs include the flavone and flavonol biosynthesis, flavonoid biosynthesis, pyrimidine metabolism, and purine metabolism pathways. Of these, the phenylpropanoid biosynthesis pathway plays an important role in regulating cold stress. This pathway was significantly enriched in both the metabolomic and transcriptomic data, with six significantly different metabolites and eight key enzymes encoded by 24 DEGs. While most metabolites, including phenylalanine, p-coumaryl alcohol, ferulic acid, coniferyl alcohol, coniferin, sinapyl alcohol, caffeic acid, and succinate, were upregulated in both cultivars under cold stress, p-coumaroyl quinic acid consistently declined. Moreover, the expression of some metabolites, such as p-coumaryl alcohol, coniferyl alcohol, coniferin, sinapyl alcohol, and succinate, was significantly higher in HX than in HS after 25 days of cold treatment (Figure 8A). Most DEGs exhibited different expression patterns in the two cultivars. Under cold stress, two genes encoding 4-coumarate–CoA ligase (4CL), four genes encoding shikimate O-hydroxycinnamoyltransferase (HCT), one gene encoding cinnamoyl-CoA reductase (CCR), three genes encoding cinnamyl-alcohol dehydrogenase (CAD), and five genes encoding coniferyl-aldehyde dehydrogenase (CALDH) exhibited higher transcription levels in HS than in HX. However, the expression of some genes was higher in the LT-HX group than in the LT-HS group, including two genes encoding phenylalanine ammonia-lyase (PAL), which is the entry enzyme for phenylpropanoids, one gene encoding caffeic acid 3-O-methyltransferase/acetylserotonin O-methyltransferase (COMT), two genes encoding ferulate-5-hydroxylase (F5H), and two genes encoding CAD (Figure 8B). The results of this analysis suggest that these upregulated DEMs and DEGs, which are related to phenylpropanoid biosynthesis, may improve cold tolerance in *C. oleifera*.

4 Discussion

Temperature affects plant growth and development and limits species' geographical distributions, particularly during the critical stages of reproduction. To survive in extreme environments, plants have evolved adaptive responses whereby they regulate developmental processes, such as flowering or dormancy, in response to variation in temperature (Susila et al., 2018). Previous studies have demonstrated that the cultivation of *Camellia sinensis* has been constrained by extreme temperatures, and that cold stress might cause flowers and fruits to drop, leading to reduced yields (Ru and Ju, 2010). The risk for cold damage can be reduced, and the production of *C. oleifera* increased, by developing new varieties. In this study, we sought to unravel the different regulatory mechanisms involved in cold stress responses in HS and HX through a combination of physiological, transcriptomic, and metabolomic analyses. The results indicate that, in addition to physiological changes, numerous DEGs and DEMs were present in the flower buds of the two cultivars after the cold treatments. We found that, in contrast to the trend for auxin, the contents of D-fructose, glucose, sucrose, and ABA in the flower buds of HX were significantly higher than in HS at after 25 days of cold treatment. Furthermore, KEGG enrichment analysis indicated that some DEGs are related to the starch and sucrose metabolism, phenylpropanoid biosynthesis, and plant hormone signal transduction pathways; most of these DEGs exhibited higher expression in HX than in HS after 25 days of cold stress. These results are consistent with our phenotypic observations: HX bloomed first and produced a sticky, nectar-like material at the base of the

flowers and columns. By contrast, some flower buds in HS entered dormancy, whereas others remained closed, and flowering was delayed.

4.1 Transcriptional regulation-related genes

Stress resistance in plants is regulated by multiple genes. Among these, TFs play the important role of a molecular switch in the transcriptional regulation network of stress responses, mainly regulating the function of downstream genes *via* interactions with cis acting elements in downstream target gene promoters (Liu et al., 2013). Several TFs, including MYB, WRKY, NAC, and AP2, are associated with cold tolerance in plants (Khan et al., 2018). Of these, MYBs play key roles in regulating responses to a variety of abiotic stresses (e.g., low temperatures, drought, and high salt). For example, in a previous study, overexpression of the gene *OsMYB4* in *Arabidopsis thaliana* led to markedly enhanced cold tolerance (Vannini et al., 2004). By contrast, overexpression of *VcMYB4a* in blueberry led to increased cold sensitivity, suggesting that MYB genes may negatively regulate cold tolerance through different signaling pathways (Zhang et al., 2020). *ZmWRKY106* in maize (*Zea mays*) is induced by drought stress, and its promoter region contains C-repeat/dehydration response element (DRE), low-temperature response element (LTR), and other important stress response elements. Both high-temperature stress and exogenous ABA can significantly induce the expression of *ZmWRKY106* through the ABA signal pathway (Wang et al., 2018). In addition, the AP2/ERF genes, such as *BpERF13*, are also related to cold stress; overexpressed (OE) transgenic lines of *Betula platyphylla* upregulate CBF genes and mitigate reactive oxygen species (ROS) under cold treatment (Lv et al., 2020). *CaNAC064*, in the NAC family, is a crucial regulator of cold stress tolerance in peppers, whereas *PbeNAC1* plays an important role in improving the cold tolerance of *Pyrus betulifolia* in cold environments (Jin et al., 2017; Hou et al., 2020). Our transcriptomic analysis revealed that the majority of DEGs between HX and HS during the later stages of cold stress belong to the MYB, WRKY, NAC, and AP2 families, confirming that cold resistance in *C. oleifera* is regulated by multiple TF genes, and that genes from these four families play key roles in cold resistance. Further research is required to understand the cold resistance mechanisms of TFs in *C. oleifera*.

4.2 Plant hormone signal transduction pathways involved in cold resistance

Among the complex mechanisms by which plants adapt to cold stress, the hormone system is critical (Zhao et al., 2021). Growth regulators include auxins, GA, cytokinin, ABA, ET, SA, JA, and BR. Of these, ABA and auxin play a key role in the signal transduction processes involved in plant resistance to abiotic stress (Eremina et al., 2016). ABA plays an important role in mediating cold perception and promoting cold tolerance in *Populus euphratica* (Chen et al., 2014) Furthermore, exogenous ABA can induce various cold tolerance mechanisms in the cells and seedlings of cold-sensitive rice (Shinkawa et al., 2013). Our hormone content measurements indicate that ABA is a significant component of the cold stress responses of *C. oleifera*. ABA is essential for plant growth and development, and response mechanisms to cold stress may be either ABA-dependent or ABA-independent (Tuteja, 2007). The ABA receptors *PP2C* and *SnRK2* are important regulators that can play either positive or negative regulatory roles in different plant species (Yang et al., 2017). Both *PP2C* and *SnRK2* were differentially expressed in *C. oleifera*, indicating that the species' cold signal transduction pathway may be ABA-dependent. Moreover, the expression levels of *PYR/PYL*, *PP2C*, and *SnRK2* and the content of ABA were higher in HX than in HS after 25 days of cold stress, indicating that the ABA signal transduction pathway plays a larger role in cold signal transmission in HX.

Recent studies have found that auxin is heavily involved in the cold stress responses of plants. For example, a large number of auxin-regulated genes in *Brassica napus* are induced by cold stress (Guan et al., 2019). Genes encoding Aux1, GH3, and SaUR are significantly upregulated in both *Capsicum pubescens* (a cold-tolerant pepper) and *C. chinense* (a cold-sensitive pepper) after 12 h of cold treatment; however, the expression of these genes is significantly higher in *C. pubescens* (Gao et al., 2022). Similarly, the DEGs encoding AUX/IAAs are both up- and downregulated at 4°C in a cold-tolerant rice cultivar (Zhao et al., 2015). In our analyses, two AUX/IAA genes were downregulated during the long-term cold treatment, and five were upregulated. These results corroborate those of other studies, and imply that auxin-related genes are associated with cold tolerance in *C. oleifera*. Furthermore, most genes exhibited higher expression in HX than in HS after 25 days of cold stress. Conversely, context of auxins such as IAA exhibited contrasting trends in the two cultivars, suggesting that auxin-responsive genes may negatively regulate the synthesis of auxin, and that auxin hormone signal transduction is more active in HX than HS.

4.3 Starch and sucrose metabolism in response to cold stress

Carbohydrate metabolism is critical to plant growth and tolerance of environmental stress (Sharma et al., 2021). Remobilization of starch and sucrose releases energy, sugars, and derived metabolites, which helps to alleviate abiotic stress, and is a fundamental process in plant adaptation (Thalmann and Santelia, 2017). In this process, key enzyme genes involved in starch and sucrose metabolism, including β-glucosidase genes (*Bglus*), β-mylase genes (*BAMYs*), hexokinase genes (*HXKs*), sucrose phosphate synthetase genes (*SPSs*), and sucrose invertase genes (*INVs*), contribute to cold stress responses by regulating osmotic adjustment ability *via* the accumulation of soluble sugars, or by inducing the expression of cold resistance genes and key enzyme genes associated with the antioxidant system (Cao et al., 2014). For example, the beta-glucosidase gene in kiwi fruit is induced by cold stress and exhibits significantly different expression patterns under cold treatments. Similarly, the transcription levels of the glucosidase gene in chickpea (*Cicer arietinum*) increase considerably under cold stress, thus conferring significantly increased resistance (Khazaei et al., 2015; Sun et al., 2021).

Furthermore, *BAMY* genes play a key role in plant responses to cold stress by degrading starch and regulating the accumulation of soluble sugars (Ma et al., 2022). For example, overexpression of *PbrBAM3* in tobacco (*Nicotiana tabacum*) and pear (*Pyrus ussuriensis* and *P. betulifolia*) increase *BAM* activity and thereby have a positive effect on cold tolerance (Zhao et al., 2019; Liang et al., 2021). *HXK*-related genes play important roles in sugar sensing and signal transduction. *HXK*s are more significantly induced in a cold-tolerant cultivar of *Jatropha curcas* (SCZ) than in the cold-sensitive cultivar YH9. In addition, overexpression of *SPS* and *INV* contribute to cold tolerance by influencing sugar accumulation (Bhowmik et al., 2006; Dahro et al., 2016).

In our analyses, several key enzyme genes related to starch and sucrose metabolism, including *INV*s, *HXK*s, *FRK*s, *SPS*s, *Bglu*s, and *BAMY*s, were involved in cold responses. Most of these genes were induced in HX, and their expression was higher in HX than in HS after 25 days of cold stress, which is consistent with changes observed in the contents of sucrose, D-fructose, and glucose. These results indicate that differences in the cold responses of HS and HX may be attributable to differences in sugar accumulation that are regulated by multiple genes. Moreover, the accumulation of sugars in the flower buds was more pronounced in HX than in HS after 25 days of treatment, potentially providing energy to the flowers of HX and preventing cold stress from affecting flowering. This may explain why the buds of HX opened normally whereas those of HS opened more slowly or not at all, and indicates that HX flower buds are less affected by cold stress. This finding may contribute to improving cold tolerance in *C. oleifera*.

4.4 Phenylpropanoid biosynthesis involved in cold resistance

Our integrated transcriptomic and metabolomic analysis revealed significant enrichment of DEGs and DEMs in the phenylpropanoid biosynthesis pathway. Phenylpropanoid metabolism is among the most important metabolic pathways in plants, and the metabolites from this pathway affect development and plant–environment interactions (Dong and Lin, 2021). Cold stress causes excess accumulation of ROS in the cell membrane, leading to cell damage. Phenolic compounds use electrons and hydrogen atoms to flush ROS and prevent lipid peroxidation (Rezaie et al., 2020; Devireddy et al., 2021). Phenolic biosynthesis responds to cold stress by enhancing the expression of key genes encoding phenylalanine ammonia-lyase (PAL), cinnamyl-alcohol dehydrogenase (CAD), and hydroxycinnamate transferase (HCT) (Sharma et al., 2019). Integrated transcriptomic and metabolomic analyses have demonstrated that genes related to PAL and coniferin exhibit higher expression in the cold-tolerant peanut (*Arachis hypogaea*) cultivar SLH than in the cold-sensitive cultivar ZH12 (Wang et al., 2021). In addition, the expression of *PAL* genes is higher in a cold-tolerant variety of Tartary buckwheat (TM) than in the sensitive variety RG (Raza et al., 2021). Sinapyl alcohol is among the precursors of lignin, which is associated with water and solute transport and cell wall rigidity, and contributes to abiotic stress resistance (Lee et al., 2021). *Hosta ventricosa* enhances its cold resistance by adjusting the ratio of sinapyl alcohol to coniferyl alcohol, thereby altering the morphological structure of the cell wall (Zhuang et al., 2021). Furthermore, sinapyl alcohol is significantly upregulated in the tolerant *Zea mays* cultivar B144 (Yu et al., 2022). In addition, succinate accumulates when plants are exposed to cold stress (Song et al., 2016; Xie et al., 2022). We found that DEMs, including coniferin, coniferyl alcohol, sinapyl alcohol, and succinate, along with related genes, exhibited similar positive trends under cold stress, and that these genes were more strongly expressed in HX than in HS after 25 days of cold treatment. The results indicate that the aforementioned genes and metabolites are associated with improved cold tolerance in the flowers buds of *C. oleifera*. Moreover, transcriptional regulation and metabolism were more active in HX flower buds under long-term cold stress, which might explain why HX flower buds opened normally whereas most HS flower buds delayed opening or did not open at all. This suggests that the flower buds of HX are less affected by cold stress than those of HS. While this contradicts the results of a previous study (Wu et al., 2020a), the contradiction is likely attributable to differences in test materials: that study used *C. oleifera* leaves, whereas we used flower buds, and vegetative organs (leaves) and reproductive organs (flower buds) differ in their responses to cold stress. Therefore, cold resistance in different *C. oleifera* cultivars may vary among tissues and organs.

5 Conclusion

Cold stress negatively impacts plant growth and development and leads to changes at the phenotypic, physiological, and molecular levels. In this study, differences in response mechanisms in the flower buds of two *C. oleifera* cultivars (HS and HX) under cold stress were investigated from physiological, transcriptomic, and metabonomic perspectives. Metabolites, including coniferin, coniferyl alcohol, succinate, and sinapyl alcohol, accumulated significantly in response to the cold treatments, and expression of related genes increased. Effects were more pronounced in HX than in HS during the later stages of the treatments, which may explain why the flower buds of HX opened normally whereas most HS flower buds either delayed opening or failed to open at all. Our results indicate that HX maintains higher levels of carbohydrate metabolism in the flower buds than does HS, and that the flower buds of HX are less strongly affected by cold stress. The cold resistance of *C. oleifera* cultivars varies among tissues and organs. The genes and metabolic processes identified in this study provide valuable information for future molecular breeding.

Data availability statement

The data presented in the study are deposited in the Figshare repository, and the link is 10.6084/m9.figshare.21775748.

Author contributions

Y-JW and L-LW have contributed equally to this work and share first authorship. Y-JW analyzed the data as well as and writing the original draft, L-LW designed and performed the experiments,

writing review and revising, as well as funding acquisition. M-HS and ZL writing review and edited the manuscript and J-AL and X-FT revised and proofread the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Hunan Provincial Natural Science Fund [grant number 2021JJ41073], the Scientific research projects of Education Department of Hunan Province [grant number 21B0263] and the Scientific research projects of Education Department of Hunan Province [grant number 19B583].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1126660/full#supplementary-material

SUPPLEMENTARY TABLE 1
Absolute quantification of 13 sugars and three plant hormones detected in HS and HX at different times of cold stress.

SUPPLEMENTARY TABLE 2
DEGs detected in HS and HX at different times of cold stress.

References

Bhowmik, P. K., Tamura, K., Sanada, Y., Tase, K., and Yamada, T. (2006). Sucrose metabolism of perennial ryegrass in relation to cold acclimation. *Z Naturforsch. C J. Biosci.* 61 (1-2), 99–104. doi: 10.1515/znc-2006-1-218

Cao, Y.-Y., Yang, M.-T., Li, X., Zhou, Z.-Q., Wang, X.-J., and Bai, J.-G. (2014). Exogenous sucrose increases chilling tolerance in cucumber seedlings by modulating antioxidant enzyme activity and regulating proline and soluble sugar contents. *Scientia Hortic.* 179, 67–77. doi: 10.1016/j.scienta.2014.09.016

Chen, J., Tian, Q., Pang, T., Jiang, L., Wu, R., Xia, X., et al. (2014). Deep-sequencing transcriptome analysis of low temperature perception in a desert tree, populus euphratica. *BMC Genomics* 15, 326. doi: 10.1186/1471-2164-15-326

Dahro, B., Wang, F., Peng, T., and Liu, J. H. (2016). *PtrA/NINV*, an alkaline/neutral invertase gene of *Poncirus trifoliata*, confers enhanced tolerance to multiple abiotic stresses by modulating ROS levels and maintaining photosynthetic efficiency. *BMC Plant Biol.* 16, 76. doi: 10.1186/s12870-016-0761-0

Devireddy, A. R., Tschaplinski, T. J., Tuskan, G. A., Muchero, W., and Chen, J. G. (2021). Role of reactive oxygen species and hormones in plant responses to temperature changes. *Int. J. Mol. Sci.* 22 (16), 8843. doi: 10.3390/ijms22168843

Dong, N. Q., and Lin, H. X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *J. Integr. Plant Biol.* 63 (1), 180–209. doi: 10.1111/jipb.13054

Dong, W., Ma, X., Jiang, H., Zhao, C., and Ma, H. (2020). Physiological and transcriptome analysis of *Poa pratensis* var. *anceps* cv. qinghai in response to cold stress. *BMC Plant Biol.* 20 (1), 362. doi: 10.1186/s12870-020-02559-1

Eremina, M., Rozhon, W., and Poppenberger, B. (2016). Hormonal control of cold stress responses in plants. *Cell Mol. Life Sci.* 73 (4), 797–810. doi: 10.1007/s00018-015-2089-6

Gao, C., Mumtaz, M. A., Zhou, Y., Yang, Z., Shu, H., Zhu, J., et al. (2022). Integrated transcriptomic and metabolomic analyses of cold-tolerant and cold-sensitive pepper species reveal key genes and essential metabolic pathways involved in response to cold stress. *Int. J. Mol. Sci.* 23 (12), 6683. doi: 10.3390/ijms23126683

Gao, C., Yuan, D. Y., Wang, B. F., Yang, Y., Liu, D. M., and Han, Z. Q. (2015). A cytological study of anther and pollen development in *Camellia oleifera*. *Genet. Mol. Res.* 14 (3), 8755–8765. doi: 10.4238/2015.July.31.24

Guan, S., Xu, Q., Ma, D., Zhang, W., Xu, Z., Zhao, M., et al. (2019). Transcriptomics profiling in response to cold stress in cultivated rice and weedy rice. *Gene* 685, 96–105. doi: 10.1016/j.gene.2018.10.066

Hou, X. M., Zhang, H. F., Liu, S. Y., Wang, X. K., Zhang, Y. M., Meng, Y. C., et al. (2020). The NAC transcription factor *CaNAC064* is a regulator of cold stress tolerance in peppers. *Plant Sci.* 291, 110346. doi: 10.1016/j.plantsci.2019.110346

Jin, C., Li, K. Q., Xu, X. Y., Zhang, H. P., Chen, H. X., Chen, Y. H., et al. (2017). A novel NAC transcription factor, *PbeNAC1*, of *Pyrus betulifolia* confers cold and drought tolerance *via* interacting with *PbeDREBs* and activating the expression of stress-responsive genes. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01049

Khan, S. A., Li, M. Z., Wang, S. M., and Yin, H. J. (2018). Revisiting the role of plant transcription factors in the battle against abiotic stress. *Int. J. Mol. Sci.* 19 (6), 1634. doi: 10.3390/ijms19061634

Khazaei, M., Maali-Amiri, R., Talei, A. R., and Ramezanpour, S. (2015). Differential transcript accumulation of *dhydrin* and *beta-glucosidase* genes to cold-induced oxidative stress in chickpea. *J. Agric. Sci. Technol.* 17 (3), 725–734. Available at: https://jast.modares.ac.ir/article-23-6072-en.html.

Kumar, A. A., Mishra, P., Kumari, K., and Panigrahi, K. C. (2012). Environmental stress influencing plant development and flowering. *Front. Biosci. (Schol Ed)* 4 (4), 1315–1324. doi: 10.2741/s333

Lee, C. J., Kim, S. E., Park, S. U., Lim, Y. H., Choi, H. Y., Kim, W. G., et al. (2021). Tuberous roots of transgenic sweetpotato overexpressing *IbCAD1* have enhanced low-temperature storage phenotypes. *Plant Physiol. Biochem.* 166, 549–557. doi: 10.1016/j.plaphy.2021.06.024

Li, N. N., Yue, C., Cao, H. L., Qian, W. J., Hao, X. Y., Wang, Y. C., et al. (2018). Transcriptome sequencing dissection of the mechanisms underlying differential cold sensitivity in young and mature leaves of the tea plant (*Camellia sinensis*). *J. Plant Physiol.* 224-225, 144–155. doi: 10.1016/j.jplph.2018.03.017

Liang, G., He, H., Nai, G., Feng, L., Li, Y., Zhou, Q., et al. (2021). Genome-wide identification of BAM genes in grapevine (*Vitis vinifera* l.) and ectopic expression of *VvBAM1* modulating soluble sugar levels to improve low-temperature tolerance in tomato. *BMC Plant Biol.* 21 (1), 156. doi: 10.1186/s12870-021-02916-8

Liu, L., Cheng, X., Zhao, W., Wang, Y., Dong, X., Chen, L., et al. (2018). Systematic characterization of volatile organic components and pyrolyzates from *Camellia oleifera* seed cake for developing high value-added products. *Arabian J. Chem.* 11 (6), 802–814. doi: 10.1016/j.arabjc.2017.12.031

Liu, J.-H., Peng, T., and Dai, W. (2013). Critical cis-acting elements and interacting transcription factors: Key players associated with abiotic stress responses in plants. *Plant Mol. Biol. Rep.* 32 (2), 303–317. doi: 10.1007/s11105-013-0667-z

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Lv, K., Li, J., Zhao, K., Chen, S., Nie, J., Zhang, W., et al. (2020). Overexpression of an *AP2/ERF* family gene, *BpERF13*, in birch enhances cold tolerance through upregulating *CBF* genes and mitigating reactive oxygen species. *Plant Sci.* 292, 110375. doi: 10.1016/j.plantsci.2019.110375

Ma, Y., Han, Y., Feng, X., Gao, H., Cao, B., and Song, L. (2022). Genome-wide identification of *BAM* (beta-amylase) gene family in jujube (*Ziziphus jujuba* mill.) and expression in response to abiotic stress. *BMC Genomics* 23 (1), 438. doi: 10.1186/s12864-022-08630-5

Proebsting, E. L., and Mills, H. H. (1978). Low temperature resistance of developing flower buds of six deciduous fruit Species1. *J. Am. Soc. Hortic. Sci.* 103 (2), 192–198. doi: 10.21273/jashs.103.2.192

Qin, S. Y., Rong, J., Zhang, W. J., and Chen, J. K. (2018). Cultivation history of *Camellia oleifera* and genetic resources in the Yangtze river basin. *Biodiversity Sci.* 26 (4), 384–395. doi: 10.17520/biods.2017254

Quan, W., Wang, A., Gao, C., and Li, C. (2022). Applications of Chinese *Camellia oleifera* and its by-products: A review. *Front. Chem.* 10. doi: 10.3389/fchem.2022.921246

Raza, A., Su, W., Hussain, M. A., Mehmood, S. S., Zhang, X., Cheng, Y., et al. (2021). Integrated analysis of metabolome and transcriptome reveals insights for cold tolerance in rapeseed (*Brassica napus* l.). *Front. Plant Sci.* 12, 5290. doi: 10.3389/fpls.2021.721681

Rezaie, R., Abdollahi Mandoulakani, B., and Fattahi, M. (2020). Cold stress changes antioxidant defense system, phenylpropanoid contents and expression of genes involved in their biosynthesis in *Ocimum basilicum* l. *Sci. Rep.* 10 (1), 5290. doi: 10.1038/s41598-020-62090-z

Ru, Z. Y., and Ju, L. Z. (2010). Effects of the climate during flowering period on post-flowering fruit setting in *Camellia oleifera*. *J. Zhejiang Forestry Coll.* 27(3), 323–328. Available at: https://kns.cnki.net/KXReader/Detail?invoice.

Salazar-Gutiérrez, M. R., Chaves, B., and Hoogenboom, G. (2016). Freezing tolerance of apple flower buds. *Scientia Hortic.* 198, 344–351. doi: 10.1016/j.scienta.2015.12.003

Sharma, K. D., Patil, G., and Kiran, A. (2021). Characterization and differential expression of sucrose and starch metabolism genes in contrasting chickpea (*Cicer arietinum* l.) genotypes under low temperature. *J. Genet.* 100 (2), 71. doi: 10.1007/s12041-021-01317-y

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules* 24 (13), 2452. doi: 10.3390/molecules24132452

Shinkawa, R., Morishita, A., Amikura, K., Machida, R., Murakawa, H., Kuchitsu, K., et al. (2013). Abscisic acid induced freezing tolerance in chilling-sensitive suspension cultures and seedlings of rice. *BMC Res. Notes* 6, 351. doi: 10.1186/1756-0500-6-351

Song, Y., Liu, L., Wei, Y., Li, G., Yue, X., and An, L. (2016). Metabolite profiling of *adh1* mutant response to cold stress in *Arabidopsis*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.02072

Soualiou, S., Duan, F., Li, X., and Zhou, W. (2022). CROP PRODUCTION UNDER COLD STRESS: An understanding of plant responses, acclimation processes, and management strategies. *Plant Physiol. Biochem.* 190, 47–61. doi: 10.1016/j.plaphy.2022.08.024

Sun, S., Lin, M., Qi, X., Chen, J., Gu, H., Zhong, Y., et al. (2021). Full-length transcriptome profiling reveals insight into the cold response of two kiwifruit genotypes (A. *arguta*) with contrasting freezing tolerances. *BMC Plant Biol.* 21 (1), 365. doi: 10.1186/s12870-021-03152-w

Susila, H., Nasim, Z., and Ahn, J. H. (2018). Ambient temperature-responsive mechanisms coordinate regulation of flowering time. *Int. J. Mol. Sci.* 19 (10), 3196. doi: 10.3390/ijms19103196

Tan, X. F., Yuan, D. Y., Yuan, J., Zou, F., Xie, P., Su, Y., et al. (2011). An elite variety: *Camellia oleifera* 'Huashuo'. *Scientia Silvae Sinicae* 47 (12), 184–209.

Tan, X. F., Yuan, D. Y., Zou, F., Yuan, J., Xie, P., Su, Y., et al. (2012). An elite variety of oiltea: *Camellia oleifera* 'Huaxin'. *Scientia Silvae Sinicae* 48 (3), 170–171.

Teixeira, A. M., and Sousa, C. (2021). A review on the biological activity of *Camellia* species. *Molecules* 26 (8), 2178. doi: 10.3390/molecules26082178

Thalmann, M., and Santelia, D. (2017). Starch as a determinant of plant fitness under abiotic stress. *New Phytol.* 214 (3), 943–951. doi: 10.1111/nph.14491

Tuteja, N. (2007). Abscisic acid and abiotic stress signaling. *Plant Signal Behav.* 2 (3), 135–138. doi: 10.4161/psb.2.3.4156

Tuteja, N., Gill, S. S., and Tuteja, R. (2011). Plant responses to abiotic stresses: Shedding light on salt, drought, cold and heavy metal stress. *Omics Plant abiotic Stress tolerance* 1, 39–64. doi: 10.2174/9781608050581111010039

Vannini, C., Locatelli, F., Bracale, M., Magnani, E., Marsoni, M., Osnato, M., et al. (2004). Overexpression of the rice *Osmyb4* gene increases chilling and freezing tolerance of *Arabidopsis* thaliana plants. *Plant J.* 37 (1), 115–127. doi: 10.1046/j.1365-313x.2003.01938.x

Wang, X., Liu, Y., Han, Z., Chen, Y., Huai, D., Kang, Y., et al. (2021). Integrated transcriptomics and metabolomics analysis reveal key metabolism pathways contributing to cold tolerance in peanut. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.752474

Wang, C. T., Ru, J. N., Liu, Y. W., Li, M., Zhao, D., Yang, J. F., et al. (2018). Maize WRKY transcription factor ZmWRKY106 confers drought and heat tolerance in transgenic plants. *Int. J. Mol. Sci.* 19 (10), 3046. doi: 10.3390/ijms19103046

Wu, L. L., Li, J. A., Gu, Y. Y., Zhang, F. H., Gu, L., Tan, X. F., et al. (2020b). Effect of chilling temperature on chlorophyll florescence, leaf anatomical structure, and physiological and biochemical characteristics of two *Camellia oleifera* cultivars. *Int. J. Agric. Biol.* 23 (4), 777–785. doi: 10.17957/Ijab/15.1352

Wu, L. L., Li, J. A., Wang, N., Gu, Y. Y., Zhang, F. H., and Tan, X. F. (2020a). The effects of low temperature stress on the flowering, fruiting and physiological characteristics of two camellia oleifera cultivars. *Plant Physiol. J.* 56 (4), 681–692. doi: 10.13592/j.cnki.ppj.2020.0061

Wu, L. L., Wang, Y. J., Guo, P. R., Li, Z., Li, J. A., and Tan, X. F. (2022). Metabonomic and transcriptomic analyses of *Camellia oleifera* flower buds treated with low-temperature stress during the flowering stage. *Ind. Crops Products* 189, 115874. doi: 10.1016/j.indcrop.2022.115874

Xie, H., Wang, Q., Zhang, P., Zhang, X., Huang, T., Guo, Y., et al. (2022). Transcriptomic and metabolomic analysis of the response of quinoa seedlings to low temperatures. *Biomolecules* 12 (7), 977. doi: 10.3390/biom12070977

Yang, W., Zhang, W., and Wang, X. (2017). Post-translational control of ABA signalling: the roles of protein phosphorylation and ubiquitination. *Plant Biotechnol. J.* 15 (1), 4–14. doi: 10.1111/pbi.12652

Yu, T., Zhang, J., Cao, J., Li, X., Li, S., Liu, C., et al. (2022). Metabolic insight into cold stress response in two contrasting maize lines. *Life (Basel)* 12 (2), 282. doi: 10.3390/life12020282

Zhang, C. Y., Liu, H. C., Zhang, X. S., Guo, Q. X., Bian, S. M., Wang, J. Y., et al. (2020). *VcMYB4a*, an R2R3-MYB transcription factor from *Vaccinium corymbosum*, negatively regulates salt, drought, and temperature stress. *Gene* 757, 144935. doi: 10.1016/j.gene.2020.144935

Zhang, F., Zhu, F., Chen, B., Su, E., Chen, Y., and Cao, F. (2022). Composition, bioactive substances, extraction technologies and the influences on characteristics of *Camellia oleifera* oil: A review. *Food Res. Int.* 156, 111159. doi: 10.1016/j.foodres.2022.111159

Zhao, B., Liu, Q., Wang, B., and Yuan, F. (2021). Roles of phytohormones and their signaling pathways in leaf development and stress responses. *J. Agric. Food Chem.* 69 (12), 3566–3584. doi: 10.1021/acs.jafc.0c07908

Zhao, L., Yang, T., Xing, C., Dong, H., Qi, K., Gao, J., et al. (2019). The beta-amylase *PbrBAM3* from pear (*Pyrus betulaefolia*) regulates soluble sugar accumulation and ROS homeostasis in response to cold stress. *Plant Sci.* 287, 110184. doi: 10.1016/j.plantsci.2019.110184

Zhao, J., Zhang, S., Yang, T., Zeng, Z., Huang, Z., Liu, Q., et al. (2015). Global transcriptional profiling of a cold-tolerant rice variety under moderate cold stress reveals different cold stress response mechanisms. *Physiol. Plant* 154 (3), 381–394. doi: 10.1111/ppl.12291

Zhuang, Q., Chen, S., Jua, Z., and Yao, Y. (2021). Joint transcriptomic and metabolomic analysis reveals the mechanism of low-temperature tolerance in *Hosta ventricosa*. *PloS One* 16 (11), e0259455. doi: 10.1371/journal.pone.0259455

Zinn, K. E., Tunc-Ozdemir, M., and Harper, J. F. (2010). Temperature stress and plant sexual reproduction: Usncovering the weakest links. *J. Exp. Bot.* 61 (7), 1959–1968. doi: 10.1093/jxb/erq053

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 March 2023
DOI 10.3389/fpls.2023.1143349

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Wuhan Botanical Garden (CAS), China

REVIEWED BY
Hao Chen,
Central South University Forestry and Technology, China
Yan Zhang,
Tobacco Research Institute (CAAS), China
Deng Lixi,
Moutai College, China

*CORRESPONDENCE
Long Wang
wanglong8591@hnu.edu.cn
Guoyun Xu
xgyun1018@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 12 January 2023
ACCEPTED 21 February 2023
PUBLISHED 07 March 2023

CITATION
Li C, Hong Y, Sun J, Wang G, Zhou H, Xu L, Wang L and Xu G (2023) Temporal transcriptome analysis reveals several key pathways involve in cadmium stress response in *Nicotiana tabacum* L. *Front. Plant Sci.* 14:1143349. doi: 10.3389/fpls.2023.1143349

COPYRIGHT
© 2023 Li, Hong, Sun, Wang, Zhou, Xu, Wang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Temporal transcriptome analysis reveals several key pathways involve in cadmium stress response in *Nicotiana tabacum* L.

Chenyang Li[1,2†], Yi Hong[1†], Jinhao Sun[3†], Guoping Wang[4], Huina Zhou[1], Liangtao Xu[3], Long Wang[2*] and Guoyun Xu[1*]

[1]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of China National Tobacco Corporation, Zhengzhou, China, [2]College of Biology, State Key Laboratory of Chemo/Biosensing and Chemometrics, and Hunan Province Key Laboratory of Plant Functional Genomics and Developmental Regulation, Hunan University, Changsha, China, [3]Technology Center, China Tobacco Jiangsu Industrial Co. Ltd., Nanjing, China, [4]Yuxi Zhongyan Tobacco Seed Co., Ltd., Yuxi, Yunnan, China

Tobacco has a strong cadmium (Cd) enrichment capacity, meaning that it can absorb large quantities from the environment, but too much Cd will cause damage to the plant. It is not yet clear how the plant can dynamically respond to Cd stress. Here, we performed a temporal transcriptome analysis of tobacco roots under Cd treatment from 0 to 48 h. The number of differentially expressed genes (DEGs) was found to change significantly at 3 h of Cd treatment, which we used to define the early and middle stages of the Cd stress response. The gene ontology (GO) term analysis indicates that genes related to photosynthesis and fatty acid synthesis were enriched during the early phases of the stress response, and in the middle phase biological process related to metal ion transport, DNA damage repair, and metabolism were enriched. It was also found that plants use precursor mRNA (pre-mRNA) processes to first resist Cd stress, and with the increasing of Cd treatment time, the overlapped genes number of DEGs and DAS increased, suggesting the transcriptional levels and post-transcriptional level might influence each other. This study allowed us to better understand how plants dynamically respond to cadmium stress at the transcriptional and post-transcriptional levels and provided a reference for the screening of Cd-tolerant genes in the future.

KEYWORDS

plant transcriptomics, cadmium stress, temporal transcriptome, alternative splicing, *Nicotiana tabacum* L.

1 Introduction

In recent years, heavy metal pollution of farmland soil has become a prominent environmental problem (Yang et al., 2018; Qin et al., 2021). China alone loses roughly 10 million tons of grain every year due to this heavy metal pollution (Shibao et al., 2019). According to the national survey bulletin on soil pollution, more than 19.4% of the cultivated soil was contaminated, with Cd-contaminated soil accounting for 7.0% (Yang et al., 2022). The accumulation of high levels of cadmium in a plant can result in significant toxicity, leading to a range of disruptions that can damage the plant. These disruptions may include impacts on cell division, decreased growth rate, disruption of photosynthesis, alteration of plant enzyme activity and membrane integrity, and an increase in lipid peroxidation (Haider et al., 2021).

As a very important cash economic in China and the world, the Cd enrichment capacity of tobacco (Nicotiana tabacum) is of great interest. Excessive accumulation of Cd reduces the quality and yield of tobacco. At present, research on the mechanisms of how Cd affects tobacco growth and development has made some processes, however, most of this research focuses on the antioxidant effects of Cd on tobacco (Borgo et al., 2022). There are very few studies analyzing how tobacco responds to Cd stress and which genes participate in Cd response at pre-mRNA and mRNA levels (Fu et al., 2019). In recent years, the continuous development of high-throughput sequencing technology has increased the ability to study how plants respond to heavy metal pollution. Temporal transcriptome sequencing in particular has proven to be an effective way to study the response of plants to external environmental signals. This method could dynamically monitor plant responses to different stages of stimuli exposure at the gene expression level. For example, it has previously been used to reveal how wheat can coordinate the expression of homologous genes to cope with various environmental constraints at the genome-wide level (Hao et al., 2021); how candidate genes regulate flower opening and closing in Iris, leading to an improved understanding of the regulatory network of flowering (Liu et al., 2022).

Alternative splicing (AS) is a typical post-transcriptional regulation. It is a molecular mechanism that produces multiple mRNA transcripts from a single gene locus *via* the alternative selection of splicing sites during pre-mRNA processing (Lee and Rio, 2015). Studies have shown that AS plays an important role in plant responses to a variety of abiotic stresses, such as salt (Jian et al., 2022), droughts (Ding et al., 2022), and temperature stress (Weng et al., 2021), but there are still few studies on the response to Cd stress. Therefore, it is of great significance to understand the mechanism of plant response to Cd by using technologies such as differential expression genes (DEGs) and alternative splicing analysis.

In this study, we analyzed the response of tobacco roots to Cd stress by using a temporal transcriptome analysis. It was found that the number of DEGs increase significantly after 3 hours of Cd treatment. In addition, the main signal pathways were found to differ at various time points of the Cd stress response and plants employed pre-mRNA process to initially resist Cd stress. Therefore, this study provides data to support a dynamic understanding of plant Cd stress responses at transcriptional and post-transcriptional levels.

2 Materials and methods

2.1 Plant materials and cadmium stress treatment

The experimental material was the roots of tobacco cukltivar K326, which was bred by Northup King Seed Company by crossing McNair30 and NC95. Tobacco seeds were germinated in sterile, moist vermiculite. Uniformly healthy 5-leaf stage seedlings were transplanted into 1/2 MS medium (100 mL pyridoxine flask containing 50 mL Murashige and Skoog medium liquid medium, with vitamins and sugar), and containers were randomly placed in a greenhouse under natural light irradiation at a controlled greenhouse temperature of 25 ± 3°C. Plants were treated with 50 μM $CdCl_2$ for 5 min, 15 min, 30 min, 1 h, 3 h, 6 h, 12 h, 24 h, and 48 h. At the end of the 48 h Cd treatment, tobacco roots, rinsed three times with deionized water and dried, and stored at -80°C. Samples without Cd treatment were labeled as T0 and samples under Cd treatment were labeled as T1-T9 depending on the treatment.

2.2 RNA-seq library construction

Samples from the different treatments were powdered in liquid nitrogen to extract RNA, and RNA-seq library was constructed as described previously (Wang et al., 2020). To perform RNA-seq, total RNA was first extracted from cells using a TRIzol reagent and treated with DNase to eliminate any genomic DNA contamination. The mRNA was then selectively captured using Oligo(dT) magnetic beads and fragmented using an interrupt reagent in a Thermomixer (Eppendorf, Hamburg, Germany). cDNA synthesis was carried out using the fragmented mRNA as template. Double-stranded cDNA was generated using a double-stranded cDNA synthesis reaction system and purified using a purification kit. The purified cDNA was then subjected to cohesive end repair, followed by the addition of an "A" base to the 3' end and ligation of a linker. cDNA fragments of the desired size were selected, and PCR amplification was performed. The resulting libraries were sequenced using paired-end sequencing on an Illumina Hiseq4000 (BGI, Shenzhen, China) following the manufacturer's protocol. In this study, thirty RNA libraries were generated, including three control libraries and twenty-seven treatment libraries, to investigate the effect of treatment on gene expression.

2.3 Bioinformatics analysis

Raw reads were analyzed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) prior to assembly. Trimmomatic v0.36 (http://usadellab.org/cms/?

page=trimmomatic), parameter "CROP:150 ILLUMINACLIP : TruSeq3-PE-2.fa:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:4" is used to remove low quality sequences. Cutadapt (Martin, 2011) software was used to remove low quality reads, including those with sequencing adapters, sequencing primers and nucleotides with mass fractions below 20. At the same time, Q20, Q30 and GC content were also calculated for clean data. All downstream analyses were based on high quality clean data. The raw sequence data was submitted to the NCBI database under the registration number PRJNA93401.

Reference genome and gene model annotation files were downloaded from the Nicotiana tabacum K326 Flue-cured (https://solgenomics.net/organism/Nicotiana_tabacum/genome). The index of the reference genome was built using Hisat2 (v2.2.1) (Kim et al., 2019) and paired-end clean reads were aligned to the reference genome using Hisat2 (v2.2.1). Gene expression analyses were constructed as described previously (Wang et al., 2020). We used DEGseq2 R package (1.36.0) (Love et al., 2014) and using the normalization method of quantiles with fold change ≥2, and *q-value* < 0.05 as the threshold for determining whether the gene was differentially expressed to obtain DEGs, normalization of data using Z-score.

Alternative splicing events were identified using rMATS v4.1.1 (Shen et al., 2014) with parameters "-t paired –len 150 –c 0.0001 –analysis U –libType fr-firststrand –novelSS 1" and with the updated GTF file. Five types of AS events were then classified, including ES, A5SS, A3SS, MXE, and IR. Inclusion level (ψ) was a quantitative measurement of AS. With exon skipping as an example, inclusion level is estimated by the proportion of exon-exon junction counts supporting the exon-inclusion isoform. Inclusion level can be applied to all AS categories (exon skipping, intron retention, alternative 5′/3′ splice site, mutually exclusive exons), with minor changes made when defining "inclusion". Parameters FDR < 0.05 and ΔPSI ≥ 0.1 were used to identify differential alternate splicing (DAS) events between the control and cadmium treatment groups.

2.4 Gene annotation and metabolic pathway analysis

Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the clusterProfiler (4.4.4) R package (Wu et al., 2021). GO terms with corrected *p-value* less than 0.05 were considered significantly enriched by differentially expressed genes. The clusterProfiler (4.4.4) R package was used to test the statistical enrichment analysis of differentially expressed genes in the KEGG pathway.

3 Results

3.1 Response of genes expression to cadmium stress in tobacco roots

To analyze how tobacco roots dynamically respond to cadmium stress, we treated 5-leaf stage *Nicotiana tobacco* with 50 μM cadmium ions and collected samples at different time points (0, 5 min, 30 min, 1 h, 3 h, 6 h, 24 h, and 48 h) for RNA-seq (Figure 1A). We conducted three biological replicates per sample. A time-series differential expression analyses were performed to explore the global temporal patterns of transcriptomic changes. We generated at least 10 Gb of clean bases for each sample, and more than 85% of the sequencing reads could be uniquely mapped to the *N. tabacum* K326 genome, indicating the high quality of the RNA-seq data (Table S1). Correlation analyses were used to estimate the relationships between the 30 transcriptome samples. Our results showed that all replicated data was generally well correlated within the same time point (T1-T9; Figure 1B).

We then identified the DEGs occurring in the root of K326 following exposure to Cd stress treatment. We monitored how these changes occurred over different time and compared them with the untreated control (T0). Our results identified a total of 64,472 DEGs by using the edgeR package with *q-value* < 0.05, fold change ≥2. Compared with the T0 group, within 5 minutes there were 601 up-regulated genes in the Cd stress-treated group, while 916 DEGs were down-regulated in the T1 vs T0 Cd stress-treated group (Figure 1B). The number of DEGs increased rapidly, as T1 vs T0 group contained almost 1/3 of those in the T2 vs T0 group (Figure 1B). This degree of variation suggests that the mRNA response strategy has achieved a certain scale after 15 min of Cd treatment. The number of DEGs responding to Cd stress increased significantly after 3 h compared to 1 h (Figure 1C), and the highest number of DEGs were detected in the 48 h samples. Based on these results, the treatment time points were divided into early (T0 to T4) and mid-term (T5 to T9) responses to Cd stress. Additionally, we detected 274 overlapped genes in the Venny analysis of the early stage of Cd stress, which reached 5.5-18% of the total DEGs (Figure 1D). We also detected 2,051 overlapped genes in the Venny analysis of the mid-term stage of Cd stress, which reached 15-22% of total DEGs (Figure 1E). In summary, the number of DEGs in the middle stages of Cd treatment was significantly increased compared with that in the early stage of treatment. The proportion of overlapping genes is significantly higher in the mid-stage compared to the early stage, indicating that more common genes are synergistically involved in the cadmium response in the mid-stage.

3.2 GO term and KEGG analysis in the early stage of Cd treatment

To investigate the response mechanisms in plants at the early stage of Cd stress treatment, we analyzed the DEGs from T0 to T4. The DEGs heatmap show that the number of up- and down-regulated genes from the 274 overlapped genes is similar between T1-T4 samples (Figure 2A). During our experiment, a gene ontology enrichment analysis revealed that the 274 overlapped genes were mainly involved in the rhythmic process, response to abiotic stimulus, metal ion transport, circadian rhythm, and cation transport et al. (Figure 2B). Circadian rhythms are essential for normal gene expression, indicating that Cd treatment could induce a disorder in the expression of rhythm-related genes. KEGG

10.3389/fpls.2023.1143349

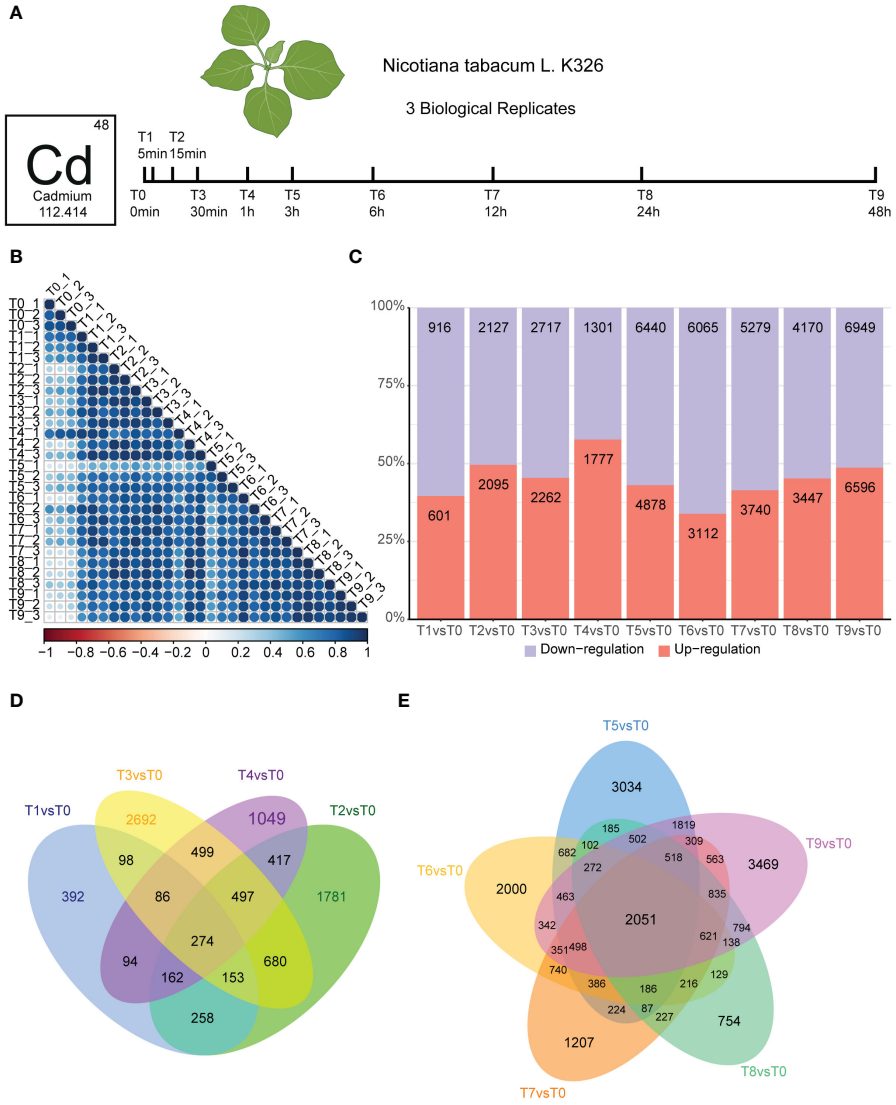


FIGURE 1
Experiment overview and effect of Cd stress response in tobacco root. **(A)**.Seedlings were incubated in a growth chamber until they grew to the five-leaf stage, then the roots of seedings were treated with 50 μM $CdCl_2$ for different times. Three biological replicates of *Nicotiana tabacum* ecotype K326 were harvested. **(B)**. Heatmap depicting pairwise Pearson correlation of gene expression values of all samples. **(C)**. Bar graph showing the total number of differentially upregulated (orange) and downregulated (blue) genes in T1vsT0, T2vsT0, T3vsT0, T4vsT0, T5vsT0, T6vsT0, T7vsT0, T8vsT0, and T9vsT0 samples. **(D)**. Venn diagram showing common and unique genes in T1vsT0, T2vsT0, T3vsT0, and T4vsT0 samples. The number of genes intersecting all time points was 274. **(E)**. Venn diagrams showing differential gene statistics for T5vsT0, T6vsT0, T7vsT0, T8vsT0, and T9vsT0 samples. The number of genes intersecting all time points was 2,051.

analysis indicated that in response to the early stage of Cd exposure, the overlapped DEGs enriched photosynthesis, MAPK signaling pathway, TCA cycle, and carbon metabolism (Figure 2C). These results demonstrate that Cd stress affects the photosynthesis of plants, leading to a decline in the overall photosynthetic rate. Moreover, other corresponding energy metabolism processes must also be affected. Therefore, the continuously changing, overlapping genes are mostly related to energy metabolism, metal ion transport, and abiotic stimulus response.

Next, we analyzed the GO term of all multiple specific response pathways at each time point in the early stage. The results showed that the photosynthesis, lipid biosynthetic process, and fatty acid biosynthetic processes were enriched after 5 minutes of Cd treatment (Figures 2D, E). After 15 minutes of Cd treatment, the pathways related to cell wall organization and response to biotic stimulus were affected, and after an hour, the plant began to respond to oxidative stress (Figures 2D, E). A similar conclusion was found by analyzing the KEGG pathway which affected the photosynthesis and fatty acid elongation after just 5 minutes of treatment. After 15 min of treatment, plant hormone signal transduction, MAPK signaling pathway, and phenylpropanoid biosynthesis were also enriched (Figure S1A). In summary, the main pathways involved in Cd stress response occur at different stages. Several signal pathways are continuously affected after Cd treatment such as photosynthesis and lipid biosynthesis. However, more signal pathways have a spatio-temporal specific response.

10.3389/fpls.2023.1143349

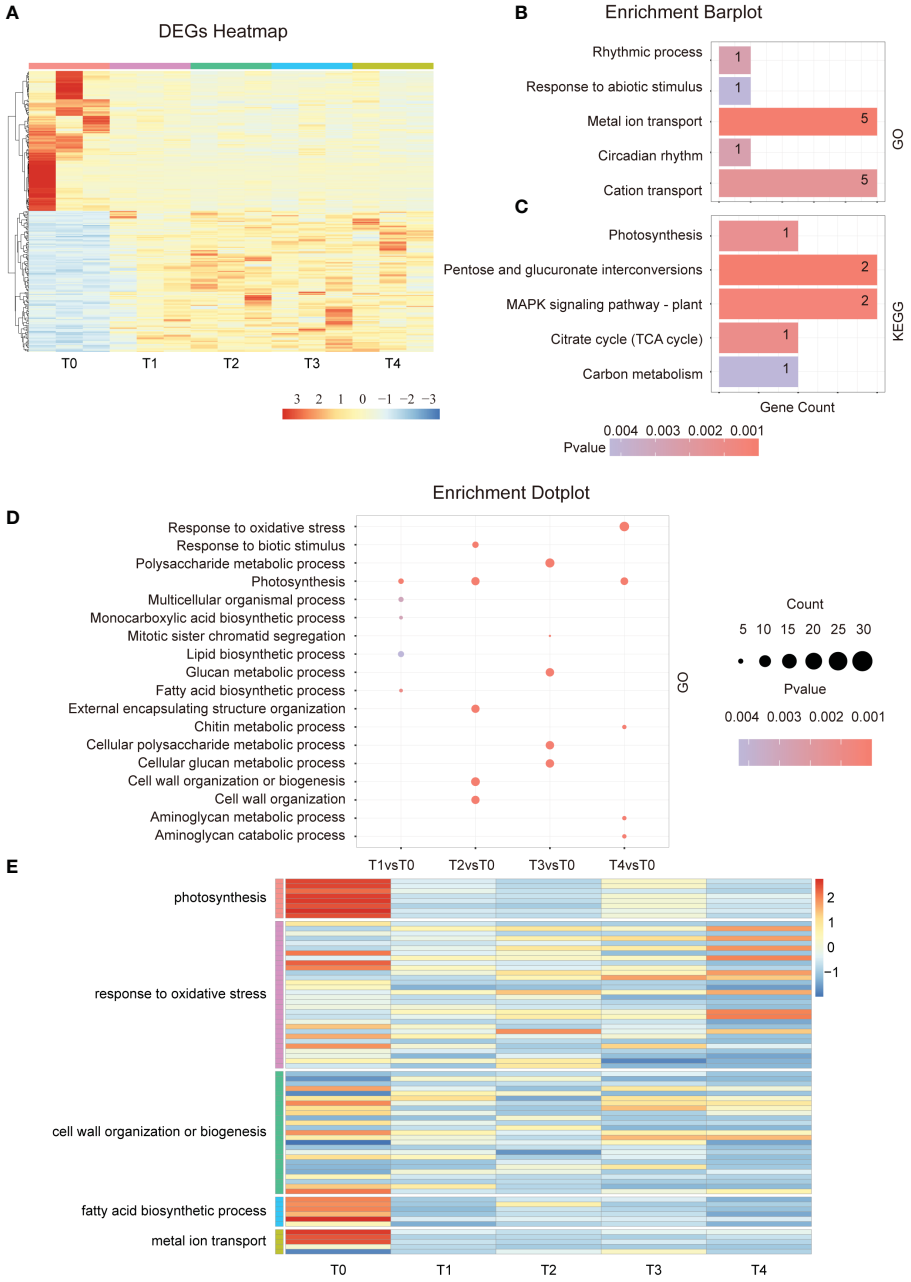


FIGURE 2
Differential gene expression in the early stage of Cd treatment. **(A)**. Hierarchical clustering heat map showing the gene expressions of a total of 15 samples at each time point during the early phase of Cd stress treatment (T0-T4). **(B)**. GO term enrichment analysis of common DEGs after cadmium treatment (T1vsT0, T2vsT0, T3vsT0, and T4vsT0). The top 5 entries were selected for shown. **(C)**. KEGG analysis of common DEGs after cadmium treatment (T1vsT0, T2vsT0, T3vsT0, and T4vsT0). The top 5 entries were selected for statistical analysis. **(D)**. GO enrichment results of differentially expressed genes at each time point (T1-T4) in the early phase of the Cd stress response. The top 3 entries for each time significance were selected for statistical analysis. **(E)**. The heat map shows the gene expression in photosynthesis, response to oxidative stress, cell wall organization or biogenesis, fatty acid biosynthetic process, and metal ion transport at the early stage of cadmium stress (T0-T4).

3.3 GO term and KEGG analysis in the middle stage of Cd treatment

To investigate the response mechanisms at the middle stage of the Cd stress response, we analyzed the RNA-seq transcriptomic data of tobacco under Cd treatment from T5 to T9. The results displayed in the DEGs heatmap showed that the number of up-regulated genes is about three times that of down-regulated genes in the 2,051 overlapped genes (Figure 3A). Among these genes, GO enrichment analysis showed that cell wall modification, DNA integration, cell wall organization, polysaccharide biosynthetic process, and metal ion transport were enriched (Figure 3B). Likewise, the KEGG analysis showed that the RNA transport, nucleotide excision repair, MAPK signaling pathway, and endocytosis pathway were also enriched (Figure 3C). Moreover, the GO term analysis revealed that nuclear division, chromosome

10.3389/fpls.2023.1143349

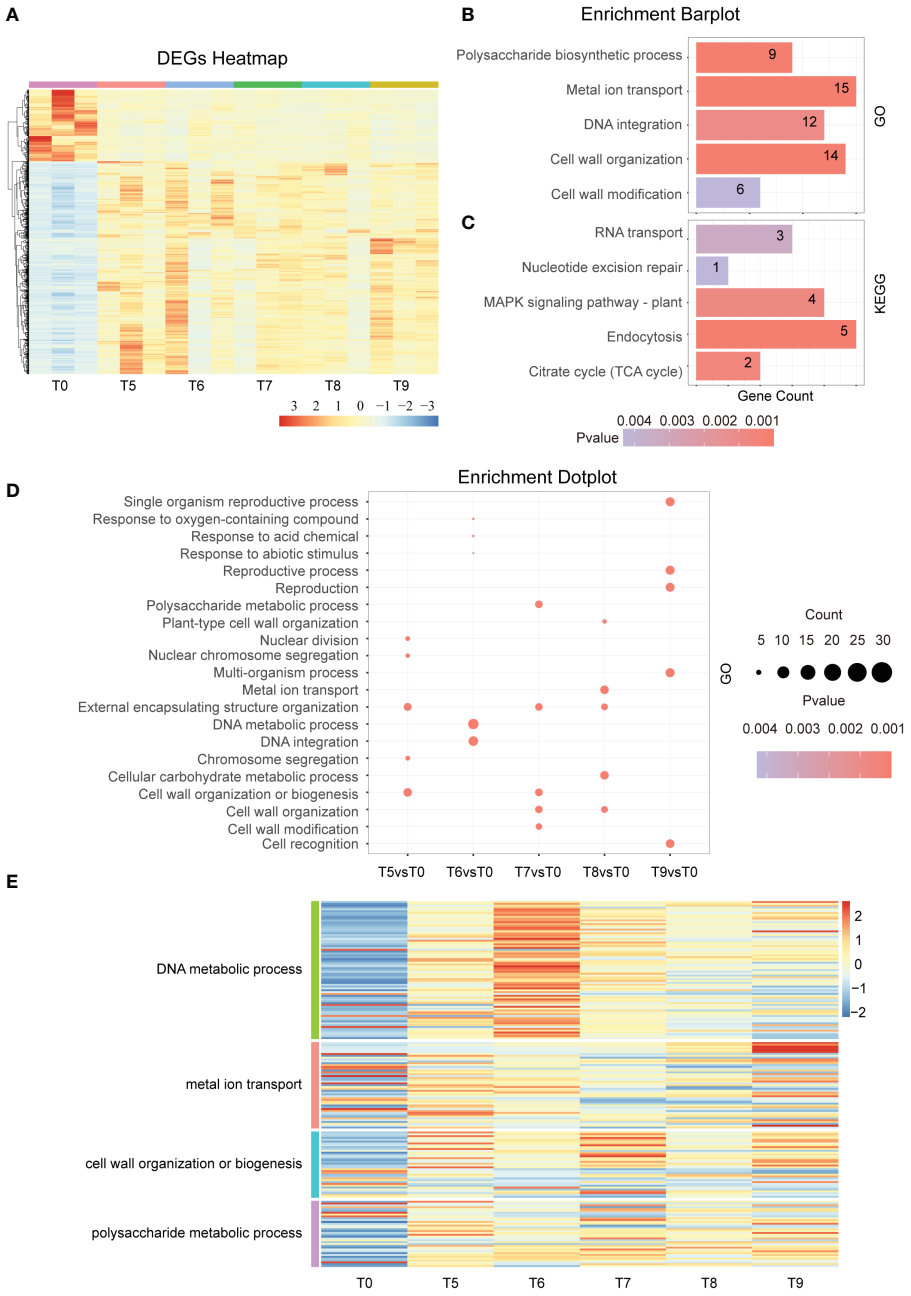


FIGURE 3
Differential gene expression in the middle stage of Cd treatment. **(A)**. Hierarchical clustering heat map showing the gene expression of a total of 18 samples at each time point during the middle phase of Cd stress treatment (T0 and T5-T9). **(B)**. GO term enrichment analysis of common DEGs after Cd treatment (T5vsT0, T6vsT0, T7vsT0, T8vsT0, and T9vsT0). The top 5 entries for each time significance were selected for statistical analysis. **(C)**. KEGG analysis of common DEGs after cadmium treatment (T5vsT0, T6vsT0, T7vsT0, T8vsT0, and T9vsT0). The top 5 entries for each time significance were selected for statistical analysis. **(D)**. GO enrichment results of differentially expressed genes at each time point (T5-T9) in the middle phase of the Cd stress response. The top 3 entries for each time significance were selected for statistical analysis. **(E)**. The heat map shows the gene expression in DNA metabolic process, metal ion transport, cell wall organization or biogenesis, and polysaccharide metabolic process at the middle stage of cadmium stress (T5-T9).

segregation, and cell wall organization or biogenesis were enriched at T5 vs T0. The GO term related to DNA metabolic processes and DNA integration were enriched in T6 vs T0 (Figures 3D, E). Similarly, signal pathways changed as time continued, especially at 48h of Cd treatment where it was noted that the main enrichment pathway becomes more involved in reproductive processes and cell recognition (Figures 3D, E). This led to a significant difference between the main KEGG pathways after 48 h and all other time points (Figure S1B). These results indicate that cell walls are disrupted by continuous Cd stress and that the early responses to exogenous Cd and abiotic stimuli are gradually converted into pathways related to DNA damage and cell well repair.

10.3389/fpls.2023.1143349

3.4 Transcriptional dynamics of cadmium stress response in tobacco roots

To analyze gene expression patterns, a time-course analysis was conducted by clustering all genes from different time points to investigate their expression dynamics. Thousands of Cd-response genes were classified into 8 different expression patterns based on their response to the Cd treatment (Figures 4A–H). The expression of genes in cluster 1 was the highest at T1 and gradually decreased over time. GO analysis indicated that these genes are involved in photosynthesis, multicellular organism development, and development processes. This indicates that Cd stress first destroys a plant's photosynthesis which leads to an effect on plant growth. Genes in clusters 2 and 3 showed similar expression patterns (Figures 4B, C), exhibiting a rapid response to Cd stress at 30 min and 1h, respectively. The GO analysis indicates that these genes are involved in sucrose metabolic processes, amino acid activation, reactive oxygen species, the generation of precursor

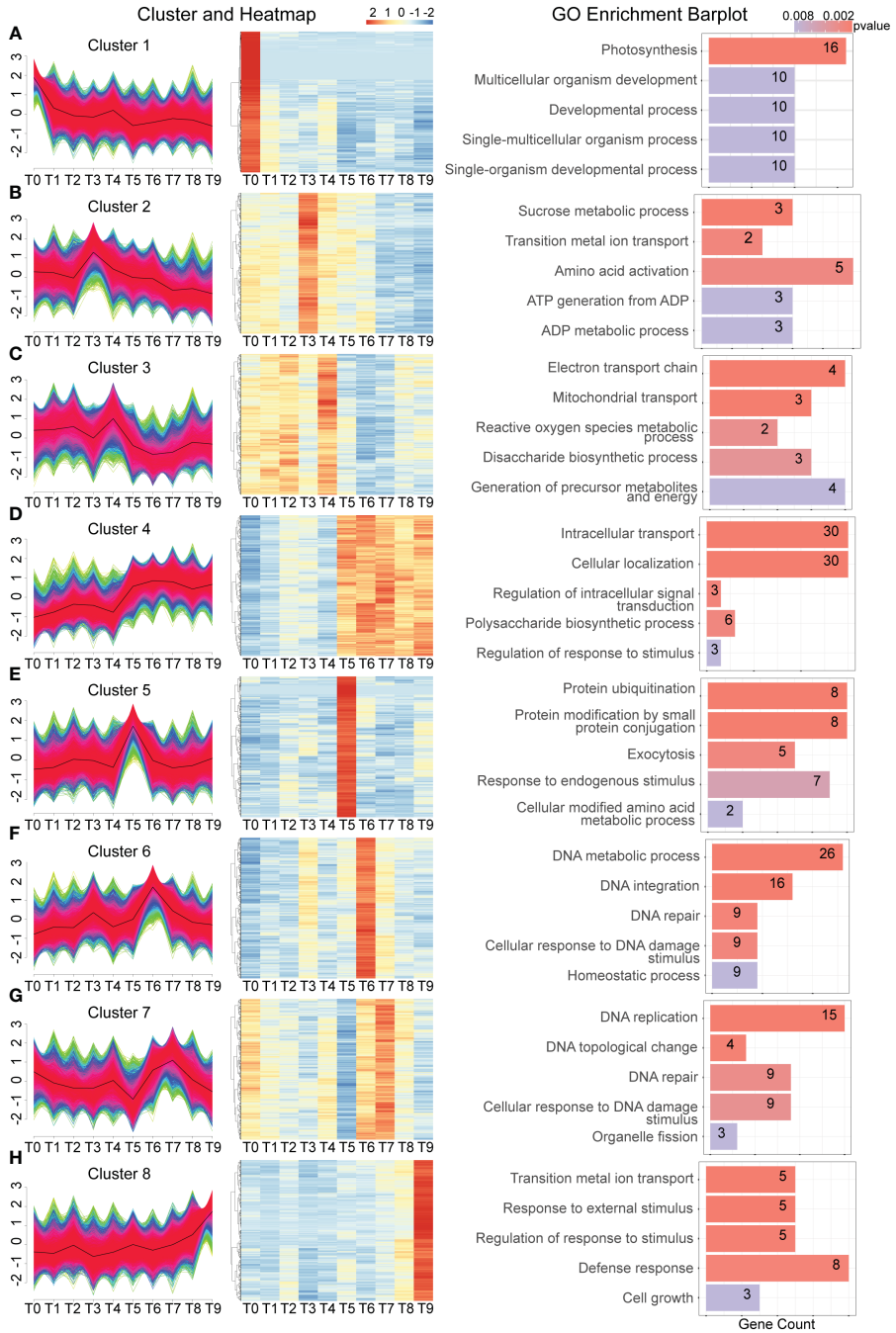


FIGURE 4
Time-course analysis of dynamic gene expression changes during Cd treatment. **(A–H)**. From left to right, gene expression is divided into 8 significant clusters based on the similarity of gene expression in each sample. The right side is a heat map of the expression of the top 1,000 genes most relevant to each cluster in each sample, graphically showing the top 5 significant GO terms.

metabolites, and energy, meaning that a plant likely produces energy to resist Cd stress after exposure.

Genes in clusters 5 and 6 also followed this expression pattern. At 3 h after Cd treatment, both the protein ubiquitination and response to endogenous stimulus were enriched. GO analysis indicated that genes enriched DNA metabolic process, DNA integration, DNA repair, and the cellular response to DNA damage stimulus at 6 h (T6) following Cd treatment (Figures 4D–F). It has been reported that in response to Cd stress, many genes with repair functions become active to mend the damaged parts. A similar function was observed at the T7 time point, where many genes with repair functions became active due to the altered cell structure (Figure 4G). Due to the persistence of cadmium stress, genes related to cell growth and defense response continued to be active at the T9 time point (Figure 4H). Collectively, these expression patterns illustrate the relationship similarities and differences between the early and middle stages of the Cd stress response.

3.5 Identification of different alternative splicing events in response to Cd stress

To identify the different alternative splicing (DAS) events that were sensitive to cadmium stress, AS analysis was performed using rMATS. A total of 3,367 DAS events from 1,883 genes were identified in tobacco roots. We found that similar to the sharp increase of DEGs in T5 vs. T0, the number of genes with AS increased significantly during this time point (Figure 5A). Among the DAS types between T1-T4, retained intron (RI) was the most abundant (38.40%), followed by alternative 3'SS (26.55%), alternative 5'SS (16.87%), exon skipping (15.24%), and mutually exclusive exons (2.94%) (Figure 5B). We identified 4,291 AS events from 3,519 genes over a total of five time periods from 3 h to 48 h. Among the AS types, RI was the most abundant (46.03%), followed by A3SS (19.70%), SE (15.68%), A5SS (14.29%), and MXE (4.31%). In addition, we identified 58 and 235 overlapped AS events in the early and middle stages of Cd stress, respectively (Figures 5C, D).

Similar enrichment analysis was performed for differential AS events in the early stage of cadmium stress treatment. It showed that cellular responses to stimulus were enriched in the T1 time point, which indicates that plants begin to respond to Cd stress by regulating the AS of stimulus-response genes 5 min after treatment. In addition, the AS events were also involved in RNA processing, RNA modification, DNA repair, and dephosphorylation processes in the early stage of Cd treatment (Figure 5E). Meanwhile, GO analysis showed enrichment in cellular responses to stress, DNA damage, and RNA modification in the T8 and T9 time points (Figure 5F). The same enrichment results are reflected in the KEGG pathway analysis, where the differential AS genes were mostly expressed in rhythm-related pathways and various amino acid synthesis pathways in the early stage of the stress response (Figure S2A). In the middle stage of the stress response, there was an enrichment of the mRNA surveillance pathway, N-Glycan biosynthesis, and spliceosome (Figure S2B).

To analyze the relationship between DEGs and DAS, we conducted a combined analysis. For the early stage (T0-T4) Cd stress response, DE and DAS genes differed greatly, with very few intersections. 77 intersecting genes were present in the middle phase (T5-T9). GO terms such as cellular response to stimulus, signal transduction, single organism signaling, and cell communication as well as some terms related to signaling and modifications were common to both DE and DAS genes. The two time points (T5 and T6) with the highest percentage of RI events were analyzed (Figures S3A, B). At 3 h of Cd stress treatment, the GO term showed gene enrichment in dephosphorylation which was associated with nucleotide anabolism (Figure S3C). After 6 h, the GO term showed genes associated with amino acid activation, polysaccharide synthesis, and tRNA anabolism (Figure S3D). In summary, we found that the response of AS to Cd stress is faster than that at transcriptional levels due to AS event that response to stimulus appears in the process of pre-mRNA processing within 5 min, and the transcriptional levels and post-transcriptional level might influence each other.

4 Discussion

To cope with the various environmental stresses that adversely affect normal plant growth and development, plants usually employ a variety of different strategies to respond to different durations of exposure (Chaturvedi et al., 2021). In this study, we chose to analyze the root response, since they are one of the main organs that sense environmental cadmium concentration and absorb it. To this end, we used a multi-time-point transcriptome analysis of tobacco from 5 min to 48 h of cadmium stress treatment. We observed changes in the number of differentially expressed genes in the root transcriptome at different stages of the treatment. A significant increase in the number of DEGs was observed in the T5 time point compared with T4 time point, and we used this as a boundary to classify 0 minutes to 1 hour as the early-stress response and 3 hours to 48 hours as the mid-stress response. Therefore, the effect of cadmium treatment on different genes expression varies at early and mid-stages.

Photosynthesis is an important component of plant metabolism and its integrity is crucial for normal plant growth and development (Dubey, 2018). Photosynthesis was affected after only 5 min of stress due to the massive translocation of metal cations into the plant. There were also many differentially expressed genes that were enriched, causing an effect on many functions related to photosynthesis. This is consistent with the reported toxic effect of the heavy metal Cd on plant photosynthesis (Dias et al., 2012). Changes in the photosynthetic capacity also affect primary carbon metabolism processes, which in turn causes an effect on the tricarboxylic acid cycle (Heyneke and Fernie, 2018).

In the early-response period, fatty acid synthesis-related pathways are enriched. Fatty acids play an important role in land plant responses to abiotic stress. They serve as reserves of carbon and energy and provide extracellular barrier components, and stress

10.3389/fpls.2023.1143349

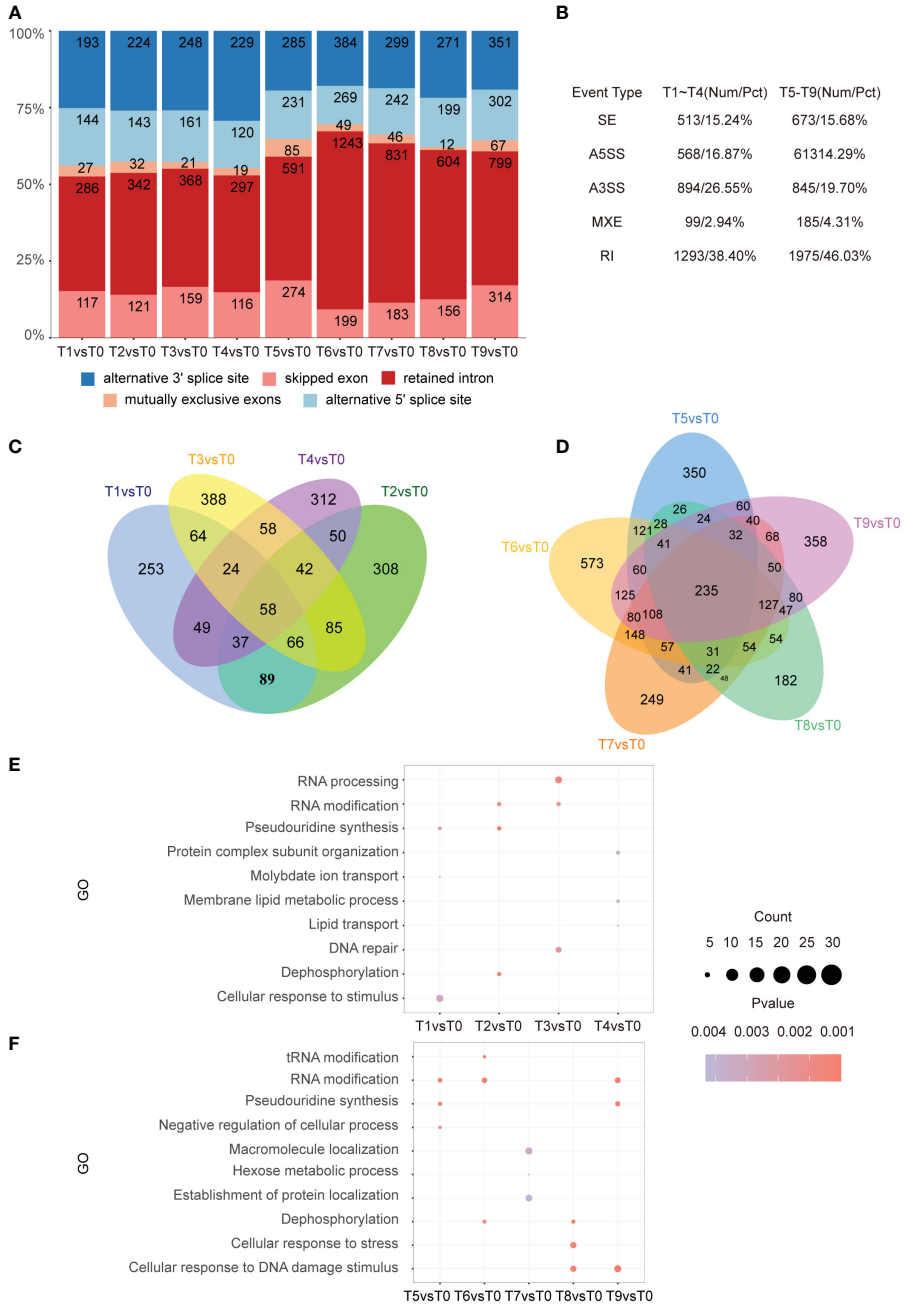


FIGURE 5
Distribution of AS types. **(A)**. Bar graph showing the number of AS events at each time point from T1 to T9. **(B)**. Five different types of AS events and their frequency. The first column is the AS event category, the second column is the corresponding early stage of the stress with its proportion, and the third column is the corresponding mid-stress event with its proportion. **(C, D)**. From left to right, the Venn diagram shows the DAS gene differences at each time point in the pre-response (T1-T4) and mid-response (T5-T9) periods to Cd stress. In the pre-response period, there were 58 intersecting genes. In the mid-response period, there were 235 intersecting genes. **(E, F)**. From top to bottom, the results of GO enrichment analysis of DAS genes in each time point in early (T1-T4), as well as mid-response (T5-T9) are shown. The top 3 entries for each time significance were selected for statistical analysis.

signal regulators (He and Ding, 2020). The plant cell wall is one of the most important physical factors for resisting abiotic stress, the accumulation of lignin or corkiness in the primary cell wall prevents the entry of foreign substances (Keegstra, 2010). It has been shown that cadmium causes some degree of damage to the cell wall in plants (Gutsch et al., 2018), and genes associated with cell wall repair are active under the continuation of the stress process. Cadmium can cause RNA single-strand breaks that line into DNA base modification products affecting DNA repair (Zhang et al., 2019). As cadmium continues to accumulate in the middle of

the stress response, apoptosis and DNA damage occur, and in response, the genes related to DNA repair are enriched.

It has been shown that alternative splicing may be associated with transcriptional regulation in response to environmental stress in plant. This association is related to different species or different stress (Staiger and Brown, 2013). In rice (Dong et al., 2018), there was little overlap between DASs and DEGs under different mineral-deficient conditions, whereas there was a strong overlapping relationship between the two under drought stress in tea tree (Ding et al., 2020). In wheat samples heat-treated for one-hour, pre-mRNA processing was significantly associated with transcriptional regulation (Liu et al., 2018). Early in the stress response, differentially expressed genes rarely intersected with differentially AS genes. As the treatment time of cadmium stress prolonged, the overlapped genes number of DEGs and DAS increased gradually. One possibility is that Cd stress might affect the relationship between transcription and post-transcriptional level, which needs to be further explored in the future.

Additionally, we found that several receptor kinases are induced to express in response to stress, such as gene_84425, and receptor kinases are widely involved in the process of plant stress resistance (Zhao et al., 2021), therefore this kind of genes may be used.

5 Conclusions

In this study, we investigated the transcriptome of tobacco in response to Cd stress using RNA-seq. Based on the results, we identified stage-specific differences in tobacco during prolonged treatment and divided samples into two response periods. The early phase of the stress response was characterized by a large translocation of metallic cadmium ions which affected the rhythm-related pathways, while in the middle phase of the stress response there was a shift from receiving stimuli toward repairing damage. We found that plants tend to change and increase the number of splicing patterns in response to long-term Cd stress, and alternative splicing regulation of stimulus response occurs prior to and influences transcriptional regulation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI accession PRJNA913401.

Author contributions

GX and LW conceived the project and designed research. CL and YH performed research. JS, HZ, GW, LX, and CL contributed new reagents/analytic tools. CL and LW wrote the paper. All authorscontributed to the article and approved the submitted version.

Funding

Author JS and LX were employed by Technology Center, China Tobacco Jiangsu Industrial Co. Ltd. The work was supported by Yunnan Company of China National Tobacco Corporation (NO. 2021530000242033). The study received funding from Key Scientific and Technological Project of Henan Province (grant number 212102110446) for GX, and the Project of Tobacco Genome (grant number 110202001026 (JY-09)) for GX. The funder had the following involvement with data collection and analysis of data.

Acknowledgments

We thank AJE company for helping us modify the language grammar.

Conflict of interest

Author JS and LX were employed by Technology Center, China Tobacco Jiangsu Industrial Co. Ltd. Author GW was employed by Yuxi Zhongyan Tobacco Seed Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1143349/full#supplementary-material

References

Borgo, L., Rabêlo, F. H. S., Budzinski, I. G. F., Cataldi, T. R., Ramires, T. G., Schaker, P. D. C., et al. (2022). Proline exogenously supplied or endogenously overproduced induces different nutritional, metabolic, and antioxidative responses in transgenic tobacco exposed to cadmium. *J. Plant Growth Regul.* 41 (7), 2846–2868. doi: 10.1007/s00344-021-10480-6

Chaturvedi, P., Wiese, A. J., Ghatak, A., Záveská Drábková, L., Weckwerth, W., and Honys, D. (2021). Heat stress response mechanisms in pollen development. *New Phytol.* 231 (2), 571–585. doi: 10.1111/nph.17380

Dias, M. C., Monteiro, C., Moutinho-Pereira, J. M., Correia, C. M., Gonçalves, B., and Santos, C. (2012). Cadmium toxicity affects photosynthesis and plant growth at different levels. *Acta Physiologiae Plantarum* 351281– 1289. doi: 10.1007/s11738-012-1167-8

Ding, Y. Q., Fan, K., Wang, Y., Fang, W. P., Zhu, X. J., Chen, L., et al. (2022). [Drought and heat stress-mediated modulation of alternative splicing in the genes involved in biosynthesis of metabolites related to tea quality]. *Mol. Biol. (Mosk)* 56 (2), 321–322. doi: 10.31857/s0026898422020057

Ding, Y., Wang, Y., Qiu, C., Qian, W., Xie, H., and Ding, Z. (2020). Alternative splicing in tea plants was extensively triggered by drought, heat and their combined stresses. *PeerJ* 8, e8258. doi: 10.7717/peerj.8258

Dong, C., He, F., Berkowitz, O., Liu, J., Cao, P., Tang, M., et al. (2018). Alternative splicing plays a critical role in maintaining mineral nutrient homeostasis in rice (Oryza sativa). *Plant Cell* 30 (10), 2267–2285. doi: 10.1105/tpc.18.00051

Dubey, R. S. (2018). "Photosynthesis in plants under stressful conditions," in *Handbook of photosynthesis* (CRC Press), 629–649.

Fu, Y., Mason, A. S., Zhang, Y., Lin, B., Xiao, M., Fu, D., et al. (2019). MicroRNA-mRNA expression profiles and their potential role in cadmium stress response in brassica napus. *BMC Plant Biol.* 19 (1), 570. doi: 10.1186/s12870-019-2189-9

Gutsch, A., Keunen, E., Guerriero, G., Renaut, J., Cuypers, A., Hausman, J. F., et al. (2018). Long-term cadmium exposure influences the abundance of proteins that impact the cell wall structure in medicago sativa stems. *Plant Biol. (Stuttg)* 20 (6), 1023–1035. doi: 10.1111/plb.12865

Haider, F. U., Liqun, C., Coulter, J. A., Cheema, S. A., Wu, J., Zhang, R., et al. (2021). Cadmium toxicity in plants: Impacts and remediation strategies. *Ecotoxicol Environ. Saf.* 211, 111887. doi: 10.1016/j.ecoenv.2020.111887

Hao, X., Wang, Q., Hou, J., Liu, K., Feng, B., and Shao, C. (2021). Temporal transcriptome analysis reveals dynamic expression profiles of gametes and embryonic development in Japanese flounder (Paralichthys olivaceus). *Genes (Basel)* 12 (10):1561. doi: 10.3390/genes12101561

He, M., and Ding, N. Z. (2020). Plant unsaturated fatty acids: Multiple roles in stress response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.562785

Heyneke, E., and Fernie, A. R. (2018). Metabolic regulation of photosynthesis. *Biochem. Soc. Trans.* 46 (2), 321–328. doi: 10.1042/bst20170296

Jian, G., Mo, Y., Hu, Y., Huang, Y., Ren, L., Zhang, Y., et al. (2022). Variety-specific transcriptional and alternative splicing regulations modulate salt tolerance in rice from early stage of stress. *Rice (N Y)* 15 (1), 56. doi: 10.1186/s12284-022-00599-9

Keegstra, K. (2010). Plant cell walls. *Plant Physiol.* 154 (2), 483–486. doi: 10.1104/pp.110.161240

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi: 10.1038/s41587-019-0201-4

Lee, Y., and Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323. doi: 10.1146/annurev-biochem-060614-034316

Liu, R., Gao, Y., Guan, C., Ding, L., Fan, Z., and Zhang, Q. (2022). The comparison of temporal transcriptome changes between morning-opening and afternoon-opening iris flowers reveals the candidate genes regulating flower opening and closing. *J. Plant Biol.* 1–19. doi: 10.1007/s12374-022-09363-4

Liu, Z., Qin, J., Tian, X., Xu, S., Wang, Y., Li, H., et al. (2018). Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (Triticum aestivum l.). *Plant Biotechnol. J.* 16 (3), 714–726. doi: 10.1111/pbi.12822

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Qin, G., Niu, Z., Yu, J., Li, Z., Ma, J., and Xiang, P. (2021). Soil heavy metal pollution and food safety in China: Effects, sources and removing technology. *Chemosphere* 267, 129205. doi: 10.1016/j.chemosphere.2020.129205

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl. Acad. Sci. U.S.A.* 111 (51), E5593–E5601. doi: 10.1073/pnas.1419161111

Shibao, C., Meng, W., Shanshan, L., Han, Z., Xiaoqin, L., Xiaoyi, S., et al. (2019). Current status of and discussion on farmland heavy metal pollution prevention in China. *Earth Sci. Front.* 26 (6), 35. doi: 10.13745/j.esf.sf.2019.8.6

Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25 (10), 3640–3656. doi: 10.1105/tpc.113.113803

Wang, L., Yang, T., Wang, B., Lin, Q., Zhu, S., Li, C., et al. (2020). RALF1-FERONIA complex affects splicing dynamics to modulate stress responses and growth in plants. *Sci. Adv.* 6 (21), eaaz1622. doi: 10.1126/sciadv.aaz1622

Weng, X., Zhou, X., Xie, S., Gu, J., and Wang, Z. Y. (2021). Identification of cassava alternative splicing-related genes and functional characterization of MeSCL30 involvement in drought stress. *Plant Physiol. Biochem.* 160, 130–142. doi: 10.1016/j.plaphy.2021.01.016

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2 (3), 100141. doi: 10.1016/j.xinn.2021.100141

Yang, Q., Li, Z., Lu, X., Duan, Q., Huang, L., and Bi, J. (2018). A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. *Sci. Total Environ.* 642, 690–700. doi: 10.1016/j.scitotenv.2018.06.068

Yang, W., Wang, S., Zhou, H., Zeng, M., Zhang, J., Huang, F., et al. (2022). Combined amendment reduces soil cd availability and rice cd accumulation in three consecutive rice planting seasons. *J. Environ. Sci. (China)* 111, 141–152. doi: 10.1016/j.jes.2021.03.027

Zhang, S., Hao, S., Qiu, Z., Wang, Y., Zhao, Y., Li, Y., et al. (2019). Cadmium disrupts the DNA damage response by destabilizing RNF168. *Food Chem. Toxicol.* 133, 110745. doi: 10.1016/j.fct.2019.110745

Zhao, C., Jiang, W., Zayed, O., Liu, X., Tang, K., Nie, W., et al. (2021). The LRXs-RALFs-FER module controls plant growth and salt stress responses by modulating multiple plant hormones. *Natl. Sci. Rev.* 8 (1), nwaa149. doi: 10.1093/nsr/nwaa149

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 March 2023
DOI 10.3389/fpls.2023.1147946

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Wuhan Botanical Garden (CAS), China

REVIEWED BY
Lin Zhang,
Hubei University of Chinese Medicine, China
Sajid Fiaz,
The University of Haripur, Pakistan

*CORRESPONDENCE
Jinping Guo
jinpguo@126.com
Baopeng Ding
dingbaopeng2006@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 19 January 2023
ACCEPTED 06 March 2023
PUBLISHED 21 March 2023

CITATION
Wang J, Hu H, Liang X, Tahir ul Qamar M, Zhang Y, Zhao J, Ren H, Yan X, Ding B and Guo J (2023) High-quality genome assembly and comparative genomic profiling of yellowhorn (*Xanthoceras sorbifolia*) revealed environmental adaptation footprints and seed oil contents variations.
Front. Plant Sci. 14:1147946.
doi: 10.3389/fpls.2023.1147946

COPYRIGHT
© 2023 Wang, Hu, Liang, Tahir ul Qamar, Zhang, Zhao, Ren, Yan, Ding and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

High-quality genome assembly and comparative genomic profiling of yellowhorn (*Xanthoceras sorbifolia*) revealed environmental adaptation footprints and seed oil contents variations

Juan Wang[1,2†], Haifei Hu[3,4†], Xizhen Liang[1,2], Muhammad Tahir ul Qamar[5], Yunxiang Zhang[1,2], Jianguo Zhao[6], Hongqian Ren[1,2], Xingrong Yan[1,2], Baopeng Ding[1,6*] and Jinping Guo[1,2*]

[1]College of Forestry, Shanxi Agricultural University, Taigu, Shanxi, China, [2]Shanxi Key Laboratory of Functional Oil Tree Cultivation and Research, Shanxi Agricultural University, Taigu, Shanxi, China, [3]Rice Research Institute, Guangdong Key Laboratory of New Technology in Rice Breeding, Guangzhou, China, [4]Guangdong Rice Engineering Laboratory, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [5]Integrative Omics and Molecular Modeling Laboratory, Department of Bioinformatics and Biotechnology, Government College University Faisalabad (GCUF), Faisalabad, Pakistan, [6]Engineering Research Center of Coalbased Ecological Carbon Sequestration Technology of the Ministry of Education, Datong University, Taigu, Shanxi, China

Yellowhorn (*Xanthoceras sorbifolia*) is a species of deciduous tree that is native to Northern and Central China, including Loess Plateau. The yellowhorn tree is a hardy plant, tolerating a wide range of growing conditions, and is often grown for ornamental purposes in parks, gardens, and other landscaped areas. The seeds of yellowhorn are edible and contain rich oil and fatty acid contents, making it an ideal plant for oil production. However, the mechanism of its ability to adapt to extreme environments and the genetic basis of oil synthesis remains to be elucidated. In this study, we reported a high-quality and near gap-less yellowhorn genome assembly, containing the highest genome continuity with a contig N50 of 32.5 Mb. Comparative genomics analysis showed that 1,237 and 231 gene families under expansion and the yellowhorn-specific gene family NB-ARC were enriched in photosynthesis and root cap development, which may contribute to the environmental adaption and abiotic stress resistance of yellowhorn. A 3-ketoacyl-CoA thiolase (*KAT*) gene (*Xso_LG02_00600*) was identified under positive selection, which may be associated with variations of seed oil content among different yellowhorn cultivars. This study provided insights into environmental adaptation and seed oil content variations of yellowhorn to accelerate its genetic improvement.

KEYWORDS

yellowhorn, pan-genomics, genomic profiling, adaptation, oil contents

Introduction

Yellowhorn (*Xanthoceras sorbifolia*), belonging to the *Xanthoceras* genus (*Sapindaceae* family), is a unique woody tree plant species widely growing in Northern and Central China (Bi et al., 2019; Liu et al., 2021a). Yellowhorn shows strong abiotic stress resistance ability and can grow under extreme environmental conditions, including extreme temperature, drought conditions, saline, and alkaline land (Ruan et al., 2017). Furthermore, yellowhorn is easy to reproduce, sowing, root cutting, and grafting and is now considered a promising afforestation species for many arid areas. This oil-rich tree produces capsular fruits from hermaphrodites, with about 60% of its seed kernel containing edible seed oil for the human diet and around 4% nervonic acid essential for nerve and brain development with high medicinal and ornamental value (Liang et al., 2019; Liang et al., 2022). However, yellowhorn also contains moderate erucic acid (about 9% of the total fatty acid) that can damage the heart at high doses (Liu et al., 2021b). Therefore, to make yellowhorn a more desirable species for oil production, it is essential to underly the genetic basis of its oil synthesis pathway and design and cultivate new species with a lower level of erucic acid and a higher level of nervonic acid.

The 3-ketoacyl-CoA thiolase (KAT) is a member of thiolase and can catalyze the final step of fatty acid β-oxidation and the claisen condensation reaction between two Acetyl-CoAs and lead to carbon chain elongation, which is a key step in the fatty acid biosynthetic pathways (Footitt et al., 2007). So far, KAT has been reported to play an important role in producing various energy-storage molecules, such as fatty acids and affecting seed oil content and synthesis in *Arabidopsis thaliana* (Germain et al., 2001) and *Jatropha curcas* (Gomes et al., 2010). However, the mechanism of KAT regulation in yellowhorn and how it underly the fatty acid synthesis remains to be elucidated.

The rapid development of sequencing technologies has facilitated the development of yellowhorn genomes, with two good-quality yellowhorn genome assemblies being published recently (Liang et al., 2019; Liang et al., 2022). These published yellowhorn genomes were sampled and collected from a valley terrain environment with mountains and rivers in Shandong Province. However, yellowhorn also grew and adapted to the loess plateau with a more extreme climate. Therefore, in this study, using long-read sequencing, we sequenced and assembled a gapless *Xanthoceras sorbifolia* genome of the superior line "G11" (Data named *XsoG11*), which was collected from the loess plateau located in Shanxi province. By performing the comparative genomic analysis among representative angiosperms, we revealed that gene families with functions of photosynthesis and root cap development were expanded and existed in yellowhorn, which may associate with adaptation to extreme environmental conditions. With the availability of high-quality yellowhorn reference genomes, we performed the pangenome-wide analysis among three yellowhorn genomes and identified gene content variations that may associate with environment adaptation and oil content variations of different yellowhorn cultivars. All the above results will provide new insights into genetic diversity study of yellowhorn and helps in its genetic improvement.

Materials and methods

Plant materials

Xanthoceras sorbifolia superior cultivar "XsoG11" (*Xanthoceras sorbifolia* superior G11) is a strain with highly comprehensive evaluation selected by *Xanthoceras sorbifolia* research group of Shanxi Agricultural University collected from Lvliang Mountain (Shanxi province; 111°47′17″East, 37°15′57″North) (Supplementary Figure 1), located in semi-arid area, which is extremely cold in winter. The DNA sequencing libraries of PacBio HiFi long reads, Illumina short reads, and Hi-C reads were prepared according to the standard Illumina and PacBio library construction protocol for the generation of genome assembly "XsoG11" (Liang et al., 2021).

Genome assembly

The clean PacBio HiFi reads were assembled using Hifiasm (v.0.15) (Cheng et al., 2021) with default parameters. Then, the original assembly result is polled using pilon (v1.23) (Walker et al., 2014) to get the final genome assembly result. Chromosome-length scaffolds were generated by aligning the raw HiC-reads to the draft assembly using Juicer (v.1.6) (Durand et al., 2016) with the resulting alignments processing by the 3D-DNA pipeline (v.19) (Dudchenko et al., 2017) to generate the candidate chromosome-length assemblies. This candidate assembly was reviewed and curated using Juicebox Assembly Tools (v.1.11.08) (Robinson et al., 2018). BUSCO V3 (Simao et al., 2015) with eukaryota_odb9 was used to assess the completeness of the assembly.

Repeat sequence annotation

For the repeat sequence annotation, trf (v4.09) (Benson, 1999) was used to predict tandem repeats; Microsatellite sequence uses misa Pl program prediction; LTR First use LTR separately_Finder and LTR_Harvest Identify, then use LTR_ Retriever (v2.7) (Ou and Jiang, 2018) integrates the results of the above two software to obtain the final LTR identification results; LINE, SINE, and DNA transposons were identified by RepeatMasker (v4.0.9) (Tarailo-Graovac and Chen, 2009). The two methods are combined to identify the repeat contents in our genome, homology-based and *de novo* prediction. Homology-based analysis: We identified the known TEs within the XsoG11 genome using RepeatMasker (v4.0.9) (Tarailo-Graovac and Chen, 2009) with the Repbase TE library. *De novo* prediction: We constructed a *de novo* repeat library of the XsoG11 genome using RepeatModeler, which can automatically execute two core *de novo* repeat-finding programs, namely, RECON (v1.08) (Bao and Eddy, 2002) and RepeatScout (v1.0.5) (Benson, 1999), to comprehensively conduct, refine and classify consensus models of putative interspersed repeats for the XsoG11 genome. Furthermore, we performed a *de novo* search for long terminal repeat (LTR) retrotransposons against the XsoG11 genome sequences using LTR_Finder (v1.0.7) (Xu and Wang, 2007), LTR_harvest (v1.5.11) and LTR_retriever (v2.7) (Ou and Jiang, 2018). We also identified

tandem repeats using the Tandem Repeat Finder (TRF) package and the SimpleSequence Repeats (SSR) using misa (v1.0) (Beier et al., 2017). Finally, we merged the library files of the two methods to identify and determine the repeat contents.

Gene annotation

We predicted protein-coding genes of the XsoG11 genome using three methods: ab initio gene prediction, homology-based gene prediction, and RNA-Seq-guided gene prediction. Before gene prediction, the assembled XsoG11 was hard and soft masked using RepeatMasker (v4.0.9) (Tarailo-Graovac and Chen, 2009). We adopted Augustus (v3.3.3) (Stanke et al., 2008) to perform *ab initio* gene prediction. Models used for each gene predictor were trained from a set of high-quality proteins generated from the RNA-Seq dataset. We used maker (v2.31.10) (Holt and Yandell, 2011) to conduct homology-based gene prediction. First, the protein sequences and transcripts sequences were aligned to our genome assembly and predicted coding gene using maker with the default parameters. To carry out RNA-Seq-guided gene prediction, we first aligned clean RNA-Seq reads to the genome using hisat2 (v2.0.0) (Kim et al., 2015), and the gene structure was formed using Trinity (v2.3.2) (Grabherr et al., 2011), Transdecoder (v2.01) (Haas et al., 2017) and maker (v2.31.10) (Holt and Yandell, 2011). Finally, EVidenceModeler (v1.1.1) (Haas et al., 2008) was used to integrate the prediction results of the three methods to predict gene models. Functional annotation was performed by comparing proteins with various functional databases including NR, swiss pro, KOG and TrEMBL, using BLASTP (e-value < 1e-5) (Camacho et al., 2009).

Comparative genomics analysis

Using the assembled yellowhorn genome (XsoG11) and nine other related angiosperm genomes, we performed a comparative genome analysis using OrthoFinder (v 2.4.0) (Emms and Kelly, 2019) to identify the orthologous gene families in the yellowhorn genome. The analysis process of the OrthoFinder was indicated as follows: 1) Use the diamond to input all protein sequences for all-*vs*-all comparison and detect homologous gene pairs (Evaluate < 1e-5 and the minimum coverage is > 40%). 2) Input the list of homologous gene pairs into MCL program for family clustering. A maximum likelihood phylogenetic tree of ten species was constructed based on shared single-copy genes using Mega V5 (Tamura et al., 2011). Expanded and contracted gene families were detected using CAFÉ (v4.2.1) (De Bie et al., 2006). The expanded gene families were functionally annotated on Pfam v32.0 (Mistry et al., 2021) and Swiss-Prot (UniProt, 2021) databases. The functional enrichment of each gene family was determined using a Fisher's exact test (false discovery rate < 0.05).

Pan-genomics analysis

The genome sequences and protein sequences of two published yellowhorn (WF18 and Xsv2) were downloaded from Liang and Liu study (Liang et al., 2019, Liu et al., 2021a). Orthologous genes among the yellowhorn genomes were identified by Orthovenn2 (Xu et al., 2019), a web tool used to identify orthologous and paralogous genes, with a pairwise sequence similarity cut-off of 10- 5 and inflation of 1.5 to define orthologous cluster structure. KaKs_Calculator 2.0 (Wang et al., 2010) was used to calculate orthologous gene clusters' non-synonymous/synonymous substitution ratio. Orthologous clusters and gene pairs under positive selection (Ka/ks > 1) were analyzed by UniProt search and TopGO (Alexa et al., 2006) using Fisher's exact tests for functional annotation and enrichment analysis. Furthermore, the two published genomes were compared to our genome assembly using the mumandco_ V3 program (parameter default) (O'donnell and Fischer, 2020).

Results

Genome assembly and annotation

In this study, we used the long read sequencing to *de novo* assemble a near gapless genome assembly of the Shanxi yellowhorn cultivar XsoG11 (Figure 1A; Table 1). We generated approximately 30-fold coverage of PacBio CCS (HiFi) reads and assembled the CCS (HiFi) reads using Hifiasm: a haplotype-resolved assembler for accurate HiFi reads (Cheng et al., 2021). The assembly length of the XsoG11 genome is 489.18 Mb with a contig N50 of 32.5 Mb, showing the highest contiguity than the previously published yellowhorn genomes (Table 2). The contigs were further polished using pilon, then ordered, oriented and anchored to chromosomes using *in-situ* Hi-C sequencing. We found that around 96.2% (470.79 Mb) of sequences are anchored to the chromosome and seven chromosomes in our genome do not contain any gaps (Table 1; Figure 1B). XsoG11 assembly has 95.7% complete BUSCOs (Table 2; Figure 1C), comparable to the previously published Xsv2 (Liu et al., 2021b) and WF18 yellowhorn genomes (Liang et al., 2019). In addition, we identified approximately 68.71% repeat sequences in the assembled genome, in which the long terminal repeat (LTR) retrotransposon element represents the most abundant transposable elements (TEs) class, accounting for 35.8% TEs (Table 3). Using RNA-seq transcript mapping combined with ab initio prediction and homologous protein searches, we predicted 35,039 protein-coding genes with an average gene length of 2,662 bp in the XsoG11 genome (Supplementary Table 1), in which 27,082 (85%) genes have functional annotation from at least one functional protein database, including nr (84%), TrEMBL (84%), KOG (41%), Swiss-Prot (57%), Pfam (69%), Gene Ontology (GO) (42%), and KEGG (23%) (Supplementary Table 2). We also identified 1,250 tRNAs, 770 small nucleolar RNAs and 4,691 small nuclear RNAs (Supplementary Table 3).

Yellowhorn phylogenetics and gene family expansion analyses

The change in gene family size plays an important role in the evolution of angiosperms' environmental adaptation and trait

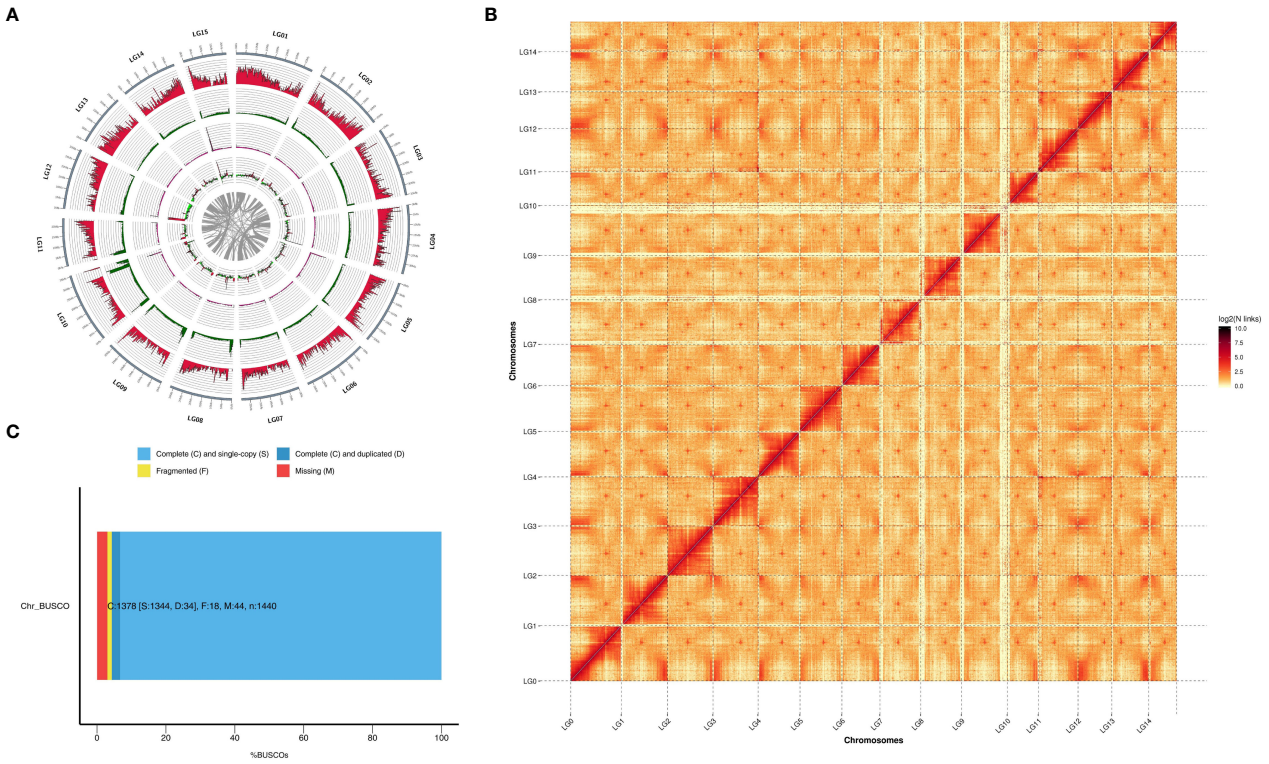


FIGURE 1
The characteristics of yellowhorn XsoG11 genome assembly. **(A)** The yellowhorn genome feature. From outer-most track to innermost track: gene density, transposable element density, repeat sequence density, GC content, and Intra-genome collinear blocks. **(B)** Buscos of complete genome. **(C)** Contact map of Hi-C links among 14 pseudochromosomes.

TABLE 1 Genome Statistic of *Xanthoceras sorbifolia* superior G11.

Chr	Chr_size(bp)	Contig_number	Contig_size(bp)	GC_content(%)
LG01	39494243	1	39494243	34.9
LG02	35873774	1	35873774	35.5
LG03	35257883	4	35257583	35.05
LG04	34944601	2	34944501	35.14
LG05	32593928	4	32593628	34.99
LG06	32495627	1	32495627	35.18
LG07	29437501	1	29437501	35.28
LG08	31632995	1	31632995	35.22
LG09	31780288	3	31780088	35.34
LG10	35671781	2	35671681	35.4
LG11	24152185	3	24151985	35.44
LG12	30754806	7	30754206	36.83
LG13	26198562	1	26198562	35.16
LG14	28601909	1	28601909	35.09
LG15	21899742	2	21899642	35.21
ChrAll	470789825	34	470787925	35.31

TABLE 2 Statistic of different yellowhorn genome assemblies.

TypeParameter	WF18v1	Xsv2	ZS4	WF18	XsoG11
Assembly Genome size(Mb)	490.44	470	504.2	440	489.18
Chromosome-scale scaffolds(Mp)	490.24 (99.96%)	446.2 (94.9%)	489.29 (97.04%)	420 (95.4%)	470.79 (96.24%)
Total num. of scaffolds	22	988	297	267	417
Total num. of chromosomes	15	15	15	15	15
ScaffoldN50(Mb)	34	30.8	32.17	29.4	31.6
Total num. of Contigs	2,428	3,302	3,035	2,002	417
Contig N50(Mb)	0.42	0.42	1.04	0.64	31.6
Complete BUSCOs	98.70%	97.50%	98.70%	84.60%	95.70%
GC content of the genome(%)	34.71	34.94	36.95	32.75	35.70%
Protein-coding genes	29,888	22,049	24,672	21,059	35,039
Reference	(Liang et al., 2022)	(Liu et al., 2021a)	(Bi et al., 2019)	(Liang et al., 2019)	This study

formation during evolution (Van de Peer et al., 2009). To further dissect the genetic basis of high seed oil content and the ability to adapt to extreme environments of yellowhorn, we performed the comparative genomics and gene family expansion analysis in yellowhorn and eight other representative angiosperms and the outgroup species (*Taxus chinensis*). We first determined the phylogeny position of yellowhorn by constructing a phylogenetic tree using 17 single-copy orthologous genes conserved in 10 representative angiosperms (Figure 2A; Supplementary Table 4). Our result inferred that *Acer yangbiense* was the most recent common ancestor of yellowhorn, which diverged around 76.5 million years ago (Figure 2B).

We further performed the collinear analysis by comparing the yellowhorn (XsoG11) genome with *Acer yangbiense* and *Sapindus mukorossi* genomes (Figure 2C). Although these species belong to the *Sapindaceae* family, we identified significant structural variations in yellowhorn and *Acer yangbiense* and *Sapindus mukorossi*, suggesting that significant chromosomal differentiation occurred since they derived from the last common ancestor. A total of 4,622 gene families were shared by ten studied species

TABLE 3 Repeat elements of the XsoG11 yellowhorn genome.

repeat_type	total_size	Percentage of genome (%)
DNA/CMC-EnSpm	1456919	0.30%
DNA/hAT-Ac	5043893	1.03%
DNA/hAT-Tag1	1039259	0.21%
DNA/hAT-Tip100	1516665	0.31%
DNA/MuLE-MuDR	4508774	0.92%
DNA/PIF-Harbinger	887491	0.18%
DNA/TcMar-Pogo	209047	0.04%
LINE/L1	19514034	3.99%
LINE/RTE-X	67436	0.01%
Low_complexity	1285729	0.26%
LTR	1516161	0.31%
LTR/Caulimovirus	1471510	0.30%
LTR/Copia	46538428	9.51%
LTR/Gypsy	62413913	12.76%
Simple_repeat	8520634	1.74%
Unknown	180122927	36.82%
Total	336112820	68.71%

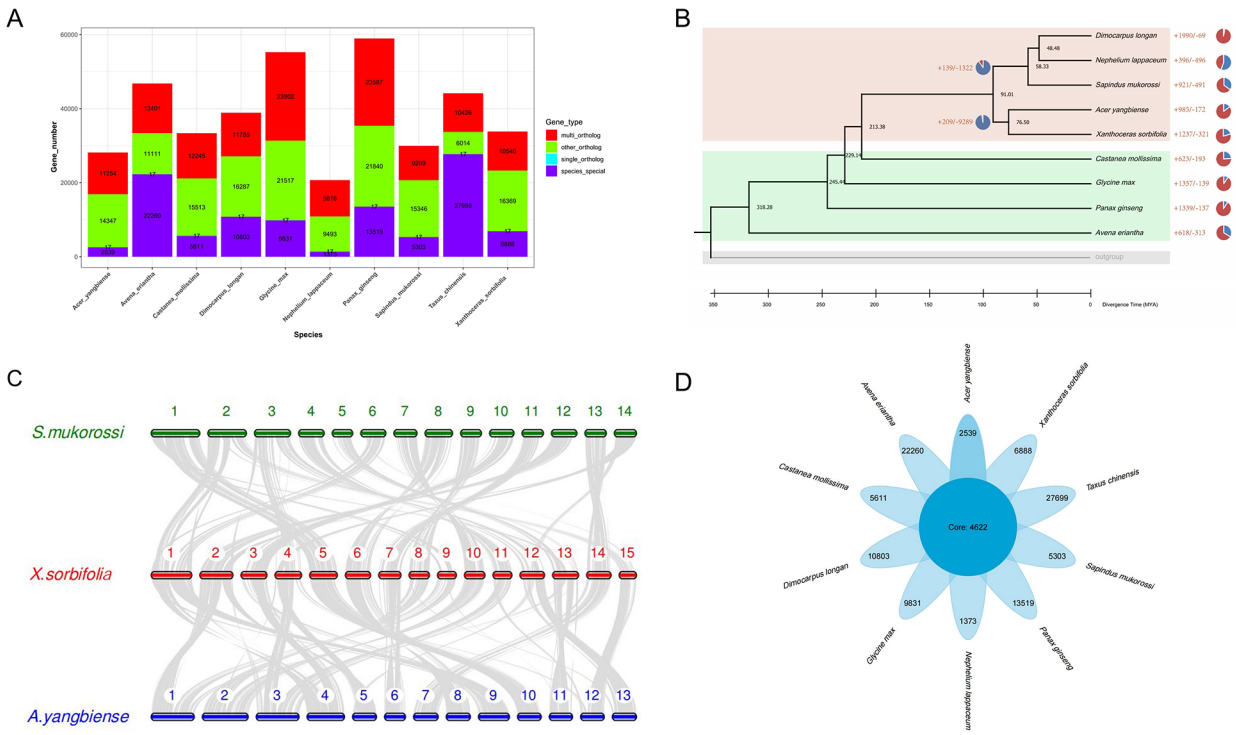


FIGURE 2
Comparative analysis yellowhorn and nine other representative angiosperms. **(A)** Gene family clusters. Multi_ortholog: This family exists in all species, and the number of members in at least one species is greater than or equal to two; single_ortholog: a single copy orthologous gene family, which exists in all species and has 1 member in all species; other_ortholog: The family exists in at least two species and does not exist in at least one species; species_special: the unique gene family of each species. **(B)** Maximum-likelihood phylogeny of yellowhorn and nine other representative angiosperms. **(C)** Collinear analysis by comparing the yellowhorn (*X.*sorbifolia, XsoG11) genome with *Acer yangbiense* (*A. yangbiense*) and *Sapindus mukorossi* genomes (*S. mukorossi*). **(D)** The comparative genomic analysis shows the number of the core and species specific gene families.

(Figure 2D), with yellowhorn having 6,888 species-specific gene families. GO enrichment analysis indicates that these species-specific gene families are significantly enriched in functions associated with photosynthesis (Supplementary Table 5). Additionally, the gene family size analysis showed that 1,237 and 231 gene families were found to be expanding and contracting (Supplementary Table 6). Enrichment analysis showed that the expanding gene families were enriched for functions associated with disease resistance (Pfam: NB-ARC domain) (Supplementary Table 7) and photosynthesis (GO: "photosynthetic electron transport in photosynthesis", "photosynthetic electron transport chain", "photosynthesis, light reaction" and "photosynthesis"; KEGG: ko00195: Photosynthesis) (Supplementary Tables 8, 9). These results may suggest that gene family expansion was associated with photosynthesis and biotic stress resistance in yellowhorn.

Pan-genome analysis of yellowhorn genomes

High-quality genome assemblies enable the accurate discovery of structural variations and genetic variations among genomes. Using the high-quality genome assembly (XsoG11) as the reference genome, we further discovered abundant structural variations (SVs), including inversions, translocations, insertions and deletions between the XsoG11 genome and the other two published yellowhorn genomes (Xsv2 and WF18) (Figure 3A; Supplementary Tables 10, 11). We identified 3,515 and 2,262 sequences uniquely present in XsoG11 by comparing this assembly with Xsv2 and WF18 genomes, respectively, in which 1,005 sequences are present in the XsoG11 genome but missing in both Xsv2 and WF18 genomes (Figure 3B). In addition, we further performed the pan-genome wide analysis of gene families (Zia et al., 2022). The pangenome-wide gene family clustering analysis revealed that these three genomes shared 11,750 core orthologous clusters, whereas at least one genome (but less than three) shared 8,844 dispensable orthologous clusters, with 677 XsoG11-specific dispensable orthologous clusters (Figure 3C). The evolutionary analysis of the three yellowhorn showed that they are clustered together (Supplementary Figure 2) and the results of collinearity is consistent (Supplementary Figure 3). Functional annotation of genes located in these 1,005 sequences found oxidative phosphorylation (ko00190, $P < 2.19E-83$), ribosome (ko03010, $P < 1.06E-61$), RNA polymerase genes (ko03020, $P < 1.44E-13$) were the first three most significantly enriched in KEGG pathyway. In addition, photosynthetic pathway genes were also significantly enriched (ko00195, $P < 0.003$) (Figure 4). Gene ontology (GO) enrichment analysis shows that genes with essential biological functions, including RNA-DNA hybrid ribonuclease activity,

10.3389/fpls.2023.1147946

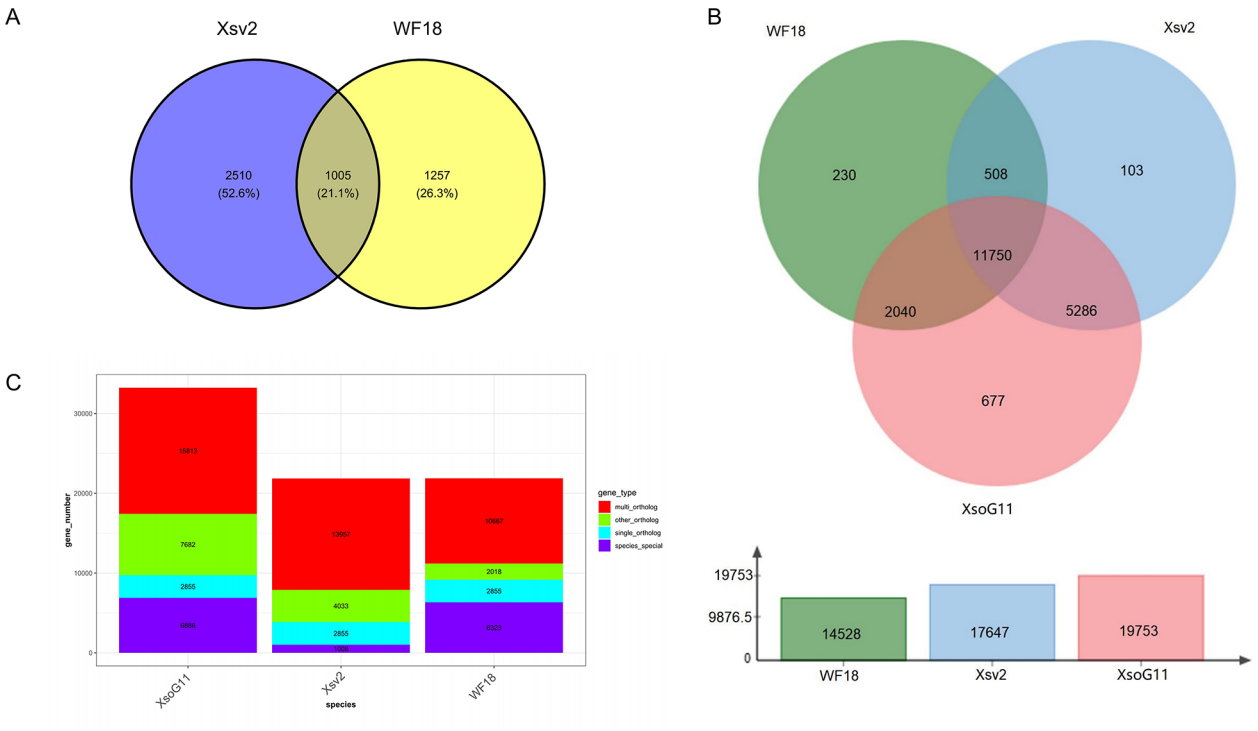


FIGURE 3
Pan-genome analysis of yellowhorn genomes. **(A)** The Venn diagram shows the number of SVs that are uniquely and commonly absent in the Xsv2 and WF18 genome, by comparing with the XsoG11 genome. **(B)** Pangenome-wide analysis of the core and dispensable gene families among three different yellowhorn genomes. **(C)** Gene family clustering,.

DNA recombination, oxidation-reduction process and DNA integration, were enriched in core orthologous clusters (Supplementary Table 12). By contrast, genes with functions potentially associated with fatty acid synthesis and abiotic stress responses, such as photosynthesis and root cap development, are enriched in XsoG11-specific dispensable orthologous clusters (Supplementary Table 13). We further examined the non-synonymous/synonymous substitution ratio (Ka/Ks) of homologous gene pairs of XsoG11 and two other published yellowhorn genomes (Xsv2 and WF18). The result showed that a total of 364 genes in the XsoG11 are under positive selection (Ka/Ks > 1) (Supplementary Table 14), including a gene (*Xso_LG02_00600*) encoding 3-ketoacyl-CoA thiolase associated with the formation of fatty acid (ko01040: Biosynthesis of unsaturated fatty acids; ko00592, alpha-Linolenic acid metabolism).

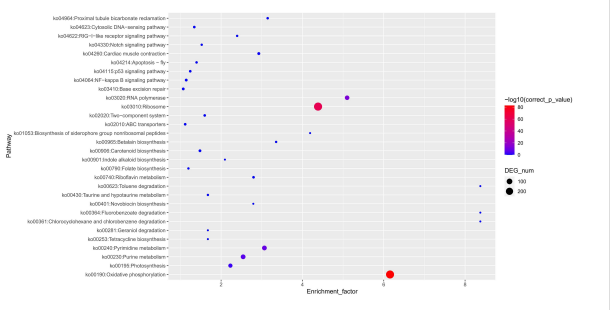


FIGURE 4
XsoG11 special gene enrichment in KEGG.

Discussion

A high-quality yellowhorn genome assembly is key to underlying the genetic basis of its ability to adapt to extreme environments and produce high oil seed content. In this study, using Circular consensus sequencing (CCS) long-read sequencing, we present a high-quality chromosome-scale genome assembly of Shanxi's yellowhorn cultivar (XsoG11). Compared with the previously published yellowhorn Shandong's cultivar "Shanyou 1" genome (WF18) (Liang et al., 2021), we assembled a higher contiguity (Contig N50: XsoG11: 31.6 Mb *vs* WF18v1: 0.42 Mb) and a near-gapless yellowhorn genome, with seven out of 15 chromosomes having no gaps. This additional high-quality genome can provide novel genomic resources for future yellowhorn improvement.

Photosynthesis is an important physiological process that converts light energy into chemical energy, affecting plant growth and development, respiration and transpiration (Goudriaan et al., 1985). Recent studies suggest that genes involved in the photosynthesis pathway are essential for environmental adaptation to different light regimes and coping with climate change by regulating the circadian clock and light perception (Kreps and Kay, 1997; Quint et al., 2016). For example, a pan-genome study of mung beans indicates that the presence/absence

variation (PAV) of genes regulating the photosynthesis pathway enables mung beans to adapt to different environments (Liu et al., 2022). Gene expansion through tandem duplication is important for stress response (Hanada et al., 2008). Photosynthesis-related genes, early light-induced proteins (ELIPs), were found to be expanded in plants showing drought resistance (Liu et al., 2020). Combined with the public genome assemblies of representative angiosperms and yellowhorn, this high-quality assembly enabled us to identify gene family differences and the expansion/contraction of yellowhorn during speciation and divergence and among yellowhorn cultivars growing in a different climate region. Compared with other angiosperms and yellowhorn cultivars collected in Shandong Province, the yellowhorn assembled in this study grows in the Loess Plateau of the middle of Shanxi Province with poor and dry soil (Yao et al., 2013). Both GO enrichment analyses of yellowhorn's expanding gene families and pangenome-wide XsoG11-specific dispensable gene families showed that gene families are significantly enriched in functions associated with photosynthesis. The enrichment of photosynthesis can lead to the accumulation of plant carbohydrates and other carbon sources, resulting in the generation of more energy to cope with the adverse stress of environmental factors in the Loess Plateau (Farrar and Williams, 1991). In addition, we also found that genes associated with root cap development were also enriched in the XsoG11-specific dispensable gene families that were missing in other yellowhorn cultivars. This may reflect that the yellowhorn cultivar growing in the drought Loess Plateau requires a more robust and extended root cap system (Du et al., 2011).

We also revealed that a gene (*Xso_LG02_00600*) encoding 3-ketoacyl-CoA thiolase (KAT) was under positive selection by comparing the value of Ka/Ks of homologous gene pairs of our yellowhorn cultivar with the Shandong's yellowhorn cultivars. KAT is an important catalyst for the process of fatty acid beta-oxidation. In *Arabidopsis thaliana*, the *KAT* gene was demonstrated to be activated in the early germination and seedling stage and led to fatty acyl-CoAs accumulation and the form of triacylglycerol, facilitating lipid storage in the oil seed (Germain et al., 2001; Footitt et al., 2007). A similar finding was reported in *Ophiocordyceps sinensis* that KAT was regulated and participated in the fatty acid pathway and provided sufficient energy for organisms by catalyzing the tricarboxylic acid cycle and electronic respiratory chain (Wang et al., 2022). The positive selection of the *KAT* gene in our Shanxi's yellowhorn cultivar may associate with its seed oil content variations among the comparative yellowhorn cultivars.

Conclusion

In a nutshell, we assembled a high-quality and near gap-less yellowhorn genome collected from Loess Plateau, providing valuable genomic resources for future yellowhorn genetic improvement. The functional analysis shows that the gene family under expansion and the yellowhorn-specific gene family are enriched in functions associated with photosynthesis and root cap development, which may relate to the environmental adaption of yellowhorn. A *KAT* gene under positive selection was identified, reflecting variations of seed oil content among different yellowhorn cultivars. This study provide a foundation for further genetic improvement of yellowhorn.

Data availability statement

All sequencing data and genome assembly generated in this study are available on SRA at the NCBI with the accession numbers PRJNA924504.

Author contributions

JG and BD designed the experiment. JW and HH performed research and did the methodology, finish writing and editing. XL, YZ and JZ helped with the part of data and results. MTQ, HR and XY critically revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by Shanxi Agricultural University Doctoral Research Initiation Project (2021BQ17); Basic Research Program of Shanxi Province (202103021224168); The Shanxi Postdoctoral Research Fund Project (K462102907), Shanxi Province doctoral graduates and postdoctoral researchers come to work in Shanxi Province to reward the fund scientific research project (SXBYKY2021061).

Acknowledgments

We would like to acknowledge The Engineering Research Center of Coal-based Ecological Carbon Sequestration Technology of the Ministry of Education for suggestion.We would also like to acknowledge Beijing Biomarker Technology Co. Ltd. and Wuhan Carboncode Biotechnologies Co.,Ltd. for performing DNA sequencing and genome assembly.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1147946/full#supplementary-material

SUPPLEMENTARY FIGURE 1
XsoG11 yellowhorn tree used in this study.

SUPPLEMENTARY FIGURE 2
Evolution tree of three yellowhorn ("WF18", "XsoG11" and "Xsv2").

SUPPLEMENTARY FIGURE 3
The collinearity of three yellowhorn ("WF18","XsoG11" and "Xsv2").

References

Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2016). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600–1607. doi: 10.1093/bioinformatics/btl140

Bao, Z., and Eddy, S. R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276. doi: 10.1101/gr.88502

Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bi, Q., Zhao, Y., Du, W., Lu, Y., Gui, L., Zheng, Z., et al. (2019). Pseudomolecule-level assembly of the Chinese oil tree yellowhorn (*Xanthoceras sorbifolium*) genome. *GigaScience* 8 (6), giz070. doi: 10.1093/gigascience/giz070

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 1–9. doi: 10.1186/1471-2105-10-421

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175. doi: 10.1038/s41592-020-01056-5

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097

Du, S., Wang, Y. L., Kume, T., Zhang, J. G., Otsuki, K., Yamanaka, N., et al. (2011). Sapflow characteristics and climatic responses in three forest species in the semiarid loess plateau region of China. *Agric. For. Meteorol.* 151, 1–10. doi: 10.1016/j.agrformet.2010.08.011

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-c yields chromosome-length scaffolds. *Science* 356, 92–95. doi: 10.1126/science.aal3327

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-c experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y

Farrar, J., and Williams, M. (1991). The effects of increased atmospheric carbon dioxide and temperature on carbon partitioning, source-sink relations and respiration. *Plant Cell Environ.* 14, 819–830. doi: 10.1111/j.1365-3040.1991.tb01445.x

Footitt, S., Cornah, J. E., Pracharoenwattana, I., Bryce, J. H., and Smith, S. M. (2007). The *Arabidopsis 3-ketoacyl-CoA thiolase-2* (*kat2-1*) mutant exhibits increased flowering but reduced reproductive success. *J. Exp. Bot.* 58, 2959–2968. doi: 10.1093/jxb/erm146

Germain, V., Rylott, E. L., Larson, T. R., Sherson, S. M., Bechtold, N., Carde, J. P., et al. (2001). Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid β-oxidation and breakdown of triacylglycerol in lipid bodies of *Arabidopsis* seedlings. *Plant J.* 28, 1–12. doi: 10.1046/j.1365-313X.2001.01095.x

Gomes, K. A., Almeida, T. C., Gesteira, A. S., Lôbo, I. P., Guimarães, A. C. R., De Miranda, A. B., et al. (2010). ESTs from seeds to assist the selective breeding of jatropha curcas l. for oil and active compounds. *Genomics Insights* 3, GEI. S4340. doi: 10.4137/GEI.S4340

Goudriaan, J., Van Laar, H., and H.v. and Louwerse, W. (1985). Photosynthesis, CO_2 and plant production. *Wheat Growth Model.* 86, 107–122. doi: 10.1007/978-1-4899-3665-3_10

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* 29, 644. doi: 10.1038/nbt.1883

Haas, B., Papanicolaou, A., and Yassour, M. (2017). TransDecoder. Available at: https://github.com/TransDecoder/TransDecoder.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S. H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12, 491. doi: 10.1186/1471-2105-12-491

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kreps, J. A., and Kay, S. A. (1997). Coordination of plant metabolism and development by the circadian clock. *Plant Cell* 9, 1235-1244. doi: 10.1105/tpc.9.7.1235

Liang, Q., Fang, H., Liu, J., Zhang, B., Bao, Y., Hou, W., et al. (2021). Analysis of the nutritional components in the kernels of yellowhorn (*Xanthoceras sorbifolium* bunge) accessions. *J. Food Compos. Anal.* 100, 103925. doi: 10.1016/j.jfca.2021.103925

Liang, Q., Li, H., Li, S., Yuan, F., Sun, J., Duan, Q., et al. (2019). The genome assembly and annotation of yellowhorn (*Xanthoceras sorbifolium* bunge). *GigaScience* 8, giz071. doi: 10.1093/gigascience/giz071

Liang, Q., Liu, J., Fang, H., Dong, Y., Wang, C., Bao, Y., et al. (2022). Genomic and transcriptomic analyses provide insights into valuable fatty acid biosynthesis and environmental adaptation of yellowhorn. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.991197

Liu, C., Wang, Y., Peng, J., Fan, B., Xu, D., Wu, J., et al. (2022). High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. *Plant Commun.* 3, 100352. doi: 10.1016/j.xplc.2022.100352

Liu, F., Wang, P., Xiong, X., Zeng, X., Zhang, X., and Wu, G. (2021b). A review of nervonic acid production in plants: Prospects for the genetic engineering of high nervonic acid cultivars plants. *Front. Plant Sci.* 12, 626625. doi: 10.3389/fpls.2021.626625

Liu, H., Yan, X., Wang, X., Zhang, D., Zhou, Q., Shi, T., et al. (2021a). Centromere-specific retrotransposons and very-long-chain fatty acid biosynthesis in the genome of yellowhorn (*Xanthoceras sorbifolium*, sapindaceae), an oil-producing tree with significant drought resistance. *Front. Plant Sci.* 12, 766389. doi: 10.3389/fpls.2021.766389

Liu, X., Zhang, Y., Yang, H., Liang, Y., Li, X., Oliver, M. J., et al. (2020). Functional aspects of early light-induced protein (ELIP) genes from the desiccation-tolerant moss *Syntrichia caninervis*. *Int. J. Mol. Sci.* 21, 1411-1429. doi: 10.3390/ijms21041411

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

O'donnell, S., and Fischer, G. (2020). MUM&Co: accurate detection of all SV types through whole-genome alignment. *Bioinformatics* 36, 3242–3243. doi: 10.1093/bioinformatics/btaa115

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Quint, M., Delker, C., Franklin, K. A., and Wigge, P. A. (2016). Molecular and genetic control of plant thermomorphogenesis. *Nat. Plants* 2, 15190. doi: 10.1038/nplants.2015.190

Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdottir, H., Mesirov, J. P., and Aiden, E. L. (2018). Juicebox.js provides a cloud-based visualization system for Hi-c data. *Cell Syst.* 6, 256–258 e251. doi: 10.1016/j.cels.2018.01.001

Ruan, C. J., Yan, R., Wang, B. X., Mopper, S., Guan, W. K., and Zhang, J. (2017). The importance of yellow horn (*Xanthoceras sorbifolia*) for restoration of arid habitats and production of bioactive seed oils. *Ecol. Eng.* 99, 504–512. doi: 10.1016/j.ecoleng.2016.11.073

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf. Chapter 4 Unit* 4, 10. doi: 10.1002/0471250953.bi0410s25

UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9, e112963. doi: 10.1371/journal.pone.0112963

Wang, Y., Li, C., Zhang, J., Xu, X., Fu, L., Xu, J., et al. (2022). Polyunsaturated fatty acids promote the rapid fusion of lipid droplets in caenorhabditis elegans. *J. Biol. Chem.* 298, 102179. doi: 10.1016/j.jbc.2022.102179

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 47, W52–W58. doi: 10.1093/nar/gkz333

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Yao, Z.-Y., Qi, J.-H., and Yin, L.-M. (2013). Biodiesel production from *Xanthoceras sorbifolia* in China: Opportunities and challenges. *Renew. Sust. Energ. Rev.* 24, 57–65. doi: 10.1016/j.rser.2013.03.047

Zia, K., Rao, M., Sadaqat, M., Azeem, F., Fatima, K., Tahir Ul Qamar, M., et al. (2022). Pangenome-wide analysis of cyclic nucleotide-gated channel (CNGC) gene family in citrus spp. revealed their intraspecies diversity and potential roles in abiotic stress tolerance. *Front. Genet.* 13, 1034921. doi: 10.3389/fgene.2022.1034921

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 April 2023
DOI 10.3389/fpls.2023.1161539

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Suwen Lu,
Nanjing Agricultural University, China
Jing Zhao,
Anhui Agricultural University, China

*CORRESPONDENCE
Qiang Zhao
zhaoqiang000666@163.com
Yongzhang Wang
qauwyz@163.com

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 08 February 2023
ACCEPTED 22 March 2023
PUBLISHED 03 April 2023

CITATION
Zhu Y, Kuang W, Leng J, Wang X, Qiu L, Kong X, Wang Y and Zhao Q (2023) The apple 14-3-3 gene *MdGRF6* negatively regulates salt tolerance. *Front. Plant Sci.* 14:1161539. doi: 10.3389/fpls.2023.1161539

COPYRIGHT
© 2023 Zhu, Kuang, Leng, Wang, Qiu, Kong, Wang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The apple 14-3-3 gene *MdGRF6* negatively regulates salt tolerance

Yuqing Zhu[1,2,3,4], Wei Kuang[1,2,3,4], Jun Leng[1,2,3,4], Xue Wang[1,2,3,4], Linlin Qiu[1,2,3,4], Xiangyue Kong[1,2,3,4], Yongzhang Wang[1,2,3,4]* and Qiang Zhao[1,2,3,4]*

[1]College of Horticulture, Qingdao Agricultural University, Qingdao, Shandong, China, [2]Engineering Laboratory of Genetic Improvement of Horticultural Crops of Shandong Province, Qingdao Agricultural University, Qingdao, Shandong, China, [3]Laboratory of Quality & Safety Risk Assessment for Fruit (Qingdao), Ministry of Agriculture and Rural Affairs, Qingdao Agricultural University, Qingdao, Shandong, China, [4]Qingdao Key Laboratory of Modern Agriculture Quality and Safety Engineering, Qingdao Agricultural University, Qingdao, Shandong, China

The 14-3-3 (GRF, general regulatory factor) regulatory proteins are highly conserved and are widely distributed throughout the eukaryotes. They are involved in the growth and development of organisms *via* target protein interactions. Although many plant 14-3-3 proteins were identified in response to stresses, little is known about their involvement in salt tolerance in apples. In our study, nineteen apple 14-3-3 proteins were cloned and identified. The transcript levels of *Md14-3-3* genes were either up or down-regulated in response to salinity treatments. Specifically, the transcript level of *MdGRF6* (a member of the *Md14-3-3* genes family) decreased due to salt stress treatment. The phenotypes of transgenic tobacco lines and wild-type (WT) did not affect plant growth under normal conditions. However, the germination rate and salt tolerance of transgenic tobacco was lower compared to the WT. Transgenic tobacco demonstrated decreased salt tolerance. The transgenic apple calli overexpressing *MdGRF6* exhibited greater sensitivity to salt stress compared to the WT plants, whereas the *MdGRF6-RNAi* transgenic apple calli improved salt stress tolerance. Moreover, the salt stress-related genes (*MdSOS2*, *MdSOS3*, *MdNHX1*, *MdATK2/3*, *MdCBL-1*, *MdMYB46*, *MdWRKY30*, and *MdHB-7*) were more strongly down-regulated in *MdGRF6-OE* transgenic apple calli lines than in the WT when subjected to salt stress treatment. Taken together, these results provide new insights into the roles of 14-3-3 protein MdGRF6 in modulating salt responses in plants.

KEYWORDS

apple, 14-3-3 proteins, *MdGRF6*, salt stress, negative regulation

1 Introduction

Plants are often affected by environmental stresses such as high salinity, drought, and extreme temperatures, which can adversely inhibit their growth and development (Golldack et al., 2014). High salinity is a severe abiotic stressor that inhibits plant development and productivity (Yoshida et al., 2014). Many genes are involved in salt stress tolerance, including antioxidant protective enzymes, signal transduction, and transcription factors (TFs) (Pardo et al., 1998; Mao et al., 2017; Jiroutova et al., 2021). Among these, 14-3-3 proteins primarily act as molecular chaperones and are extensively involved in abiotic and biotic stress responses *via* regulation of the conformation, activity, stability, and subcellular localization of target proteins (Campo et al., 2012).

The 14-3-3 proteins are a family of highly conserved proteins found in all eukaryotes, and can be classified into two categories: the non-ϵ (non-epsilon) and ϵ (epsilon) types (Brennan et al., 2013; Cao et al., 2016). The target proteins of the 14-3-3 family are phosphorylated at certain sites, causing a conformational shift that allows the family to exist as homo or heterodimers (Yasuda et al., 2014). In plants, protein interactions allow the 14-3-3 protein to attach to the plasma membrane H^+-ATP (Jahn et al., 1997). This interaction, which mediates the ATP-driven transport of H^+ across membranes, is involved in the active transport to manage both the osmotic and ionic stresses under high-salinity conditions (Wang et al., 2021). The 14-3-3 protein characteristics allow for the control of a variety of environmental signaling pathways, including those related to drought, excessive salinity, and extreme temperatures (Li et al., 2013; He et al., 2015; Tian et al., 2015).

In recent years, a growing number of studies have investigated the molecular functions of the 14-3-3 proteins in plants. To date, 13, 8, 12, 18, 5, 6, and 11 14-3-3 proteins have been identified in *Arabidopsis* (*Arabidopsis thaliana*), rice (*Oryza sativa*), tomato (*Solanum lycopersicum*), soybean (*Glycine max*), barley (*Hordeum vulgare*), cotton (*Gossypium hirsutum*), and grape (*Vitis vinifera L.*), respectively (Xu and Shi, 2006; Schoonheim et al., 2007; Yao et al., 2007; Zhang et al., 2010; Li and Dhaubhadel, 2011). In tobacco, 14-3-3 proteins interact with RSG and CDPK1 to form a heterodimeric trimer, which is used to control the gibberellin pathway (Ito et al., 2014). In tobacco, A 14-3-3 protein is involved in biological stress response *via* the interaction with the plasma membrane oxidase (NtrbohD16) to regulate ROS production in plants (Elmayan et al., 2007). In response to heat, cold, and salt stresses, rice exhibits significant variations in *OsGRF* expression (Yao et al., 2007). The 14-3-3λ and 14-3-3κ proteins in *Arabidopsis* participate in response to salt stress by regulating SOS2 activity in the SOS (salt overly sensitive) pathway (Zhou et al., 2014). 14-3-3 proteins are involved in the browning pathway of potato tubers by regulating the antioxidant enzyme activity (Łukaszewicz et al., 2002). Additionally, 14-3-3 proteins demonstrate considerable up- or down-regulation in *Vitis vinifera L.* under cold and heat stresses, suggesting their possible function in the regulation of abiotic stress response (Cheng et al., 2018).

Plants cope with salt stress through complex physiological responses and molecular regulatory mechanisms. However, the precise role of apple 14-3-3 proteins in regulating salt stress remains largely unclear. Therefore, it is very valuable to clarify the function of 14-3-3 protein under salt stress. To identify the function of *Md14-3-3s*, a total of nineteen 14-3-3 genes in apples were analyzed in this study. Functional characterization demonstrated that *MdGRF6* negatively regulates salt tolerance. Therefore, this study provides the basis for future research examining the molecular processes of *MdGRF6* in controlling the salt stress response in apples.

2 Materials and methods

2.1 Plant materials and treatments

The apple (*Malus×domestica*) seedlings were cultured in MS medium with 1 mg/L 6-BA, 0.1 mg/L NAA, and 0.1 mg/L GA3 at 23 ± 1°C and 16 h light/8 h dark photoperiod. The apple calli were cultured for 20 days in the dark at room temperature (24°C) on MS medium with 3 mg/L 2, 4-D, and 0.4 mg/L 6-BA. Tobacco seeds were surface-sterilized with 2% sodium hypochlorite (NaClO) for 10 min followed by 75% ethanol wash for 1 min. Then the seeds were washed with sterile water five times and cultured on MS medium with 0.8% agar and 2.5% sucrose at 23°C under 16 h light/8 h dark photoperiod.

For tissue expression analysis, the samples (roots, stems, leaves, fruits, and seeds) were collected from ten five-year-old apple trees grown in the experimental field of the horticulture orchard of the Qingdao Agricultural University (Shandong Province, China) in October 2022. For gene expression, the apple seedlings were treated with 100 mmol/L NaCl and harvested at 0, 1, 3, 6, and 12 h after treatment. Following the collection of each sample, the seedlings were immediately flash-frozen in liquid nitrogen and stored at -80°C until future use.

2.2 RT-qPCR analysis

Total RNA was extracted using the RNA Plant Plus Reagent (Tiangen, Beijing, China) and the first-strand cDNA was synthesized using the PrimeScript First Strand cDNA Synthesis Kit (Takara, Dalian, China). The ChamQ SYBR Mixture (Vazyme) was used for RT-qPCR reactions using a LightCycler 480 II RT-qPCR system. The cycle threshold (Ct) $2^{-\Delta\Delta CT}$ approach was used to analyze relative gene expression (Livak and Schmittgen, 2001). *MdACTIN* (apple) and *NtACTIN* (tobacco) were used as internal controls. Each RT-qPCR sample was done at least in triplicate. The primers are listed in Supplemental Table S2.

2.3 Bioinformatic analysis

Sequences of the 14-3-3 protein from apple, rice, and *Arabidopsis* were aligned using ClustalW in MAGEX. Neighbor-joining was used to construct the molecular phylogenetic trees with 1,000 reiterations. iTOL (https://itol.embl.de/itol.cgi) was used to annotate the 14-3-3

protein sequence information. TBtools was used to identify the chromosomal locations of *Md14-3-3* genes, using the apple genome annotation file (gene_models_20170612.gff3.gz) (Chen et al., 2020). Plant-CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was then used to analyze the *cis*-elements in the promoters of *Md14-3-3* genes (2,000-bp fragment at the upstream of the start codon) (Wang et al., 2014). Protein sequences of the *Md14-3-3* genes family were submitted to MEME (https://meme-suite.org/meme/doc/meme.html) for conserved motif analysis (Tripathi et al., 2016).

2.4 Plasmid construction and genetic transformation

The full-length of *MdGRF6* was cloned into the overexpression vector *pRI101-EGFP* to obtain the *35S::MdGRF6-EGFP* recombinant plasmids. The *MdGRF6-RNAi* were constructed using the RNAi vector. To generate the P_{MdGRF6}*::GUS* reporter construct, ~2 kb *MdGRF6* promoter fragments were cloned from apple genomic DNA and inserted into the *pCAMBIA1391::GUS* vector containing the GUS reporter gene. The primers in this study are listed in Supplemental Table S2.

The recombinant vectors *35S::MdGRF6-EGFP*, *MdGRF6-RNAi*, and *35S::EGFP* were introduced into *Agrobacterium rhizogenes* strain EHA105. The transgenic apple calli or tobacco plants were generated by *Agrobacturium*-mediated transformation method as described previously (Gallois and Marinho, 1995; An et al., 2017). The calli were subcultured on the MS medium containing 30 mg/L kanamycin, and the rooted transgenic tobacco plants were planted into the soil.

2.5 Subcellular localization analysis of *MdGRF6*

For subcellular localization, the *35S::MdGRF6-EGFP* or *35S::EGFP* vector was transformed into an *Agrobacterium* EHA105 strain and used for transient expression in *Nicotiana benthamiana* leaves. The fluorescence was then observed after 2-3 days using a laser confocal microscope.

2.6 The histochemical GUS staining analysis

The 30-day P_{MdGRF6}*::GUS* transgenic tobacco was used for tissue expression analysis. Young leaf, stem, and root tissues were visualized using the GUS Staining Kit (Solarbio, Beijing, China).

2.7 Salt stress treatment

The WT and transgenic apple calli were treated with 100 mM NaCl or 150 mM NaCl for 15 days. Following treatment, the apple calli were collected and the fresh weight was measured. Relative conductivity was determined according to Yang et al. method (Yang et al., 2022), using a DDSJ-318 conductometer (Yidian Scientific Instrument Co., Ltd., Shanghai, China). The thiobarbituric acid-based method was used to determine the malondialdehyde (MDA) content and the absorbance value of the reaction solution was determined using a spectrophotometer at 600, 532, and 450nm (Yoke Istrument Co., Ltd., Shanghai, China).

WT and transgenic tobacco seeds were sterilized and sown onto MS, MS + 50 mmol/L NaCl, or MS + 100 mmol/L NaCl plates to detect the germination rates. Following seed germination, consistently growing transgenic and WT tobacco were transferred to MS medium with 50 mM NaCl or 100 mM NaCl for 10 days. After salt treatment, the fresh weight, root length, chlorophyll, and MDA content were measured. As previously described, the MDA content was measured using the thiobarbituric acid method. Chlorophyll was extracted with 95% ethanol, and the supernatant was used to determine the absorbance values at 665 and 649 nm (Yoke Instrument Co., Ltd., Shanghai, China).

Uniformly-developed seedlings of transgenic and WT tobacco plants were treated with 350 mM NaCl for four weeks. Water was used as the control. After four weeks, the chlorophyll content and relative conductivity were determined as above described. Diaminobenzidine (DAB) and p-nitroblue tetrazolium chloride (NBT) histochemical staining were used to determine H_2O_2 and O_2^- contents in tobacco leaves, respectively. The activities of enzymes SOD and POD were measured using the SOD and POD assay kits (Solarbio, Beijing, China), respectively. The above experiments were replicated three times.

3 Results

3.1 Identification and phylogenetic tree analysis of *MdGRF* genes

To identify the 14-3-3 genes in the apple genome, the BLASTp and the Pfam tool were used. A total of twenty 14-3-3 genes were identified in the apple genome based on their homology to the 14-3-3 protein sequences in *Arabidopsis* from the TAIR database (http://www.arabidopsis.org). However, only nineteen genes were cloned and labeled as *MdGRF1-MdGRF19* based on their chromosomal location (Table S1). The phylogenetic tree demonstrated that the Md14-3-3 proteins were divided into two groups (ϵ and non-ϵ group) (Figure 1A). Following the protein 3D models, the 14-3-3 secondary structures were classified into three divisions. The middle position domain structure was similar, while the differences were the N-terminal and C-terminal structures (Figure 1B).

3.2 Analysis of *cis*-acting elements in the promoters of *Md14-3-3* genes

The *cis*-acting elements in the promoter regions of *Md14-3-3* genes were examined using the PlantCARE database (Figure 2). Based on their functions, the *cis*-acting elements were divided into

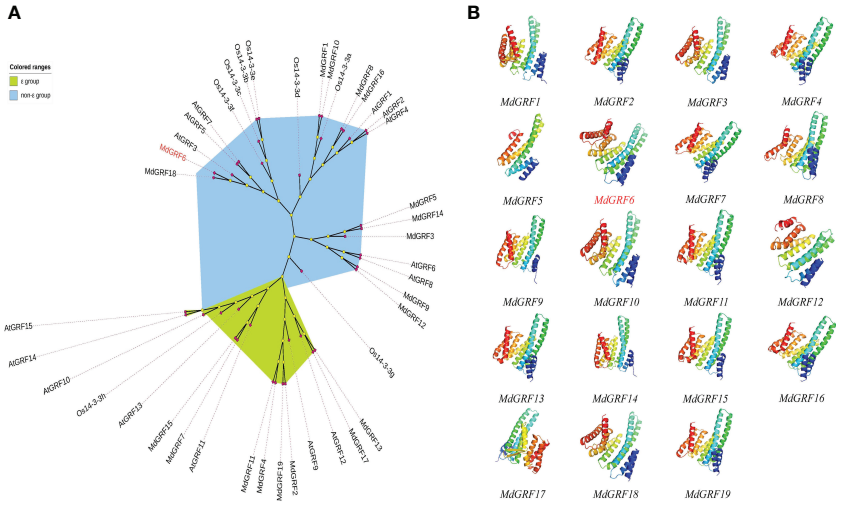


FIGURE 1
Phylogenetic tree and protein structures of Md14-3-3s. **(A)** The NJ phylogenetic tree (1,000 bootstrap replicates) was generated with 14-3-3 proteins from apples, rice, and *Arabidopsis*. **(B)** Phyre2 in 'Normal' mode and Pymol were used to generate all 3D models of MdGRFs. Images were colored by rainbow N → C terminus. All models were based on template d1o9da or c3e6yB.

three groups: biotic/abiotic stress, growth and development, and phytohormone response. The *cis*-acting elements of plant biotic/abiotic stress, including salicylic acid (TCA-element: CCATCTTTTT), defense and stress (TC-rich repeats: GTTTTCTTAC), low temperature (LTR: CCGAAA), MYB-binding site involving in drought-inducibility (MBS: CAACTG), were identified in the promoters of the *Md14-3-3* genes. A large number of core *cis*-acting elements (such as TATA box, CAAT box, and others) were also identified in the *Md14-3-3s* promoter regions. In addition, multiple elements responding to phytohormones were identified. The promoters of *Md14-3-3s* also contained *cis*-acting elements responding to TGA-element (AACGAC), ABRE (ACGTG), GARE (TCTGTTG), and CGTCA that involved in auxin (IAA), abscisic acid (ABA), gibberellin (GA), and methyl jasmonate (MeJA) response.

3.3 Chromosome localization, gene structure, and conserved *Md14-3-3* motifs analysis

The *Md14-3-3* genes were identified on twelve apple chromosomes (Chr): Chr00, 01, 05, 06, 07, 08, 10, 12, 13, 15, 16, and 17 (Figure 3A). The intron/exon analysis showed that the *Md14-3-3* genes contained from 0 to 6 introns (Figure 3B). Among these, *MdGRF4*, *7*, *11*, *13*, *15*, *17*, and *19* contained the most introns (6 introns), while *MdGRF5* had none. Additionally, the phylogenetic tree demonstrated that *MdGRF6* and *18*, which were both situated on the same branch, had similar intron/exon distribution (Figure 3B). Eight motifs in the Md14-3-3 proteins were predicted using the Multiple Em for Motif Elicitation (MEME) program. Most of the conserved motifs were similar in distribution (Figure 3C). However,

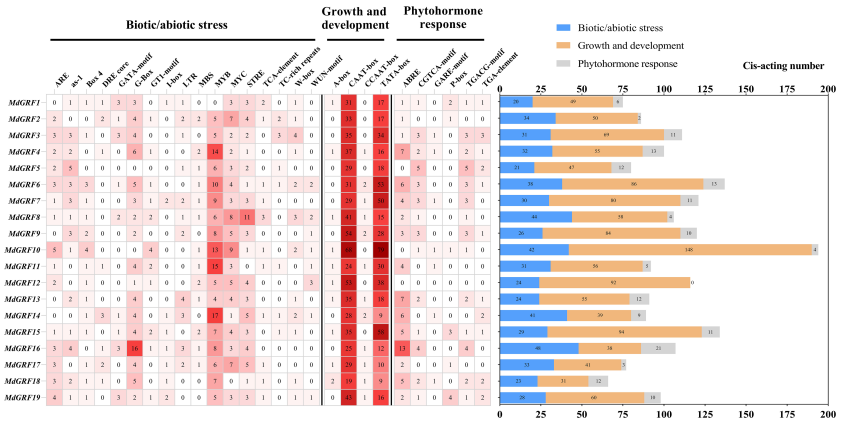


FIGURE 2
Promoter analysis of *Md14-3-3s*. The amount of various *cis*-acting elements are presented in the grid by the different numbers and color tones. The sum of the *cis*-acting elements in the three classes of each gene is displayed in the histogram.

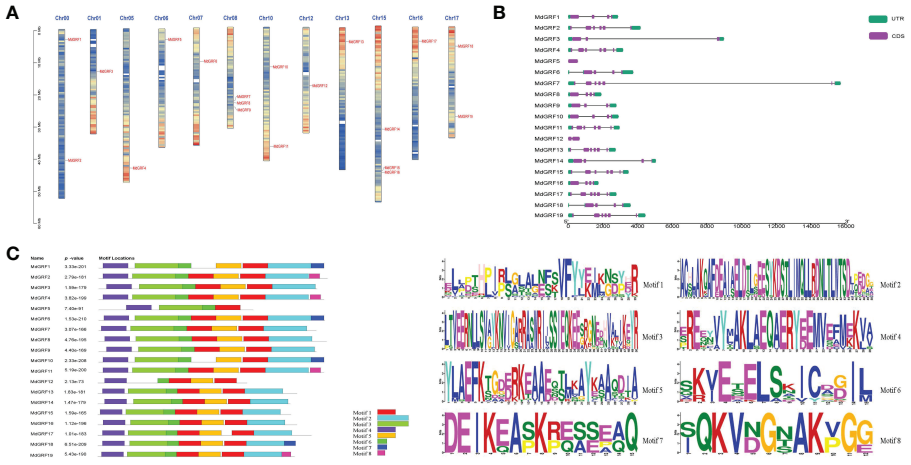


FIGURE 3
Characterization and chromosomal localization of the *Md14-3-3* genes. **(A)** Chromosomal localization of 14-3-3 genes in apple. **(B)** Structural analysis of the *Md14-3-3* genes. Different blocks represent UTRs (green), CDs (purple), and introns (black lines). **(C)** The motif details of the identified conserved domains of the Md14-3-3 proteins. Each motif is represented by a different colored box.

these conserved motif differences may be responsible for different gene functions.

3.4 Expression patterns of *MdGRF6* and subcellular localization

Previous studies have shown that 14-3-3 genes play an important role in plant response to salt stress (Zhou et al., 2014). Here, the transcript levels of the *MdGRFs* family members were up or down-regulated in response to salinity treatments (Figure S1). One of the members, *MdGRF6* was chosen for further analysis. *MdGRF6* was down-regulated under salt treatment, with a 0.42-fold down-regulation at the sixth hour. However, there was no significant change in the control group (Figures 4A, B). Next, gene expression in apple roots, stems, leaves, fruits, and seeds were analyzed. The results demonstrated that *MdGRF6* was highly expressed in seeds, with the lowest expression in leaves (Figure 4C). To confirm *MdGRF6* these results, the P_{MdGRF6}*::GUS* transgenic tobacco plants were then developed. GUS signals were detected in the roots, stems, and leaves of the transgenic tobacco seedlings (Figure 4D).

In addition, the constructed expression vectors *35S::MdGRF6-EGFP* or *35S::EGFP* were injected into *Nicotiana benthamiana* leaves to examine the subcellular localization of MdGRF6. As a result, we found that MdGRF6 was localized in the cytoplasm and cell membrane (Figure S2).

3.5 Overexpression of *MdGRF6* increased sensitivity to salt stress in transgenic tobacco

To characterize the function of *MdGRF6* under salt stress, three independent transgenic tobacco lines (#1, #2, and #3) were selected for further analyses using RT-qPCR and western blotting (Figure S3). The germination rates of the WT and *MdGRF6-OE* lines on MS medium with or without NaCl were first determined. The germination rates were comparable on MS-only media (Figures 5A, B). However, supplementation with 50mM or 100mM NaCl significantly reduced the germination rates of all lines, but had the most significant effects in the transgenic plants (Figures 5A, B). Furthermore, the roots were significantly shorter in the *MdGRF6-OE* transgenic plants compared to WT (Figures 5C, D). Under salt stress, *MdGRF6-OE* transgenic plants exhibited lower fresh weights and chlorophyll content compared to WT (Figures 5E, F), and exhibited significantly higher electrolyte leakage and MDA content compared to WT (Figures 5G, H).

In addition, 30-day-old soil-grown seedlings of WT and transgenic plants were supplied with water or 350 mM NaCl for four weeks. The results of observed phenotypes were consistent with result of germination rates (Figure 6A). ROS levels may increase in response to abiotic stress, and their accumulation is harmful to plant cells (Sun et al., 2010). Under salt treatment, DAB and NBT histochemical staining of transgenic leaves were more intense than in WT than *MdGRF6-OE* plants, and *MdGRF6* transgenic lines accumulated more H_2O_2 and $O_2^{\cdot-}$ contents compared to WT plants (Figures 6B, C). Consistent with this data, the activities of important antioxidant enzymes (POD and SOD) displayed lower in *MdGRF6-OE* plants compared to WT (Figures 6F, G). Furthermore, the increased electrical conductivity and decreased chlorophyll content further confirmed that *MdGRF6* overexpression reduced salt tolerance (Figures 6D, E). According to the above results, the ectopic expression of *MdGRF6* decreased tolerance to salt stress by impeding the activity of several antioxidant enzymes and causing H_2O_2 and $O_2^{\cdot-}$ accumulation.

3.6 *MdGRF6* negatively regulates salt stress tolerance in transgenic calli

To further investigate *MdGRF6* function under salt stress, the *MdGRF6-OE* and *MdGRF6-RNAi* transgenic apple calli were

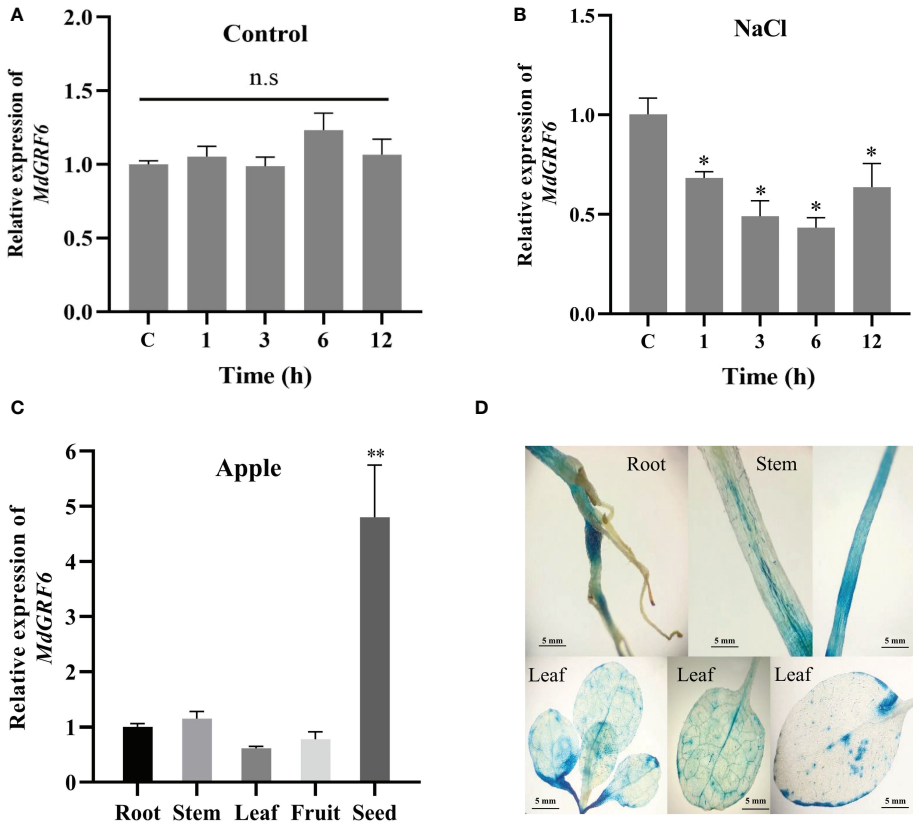


FIGURE 4
Expression patterns of *Md14-3-3* genes. **(A, B)** RT-qPCR analysis of *MdGRF6* expression in apple seedlings treated without or with 100 mM NaCl for the indicated time. **(C)** Expression analysis of *MdGRF6* in different apple tissues. **(D)** P_{MdGRF6}::*GUS* expression pattern in transgenic tobacco in different tissues. Data are presented as means ± SD (n = 3). Asterisks indicate significant differences LSD test, $^{**}P < 0.01$; $^{*}P < 0.05$; ns, no significant difference.

obtained. RT-qPCR demonstrated that *MdGRF6-OE* and *MdGRF6-RNAi* transgenic calli generated significantly higher or lower expression levels compared to WT, respectively (Figure S4). Then, the 15-day-old WT, *MdGRF6-OE*, and *MdGRF6-RNAi* transgenic calli were placed on MS medium containing 100 mM or 150 mM NaCl. The results demonstrate that *MdGRF6-OE* transgenic calli grew significantly slower than WT, and *MdGRF6-RNAi* transgenic calli grew much stronger than WT (Figure 7A). In agreement with the observed phenotype, *MdGRF6-OE* calli exhibited higher MDA level, lower fresh weight, and higher electrical conductivity compared to the WT, whereas *MdGRF6-RNAi* transgenic calli reduced MDA content, electrical conductivity, and had higher fresh weight under salt stress (Figures 7B–D). Overall, the results suggest that *MdGRF6* overexpression improved sensitivity to salt in transgenic apple calli.

3.7 Overexpression of *MdGRF6* downregulates salt stress-related gene expression

To investigate the molecular mechanisms of *MdGRF6*-mediated salt stress tolerance, RT-qPCR was used to assess the expression levels of salt stress-related genes (*MdSOS2*, *MdSOS3*, *MdNHX1*, *MdATK2/3*, *MdCBL-1*, *MdMYB46*, *MdWRKY30*, and *MdHB-7*). Under normal conditions, the transcript abundance did not differ between the *MdGRF6-OE* transgenic calli and WT. However, the expression of these genes was considerably lower in *MdGRF6-OE* transgenic calli, and was higher in *MdGRF6-RNAi* transgenic calli compared to WT calli under salt stress (Figures 8A–H). The above results suggested that *MdGRF6* may negatively regulate these genes under salt stress.

4 Discussion

Salinization is one of the most increasingly severe environmental and ecological issues that threaten the limited soil resources on which humans depend and poses a significant constraint on the sustainability of crop yields (Jiao et al., 2019; Hassani et al., 2021; Li et al., 2021). As crucial regulatory proteins in signaling networks involved in adaptation to various abiotic pressures, 14-3-3 proteins have recently attracted considerable interest. In the current study, a 14-3-3 gene (*MdGRF6*) that negatively regulated salt stress tolerance in apple was identified (Figures 6, 7).

The 14-3-3 proteins contained highly conserved proteins that are widely expressed in all eukaryotes. The family has fifteen and

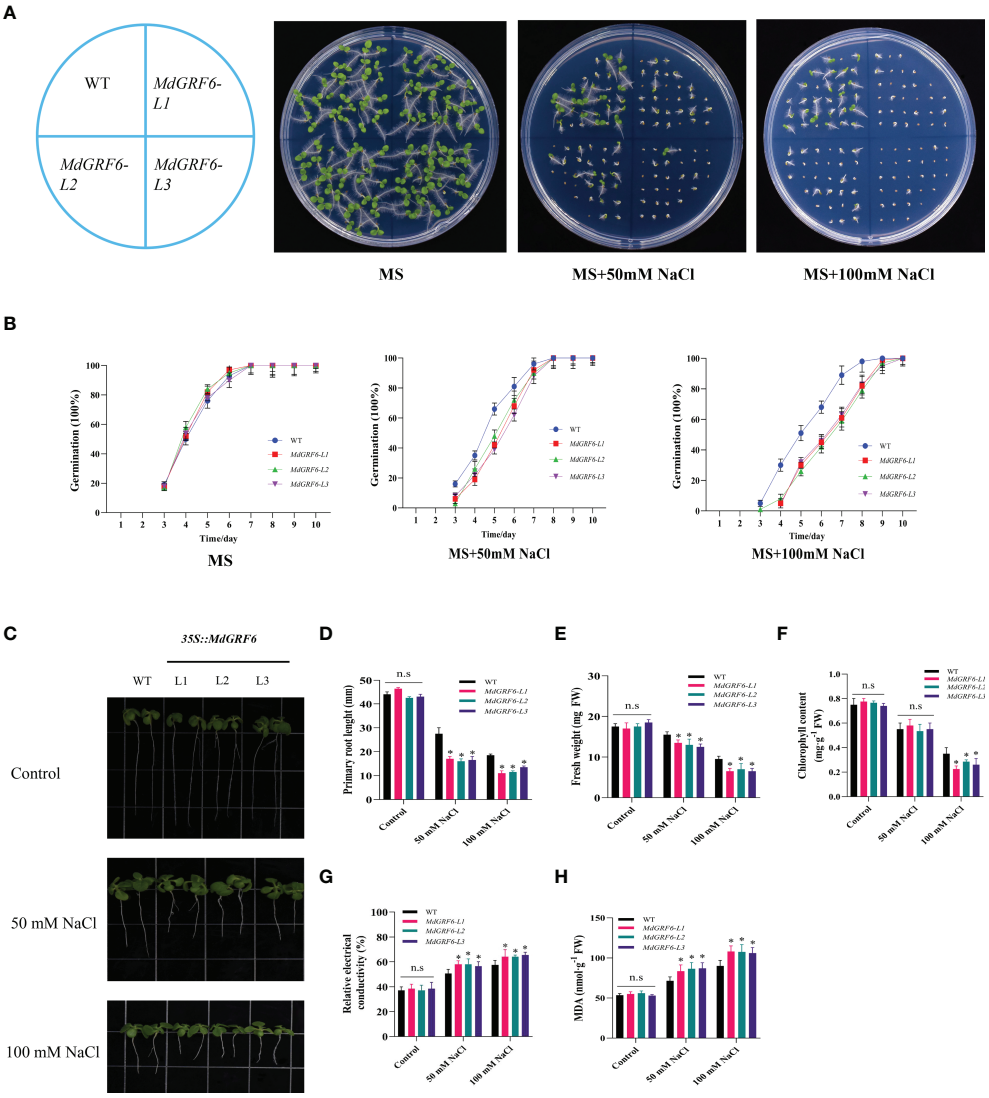


FIGURE 5
Salt response of *35S::MdGRF6* transgenic tobacco plants during germination. **(A)** Effects of salt treatment on the germination of WT and *35S::MdGRF6* transgenic tobacco lines. Germination was assessed at the indicated time. **(B)** The WT and *35S::MdGRF6* transgenic tobacco lines were grown on MS medium supplemented with 0, 50, or 100 mM NaCl. The seeds germinated after 10 days. Representative seedlings are presented in the images. **(C)** The seedlings were grown vertically for 4 days on MS medium and then moved to the medium containing different doses of 50 or 100 mM NaCl for an additional 10 days in a vertical position. **(D)** Root elongation of WT and *35S::MdGRF6* transgenic tobacco seedlings in response to salt stress. **(E–H)** Fresh weight **(E)**, chlorophyll content **(F)**, relative electrical conductivity **(G)**, and MDA content **(H)** in WT and *35S::MdGRF6* transgenic tobacco seedlings under control or salt stress conditions. Error bars indicate the means ± SD (n = 3). The asterisks indicate significant differences (LSD test, $*P < 0.05$; ns, no significant difference).

eight members in the *Arabidopsis* and rice genomes, respectively (DeLille et al., 2001; Rosenquist et al., 2001; Yao et al., 2007). Previously, eighteen or twenty *Md14-3-3* gene family members have been identified in the apple genome (Ren et al., 2019; Zuo et al., 2021). Here, twenty *Md14-3-3* genes were identified in the apple (Table S1), similar to Ren et al. (2019). However, the *MD17G1105100* could not be cloned due to either its low expression level or due its non-existence. The variations in the raw high-throughput genomic sequences were likely due to splicing errors of the DNA fragments. Additionally, variations in apple germplasm resources cannot be completely ruled out (Herndon et al., 2020; Kuo et al., 2020). Previous studies have shown that 14-3-3 proteins have a significant impact on stress resistance (Roberts, 2003). In rice, OsGF14b improves salt tolerance by interacting with OsPCL1, inhibiting its ubiquitination for protection from degradation, and promoting its activity and stability (Wang et al., 2023). Overexpression of both heterologous *PvGF14a* and *PvGF14g* in transgenic *Arabidopsis* under salt stress reduces seed germination and fresh seedling weight, suggesting the involvement of these genes in the negative regulation of salt tolerance in seedlings (Li et al., 2018). In *Arabidopsis thaliana*, *GRF3* is crucial in osmotic stress response and root growth, as well as negatively regulating mitochondrial retrograde control of *AOX1a* expression *via* the ROS pathway (Li et al., 2022). Notably, we observed that *MdGRF6* was homologous to *AtGRF3* and both belonged to the non-ϵ group (Figure 1A). In the present study, *MdGRF6* is a

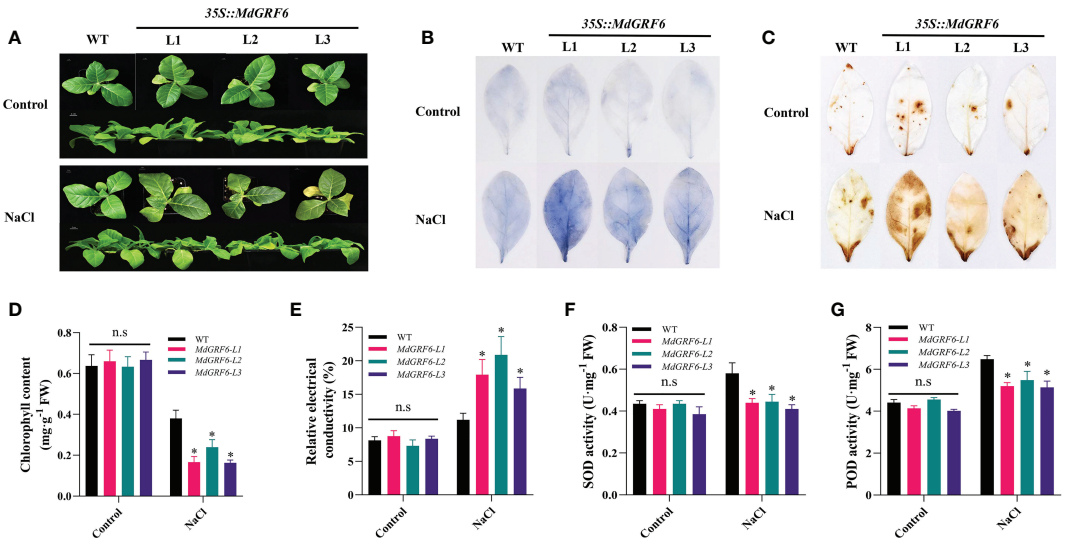


FIGURE 6
Effects of salt on the growth of germinated WT and *35S::MdGRF6* transgenic tobacco plants. **(A)** The phenotype of WT and three transgenic tobacco lines under salinity stress conditions. **(B, C)** Following salt stress treatments, leaves from *MdGRF6* transgenic lines and WT were stained for H_2O_2 using DAB staining and for $O_2^{\cdot-}$ using NBT staining. Statistical analysis of **(D)** chlorophyll content, **(E)** electrical conductivity, **(F)** SOD activity, and **(G)** POD activity in WT and transgenic lines after salt treatments. Error bars indicate the means ± SD (n = 3). The asterisks indicate significant differences (LSD test, *$P < 0.05$; ns, no significant difference).

negative regulator and its expression was reduced due to salt stress, similar to *AtGRF3*. The 14-3-3 protein family, therefore, exhibits high conservation of orthologs among different species (Zhou et al., 2014; Li et al., 2018).

Salt stress induces cell membrane damage primarily through osmotic and ion stresses (Golldack et al., 2014; Yoshida et al., 2014). Multiple evidences indicate that the 14-3-3 proteins are primarily localized at the plasma membrane to modify multiple ion channels and alleviate osmotic stress (Yang et al., 2019). Using the 14-3-3 omega (At1g78300) from *Arabidopsis* as bait, ion transport proteins (Ca^{2+}, K^+, and Cl^-) were discovered using proteomic analysis of tandem affinity purified 14-3-3 protein complexes (Chang et al., 2009). Relative electrical conductivity reflects the state of the plant membrane system. An increase in the conductivity of the external

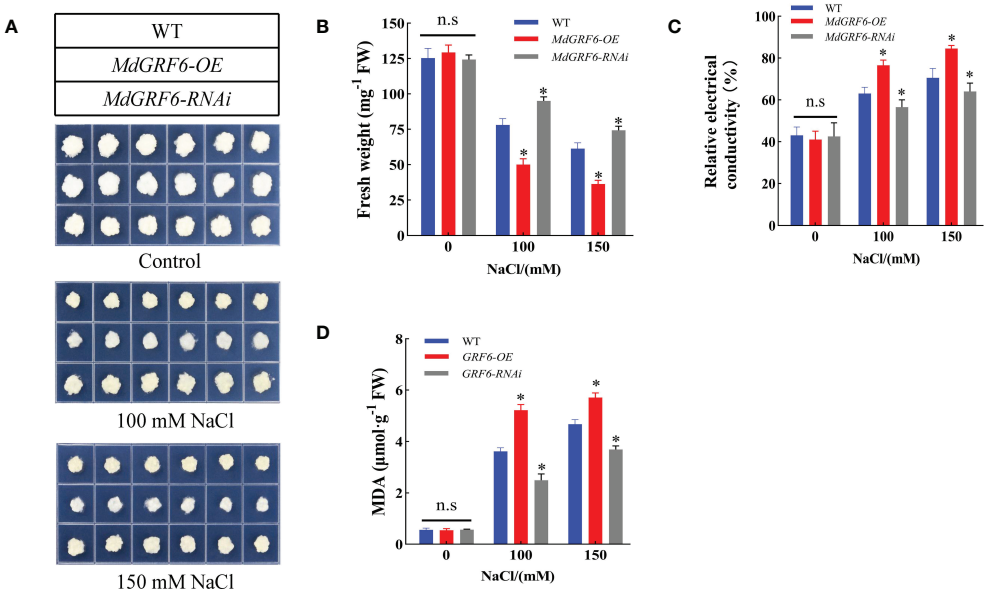


FIGURE 7
MdGRF6 enhanced sensitivity to salt stress in apple. **(A)** The phenotypes of transgenic and WT apple calli in response to NaCl treatment. The apple calli were placed on the MS medium containing 100 or 150 mM NaCl for 15 days. **(B-D)** Fresh weight, electrical conductivity, and MDA content of the WT and transgenic apple calli after 15 days under salt treatment. Error bars indicate the means ± SD (n = 3). The asterisks indicate significant differences (LSD test, *$P < 0.05$; ns, no significant difference).

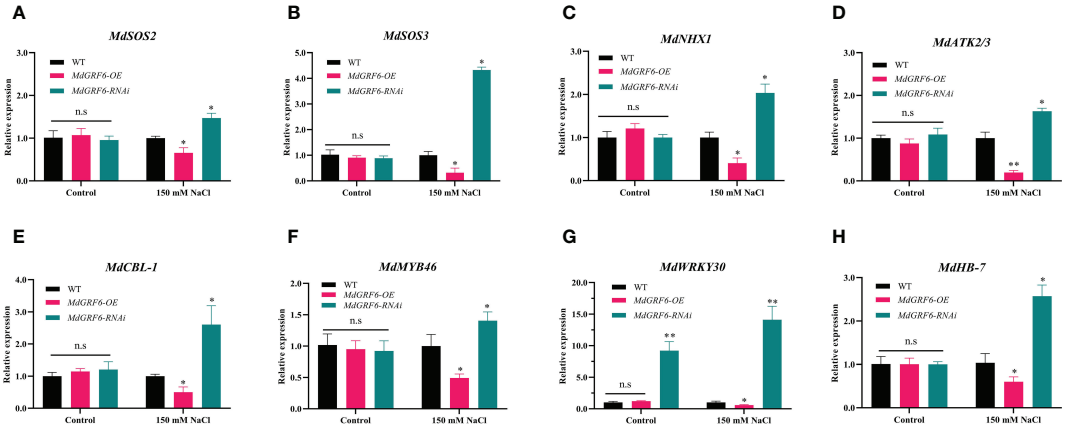


FIGURE 8
MdGRF6 affected the expression of salt stress-related genes in response to salt stress. **(A–H)** The relative expression of *MdSOS2* **(A)**, *MdSOS3* **(B)**, *MdNHX1* **(C)**, *MdATK2/3* **(D)** *MdCBL-1* **(E)**, *MdMYB46* **(F)**, *MdWRKY30* **(G)**, and *MdHB-7* **(H)** in WT, *MdGRF6-OE* and *MdGRF6-RNAi* transgenic calli under control and salt treatments. Error bars indicate the means ± SD (n = 3). The asterisks indicate significant differences LSD test, $^{**}P < 0.01$; $^{*}P < 0.05$; ns, no significant difference.

medium suggests cell membrane damage and ion outflow (Ilik et al., 2018). Similarly, MDA content is an essential indicator of lipid peroxidation in plants and reflects their resistance to external adversity. When various enzymes and membrane systems in plant tissues are disrupted, MDA content increases (Hernandez et al., 2010; Mo et al., 2016). In the present study, the MDA content and relative electrical conductivity were higher in *MdGRF6-OE* calli and tobacco tissues compared to WT under salt stress. Thus, this study further confirmed that overexpression of *MdGRF6* accelerated membrane damage at high salinity.

Under abiotic stresses, plants commonly produce a large amount of ROS, which can damage the mitochondria, chloroplasts, and cell membranes unless the ROS are promptly removed (Li et al., 2015; Jia et al., 2019). Protective substances, such as SOD and POD, can scavenge ROS to alleviate salt stress and protect plant cell membrane structure (Liang et al., 2017; Guo et al., 2020). Many studies have shown that 14-3-3 proteins play a significant role in protection from osmotic and oxidative stress in plants (Yan et al., 2004). For example, RBOHD-dependent H_2O_2 operates upstream of H^+-ATPase and 14-3-3 proteins, required for salt tolerance in pumpkin (Huang et al., 2019). Here, we observed that the activities of antioxidant enzymes SOD and POD were inhibited in the *MdGRF6-OE* transgenic tobacco (Figures 6F, G). In addition, ROS-induced cellular damage leads to the destruction of photosynthetic machinery including chlorophyll and carotenoids (Cardenas-Perez et al., 2020). Just like this, the chlorophyll contents were markedly decreased in the transgenic tobacco lines compared to WT (Figures 5F, 6D). In *Arabidopsis*, it was found that 14-3-3 proteins can respond to salt stress by regulating the activity of antioxidant enzymes (Visconti et al., 2019; Wang et al., 2019). In general, 14-3-3 proteins can interact with many TFs to regulate salt stress resistance in plants (Chang et al., 2019; Wang et al., 2023). For example, 14-3-3 proteins can interact with GmMYB173 to regulate antioxidant enzyme activity and hydrogen peroxide scavenging in soybean to regulate salt resistance (Pi et al., 2018). Just like this, MdGRF6 protein may regulate the activity of antioxidant enzymes in response to salt stress by interacting with these TFs. In addition, many studies have also confirmed that 14-3-3 proteins can regulate the expression of antioxidant enzyme gene in response to stresses (Manosalva et al., 2011; Li et al., 2014). In apple, many genes such as *MdMYB46*, *MdWRKY30*, and *MdHB-7* are important components of the gene network that is involved in the regulation of reactive oxygen species-related genes, and they regulate ROS scavenging in response to salt stress (Chen K. et al., 2019; Dong et al., 2020; Zhao et al., 2021). Here, we found that the expression of *MdMYB46*, *MdWRKY30*, and *MdHB-7* was decreased in *MdGRF6-OE* plants, and increased in *MdGRF6-RNAi* plants (Figures 8F–H). This may be due to the fact that MdGRF6 interacts with other proteins to indirectly regulate their expression to modulate salt stress resistance.

Previous studies have shown that many genes can also affect ionic balance (Na^+/K^+ ratio) pathways in response to salt stress (Chen Z. X. et al., 2019). For example, the *MdSOS2*, *MdSOS3*, *MdNHX1*, *MdATK2/3*, and *MdCBL-1* were reported to involve in salt stress response (Shabala, 2013; Hu et al., 2016; Borkiewicz et al., 2020; Su et al., 2020). We found that the *MdSOS2*, *MdSOS3*, *MdNHX1*, *MdATK2/3*, and *MdCBL-1* genes were upregulated in *MdGRF6-RNAi* transgenic calli, but decreased in *MdGRF6-OE* lines under salt stress conditions (Figures 8A–E). Many functional studies have demonstrated that 14-3-3 proteins interact with the WRKY family (Chang et al., 2009; Rushton et al., 2010). In *Fortunella crassifolia*, FcWRKY40 plays an active role in salt tolerance by directly regulating *SOS2* and *P5CS1* to control ion homeostasis (Dai et al., 2018). 14-3-3 interacts with the SOS pathway proteins to cope with high salinity, attenuating interactions with AtSOS2 and activating AtSOS2 kinase activity under salt stress conditions in *Arabidopsis* (Yang et al., 2019). Therefore, we speculated that *MdGRF6* may directly regulate the expression of those genes, or indirectly regulate them by interacting with WRKY-TFs.

In this study, the mechanism by which *MdGRF6* elevates the salt sensitivity of apple possibly by regulating the activity of the

antioxidant enzymes and expression of salt stress-responsive genes were investigated. In conclusion, the molecular biological functions of *MdGRF6* were investigated to provide a basis for a more in-depth exploration of the molecular mechanisms of salt stress in apples. Moreover, this study enhanced the understanding of the biological functions of the 14-3-3 gene family in apples.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

QZ and YW conceived and designed the research. YZ provided experimental materials. YZ, WK, JL, XW, LQ, and XK performed the experiments. YZ and QZ analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (2018YFD1000300), the Shandong Province Government (SDAIT-06-05), and the Talents of High-Level Scientific Research Foundation of Qingdao Agricultural University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1161539/full#supplementary-material

References

An, J. P., Qu, F. J., Yao, J. F., Wang, X. N., You, C. X., Wang, X. F., et al. (2017). The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. *Hortic. Res.* 4, 17023. doi: 10.1038/hortres.2017.23

Borkiewicz, L., Polkowska-Kowalczyk, L., Ciesla, J., Sowinski, P., Jonczyk, M., Rymaszewski, W., et al. (2020). Expression of maize calcium-dependent protein kinase (ZmCPK11) improves salt tolerance in transgenic *Arabidopsis* plants by regulating sodium and potassium homeostasis and stabilizing photosystem II. *Physiol. Plant* 168, 38–57. doi: 10.1111/ppl.12938

Brennan, G. P., Jimenez-Mateos, E. M., McKiernan, R. C., Engel, T., Tzivion, G., and Henshall, D. C. (2013). Transgenic overexpression of 14-3-3 zeta protects hippocampus against endoplasmic reticulum stress and status epilepticus *in vivo*. *PloS One* 8, e54491. doi: 10.1371/journal.pone.0054491

Campo, S., Peris-Peris, C., Montesinos, L., Peñas, G., Messeguer, J., and Blanca, S. S. (2012). Expression of the maize ZmGF14-6 gene in rice confers tolerance to drought stress while enhancing susceptibility to pathogen infection. *J. Exp. Bot.* 63, 983–999. doi: 10.1093/jxb/err328

Cao, H., Xu, Y., Yuan, L., Bian, Y., Wang, L., Zhen, S., et al. (2016). Molecular characterization of the 14-3-3 gene family in *Brachypodium distachyon* l. reveals high evolutionary conservation and diverse responses to abiotic stresses. *Front. Plant Sci.* 7, 1099. doi: 10.3389/fpls.2016.01099

Cardenas-Perez, S., Piernik, A., Ludwiczak, A., Duszyn, M., Szmidt-Jaworska, A., and Chanona-Perez, J. J. (2020). Image and fractal analysis as a tool for evaluating salinity growth response between two salicornia europaea populations. *BMC Plant Biol.* 20, 467. doi: 10.1186/s12870-020-02633-8

Chang, I. F., Curran, A., Woolsey, R., Quilici, D., Cushman, J. C., Mittler, R., et al. (2009). Proteomic profiling of tandem affinity purified 14-3-3 protein complexes in *Arabidopsis thaliana*. *Proteomics* 9, 2967–2985. doi: 10.1002/pmic.200800445

Chang, H. C., Tsai, M. C., Wu, S. S., and Chang, I. F. (2019). Regulation of ABI5 expression by ABF3 during salt stress responses in *Arabidopsis thaliana*. *Botanical Stud.* 60, 16. doi: 10.1186/s40529-019-0264-z

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, K., Song, M., Guo, Y., Liu, L., Xue, H., Dai, H., et al. (2019). MdMYB46 could enhance salt and osmotic stress tolerance in apple by directly activating stress-responsive signals. *Plant Biotechnol. J.* 17, 2341–2355. doi: 10.1111/pbi.13151

Chen, Z. X., Zhao, X. Q., Hu, Z. H., and Leng, P. S. (2019). Nitric oxide modulating ion balance in hylotelephium erythrostictum roots subjected to NaCl stress based on the analysis of transcriptome, fluorescence, and ion fluxes. *Sci. Rep.* 9, 18317. doi: 10.1038/s41598-019-54611-2

Cheng, C., Wang, Y., Chai, F., Li, S., Xin, H., and Liang, Z. (2018). Genome-wide identification and characterization of the 14-3-3 family in vitis vinifera l. during berry development and cold- and heat-stress response. *BMC Genomics* 19, 579. doi: 10.1186/s12864-018-4955-8

Dai, W. S., Wang, M., Gong, X. Q., and Liu, J. H. (2018). The transcription factor FcWRKY40 of fortunella crassifolia functions positively in salt tolerance through modulation of ion homeostasis and proline biosynthesis by directly regulating SOS2 and P5CS1 homologs. *New Phytol.* 219, 972–989. doi: 10.1111/nph.15240

DeLille, J. M., Sehnke, P. C., and Ferl, R. J. (2001). The *arabidopsis* 14-3-3 family of signaling regulators. *Plant Physiol.* 126, 35–38. doi: 10.1104/pp.126.1.35

Dong, Q., Zheng, W., Duan, D., Huang, D., Wang, Q., Liu, C., et al. (2020). MdWRKY30, a group IIa WRKY gene from apple, confers tolerance to salinity and osmotic stresses in transgenic apple callus and *Arabidopsis* seedlings. *Plant Sci.* 299, 110611. doi: 10.1016/j.plantsci.2020.110611

Elmayan, T., Fromentin, J., Riondet, C., Alcaraz, G., Blein, J. P., and Simon-Plas, F. (2007). Regulation of reactive oxygen species production by a 14-3-3 protein in elicited tobacco cells. *Plant Cell Environ.* 30, 722–732. doi: 10.1111/j.1365-3040.2007.01660.x

Gallois, P., and Marinho, P. (1995). Leaf disk transformation using agrobacterium tumefaciens-expression of heterologous genes in tobacco. *Methods Mol. Biol.* 49, 39–48. doi: 10.1385/0-89603-321-X:39

Golldack, D., Li, C., Mohan, H., and Probst, N. (2014). Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front. Plant Sci.* 5, 151. doi: 10.3389/fpls.2014.00151

Guo, H., Huang, Z., Li, M., and Hou, Z. (2020). Growth, ionic homeostasis, and physiological responses of cotton under different salt and alkali stresses. *Sci. Rep.* 10, 21844. doi: 10.1038/s41598-020-79045-z

Hassani, A., Azapagic, A., and Shokri, N. (2021). Global predictions of primary soil salinization under changing climate in the 21st century. *Nat. Commun.* 12, 6663. doi: 10.1038/s41467-021-26907-3

He, Y., Wu, J., Lv, B., Li, J., Gao, Z., Xu, W., et al. (2015). Involvement of 14-3-3 protein GRF9 in root growth and response under polyethylene glycol-induced water stress. *J. Exp. Bot.* 66, 2271–2281. doi: 10.1093/jxb/erv149

Hernandez, M., Fernandez-Garcia, N., Diaz-Vivancos, P., and Olmos, E. (2010). A different role for hydrogen peroxide and the antioxidative system under short and long salt stress in *Brassica oleracea* roots. *J. Exp. Bot.* 61, 521–535. doi: 10.1093/jxb/erp321

Herndon, N., Shelton, J. M., Gerischer, L., Ioannidis, P., Ninova, M., Donitz, J., et al. (2020). Enhanced genome assembly and a new official gene set for tribolium castaneum. *BMC Genomics* 21, 47. doi: 10.1186/s12864-019-6394-6

Hu, D. G., Ma, Q. J., Sun, C. H., Sun, M. H., You, C. X., and Hao, Y. J. (2016). Overexpression of MdSOS2L1, a CIPK protein kinase, increases the antioxidant metabolites to enhance salt tolerance in apple and tomato. *Physiologia Plantarum* 156, 201–214. doi: 10.1111/ppl.12354

Huang, Y., Cao, H., Yang, L., Chen, C., Shabala, L., Xiong, M., et al. (2019). Tissue-specific respiratory burst oxidase homolog-dependent H2O2 signaling to the plasma membrane h^+-ATPase confers potassium uptake and salinity tolerance in cucurbitaceae. *J. Exp. Bot.* 70, 5879–5893. doi: 10.1093/jxb/erz328

Ilik, P., Spundova, M., Sicner, M., Melkovicova, H., Kucerova, Z., Krchnak, P., et al. (2018). Estimating heat tolerance of plants by ion leakage: A new method based on gradual heating. *New Phytol.* 218, 1278–1287. doi: 10.1111/nph.15097

Ito, T., Nakata, M., Fukazawa, J., Ishida, S., and Takahashi, Y. (2014). Scaffold function of $Ca2^+$-dependent protein kinase: Tobacco $Ca2^+$-DEPENDENT PROTEIN KINASE1 transfers 14-3-3 to the substrate REPRESSION OF SHOOT GROWTH after phosphorylation. *Plant Physiol.* 165, 1737–1750. doi: 10.1104/pp.114.236448

Jahn, T., Fuglsang, A. T., Olsson, A., Bruntrup, I. M., Collinge, D. B., Volkmann, D., et al. (1997). The 14-3-3 protein interacts directly with the c-terminal region of the plant plasma membrane h^+-ATPase. *Plant Cell* 9, 1805–1814. doi: 10.1105/tpc.9.10.1805

Jia, T., Wang, J., Chang, W., Fan, X., Sui, X., and Song, F. (2019). Proteomics analysis of *E. angustifolia* seedlings inoculated with arbuscular mycorrhizal fungi under salt stress. *Int. J. Mol. Sci.* 20, 788. doi: 10.3390/ijms20030788

Jiao, X., Li, Y., Zhang, X., Liu, C., Liang, W., Li, C., et al. (2019). Exogenous dopamine application promotes alkali tolerance of apple seedlings. *Plants (Basel)* 8, 580. doi: 10.3390/plants8120580

Jiroutova, P., Kovalikova, Z., Toman, J., Dobrovolna, D., and Andrys, R. (2021). Complex analysis of antioxidant activity, abscisic acid level, and accumulation of osmotica in apple and cherry *in vitro* cultures under osmotic stress. *Int. J. Mol. Sci.* 22, 7922. doi: 10.3390/ijms22157922

Kuo, R. I., Cheng, Y. Y., Zhang, R. X., Brown, J. W. S., Smith, J., Archibald, A. L., et al. (2020). Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics* 21, 751. doi: 10.1186/s12864-020-07123-7

Li, Y. Q., Chai, Y. H., Wang, X. S., Huang, L. Y., Luo, X. M., Qiu, C., et al. (2021). Bacterial community in saline farmland soil on the Tibetan plateau: Responding to salinization while resisting extreme environments. *BMC Microbiol.* 21, 119. doi: 10.1186/s12866-021-02190-6

Li, X., and Dhaubhadel, S. (2011). Soybean 14-3-3 gene family: Identification and molecular characterization. *Planta* 233, 569–582. doi: 10.1007/s00425-010-1315-6

Li, X. Y., Gao, G. L., Li, Y. J., Sun, W. K., He, X. Y., Li, R. H., et al. (2018). Functional roles of two 14-3-3s in response to salt stress in common bean. *Acta Physiol. Plant* 40, 209. doi: 10.1007/s11738-018-2787-4

Li, H., Liu, D., He, H., Zhang, N., Ge, F., and Chen, C. (2013). Overexpression of Pp14-3-3 from pyrus pyrifolia fruit increases drought and salt tolerance in transgenic tobacco plant. *Mol. Biol.* 47, 591–601. doi: 10.1016/j.ieri.2013.11.077

Li, H. L., Liu, D. Q., He, H., Zhang, N. N., Ge, F., and Chen, C. Y. (2014). Molecular cloning of a 14-3-3 protein gene from lilium regale Wilson and overexpression of this gene in tobacco increased resistance to pathogenic fungi. *Scientia Hortic.* 168, 9–16. doi: 10.1016/j.scienta.2013.12.034

Li, X., Lyu, W., Cai, Q., Sha, T., Cai, L., Lyu, X., et al. (2022). General regulatory factor 3 regulates the expression of alternative oxidase 1a and the biosynthesis of glucosinolates in cytoplasmic male sterile brassica juncea. *Plant Sci.* 319, 111244. doi: 10.1016/j.plantsci.2022.111244

Li, W., Zhao, F., Fang, W., Xie, D., Hou, J., Yang, X., et al. (2015). Identification of early salt stress responsive proteins in seedling roots of upland cotton (*Gossypium hirsutum* l.) employing iTRAQ-based proteomic technique. *Front. Plant Sci.* 6, 732. doi: 10.3389/fpls.2015.00732

Liang, Q. Y., Wu, Y. H., Wang, K., Bai, Z. Y., Liu, Q. L., Pan, Y. Z., et al. (2017). Chrysanthemum WRKY gene DgWRKY5 enhances tolerance to salt stress in transgenic chrysanthemum. *Sci. Rep.* 7, 4799. doi: 10.1038/s41598-017-05170-x

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Łukaszewicz, M., Matysiak-Kata, I., Aksamit, A., Oszmiański, J., and Szopa, J. (2002a). 14-3-3 protein regulation of the antioxidant capacity of transgenic potato tubers. *Plant Sci.* 163, 125–130. doi: 10.1016/S0168-9452(02)00081-X

Manosalva, P. M., Bruce, M., and Leach, J. E. (2011). Rice 14-3-3 protein (GF14e) negatively affects cell death and disease resistance. *Plant J.* 68, 777–787. doi: 10.1111/j.1365-313X.2011.04728.x

Mao, K., Dong, Q., Li, C., Liu, C., and Ma, F. (2017). Genome wide identification and characterization of apple bHLH transcription factors and expression analysis in response to drought and salt stress. *Front. Plant Sci.* 8, 480. doi: 10.3389/fpls.2017.00480

Mo, Y., Wang, Y., Yang, R., Zheng, J., Liu, C., Li, H., et al. (2016). Regulation of plant growth, photosynthesis, antioxidation and osmosis by an arbuscular mycorrhizal fungus in watermelon seedlings under well-watered and drought conditions. *Front. Plant Sci.* 7, 644. doi: 10.3389/fpls.2016.00644

Pardo, J. M., Reddy, M. P., Yang, S., Maggio, A., Huh, G. H., Matsumoto, T., et al. (1998). Stress signaling through $Ca2^+$/calmodulin-dependent protein phosphatase calcineurin mediates salt adaptation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 95, 9681–9686. doi: 10.1073/pnas.95.16.9681

Pi, E. X., Zhu, C. M., Fan, W., Huang, Y. Y., Qu, L. Q., Li, Y. Y., et al. (2018). Quantitative phosphoproteomic and metabolomic analyses reveal GmMYB173 optimizes flavonoid metabolism in soybean under salt stress. *Mol. Cell. Proteomics* 17, 1209–1224. doi: 10.1074/mcp.RA117.000417

Ren, Y. R., Yang, Y. Y., Zhang, R., You, C. X., Zhao, Q., and Hao, Y. J. (2019). MdGRF11, an apple 14-3-3 protein, acts as a positive regulator of drought and salt tolerance. *Plant Sci.* 288, 110219. doi: 10.1016/j.plantsci.2019.110219

Roberts, M. R. (2003). 14-3-3 proteins find new partners in plant cell signalling. *Trends Plant Sci.* 8, 218–223. doi: 10.1016/S1360-1385(03)00056-6

Rosenquist, M., Alsterfjord, M., Larsson, C., and Sommarin, M. (2001). Data mining the *Arabidopsis* genome reveals fifteen 14-3-3 genes. expression is demonstrated for two out of five novel genes. *Plant Physiol.* 127, 142–149. doi: 10.1104/pp.127.1.142

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006

Schoonheim, P. J., Veiga, H., Pereira Dda, C., Friso, G., van Wijk, K. J., and de Boer, A. H. (2007). A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. *Plant Physiol.* 143, 670–683. doi: 10.1104/pp.106.090159

Shabala, S. (2013). Learning from halophytes: Physiological basis and strategies to improve abiotic stress tolerance in crops. *Ann. Bot.* 112, 1209–1221. doi: 10.1093/aob/mct205

Su, Q., Zheng, X., Tian, Y., and Wang, C. (2020). Exogenous brassinolide alleviates salt stress in malus hupehensis rehd. by regulating the transcription of NHX-type Na +(K+)/H+ antiporters. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00038

Sun, S. J., Guo, S. Q., Yang, X., Bao, Y. M., Tang, H. J., Sun, H., et al. (2010). Functional analysis of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice. *J. Exp. Bot.* 61, 2807–2818. doi: 10.1093/jxb/erq120

Tian, F., Wang, T., Xie, Y., Zhang, J., and Hu, J. (2015). Genome-wide identification, classification, and expression analysis of 14-3-3 gene family in populus. *PloS One* 10, e0123225. doi: 10.1371/journal.pone.0123225

Tripathi, P., Rabara, R. C., Reese, R. N., Miller, M. A., Rohila, J. S., Subramanian, S., et al. (2016). A toolbox of genes, proteins, metabolites and promoters for improving drought tolerance in soybean includes the metabolite coumestrol and stomatal development genes. *BMC Genomics* 17, 102. doi: 10.1186/s12864-016-2420-0

Visconti, S., D'ambrosio, C., Fiorillo, A., Arena, S., Muzi, C., Zottini, M., et al. (2019). Overexpression of 14-3-3 proteins enhances cold tolerance and increases levels of stress-responsive proteins of *Arabidopsis* plants. *Plant Sci.* 289, 110215. doi: 10.1016/j.plantsci.2019.110215

Wang, P. H., Lee, C. E., Lin, Y. S., Lee, M. H., Chen, P. Y., Chang, H. C., et al. (2019). The glutamate receptor-like protein GLR3.7 interacts with 14-3-3omega and participates in salt stress response in *Arabidopsis* thaliana. *Front. Plant Sci.* 10, 1169. doi: 10.3389/fpls.2019.01169

Wang, Q., Lu, X., Chen, X., Malik, W. A., Wang, D., Zhao, L., et al. (2021). Transcriptome analysis of upland cotton revealed novel pathways to scavenge reactive oxygen species (ROS) responding to Na_2SO_4 tolerance. *Sci. Rep.* 11, 8670. doi: 10.1038/s41598-021-87999-x

Wang, N. N., Shi, Y. Y., Jiang, Q., Li, H., Fan, W. X., Feng, Y., et al. (2023) A 14-3-3 protein positively regulates rice salt tolerance by stabilizing phospholipase C1. *Plant Cell Environment* 46, 1232–1248. doi: 10.1111/pce.14520

Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., et al. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. *Plant Cell* 26, 4782–4801. doi: 10.1105/tpc.114.131607

Xu, W. F., and Shi, W. M. (2006). Expression profiling of the 14-3-3 gene family in response to salt stress and potassium and iron deficiencies in young tomato (Solanum lycopersicum) roots: Analysis by real-time RT-PCR. *Ann. Bot.* 98, 965–974. doi: 10.1093/aob/mcl189

Yan, J., He, C., Wang, J., Mao, Z., Holaday, S. A., Allen, R. D., et al. (2004). Overexpression of the *Arabidopsis* 14-3-3 protein GF14 lambda in cotton leads to a "stay-green" phenotype and improves stress tolerance under moderate drought conditions. *Plant Cell Physiol.* 45, 1007–1014. doi: 10.1093/pcp/pch115

Yang, Z., Wang, C., Xue, Y., Liu, X., Chen, S., Song, C., et al. (2019). Calcium-activated 14-3-3 proteins as a molecular switch in salt stress tolerance. *Nat. Commun.* 10, 1199. doi: 10.1038/s41467-019-09181-2

Yang, L., Yan, C., Peng, S., Chen, L., Guo, J., Lu, Y., et al. (2022). Broad-spectrum resistance mechanism of serine protease Sp1 in bacillus licheniformis W10 *via* dual comparative transcriptome analysis. *Front. Microbiol.* 13, 974473. doi: 10.3389/fmicb.2022.974473

Yao, Y., Du, Y., Jiang, L., and Liu, J. Y. (2007). Molecular analysis and expression patterns of the 14-3-3 gene family from oryza sativa. *J. Biochem. Mol. Biol.* 40 (3), 349–357. doi: 10.5483/BMBRep.2007.40.3.349

Yasuda, S., Sato, T., Maekawa, S., Aoyama, S., Fukao, Y., and Yamaguchi, J. (2014). Phosphorylation of *Arabidopsis* ubiquitin ligase ATL31 is critical for plant carbon/nitrogen nutrient balance response and controls the stability of 14-3-3 proteins. *J. Biol. Chem.* 289, 15179–15193. doi: 10.1074/jbc.M113.533133

Yoshida, T., Mogami, J., and Yamaguchi-Shinozaki, K. (2014). ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr. Opin. Plant Biol.* 21, 133–139. doi: 10.1016/j.pbi.2014.07.009

Zhang, Z. T., Zhou, Y., Li, Y., Shao, S. Q., Li, B. Y., Shi, H. Y., et al. (2010). Interactome analysis of the six cotton 14-3-3s that are preferentially expressed in fibres and involved in cell elongation. *J. Exp. Bot.* 61, 3331–3344. doi: 10.1093/jxb/erq155

Zhao, S., Wang, H., Jia, X., Gao, H., Mao, K., and Ma, F. (2021). The HD-zip I transcription factor MdHB7-like confers tolerance to salinity in transgenic apple (Malus domestica). *Physiol. Plant* 172, 1452–1464. doi: 10.1111/ppl.13330

Zhou, H., Lin, H., Chen, S., Becker, K., Yang, Y., Zhao, J., et al. (2014). Inhibition of the *Arabidopsis* salt overly sensitive pathway by 14-3-3 proteins. *Plant Cell* 26, 1166–1182. doi: 10.1105/tpc.113.117069

Zuo, X., Wang, S., Xiang, W., Yang, H., Tahir, M. M., Zheng, S., et al. (2021). Genome-wide identification of the 14-3-3 gene family and its participation in floral transition by interacting with TFL1/FT in apple. *BMC Genomics* 22, 41. doi: 10.1186/s12864-020-07330-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1155531

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Abdullah,
Shanghai Jiao Tong University, China

REVIEWED BY
Riaz Ahmad,
Dera Ismail Khan, Pakistan
Taimoor Hassan Farooq,
Bangor College China, a joint unit of Bangor University and Central South University of Forestry and Technology, China
Sami Abou Fayssal,
University of Forestry, Bulgaria

*CORRESPONDENCE
Zhiwei Wang
wangzhiwei@hainanu.edu.cn

SPECIALTY SECTION
This article was submitted to Plant Bioinformatics, a section of the journal Frontiers in Plant Science

RECEIVED 31 January 2023
ACCEPTED 20 March 2023
PUBLISHED 14 April 2023

CITATION
Hao Y, Su X, Li W, Li L, Zhang Y, Mumtaz MA, Shu H, Cheng S, Zhu G and Wang Z (2023) The creation of autotetraploid provides insights into critical features of DNA methylome changes after genome doubling in water spinach (*Ipomoea aquatica* Forsk). *Front. Plant Sci.* 14:1155531. doi: 10.3389/fpls.2023.1155531

COPYRIGHT
© 2023 Hao, Su, Li, Li, Zhang, Mumtaz, Shu, Cheng, Zhu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The creation of autotetraploid provides insights into critical features of DNA methylome changes after genome doubling in water spinach (*Ipomoea aquatica* Forsk)

Yuanyuan Hao[1,2,3], Xiao Su[2], Wen Li[2], Lin Li[1,2], Yu Zhang[1,2], Muhammad Ali Mumtaz[1,2], Huangying Shu[1,2,3], Shanhan Cheng[1,2], Guopeng Zhu[1,2] and Zhiwei Wang[1,2,3]*

[1]Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, Sanya Nanfan Research Institute, Hainan University, Sanya, China, [2]Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, School of Horticulture, Hainan University, Haikou, China, [3]Hainan Yazhou Bay Seed Laboratory, Sanya, China

Water spinach (*Ipomoea aquatica* Forsk) is an essential green leafy vegetable in Asia. In this study, we induced autotetraploid water spinach by colchicine. Furthermore, DNA methylation and transcriptome of tetraploid and diploid were compared using Whole Genome Bisulfite Sequencing (WGBS) and RNA-sequencing techniques. Autotetraploid water spinach was created for the first time. Compared with the diploid parent, autotetraploid water spinach had wider leaves, thicker petioles and stems, thicker and shorter adventitious roots, longer stomas, and larger parenchyma cells. The whole genome methylation level of the autotetraploid was slightly higher than that of the diploid. Compared with the diploid, 12281 Differentially Methylated Regions (DMRs)were found in the autotetraploid, including 2356 hypermethylated and 1310 hypomethylated genes, mainly enriched in 'Arginine and Proline metabolism', 'beta – Alanine metabolism', 'Plant homone signal translation', 'Ribome', and 'Plant – pathgen interaction' pathways. Correlation analysis of transcriptome and DNA methylation data showed that 121 differentially expressed genes undergone differential methylation, related to four pathways 'Other types of O-glycan biosynthesis', 'Terpenoid backbone biosynthesis', 'Biosynthesis of secondary metabolites', and 'Metabolic paths'. This work obtained important autotetraploid resources of water spinach and revealed the genomic DNA methylation changes after genome doubling, being helpful for further studying the molecular mechanism of variations caused by polyploids of the *Ipomoea* genus.

KEYWORDS

water spinach, autotetraploid, DNA Methylation, phenotypic variation, differentially methylated region

1 Introduction

Polyploids are organisms that contain more than two complete chromosome sets in the same nucleus, which is frequent in nature and plays an important role (Hegarty et al., 2013; Tayalé and Parisod, 2013; Ren et al., 2018). All angiosperms have experienced at least one whole-genome duplication (WGD) event in their evolutionary history, thus, all plants are paleopolyploids (Ahmad and Anjum, 2018; Zhang et al., 2019). The genome doubling event could be either an autopolyploidy (a condition with more than two genomes of the same species) or an allopolyploidy (a condition in which complete genomes of two or more species combine) (Madlung and Wendel, 2013).

Homologous polyploids have double the number of chromosomes and a similar DNA sequence to their diploid parents. Polyploidy has significant effects on plant morphology, such as increased the general vigor and size of various plant parts or structures. Such characteristics are generally referred to as gigantism (Allario et al., 2011; Dai et al., 2015). However, dwarf phenotype caused by polyploidy has also been reported in some plants (Ma et al., 2016; Wang et al., 2018; Wen et al., 2022). It is speculated that the expression level of DEGs was changed after chromosome doubling, which further led to phenotypic changes. This speculation still lacks more evidence support. In addition, polyploidy potentiates plant tolerance to abiotic and biotic stresses (Tossi et al., 2022), such as increased resistance to cold and drought (del Pozo and Ramirez-Parra, 2014; Syngelaki et al., 2021; Abdolinejad and Shekafandeh, 2022).

Polyploidization induces epigenetic changes, including DNA methylation that can be stably inherited in allotetraploid and reversible during genome separation and merger in allohexaploid (Comai, 2005; Chen, 2007). DNA methylation can occur in multiple genomic regions and dinucleotide contexts, including CpG, CHH, and CHG (H = A, T, or C) contexts (Seymour and Becker, 2017). It is involved in many biological processes, including transcription, replication, DNA repair, gene translocation, and cell differentiation (Vanyushin, 2006; Vanyushin and Ashapkin, 2011; Gupta and Salgotra, 2022). Methylation variation in response to ploidy levels has been widely detected (Xu et al., 2009; Salmon and Ainouche, 2010; Xu et al., 2012; Xiang et al., 2023). Previous studies have focused on allopolyploids due to they are widespread and remain an important speciation mechanism (Qiu et al., 2020; Qin et al., 2021). Changes in DNA methylation (including small RNAs) in allotetraploids may affect gene expression and phenotypic variation (Jiang et al., 2021). Although there is increasing methylation as a direct reaction to autologous polyploidy have rarely been reported. Thus little is known about the physiological and molecular mechanisms underlying autopolyploid adaptations (Zhang et al., 2015; Xiao et al., 2022).

Water spinach is an annual or perennial herbaceous plant of the *Ipomoea* genus with heat stress adaptability, both aquatic and terrestrial. In this study, DNA methylation maps of autotetraploid water spinach and diploid water spinach were constructed, and DNA methylation levels, distribution and gene expression were compared. It was found that the genome-wide methylation level of the autotetraploid water spinach was slightly higher than that of the diploid, DNA methylation was involved in regulating gene expression after genome doubling. The study was designed to evaluate the relationship between DNA methylation and phenotypic differences of autotetraploid.

2 Materials and methods

2.1 Plant material and autotetraploid induction

Water spinach germplasm 'HNUWS003' (2n = 2x = 30) was used as the diploid parent. Polyploidy was induced by two methods: The first method was sterile tissue culture (Touchell et al., 2020). The young stem segments of 'HNUWS003' with 2-3 internodes were disinfected (70% ethanol, 30 s, 20% NaClO, 20 min) and inoculated in MS medium to produce sterile seedlings. After a large number of sterile seedlings are obtained through cutting propagation, the stem segment with a length of 2.0 ~ 3.0 cm and one internode was placed in 0.25% colchicine solution for 24 hours, then washed with sterile water three times, and finally inoculated on MS medium at the growth temperature of 24 ± 2 °C and the light duration of 16 h/d (1.5 ~ 2.5 Klux). The surviving seedlings with 15 cm were subcultured for 5 times (Nassar et al., 2008).

The second method consisted of a liquid culture. About 15 cm long seedlings of 'HNUWS003' with apical bud were cut and soaked in 0.15% colchicine for 18 hours. To ensure the production of adventitious roots on time, stem segments about 2cm out from the cut end cannot be soaked. Some growth points in the leaf axil would sprout new buds, and when they grew into a 15cm seedlings, these seedlings were used for cutting propagation for another 5 times. Liquid culture improved the method of polyploid induction and simplified the induction process. Subsequent analysis was based on autotetraploid water spinach produced by liquid culture.

The seedlings for transcriptome and DNA methylation analysis are propagated by cutting in a growth chamber with a photoperiod of 16 h light (30°C)/8 h dark (28°C). After 20 days of growth, fresh leaves were sampled with three biological replicates, frozen in liquid nitrogen, and stored at -80°C before DNA methylation and RNA-seq analysis.

2.2 Paraffin analysis of leaf and root

Leaves and roots of diploid and tetraploid plants were prepared as paraffin sections. The samples were fixed in FAA fixation solution (5 ml 38% formaldehyde +5 ml glacial acetic acid +90 ml 50% ethanol, 1:1:18 by volume) for 24 hours. Then the samples were dehydrated with gradient alcohol, 75% alcohol for 4 hours, 85% alcohol for 2 hours, 90% alcohol for 2 hours, 95% alcohol for 1 hour, anhydrous ethanol I for 30 min, anhydrous ethanol II for 30 min, alcohol benzene for 5~10 min, xylene II for 5~10 min, 65°C melting paraffin I for 1h, 65 °C melting paraffin II for 1h, 65°C melting paraffin III for 1 hour. The wax-soaked tissue was embedded in the embedding machine, and the section thickness was 4μm. Finally, the paraffin sections were stained with safflower O staining solution

and plant solid green staining solution, then observed under a microscope and photographed (Hewitson et al., 2010). The characteristics were measured using Image J software.

2.3 Root morphology

The root of water spinach was selected and cleaned with sterile water for the subsequent observation of root morphology. An imagery scan screen (Epson Expression 11000XL, Regent Instruments, Canada) was used for root scanning. WinRHIZO 2003a software (Regent Instruments, Canada) was used for root image analysis (Altaf et al., 2020).

2.4 Scanning electron microscope analysis

For SEM preparation: fresh leaves of water spinach were cut into squares of about 1cm^2 and immediately put into the electron microscope fixing solution for 2h. The fixed samples were rinsed with 0.1M PB (pH 7.4) for 3 times, 15 min each. The tissue blocks were transferred to 1% OsO4 and placed at 0.1 M PB (pH 7.4) at room temperature for 1-2 hours. After that, the tissue blocks were washed three times at 0.1M PB (pH 7.4) for 15 min each. The dehydration was performed using a graded series of alcohol-isoamyl acetate concentrations for 15 minutes each. The samples were put into the critical point dryer (Quorum K850) for drying and sputter-coated with gold for 30s by lon sputtering apparatus (Hitachi MC1000). Finally, the images were observed under the scanning electron microscope (Hitachi, SU8100) (Pathan et al., 2008).

2.5 Transmission electron microscope analysis

For TEM preparation, fresh leaves were cut into squares of about 1cm^3, the preparation of slices was finished after fixation, room temperature dehydration, resin penetration, embedding, polymerization, ultra-thin sectioning, and staining. and finally observed and photographed under a transmission microscope (Hitachi, HT7800) (Kuo, 2014).

2.6 Flow cytometry analysis

About 0.5cm² of fresh water spinach roots were placed in the petri dish, and then 400μl nuclear lysate were added. The roots were quickly chopped with a sharp blade to facilitate the extraction of complete nuclei. Then, the sample was filtered (30μm nylon filter) and centrifuged. 1600μl DAPI staining solution was added for 10min. The ploidy identification was performed using Sysmex CyFlow® Ploidy Analyser analyzer at a rate of 0.5-2μl/s (Dpooležel et al., 1989).

2.7 DNA extraction and qualification

Genomic DNA was extracted using a Plant Genomic DNA Purification Kit. DNA purity was checked using NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). DNA concentration was measured using Qubit® DNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA).

2.8 WGBS Library construction, data filtering, and reads mapping

A total amount of 100 ng genomic DNA spiked with 0.5 ng lambda DNA was fragmented by sonication to 200-300 bp with Covaris (S220). These DNA fragments were treated with bisulfite using EZ DNA Methylation-GoldTM Kit (Zymo Research), and the library was constructed by Novogene Corporation (Beijing, China). Subsequently, pair-end sequencing of sample was performed on Illumina platform (Illumina, CA, USA), and finally generated 150bp paired-end reads. The raw reads sequences were filtered by fastp (fastp 0.20.0), and the remaining reads were counted as clean reads. FastQC was used to perform basic statistics on the quality of clean data reads. Bismark software (version 0.16.3) (Miura et al., 2012) was used to perform alignments of bisulfite-treated reads to a reference genome (-X 700 –dovetail).

2.9 Methylation level analysis

Methylated sites were identified with a binomial test using the methylated counts (mC), totals count (mC+umC), and the non-conversion rate (r). Sites with FDR-corrected p-value<0.05 were considered as methylated ones. To calculate the methylation level of the sequence, the sequence was divided into several bins with a bin size of 10 kb. The sum of methylated and unmethylated read counts in each window was calculated. The methylation level (ML) for each window or C site showed the fraction of methylated Cs. The calculation formula was as follow: ML=mC/(mC+umC).

2.10 Differentially methylated analysis

Differentially methylated regions (DMRs) were identified using the DSS software, according to the following principles: The proportion of different loci with P value less than 1e-05 was more than 50% of the region, the number of regional loci was more than 3 and the length was more than 50. If the distance between two DMRs was less than 100bp, the two regions were merged. Based on the distribution of DMRs in the genome, DMR-related genes were classified as DMR genes: gene body (from TSS to TES) overlap DMRs, or DMR promoter-genes: promoter regions (2 kb upstream from TSS) overlap DMRs.

2.11 GO and KEGG enrichment analysis

The gene ontology (GO: http://www.geneontology.org/) analysis was conducted using GOseq software, and KOBAS (3.0) software was used to detect the statistical enrichment of genes in KEGG (KEGG: http://www.genome.jp/kegg/) pathway.

2.12 RNA-seq analysis

The total RNA was extracted from samples, and the mRNA library of each sample was constructed and sequenced in the Illumina platform. Raw data were firstly processed and then clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data. Clean reads were aligned to the reference genome using Hisat2 v2.0.5. FPKM of each gene was calculated based on the length of the gene and the reads count mapped to this gene. Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq2R package (1.20.0). Genes with an adjusted P-value ≤0.05 found by DESeq2 were assigned as differentially expressed.

3 Results

3.1 Phenotypic variation after colchicine treatment

After colchicine treatment, some seedlings had no morphological differences from their parents, while some seedlings showed leaf widening characteristics. The seedlings with broader leaves were selected for propagation and named 'HNUWS003-colchicine treated' for comparison with the diploid parent ('HNUWS003'). Here, 'HNUWS003-colchicine treated' water spinach has bigger, thicker leaves and stronger stems than the diploid parent. Also, the adventitious roots were thick and shorter (Figure 1). Histological observations showed that the leaves of water spinach plants treated with colchicine were thicker, with larger spongy parenchyma cells and palisade parenchyma cells (Figures 2C–F; Table S1), the diameter and hollow area of stem increased (Figures 2G–L; Table S1). In addition, the diameter of the roots increased after colchicine treatment and they had larger meristematic cells (Figures 2A, B; Table S1). TEM revealed colchicine treated water spinach plants has lager cells and larger chloroplasts than diploid parents. Furthermore, SEM showed that the stoma of colchicine treated water spinach was slightly larger than that of diploid (Figure 3; Table S1). These results indicated that colchicine treated water spinach plants showed obvious phenotypic variation compared with its parents. Flow cytometry was used to investigated the ploidy of colchicine-treated water spinach and control diploid water spinach. The result showed that the nuclear DNA content of the control diploids had a main flow cytometry peak at 12800, while the nuclear DNA content of induced water spinach showed that a main peak at channel 26000, so it can be considered as tetraploid (Figure 4).

3.2 Genome-wide DNA methylation

To investigate the mechanism of DNA methylation on genome replication, the WGBS sequencing analysis was performed on autotetraploid water spinach and its diploid parent. A total of 105.05 G raw bases were generated, and 93.34 G clean bases were

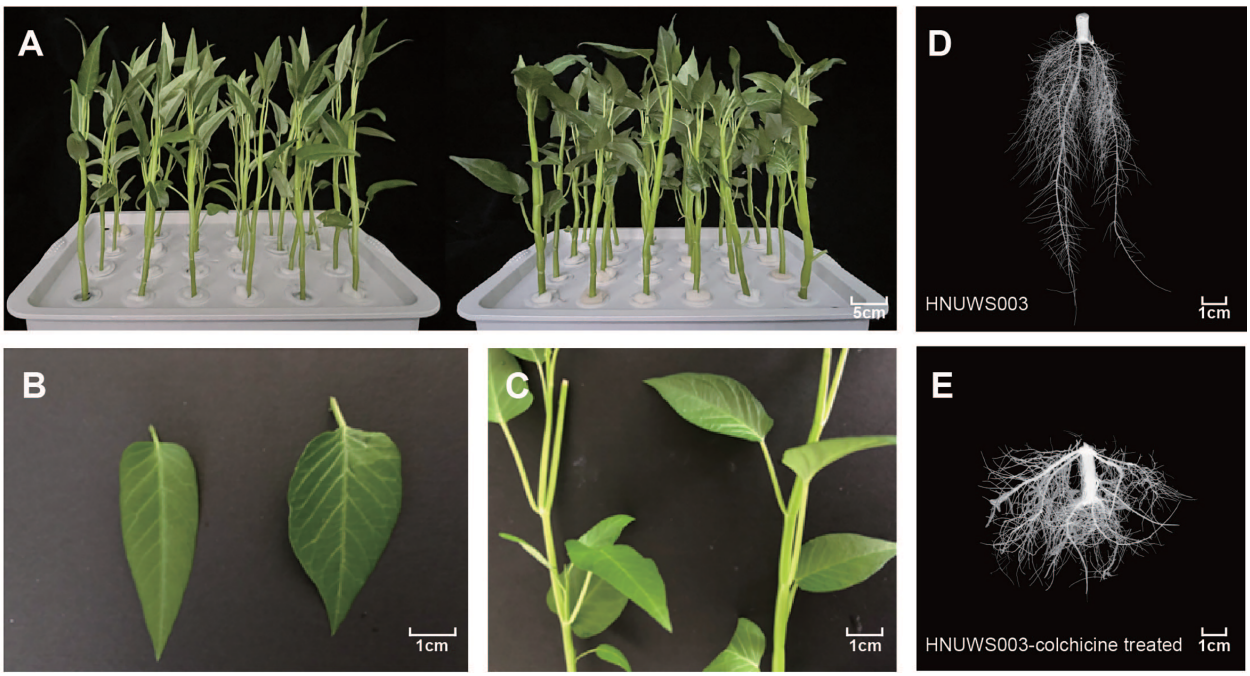


FIGURE 1
Morphological comparison of colchicine treated water spinach and its diploid parent. **(A–C)** Morphological comparison (**A**: plants; **B**: leaf; **C**: stem; left: HNUWS003; right: HNUWS003-colchicine treated). **(D)** 2X water spinach root morphology. **(E)** colchicine treated water spinach root morphology.

10.3389/fpls.2023.1155531

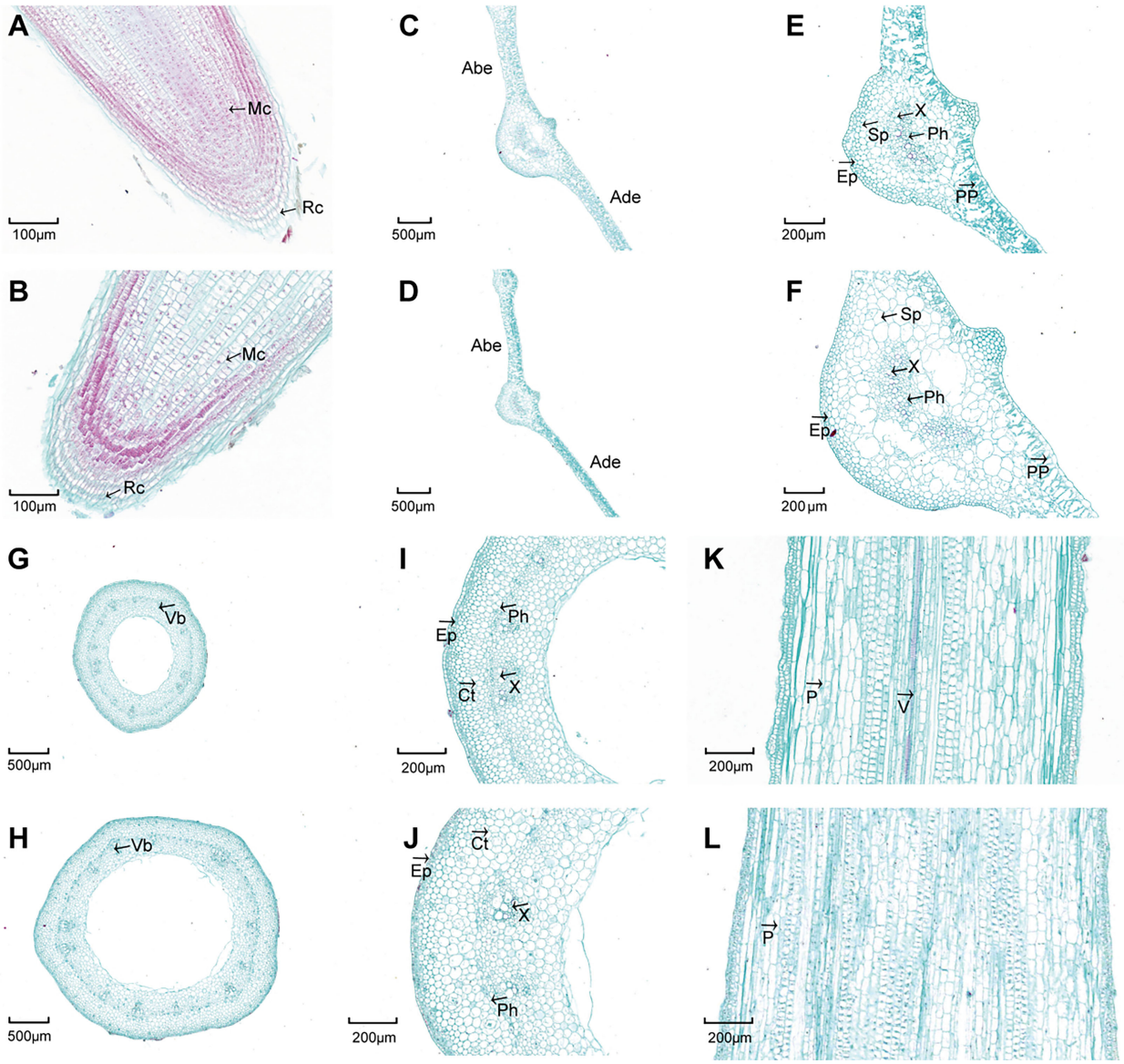


FIGURE 2
Histological observations between diploid and colchicine treated water spinach. **(A)** Longitudinal-section of 2X water spinach root tips. **(B)** Longitudinal-section of colchicine treated water spinach root tips. **(C)** Cross-section of 2X water spinach leaf (×2.0). **(D)** Cross-section of colchicine treated water spinach leaf (×2.0). **(E)** Cross-section of 2X water spinach leaf (×6.0). **(F)** Cross-section of colchicine treated water spinach leaf (×6.0). **(G)** Cross-section of 2X water spinach stem (×2.5). **(H)** Cross-section of colchicine treated water spinach stem (×2.5). **(I)** Cross-section of 2X water spinach stem (×5.0). **(J)** Cross-section of colchicine treated water spinach stem (×5.0). **(K)** Longitudinal-section of 2X water spinach stem. **(L)** Longitudinal-section of colchicine treated water spinach stem. (Rc, Root cap; Mc, Meristome cells; Ade, Adaxial epidermis; Abe, Abaxial epidermis; PP, Palisade parenchyma; Sp, Spongy parenchyma; Ph, Phloem; X, Xylem; Ep, Epidermis; Vb, Vascular bundle; Ct, Cortex; V, Vessel).

obtained after filtering, with Q20 above 96% and Q30 greater than 89%. Genome mapping analysis revealed 66.22%, 62.61%, and 64.76% clean reads from the three diploid biological replicates, and 67.19%, 67.96%, and 66.61% clean reads from the three autotetraploid biological replicates that were mapped to the reference genome (Table S2). The average read depths for diploid and autotetraploid water spinach were 14× and 15.7×, respectively (Table S3). C site coverage statistics showed that most cytosine aligned to the CHH context of the genome (Table S4).

3.3 Methylation site analysis of different ploidies water spinach

The percent of methylated C sites in various contexts was determined through methylation site analysis. The result indicated that the mC (methylated C sites) percent in autotetraploid water spinach was higher than that of diploid. In addition, the percent of methylated C sites was the highest in the CG context with about 40%, while the mCHH was the lowest at

10.3389/fpls.2023.1155531

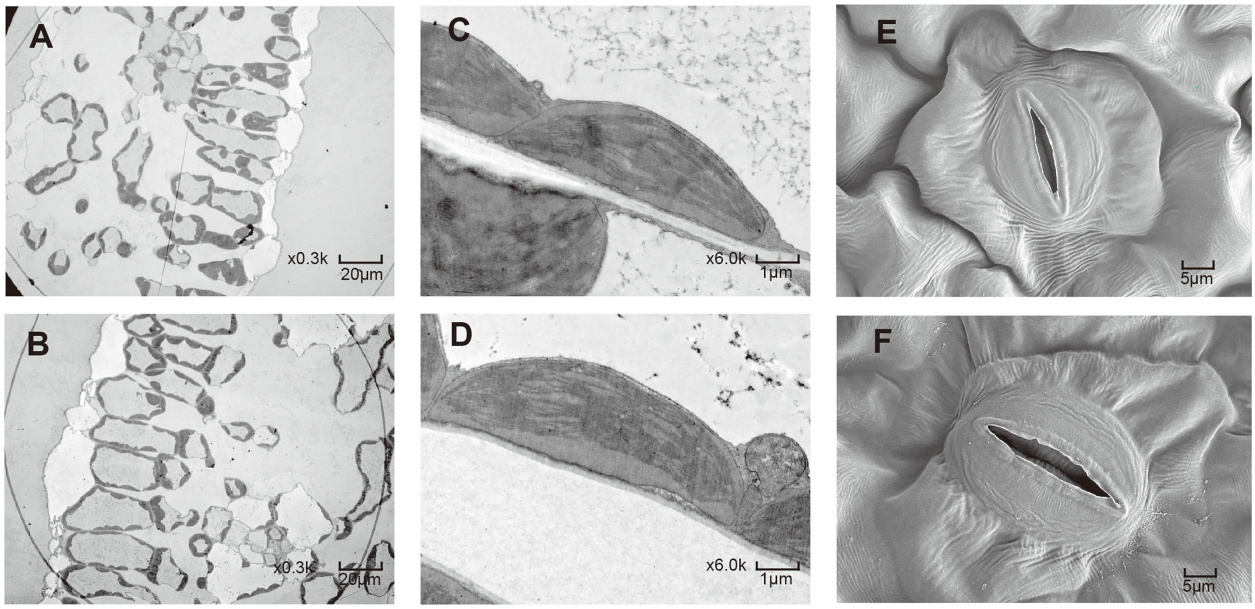


FIGURE 3
Electron microscope observation. **(A)** 2X water spinach cells under TEM. **(B)** colchicine treated water spinach cells under TEM. **(C)** 2X water spinach chloroplasts under TEM. **(D)** colchicine treated water spinach chloroplasts under TEM. **(E)** 2X water spinach stomata under SEM. **(F)** colchicine treated water spinach stomata under SEM.

around 10% (Table 1; Figure 5A). Among these mCs, more than half of them were found in CHH, then CG, and CHG, which may be due to the majority of C sites in the CHH context (Figure 5B).

The methylation level was calculated from the identified methylation sites. mCG and mCHG sites had a significant degree of methylation, with a methylation level of >80%. While mCHH sites had a wide distribution of methylation levels, mostly with slightly methylation (10-30%) (Figure 5C).

3.4 DNA methylation patterns in different ploidies water spinach genomic regions

To characterize the DNA methylation patterns in the functional regions of different ploidies of water spinach, methylation profiles were constructed by counting the average methylation levels of C sites in each context in various genomic functional regions. The methylation levels had a similar trend in CHG and CHH contexts,

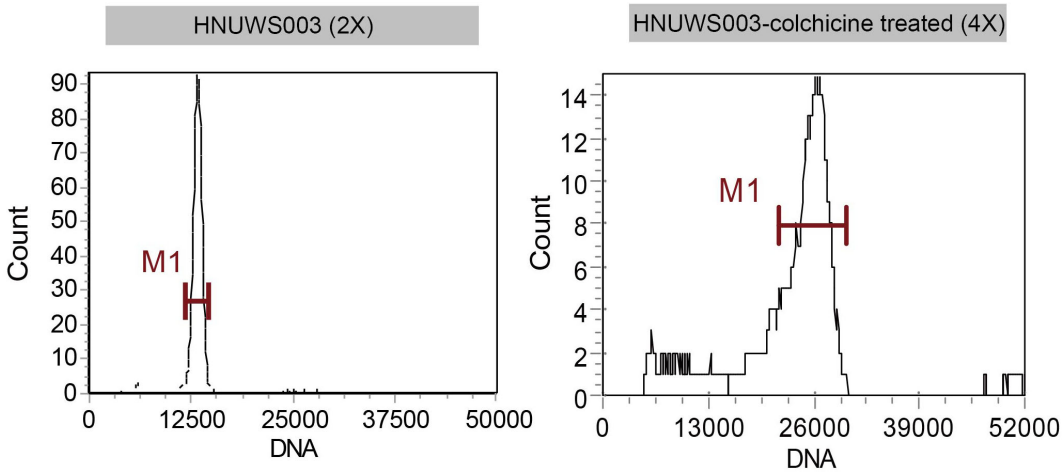


FIGURE 4
Histograms of flow cytometry finding for water spinach.

10.3389/fpls.2023.1155531

TABLE 1 Genome-wide methylation C site ratio.

Samples	mC (%)	mCpG (%)	mCHG (%)	mCHH (%)
2X iaq.-1	15.77%	38.52%	27.39%	10.40%
2X iaq.-2	15.05%	38.52%	27.52%	9.43%
2X iaq.-3	14.86%	39.21%	27.88%	9.01%
4X iaq.-1	18.68%	45.25%	32.89%	12.30%
4X iaq.-2	16.70%	40.03%	28.77%	11.16%
4X iaq.-3	18.15%	45.66%	33.12%	11.50%

with high methylation levels only in the promoter and intron regions. In contrast, exon regions had low methylation levels while intron regions had high methylation levels in CG context (Figure 6A). Further, the methylation levels in the upstream and downstream 2 kb flanking regions of the genes were analyzed. The results showed that there were a peak of CG methylation levels in the genebody regions, as well as two valleys near the transcription start and termination sites. The peak CG methylation in the Genebody was higher than that in the flanking region. In contrast, the methylation levels of CHH and CHG were significantly lower in the Genebody region than in the flanking regions (Figure 6B). The methylation levels were similar in all three contexts in 2X and 4X water spinach, but the methylation levels were higher in 4X water spinach than in 2X.

In our study, the average methylation level of TEs was much higher than that of genes, which was consistent with previous studies Notably, all three methylation contexts showed similar patterns, where TE bodies were highly methylated compared to

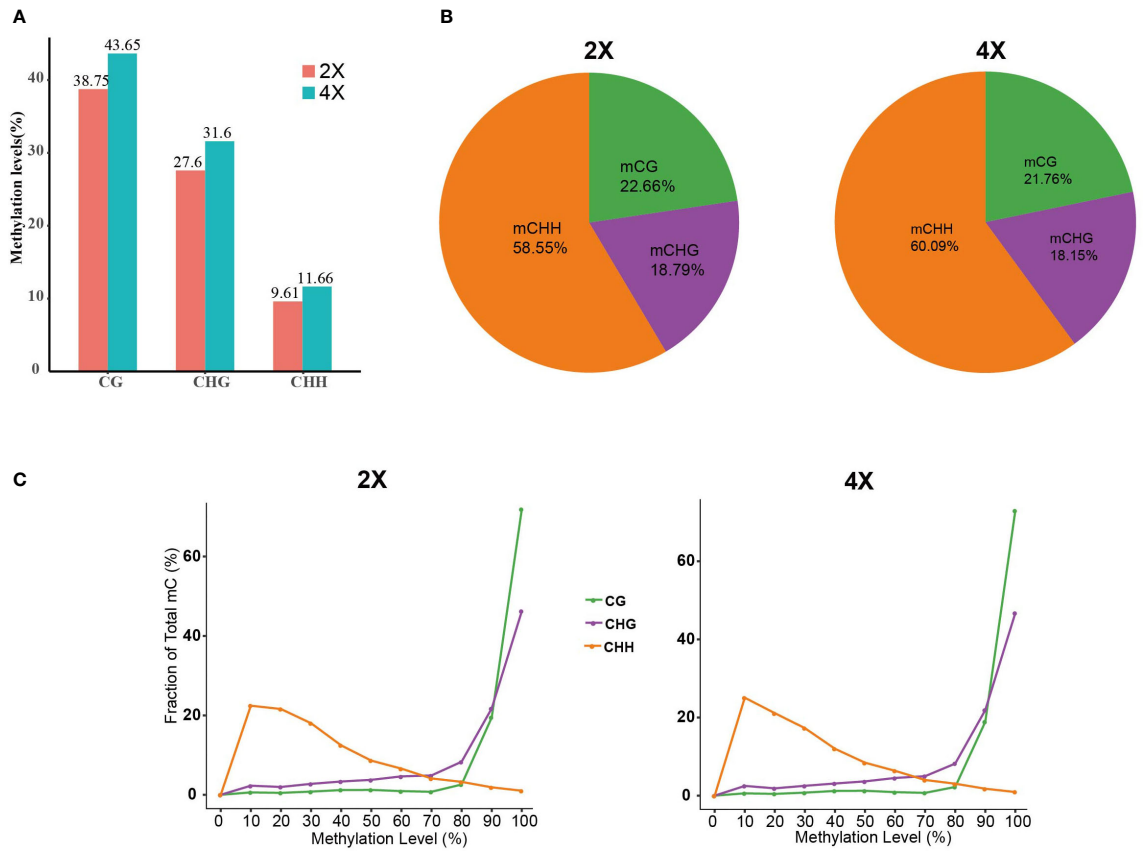


FIGURE 5
The distribution of methylated C sites. **(A)** Methylation levels in different contexts. **(B)** Map of the proportional distribution of methylated C sites. **(C)** Map of methylation site level distribution.

10.3389/fpls.2023.1155531

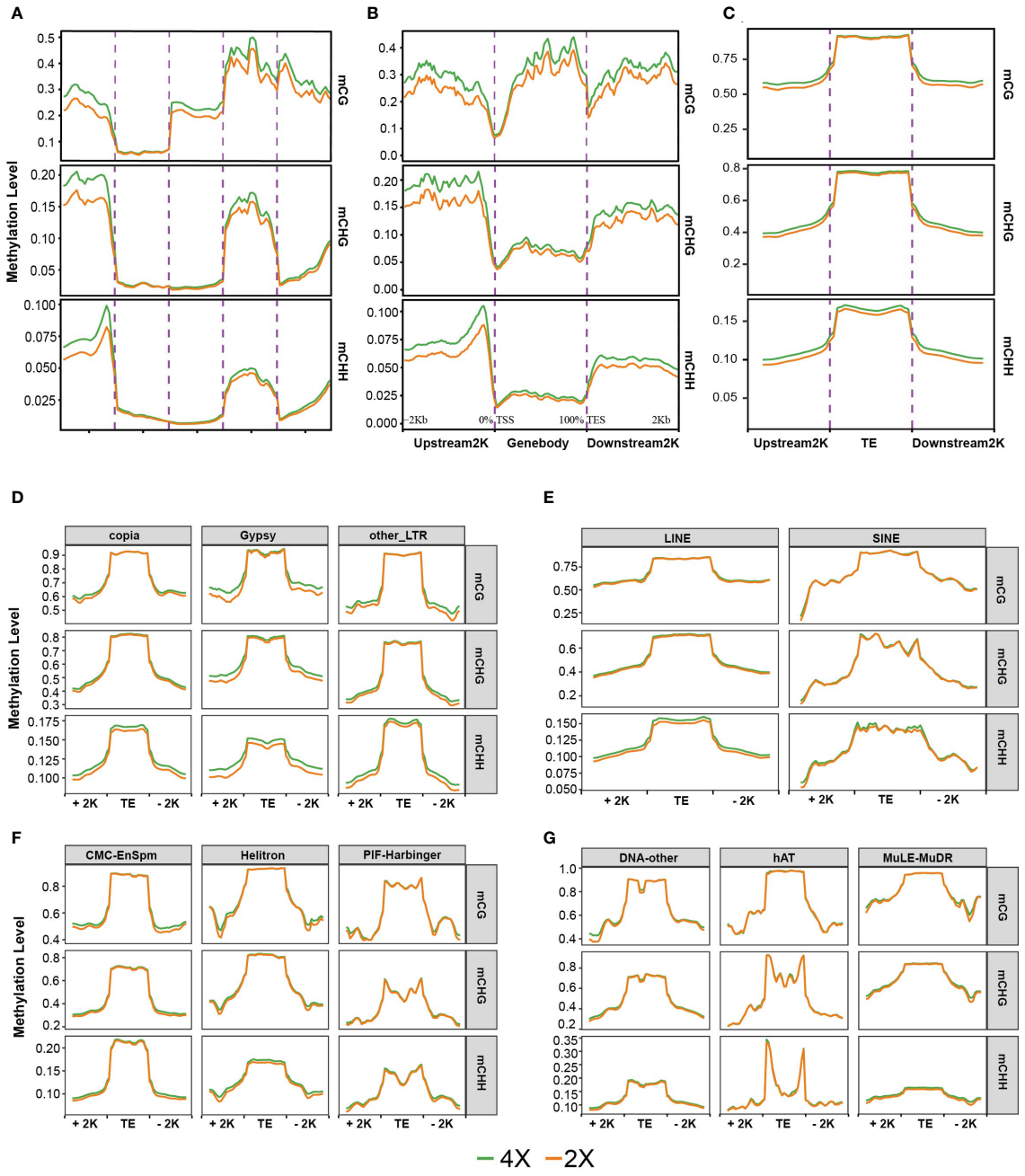


FIGURE 6
DNA methylation profiles. **(A)** Average methylation level in different regions of the genome. **(B)** Average methylation level distribution over gene body and flanking region. **(C)** Average methylation level distribution over TE and flanking region. **(D, E)** Average methylation level distribution over class I TEs. **(F, G)** Average methylation level distribution over class II TEs.

upstream and downstream regions, and 4X water spinach had increased levels of CHG and CHH methylation in TE bodies relative to diploid (Figure 6C).

TEs were divided into two classes: Class I (retrotransposons) and Class II (DNA transposons). Next, methylation levels were analyzed for each methylation context in the 11 major orders. The results displayed that all types of TE have unique methylation profiles, the methylation levels of the body regions were higher than that of the flanking regions in all types of TEs. 4X water spinach exhibited hypermethylated levels in the CHG and CHH contexts in the body regions of Copia, Gypsy, and LINE, respectively. 4X water spinach showed hypermethylation levels in the flanking regions of LTR. Class II DNA transposons showed similar methylation levels in the two genotypes of water spinach. Among them, Helitron was highly methylated in the CHH context of 4X water spinach (Figures 6D–G).

3.5 Correlation of gene expression with DNA methylation and TE insertion

The comparison of the gene-expression level between different ploidies of water spinach from transcriptome data indicated that there were 971 differentially expressed genes, of which 475 were up-regulated and 496 were down-regulated compared to 2X water spinach. The results of GO enrichment analysis showed that the down-regulated genes were mainly enriched in the BP process, 'cellular carbohydrate metabolic process' and 'disaccharide metabolic process' were the most enriched in the BP categories, 'cell wall' and 'external encapsulating structure' were the most enriched in the CC categories. In the MF category, the subcategory with the highest enrichment degree was 'amino acid binding', followed by 'carboxylic acid binding' and 'organic acid binding'. KEGG analysis of transcriptomics showed that DEGs had significantly enriched in 'Fatty acid high-temperature', 'Steroid biosynthesis', 'Circadian rhythm – plant', 'Starch and sucrose metabolism' pathways (Figure S1).

Based on the difference in methylation between 2X and 4X water spinach after genome doubling, the study tried to understand whether gene expression levels were affected by DNA methylation. Therefore, a total of 24,138 genes from RNA-seq data were classified into four quartiles of the none-expressed, low-expressed, medium-expressed, and high-expressed groups according to gene expression levels. In the genebody regions, the highest mCG methylation levels were not detected in the high-expressed genes, but were detected in those with medium-expressed, and none-expressed genes showed the lowest mCG methylation levels in 2X and 4X water spinach. There was a similar correlation between mCHH and mCHG methylation levels and gene expression, with high-expressed genes displaying the lowest methylation levels, and none-expressed genes displayed the highest methylation levels in the genebody regions. The results showed that mCG levels in the genebody region were positively correlated with gene expression levels, while CHG and CHH methylation levels were negatively correlated with gene expression in 2X and 4X water spinach (Figure 7A).

Then, the relationship between TE insertion and gene expression was analyzed. In autotetraploid and diploid water spinach, the expression level of genes inserted with TE was lower than those without TE insertion, but the expression level of genes inserted TE with flanking was higher than those without TE insertion. These results indicated that the gene expression level was affected by TE insertion (Figure 7B).

3.6 Differential methylation regions between 2X and 4X water spinach

To investigate DNA methylation variation in specific regions, the differential methylation regions (DMRs) were analyzed between 2X and 4X water spinach. In total, 554 CG, 1,162 CHG, and 10,565 CHH DMRs were identified with most of the DMRs from the CHH context genome-wide. Then, the DMRs were distinguished between hypermethylated DMR (hyper) and hypomethylated DMR (hypo), and the results showed that the number of hyper-DMRs was higher than that of hypo-DMRs for 4X compared with 2X (Figure 8A). The DMRs were inclined to localize the promoter regions rather than the gene body regions in CHH context, whereas, in the CG and CHG contexts, DMRs were distributed in the promoter, exon, and intron regions (Figure 8B). DMR identification was performed in three contexts separately, and the anchored genes (from TSS to TES), and anchored promoter region-related genes were plotted in the Venn diagram. The results showed that only a few DMR-genes and DMR-promoter genes exist in different contexts simultaneously, and most of them were discovered respectively (Figures 8D, E). In total, 3355 genes (DMR-related genes) were covered by 12281 DMRs, including 1440 DMR genes and 2057 DMR-promoter genes. Venn diagram showed that 142 genes had DMRs in both promoter and genebody regions. (Figure 8C).

To determine which biological processes and metabolic pathways were associated with DMRs, 3355 DMR-related genes were enriched and analyzed. Gene ontology (GO) analysis indicated that they were involved in the 'cellular process', 'regulation of biological process', 'regulation of biological process', 'responses to stimuli', 'metabolic process', and other processes (Figure 9A).

KEGG pathway analyses showed that hypo-DMR-related genes were enriched in 'Arginine and proline metabolism', 'beta–Alanine metabolism', 'RNA degradation', and 'Plant hormone signal transduction' pathways. Hyper-DMR-related genes showed enrichment in 'Ribosome', 'Plant–pathogen interaction', 'Basal transcription factors' and 'Biotin metabolism' pathways. Notably, both hypermethylated and hypomethylated DMR-related genes were enriched in 'Plant–pathogen interaction' and 'Ribosome' pathways (Figures 9B, C).

3.7 Combined analysis of DMR-related genes and DEGs

To assess the relationship between methylation changes and transcription changes, the overlapping genes between DEGs and DMR-related genes were identified. A total of 121 genes overlapped between DEGs and DMR-related genes, among which 43 genes were hypermethylated with downregulated expression levels, and 21 genes were hypomethylated with upregulated expression levels in 4X vs 2X (Table S5). These genes may have negative regulatory mechanisms of methylation levels and gene expression. However, 43 upregulated and 23 downregulated genes were hypermethylated and hypomethylated, respectively (Figure 10A). These findings showed that in most cases, gene expression may not be associated with differences in methylation levels.

Furthermore, 121 overlapping genes were enriched, and GO enrichment showed they were mainly involved in 'cellular process', 'metabolic process', 'binding', and 'catalytic activity' terms. At the same time, these overlapping genes were related to the four pathways of 'Other types of O-glycan biosynthesis', 'Terpenoid backbone biosynthesis', 'Biosynthesis of secondary metabolites', and 'Metabolic pathways' (Figures 10B, C).

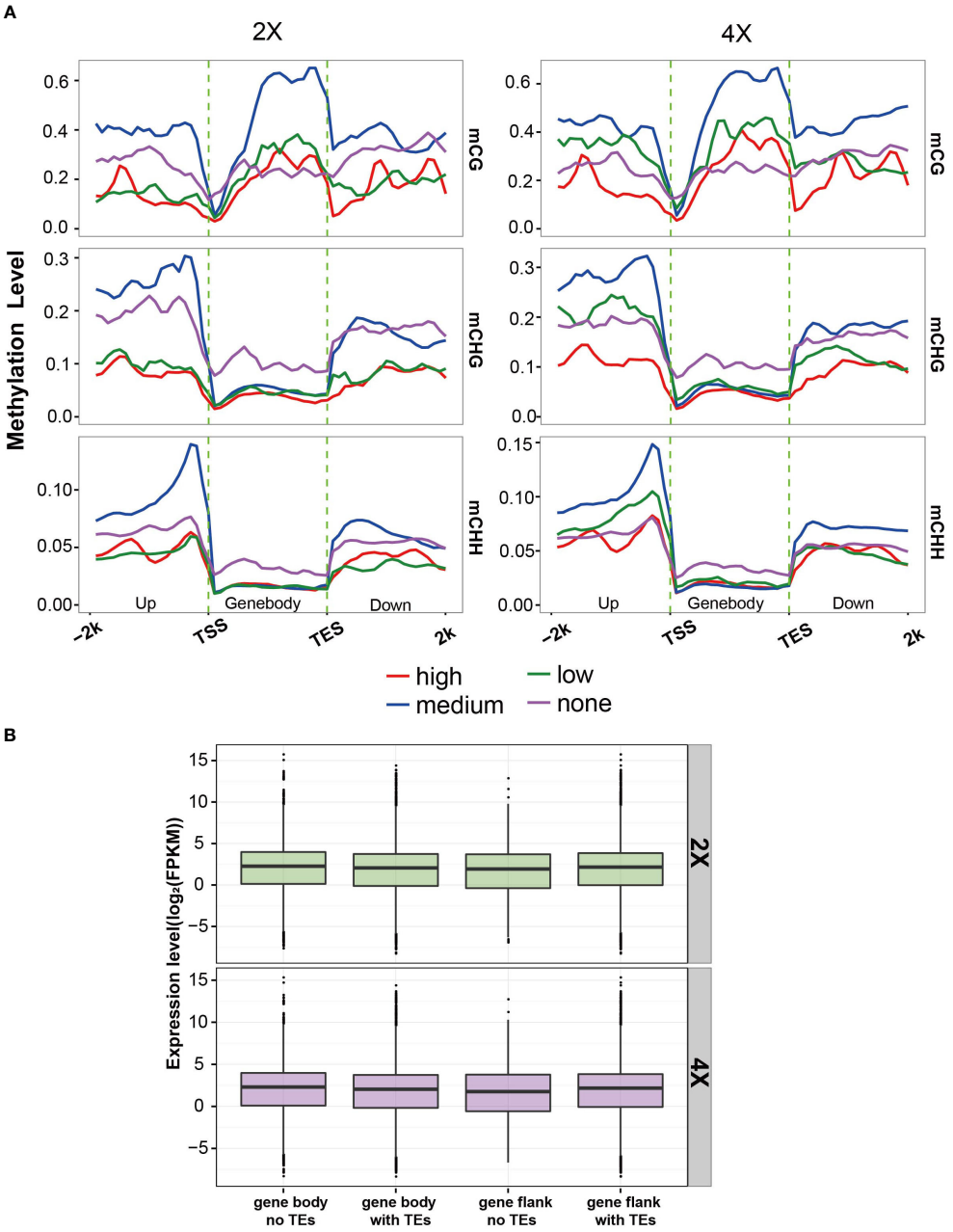


FIGURE 7
Correlation of gene expression with DNA methylation and TE insertion. **(A)** Association between DNA methylation level and gene expression in 2X and 4X water spinach. **(B)** The expression level of genes with or without TE insertion.

4 Discussion

Polyploidy was usually accompanied by morphological variation, polyploids showed more vigor and biomass in general compared to their diploid counterparts (Zhang et al., 2015; Hu et al., 2021). Polyploid plants usually increase their cell size, consequently developing large organs (Ding and Chen, 2018). The results of the present study showed autotetraploid water spinach plants had larger leaves, thicker stems, and larger stomata than the diploid parent, probably due to cell expansion caused by polyploid events as evident in previous studies (Kondorosi et al., 2000; Wang et al., 2019; Catalano et al., 2021; Wen et al., 2022). At the same time, the results of microscopic observation showed that the autotetraploid water spinach had larger spongy tissue, palisade tissue, and xylem cells. In addition, the previous study has shown that polyploidy causes morphological variation and changes flowering time (Yan et al., 2019), which needs further investigation in water spinach.

Genome doubling is usually accompanied by changes in DNA methylation. DNA methylation is an important mechanism in most biological processes, including plant growth and development, fruit

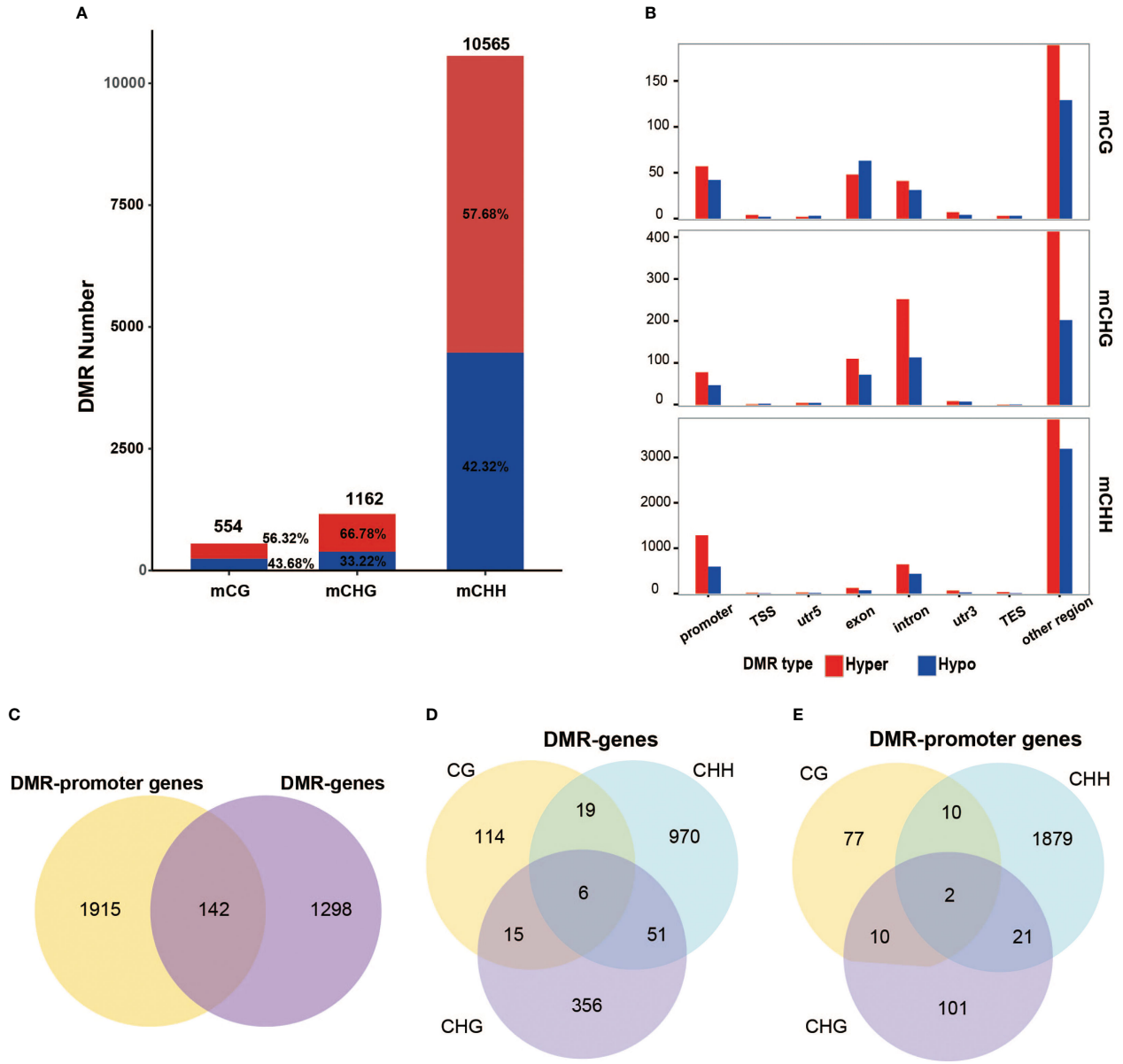


FIGURE 8
Differential methylome analysis for 4X water spinach compared with 2X. **(A)** Number of differentially hypermethylated and hypomethylated regions in all three contexts. **(B)** The distribution of DMRs in the genome. **(C)** Venn diagram of DMR-genes and DMR-promoter genes. **(D)** Venn diagrams of DMR-genes in all three contexts. **(E)** Venn diagrams of DMR-promoter genes in all three contexts.

ripening, and stress response (Huang et al., 2019; Lloyd and Lister, 2022). Previous studies revealed that genome size is positively correlated with the DNA methylation level in CG and CHG contexts but not in the CHH context (Niederhuth et al., 2016; Xu et al., 2018). This study revealed genome-wide changes in DNA methylation by WGBS sequencing technology and confirmed that genome doubling could induce genome-wide changes in DNA methylation, the methylation level of water spinach was moderate among species. In addition, the methylation level of 4X water spinach was higher than that of its diploid parent, which may be due to the expansion of the autotetraploid genome caused by genome duplication. In this study, more than half of the mCs came from CHH context, while more mCs appeared in CG context of grape and *Arabidopsis* (Zhang et al., 2015; Xiang et al., 2023). This indicates that there was a unique DNA methylation pattern in water spinach.

Previous studies have suggested that methylation of the gene body region is exclusive to angiosperms (Bewick and Schmitz, 2017). In this study, the genebody region showed hypermethylation only in CG context, while it showed hypomethylation in CHH and CHG contexts. DNA methylation levels in TE are significantly higher than in the genebody, which was consistent with studies of rice (Zhang et al., 2015), grape (Xiang et al., 2023) and switchgrass (Yan et al., 2019). These results indicated that DNA methylation was closely related to plant biological processes.

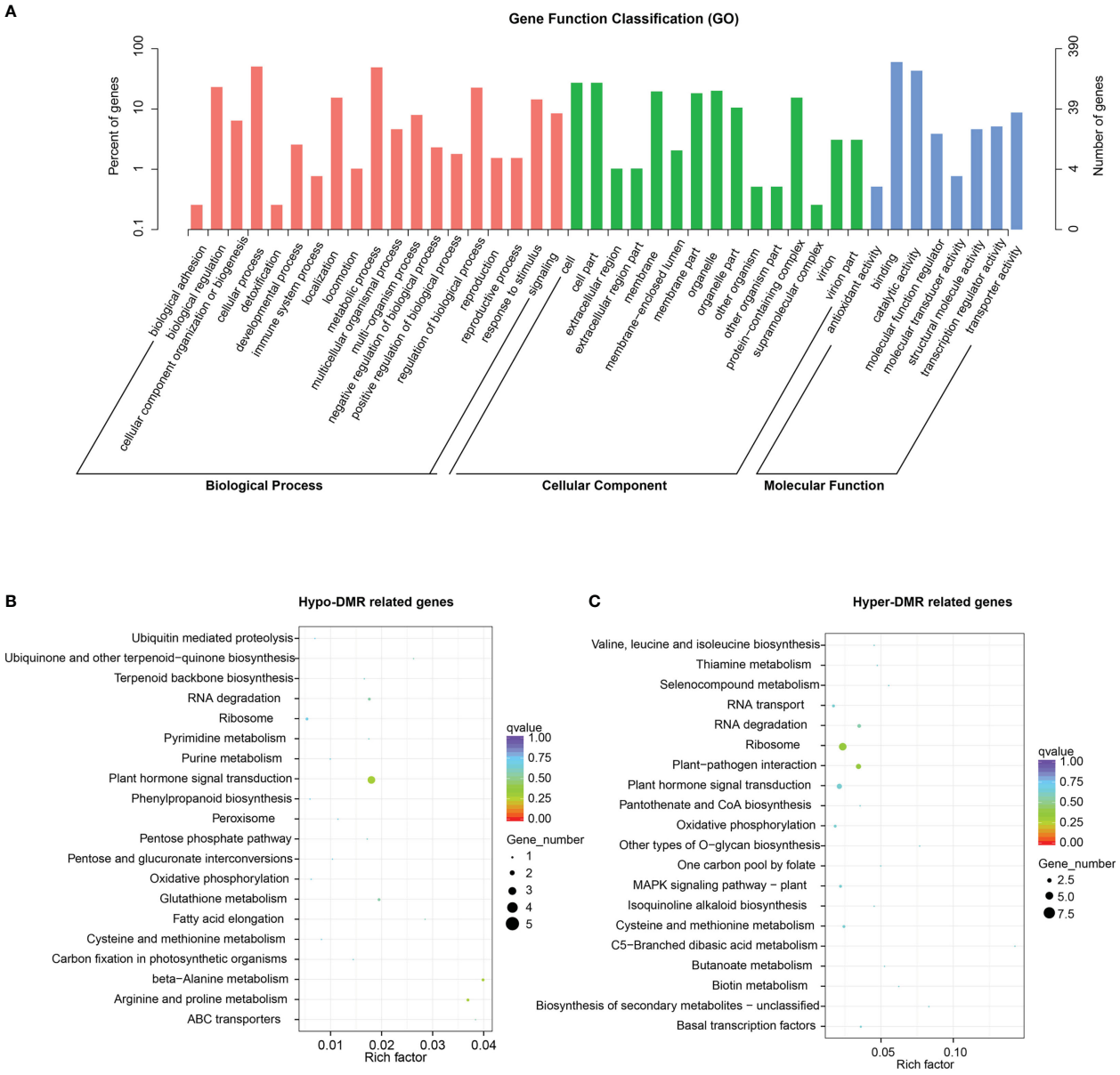


FIGURE 9
Enrichment analysis of DMR-related genes. **(A)** Gene ontology (GO) categories enriched in DMR-related genes. **(B, C)** KEGG enrichment of hypo-DMR-related genes and hyper- DMR-related genes.

DNA methylation not only maintains genomic stability but also helps regulate gene expression. The relationship between transcription and DNA methylation is complex (Seymour and Becker, 2017). Genic methylation is strongly influenced by transcription: moderately transcribed genes are hypermethylated, whereas genes at either extreme are least likely to be methylated (Zilberman et al., 2007). Promoter methylation is usually associated with gene suppression or silencing (Li et al., 2012), but the opposite situation also exists (Ding et al., 2022). Here, in autotetraploid water spinach, the high-expressed genes had low methylation levels, the medium-expressed genes have the highest methylation level in promoter region, low-expressed genes had higher methylation levels than none-expressed genes. These results suggested that methylation levels in the promoter region regulate gene expression after genome doubling. However, DNA methylation studies in many species have shown that genebody methylation level seems to be positively correlated with gene expression in CG context, whereas methylation in non-CG contexts was negatively correlated with gene expression (Hu et al., 2014; Xu et al., 2018; Xiao et al., 2022). Transposable elements (TEs) are mobile genetic elements that are ubiquitous in plant genomes and silenced by epigenetic modifications. They are generally divided into retrotransposons (class I) and DNA transposons (class II) (Feschotte et al., 2002; Underwood et al., 2017). TEs are typically silenced by epigenetic mechanisms such as histone modifications and DNA methylation, with adverse effects on the expression of nearby genes (Bucher et al., 2012; Cui and Cao, 2014; Le et al., 2015). Previous studies have confirmed the complex correlation

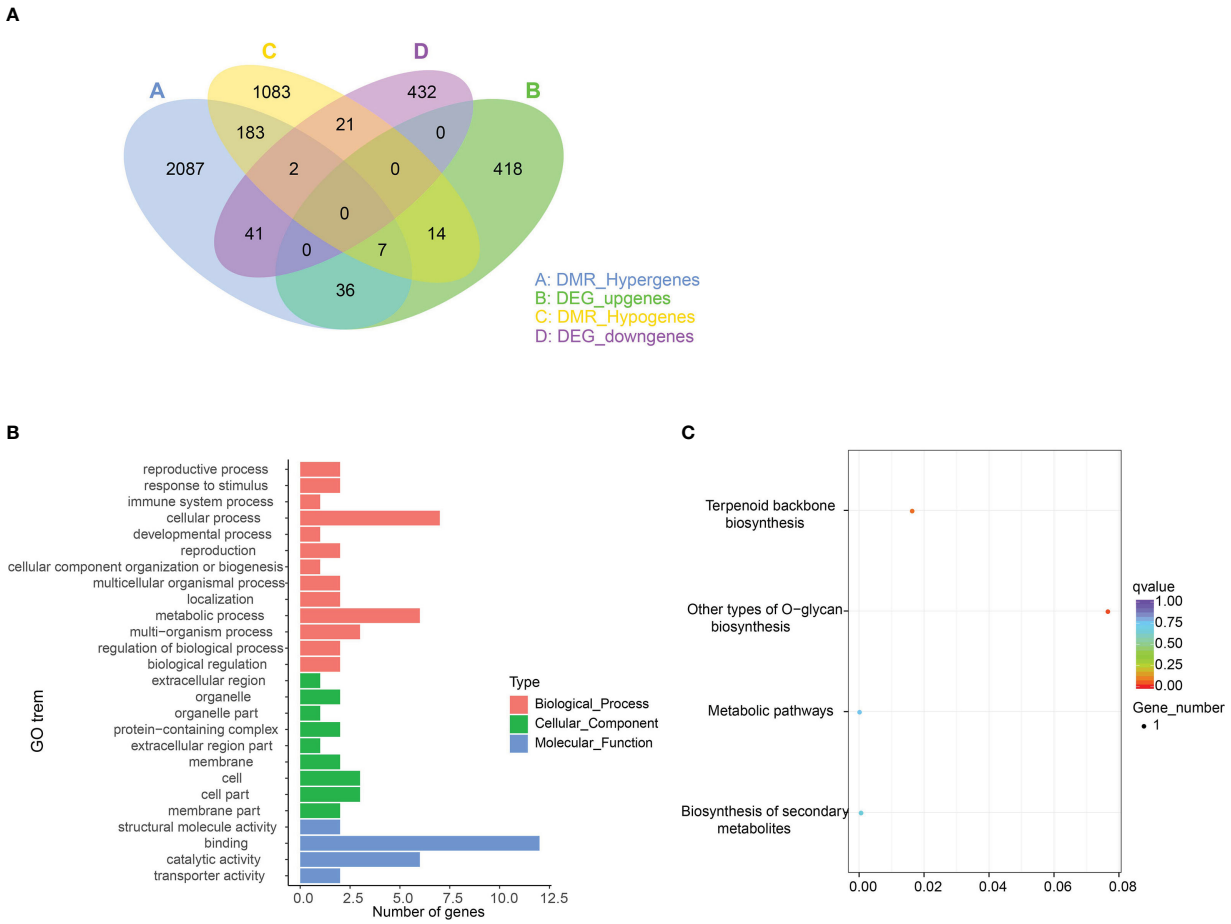


FIGURE 10
Combined analysis of DMR-related genes and DEGs. **(A)** Venn diagram of DMR-related genes and DEGs. **(B)** GO enrichment of 121 overlap genes. **(C)** KEGG enrichment of 121 overlapping genes.

between TE and DNA methylation (Inagaki, 2022; Ramakrishnan et al., 2022). The insertion of TEs within or close proximity to genes can disrupt gene expression, producing negative phenotypic and fitness consequences(Bewick and Schmitz, 2017). In the present study, the number, distance, and methylation status of TE affected neighboring gene expression(Hollister and Gaut, 2009; Hollister et al., 2011). This study revealed that This study showed that the gene expression was influenced by TE insertion or not. TE insertion in the gene body inhibited gene expression, while TE insertion in the flanking region activated gene expression.

In addition, a total of 554 CG, 1162 CHG and 10565 CHH DMRs were identified after genome doubling, among which were more hyper-DMRs. In the present study, more hypo-DMRs were produced after whole-genome double(Yan et al., 2019; Xiang et al., 2023). Moreover, in our study, the DMRs mainly come from CHH context, which is inconsistent with grape, switchgrass, and cassava (Xiao et al., 2022), and consistent with rice(Zhang et al., 2015). These results indicated that different methylation patterns exist in different species after genome-wide doubling. DNA hypermethylation may be one of the important processes regulating phenotypic changes after genome doubling. 12,281 DMRs covered 3,355 genes (DMR-related genes), and the enrichment analysis showed that they were involved in “cellular process”, “biological process regulation”, “biological process regulation”, “stimulus response”, “metabolic process” and other processes. In this study, 121 genes overlapped between DEG and DMR-related genes, which are related to signal transduction and growth and development. This suggests that the majority of DMRs were connected with phenotypic variations rather than having a direct impact on gene expression. The present study confirms the existence of indicative differences between autotetraploid and diploid water spinach, and this variation was closely related to DNA methylation.

5 Conclusions

In this study, autotetraploid water spinach was successfully created by colchicine treatment, and genome doubling caused obvious morphological variation of autotetraploid water spinach. Genome doubling was accompanied by an increase in DNA methylation level. Autopolyploidization affected the expression levels of nearby genes by inducing genome-wide variation in DNA methylation. A total of 12281 DMRs were identified, and 3355 DMR-related genes were covered. These genes are mainly involved in the metabolism and regulation of biological processes in plants. A total of 121 genes

overlapped between DEGs and DMR-related genes, they were believed to be critical genes involved in regulating gene expression through DNA methylation. The present results preliminarily reveal the apparent mechanism of phenotypic variation in autotetraploid water spinach. In the following work, we can further investigate the mechanism of DNA methylation affecting gene expression.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization, ZW and YH. Methodology, ZW, YH, XS, WL, MM. Investigation, YH, HS, LL, YZ, SC, and GZ. Writing—original draft preparation, YH. Supervision, ZW. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Hainan Province (320RC495) and the National Key Research and Development Program of China (2018YFD1000800).

Acknowledgments

Thanks to the reviewers for their suggestions on this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1155531/full#supplementary-material

References

Abdolinejad, R., and Shekafandeh, A. (2022). Tetraploidy confers superior *in vitro* water-stress tolerance to the fig tree (*Ficus carica*) by reinforcing hormonal, physiological, and biochemical defensive systems. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.796215

Ahmad, R., and Anjum, M. A. (2018). Applications of molecular markers to assess genetic diversity in vegetable and ornamental crops – a review. *Hortic. Sci. Technol.* 1, 1–7. doi: 10.46653/jhst180101001

Allario, T., Brumos, J., Colmenero-Flores, J. M., Tadeo, F., Froelicher, Y., Talon, M., et al. (2011). Large Changes in anatomy and physiology between diploid rangpur lime (*Citrus limonia*) and its autotetraploid are not associated with large changes in leaf gene expression. *J. Exp. Bot.* 62 (8), 2507–2519. doi: 10.1093/jxb/erq467

Altaf, M. A., Shahid, R., Ren, M. X., Naz, S., and Hayat, F. (2020). Exogenous melatonin enhances salt stress tolerance in tomato seedlings. *Biol. Plantar* 64, 604–615. doi: 10.32615/bp.2020.090

Bewick, A. J., and Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* 36, 103–110. doi: 10.1016/j.pbi.2016.12.007

Bucher, E., Reinders, J., and Mirouze, M. (2012). Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr. Opin. Plant Biol.* 15 (5), 503–510. doi: 10.1016/j.pbi.2012.08.006

Catalano, C., Abbate, L., Motisi, A., Crucitti, D., Cangelosi, V., Pisciotta, A., et al. (2021). Autotetraploid emergence *via* somatic embryogenesis in vitis vinifera induces marked morphological changes in shoots, mature leaves, and stomata. *Cells* 10 (6), 1336. doi: 10.3390/cells10061336

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58, 377–406. doi: 10.1146/annurev.arplant.58.032806.103835

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6 (11), 836–846. doi: 10.1038/nrg1711

Cui, X., and Cao, X. (2014). Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr. Opin. Plant Biol.* 21, 83–88. doi: 10.1016/j.pbi.2014.07.001

Dai, F., Wang, Z., Luo, G., and Tang, C. (2015). Phenotypic and transcriptomic analyses of autotetraploid and diploid mulberry (*Morus alba* l.). *Int. J. Mol. Sci.* 16 (9), 22938–22956. doi: 10.3390/ijms160922938

del Pozo, J. C., and Ramirez-Parra, E. (2014). Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. *Plant Cell Environ.* 37 (12), 2722–2737. doi: 10.1111/pce.12344

Ding, M., and Chen, Z. J. (2018). Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Curr. Opin. Plant Biol.* 42, 37–48. doi: 10.1016/j.pbi.2018.02.003

Ding, Y. Q., Zou, L. H., Wu, J. J., Ramakrishnan, M., Gao, Y. B., Zhao, L. Z., et al. (2022). The pattern of DNA methylation alteration, and its association with the expression changes of non-coding RNAs and mRNAs in *Moso bamboo* under abiotic stress. *Plant Sci.* 325, 111451. doi: 10.1016/j.plantsci.2022.111451

Dpooležel, J., Binarová, P., and Lcretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plantar* 31 (2), 113–120. doi: 10.1007/BF02907241

Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3 (5), 329–341. doi: 10.1038/nrg793

Gupta, C., and Salgotra, R. K. (2022). Epigenetics and its role in effecting agronomical traits. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.925688

Hegarty, M., Coate, J., Sherman-Broyles, S., Abbott, R., Hiscock, S., and Doyle, J. (2013). Lessons from natural and artificial polyploids in higher plants. *Cytogenet. Genome Res.* 140 (2-4), 204–225. doi: 10.1159/000353361

Hewitson, T. D., Wigg, B., and Becker, G. J. (2010). Tissue preparation for histochemistry: fixation, embedding, and antigen retrieval for light microscopy. *Methods Mol. Biol.* 611, 3–18. doi: 10.1007/978-1-60327-345-9_1

Hollister, J. D., and Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19 (8), 1419–1428. doi: 10.1101/gr.091678.109

Hollister, J. D., Smith, L. M., Guo, Y. L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence

between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. U.S.A.* 108 (6), 2322–2327. doi: 10.1073/pnas.1018222108

Hu, L., Li, N., Xu, C., Zhong, S., Lin, X., Yang, J., et al. (2014). Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc. Natl. Acad. Sci. U. S. A.* 111 (29), 10642–10647. doi: 10.1073/pnas.1410761111

Hu, Y., Sun, D., Hu, H., Zuo, X., Xia, T., and Xie, J. (2021). A comparative study on morphological and fruit quality traits of diploid and polyploid carambola (*Averrhoa carambola* l.) genotypes. *Sci. Hortic-Amsterdam* 277, 109843. doi: 10.1016/j.scienta.2020.109843

Huang, H., Liu, R., Niu, Q., Tang, K., Zhang, B., Zhang, H., et al. (2019). Global increase in DNA methylation during orange fruit development and ripening. *Proc. Natl. Acad. Sci. U. S. A.* 116 (4), 1430–1436. doi: 10.1073/pnas.1815441116

Inagaki, S. (2022). Silencing and anti-silencing mechanisms that shape the epigenome in plants. *Genes Genet. Syst.* 96 (5), 217–228. doi: 10.1266/ggs.21-00041

Jiang, X., Song, Q., Ye, W., and Chen, Z. J. (2021). Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nat. Ecol. Evol.* 5 (10), 1382–1393. doi: 10.1038/s41559-021-01523-y

Kondorosi, E., Roudier, F., and Gendreau, E. (2000). Plant cell-size control: growing by ploidy? *Curr. Opin. Plant Biol.* 3 (6), 488–492. doi: 10.1016/s1369-5266(00)00118-7

Kuo, J. (2014). Processing plant tissues for ultrastructural study. *Methods Mol. Biol.* 1117, 39–55. doi: 10.1007/978-1-62703-776-1_3

Le, T. N., Miyazaki, Y., Takuno, S., and Saze, H. (2015). Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res.* 43 (8), 3911–3921. doi: 10.1093/nar/gkv258

Li, X., Zhu, J. D., Hu, F. Y., Ge, S., Ye, M. Z., Xiang, H., et al. (2012). Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* 13, 300. doi: 10.1186/1471-2164-13-300

Lloyd, J. P. B., and Lister, R. (2022). Epigenome plasticity in plants. *Nat. Rev. Genet.* 23 (1), 55–68. doi: 10.1038/s41576-021-00407-y

Ma, Y., Xue, H., Zhang, L., Zhang, F., Ou, C., Wang, F., et al. (2016). Involvement of auxin and brassinosteroid in dwarfism of autotetraploid apple (*Malus* × *domestica*). *Sci. Rep.* 6, 26719. doi: 10.1038/srep26719

Madlung, A., and Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* 140 (2-4), 270–285. doi: 10.1159/000351430

Miura, F., Enomoto, Y., Dairiki, R., and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* 40 (17), e136. doi: 10.1093/nar/gks454

Nassar, N. M., Graciano-Ribeiro, D., Fernandes, S. D., and Araujo, P. C. (2008). Anatomical alterations due to polyploidy in cassava, *Manihot esculenta* crantz. *Genet. Mol. Res.* 7 (2), 276–283. doi: 10.4238/vol7-2gmr399

Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* 17 (1), 194. doi: 10.1186/s13059-016-1059-0

Pathan, A. K., Bond, J., and Gaskin, R. E. (2008). Sample preparation for scanning electron microscopy of plant surfaces–horses for courses. *Micron* 39 (8), 1049–1061. doi: 10.1016/j.micron.2008.05.006

Qin, J., Mo, R., Li, H., Ni, Z., Sun, Q., and Liu, Z. (2021). The transcriptional and splicing changes caused by hybridization can be globally recovered by genome doubling during allopolyploidization. *Mol. Biol. Evol.* 38 (6), 2513–2519. doi: 10.1093/molbev/msab045

Qiu, T., Liu, Z., and Liu, B. (2020). The effects of hybridization and genome doubling in plant evolution *via* allopolyploidy. *Mol. Biol. Rep.* 47 (7), 5549–5558. doi: 10.1007/s11033-020-05597-y

Ramakrishnan, M., Satish, L., Sharma, A., Kurungara Vinod, K., Emamverdian, A., Zhou, M., et al. (2022). Transposable elements in plants: Recent advancements, tools and prospects. *Plant Mol. Biol. Rep.* 40, 628–645. doi: 10.1007/s11105-022-01342-w

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11 (3), 414–428. doi: 10.1016/j.molp.2018.01.002

Salmon, A., and Ainouche, M. L. (2010). Polyploidy and DNA methylation: new tools available. *Mol. Ecol.* 19 (2), 213–225. doi: 10.1111/j.1365-294X.2009.04461.x

Seymour, D. K., and Becker, C. (2017). The causes and consequences of DNA methylome variation in plants. *Curr. Opin. Plant Biol.* 36, 56–63. doi: 10.1016/j.pbi.2017.01.005

Syngelaki, E., Paetzold, C., and Hörandl, E. (2021). Gene expression profiles suggest a better cold acclimation of polyploids in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Genes (Basel)* 12 (11), 1818. doi: 10.3390/genes12111818

Tayalé, A., and Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140 (2-4), 79–96. doi: 10.1159/000351318

Tossi, V. E., Martínez Tosar, L. J., Laino, L. E., Iannicelli, J., Regalado, J. J., Escandón, A. S., et al. (2022). Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.869423

Touchell, D. H., Palmer, I. E., and Ranney, T. G. (2020). *In vitro* ploidy manipulation for crop improvement. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00722

Underwood, C. J., Henderson, I. R., and Martienssen, R. A. (2017). Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr. Opin. Plant Biol.* 36, 135–141. doi: 10.1016/j.pbi.2017.03.002

Vanyushin, B. F. (2006). DNA Methylation in plants. *Curr. Top. Microbiol. Immunol.* 301, 67–122. doi: 10.1007/3-540-31390-7_4

Vanyushin, B. F., and Ashapkin, V. V. (2011). DNA Methylation in higher plants: past, present and future. *Biochim. Biophys. Acta* 1809 (8), 360–368. doi: 10.1016/j.bbagrm.2011.04.006

Wang, Y., Huang, S., Liu, Z., Tang, X., and Feng, H. (2018). Changes in endogenous phytohormones regulated by microRNA-target mRNAs contribute to the development of dwarf autotetraploid Chinese cabbage (*Brassica rapa* l. ssp. pekinensis). *Mol. Genet. Genomics* 293 (6), 1535–1546. doi: 10.1007/s00438-018-1480-z

Wang, L., Luo, Z., Wang, L., Deng, W., Wei, H., Liu, P., et al. (2019). Morphological, cytological and nutritional changes of autotetraploid compared to its diploid counterpart in Chinese jujube (*Ziziphus jujuba* mill.). *Sci. Hortic-Amsterdam* 249, 263–270. doi: 10.1016/j.scienta.2019.01.063

Wen, Y., Liu, H., Meng, H., Qiao, L., Zhang, G., and Cheng, Z. (2022). *In vitro* induction and phenotypic variations of autotetraploid garlic (*Allium sativum* l.) with dwarfism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.917910

Xiang, X., Gao, Y., Cui, J., Ren, G., Yin, C., and Chang, J. (2023). Methylome and transcriptome analysis of alters leaf phenotype with autotetraploid in grape. *Sci. Hortic-Amsterdam* 307, 111534. doi: 10.1016/j.scienta.2022.111534

Xiao, L., Lu, L., Zeng, W., Shang, X., Cao, S., and Yan, H. (2022). DNA Methylome and LncRNAome analysis provide insights into mechanisms of genome-dosage effects in autotetraploid cassava. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.915056

Xu, Y., Xu, H., Wu, X., Fang, X., and Wang, J. (2012). Genetic changes following hybridization and genome doubling in synthetic *Brassica napus*. *Biochem. Genet.* 50 (7-8), 616–624. doi: 10.1007/s10528-012-9505-5

Xu, Y., Zhong, L., Wu, X., Fang, X., and Wang, J. (2009). Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* 229 (3), 471–483. doi: 10.1007/s00425-008-0844-8

Xu, J., Zhou, S., Gong, X., Song, Y., van Nocker, S., Ma, F., et al. (2018). Single-base methylome analysis reveals dynamic epigenomic differences associated with water deficit in apple. *Plant Biotechnol. J.* 16 (2), 672–687. doi: 10.1111/pbi.12820

Yan, H., Bombarely, A., Xu, B., Wu, B., Frazier, T. P., Zhang, X., et al. (2019). Autopolyploidization in switchgrass alters phenotype and flowering time *via* epigenetic and transcription regulation. *J. Exp. Bot.* 70 (20), 5673–5686. doi: 10.1093/jxb/erz325

Zhang, J., Liu, Y., Xia, E. H., Yao, Q. Y., Liu, X. D., and Gao, L. Z. (2015). Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 112 (50), E7022–E7029. doi: 10.1073/pnas.1515170112

Zhang, K., Wang, X., and Cheng, F. (2019). Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic. Plant J.* 5 (6), 231–239. doi: 10.1016/j.hpj.2019.11.003

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* 39 (1), 61–69. doi: 10.1038/ng1929

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 May 2023
DOI 10.3389/fpls.2023.1169386

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Tao Pan,
Anhui Normal University, China
Muhammad Ashraf,
Center for Excellence in Molecular Plant Sciences (CAS), China
Muhammad Waheed Riaz,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Cunwu Chen
chencunwu@126.com
Yingyu Zhang
zhangyingyu0613@163.com
Cheng Song
lanniao812329218@163.com

†These authors have contributed equally to this work

RECEIVED 19 February 2023
ACCEPTED 17 April 2023
PUBLISHED 10 May 2023

CITATION
He X, Zhang W, Sabir IA, Jiao C, Li G, Wang Y, Zhu F, Dai J, Liu L, Chen C, Zhang Y and Song C (2023) The spatiotemporal profile of *Dendrobium huoshanense* and functional identification of *bHLH* genes under exogenous MeJA using comparative transcriptomics and genomics.
Front. Plant Sci. 14:1169386.
doi: 10.3389/fpls.2023.1169386

COPYRIGHT
© 2023 He, Zhang, Sabir, Jiao, Li, Wang, Zhu, Dai, Liu, Chen, Zhang and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The spatiotemporal profile of *Dendrobium huoshanense* and functional identification of *bHLH* genes under exogenous MeJA using comparative transcriptomics and genomics

Xiaomei He[1†], Wenwu Zhang[2†], Irfan Ali Sabir[3], Chunyan Jiao[4], Guohui Li[1], Yan Wang[1], Fucheng Zhu[1], Jun Dai[1], Longyun Liu[5], Cunwu Chen[1*], Yingyu Zhang[6*] and Cheng Song[1*]

[1]Anhui Engineering Laboratory for Conservation and Sustainable Utilization of Traditional Chinese Medicine Resources, Anhui Engineering Research Center for Eco-agriculture of Traditional Chinese Medicine, College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an, China, [2]School of Life Science, Anhui Agricultural University, Hefei, China, [3]Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [4]College of Life Sciences, Hefei Normal University, Hefei, China, [5]School of Bioengineering, Hefei Technology College, Hefei, China, [6]Henan Key Laboratory of Rare Diseases, Endocrinology and Metabolism Center, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China

Introduction: Alkaloids are one of the main medicinal components of *Dendrobium* species. *Dendrobium* alkaloids are mainly composed of terpene alkaloids. Jasmonic acid (JA) induce the biosynthesis of such alkaloids, mainly by enhancing the expression of JA-responsive genes to increase plant resistance and increase the content of alkaloids. Many JA-responsive genes are the target genes of bHLH transcription factors (TFs), especially the MYC2 transcription factor.

Methods: In this study, the differentially expressed genes involved in the JA signaling pathway were screened out from *Dendrobium huoshanense* using comparative transcriptomics approaches, revealing the critical roles of basic helix-loop-helix (bHLH) family, particularly the MYC2 subfamily.

Results and discussion: Microsynteny-based comparative genomics demonstrated that whole genome duplication (WGD) and segmental duplication events drove *bHLH* genes expansion and functional divergence. Tandem duplication accelerated the generation of *bHLH* paralogs. Multiple sequence alignments showed that all bHLH proteins included bHLH-zip and ACT-like conserved domains. The MYC2 subfamily had a typical bHLH-MYC_N domain. The phylogenetic tree revealed the classification and putative roles of bHLHs. The analysis of *cis*-acting elements revealed that promoter of the majority of *bHLH* genes contain multiple regulatory elements relevant to light response, hormone responses, and abiotic stresses, and the *bHLH* genes could be activated by binding these elements. The expression profiling and *q*RT-PCR results indicated that *bHLH* subgroups IIIe and IIId may have an antagonistic role

in JA-mediated expression of stress-related genes. *DhbHLH20* and *DhbHLH21* were considered to be the positive regulators in the early response of JA signaling, while *DhbHLH24* and *DhbHLH25* might be the negative regulators. Our findings may provide a practical reference for the functional study of *DhbHLH* genes and the regulation of secondary metabolites.

KEYWORDS

***Dendrobium huoshanense*, bHLH transcription factor, JA signaling, abiotic stress, genomics**

Introduction

bHLH TFs play pivotal roles in the JA signaling pathways, which are crucial for the coordination of the plant's adaptive response to abiotic stresses (Goossens et al., 2017; Altmann et al., 2020). *D. huoshanense*, a highly prized traditional Chinese medicine, has always been subjected to a variety of obstacles over the course of artificial domestication. In response, the JA-mediated *bHLH* genes are systematically tuned to regulate the resistance to abiotic stress and the biosynthesis of defensive substance (Major et al., 2017; Song et al., 2022a). bHLH TFs are involved in a variety of physiological processes, such as sexual development, nutrition, and basal metabolism (Feller et al., 2011; Zhang et al., 2020; Liu et al., 2021a). A total of 162, 167, 95, and 152 *bHLH* genes have been reported in *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*, and *Solanum lycopersicum*, respectively. The conserved domains of bHLH TFs commonly contain two functionally distinct motifs, one basic and one helix-loop-helix (HLH) region (Liu et al., 2021b). The HLH domain at the C-terminus is approximately 40 amino acids long and aids in the formation of some homodimeric and heterodimeric protein complexes (Howe et al., 2018). Earlier evolutionary analyses demonstrated that plant bHLH proteins could be subdivided into 26 subfamilies, 20 of which were present in the ancestor of vascular and bryophyte plants (Fan et al., 2021; Hao et al., 2021). In *A. thaliana, AtbHLH42* generally controls the biosynthesis of anthocyanins (Baudry et al., 2004; Song et al., 2021a; Zhou et al., 2020). *AtbHLH122* overexpression was found to have higher salt tolerance and anaerobic stress than its wild relatives (Liu et al., 2014). The expression level of *AtbHLH92* is significantly enhanced in response to NaCl, drought, and cold stresses (Zander et al., 2020; Jiang et al., 2009). *AtbHLH38* and *AtbHLH39* influenced the iron ion metabolism (Fan et al., 2021). *AtbHLH112* is a transcriptional regulator in the stress signaling pathway, despite the inhibitory effect on plant development. (Hao et al., 2021). The homologs *bHLH068* from *O. sativa* and *bHLH112* from *A. thaliana* have antagonistic effects on the mediation of flowering initiation (Chen et al., 2017).

MYC TFs are one of the subfamilies of the bHLH family, including bHLH subgroups IIId, IIIe, and IIIf (Cao et al., 2022). In *A. thaliana*, *bHLH* subgroup IIIe members include *AtMYC2*, *AtMYC3*, *AtMYC4*, and *AtMYC5*. AtMYC2-mediated expression of JA-responsive genes is finely tuned by the JAZ proteins and the coactivator mediator complex subunit MED25 (An et al., 2017; Liu et al., 2019). MYC2 could interact with the majority of JAZ proteins, depending on the JID domain (Fernández-Calvo et al., 2011; Du et al., 2017). The bHLH subclade IIIe was associated positively with the systemic JA response (Qi et al., 2015a). In contrast, bHLH subclade IIId has a negative function in numerous JA-mediated responses (Song et al., 2013). Several studies have demonstrated that the bHLH subclade IIIe is essential for JA signaling. MYC2 orchestrated a precise module of COI1/JAZ/MYC2 to participate in the JA signaling (Peñuelas et al., 2019). MYC2, MYC3, and MYC4 homologs were reported to form some homodimers or heterodimers and bind to the G-box of targeted genes, despite having some redundancy with MYC2 (Hao et al., 2021). MYC2, MYC3, and MYC4 activated JA-induced leaf senescence by binding the promoter of *SAG29* gene (Zheng et al., 2017). However, the bHLH subgroup IIIb binded to its promoter to inhibit the expression and slowed down the senescence process (Qi et al., 2015b). MYC5 modulated JA-mediated herbivore defenses and affected JA-mediated pathogen resistance in plants (Song et al., 2017).

D. huoshanense is a semi-shaded perennial orchid plant, as well as a rare and endangered traditional Chinese medicine. When *Dendrobium* plant is exposed to diverse harsh conditions, high temperature, drought, and UV-B have impacts on its yield and accumulation of medicinal substances (Song et al., 2020; Pan et al., 2023). In normal growth conditions, JA is at a low level; when subjected to external stress, the JA content of the seedlings rapidly increased, which activates JA-responsive genes to increase resistance to environmental stimuli (Zhu et al., 2017). JA also induced the expression of key enzymes in the terpenoid biosynthesis pathway and raised terpene alkaloid content in a short period of time (Song et al., 2022b). The *bHLH* genes have been widely investigated in both model plants and medicinal plants. Until now, little is known about the molecular control mechanism of the bHLH TFs in *D. huoshanense*, particularly the MYC2 in response to JA signal transduction. The response of *D. huoshanense* to stress involves intricate physiological and molecular regulatory processes. Discovering the regulatory factors associated with stress resistance is critical for understanding Dendrobium's resistance mechanism and guiding the genetic improvement of stress resistance (Wu et al., 2019). In this study, JA-induced comparative transcriptome sequencing was used to identify the crucial genes that respond to JA signaling in different spatio-temporal trajectories. It revealed the core and unique regulatory role of *bHLH* family genes, particularly the

MYC2 and JAM subfamilies, in the JA regulatory network. A total of 83 *bHLHs* were identified from the *D. huoshanense* genome. Successive analyses, including the gene mapping, the gene structure, the gene motifs, the phylogenetic tree construction, the conserved domain analysis, the *cis*-acting element analysis, the microcollinearity analysis, *etc.*, were further performed. Our study may provide a scientific reference for the functional research of *bHLH* under abiotic stress in *Dendrobium* species.

Materials and methods

Plant material and growing conditions

The cultivated seedlings were collected from the Anhui Plant Cell Engineering Center of West Anhui University. The coordinates for sampling were 31° 45′ 58′′ N and 116° 29′ 09′′ E. *D. huoshanense* has a growth temperature of 25 ± 2°C, a humidity of 60–70%, a light culture of 16 hours, and a dark culture of 8 hours. Plants with stable growth rates and growth years were chosen. The leaves were sprayed with water and a 50 mmol/L MeJA solution, respectively. Time intervals are 0.25 hours, 0.5 hours, 1 hour, 2 hours, 4 hours, 8 hours, and 16 hours. The control and MeJA-treated leaves at different time points were collected and added into liquid nitrogen for quick freezing, and then put into a -80°C refrigerator for later use.

cDNA library construction and data quality control

Total RNA was extracted with Trizol method, and the steps were referred to in the previous experimental protocol (Song et al., 2021b). Firstly, the total RNA is treated with mRNA enrichment, and those with the polyA tail is enriched with OligodT magnetic beads. The DNA probe is digested with DNaseI, and the desired RNA is obtained after purification. PCR is performed with specific primers from a cDNA library that has been amplified. The raw image data obtained by sequencing is converted into raw reads through base calling, and low-quality, adapter contamination, and reads with a high content of unknown base N are removed. The filtered data is called "clean reads." Filter with the filtering software SOAPnuke (v1.4.0). After obtaining the clean reads, HISAT (v2.1.0) was applied to align the clean reads to the reference genome sequence (Kim et al., 2015). Bowtie2 (v2.2.5) was used to map the clean reads to the reference genome, and then gene and transcript expression value was calculated based on RSEM (Li and Dewey, 2011; Langmead and Salzberg, 2012). The overall analysis of transcriptome sequencing was performed, including coverage and distribution of transcripts (Bjornson et al., 2021). The overall analysis of this transcriptome sequencing was carried out, including statistics such as sequencing saturation. All the original transcript data that was removed from redundancy was uploaded to the GSA database of the National Genomics Data Center (https://ngdc.cncb.ac.cn/gsa/), and the released accession was CRA006607.

Gene structure and variation analysis

Transcript reconstruction was performed on each sample using StringTie (v1.0.4) (Pertea et al., 2015). The Cuffmerge was used to integrate the reconstruction information of all samples. In order to identify novel transcripts, the integrated transcripts were compared to reference annotation data using Cuffcompare (Trapnell et al., 2012). The CPC (v0.9-r2) was used to predict the protein-coding new transcripts, and finally the new transcripts were combined to the reference gene sequence to obtain complete sequence (Kang et al., 2017). The rMATS (v3.2.5) was employed to detect differentially spliced genes among different samples and splicing events of different groups (Shen et al., 2014).Benjamini algorithm was applied to correct the p value to calculate the *FDR* value (Benjamini and Hochberg, 1995).

Gene annotation and quantitative analysis

Firstly, the GO, NR, and KEGG orthology databases were used for functional annotation of all genes. The getorf software was used to detect the ORF of unigene (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). Then, hmmsearch command was used to align the ORF to the transcription factor conserved domain to characterize the transcription factor family according to the PlantTFDB database (http://plntfdb.bio.uni-potsdam.de/v3.0/). The DIAMOND (v0.8.31) software was used to retrieve the plant resistance gene database PRGdb (http://prgdb.crg.eu/) (Sanseverino et al., 2009; Buchfink et al., 2014). The running parameters for DIAMOND: –evalue 1e-5 –outfmt 6 –max -target-seqs 1 –more-sensitive. Screening parameters: query coverage >= 50%, identity >= 40%. Gene expression levels of the individual samples were calculated using RSEM (v1.2.8) software (Li and Dewey, 2011). Using the cor function, the Pearson correlation between each two samples was calculated to get the correlation matrix. PCA analysis was performed using the princomp function, and the graphics were drawn using the ggplot2 package. The expression level distribution map was made based on the FPKM value of each sample. The number of co-expressed and differentially expressed genes among different samples was obtained through the Venn diagram of the expression level between groups. Time series analysis was used to figure out how genes were expressed at different points in time, and Mfuzz (v2.34.0) was used to cluster genes with similar expression patterns together (Kumar & Futschik, 2007).

Analysis of differentially expressed genes

The DEG detection between each of the two groups was undertaken using the DEseq2 method (Love et al., 2014). Hierarchical clustering analysis was performed using the pheatmap function. According to the GO annotation results and the classic classification, the differential genes by function were classified using the phyper function for enrichment analysis. Based on the KEGG annotation results and the classic classification, the

DEGs were converged into biological pathways. The phyper function in the R software was used to perform enrichment analysis, calculate the p-value, and then perform *FDR* correction on the *p*-value. Parameters with *Q*-values greater than 0.05 were considered significantly enriched.

Identification of the *bHLH* gene in *D. huoshanense*

The Hidden Markov Model of the HLH domain (PF00010) obtained from the Pfam database was used to identify bHLHs in *D. huoshanense* genome under the accession PRJNA597621 (Han et al., 2020). The Pfam database was then used to see if all candidates had the HLH domain. Compared with the databases with an e-value of 0.001, the HMM profile was used as a query to identify all HLH domains in *D. huoshanense*. Pfam (http://pfam.xfam.org/search/), SMART (http://smart.embl-heidelberg.de/), and InterPro (http://www.ebi.ac.uk/interpro/search/sequence/) were used to confirm all candidate DhbHLHs. The EnsemblPlants (http://plants.ensembl.org/index.html) database was used to obtain the genome sequence files of *A. thaliana* and *O. sativa*. The genome sequence and coding sequence file of *D. nobile* and *D. chrysotoxum* were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/genome). The physical characteristics of bHLHs, including the protein sequence length, gene ID, and chromosome location, were obtained from the *D. huoshanense* genome. The ExPASy Bioinformatics Resource Portal (https://web.expasy.org) was used to predict the molecular weight and isoelectric point of each WRKY protein (Gasteiger et al., 2003). All putative redundant sequences were discarded based on the sequence alignments.

Identification of the conserved domain and synteny analysis of *bHLH* genes

The bHLH proteins of *A. thaliana, O. sativa, D. nobile*, and *D. chrysotoxum* were compared using multiple alignment. The WebLogo (http://weblogo.berkeley.edu) service was used to investigate the conserved domains. The GeneDoc (https://www.psc.edu/biomed/genedoc) software was used to color the conserved amino acids based on protein homology. The MCScanX (https://github.com/wyp1125/MCScanX) software was used to ascertain the microsynteny between each two species, including *A. thaliana, O. sativa, D. nobile*, and *D. chrysotoxum*, based on the genome files (Wang et al., 2012).

The phylogenetic analysis of *bHLH* genes

At first, multiple sequence alignments of the bHLH proteins were performed using the ClustalX 2.11 with the default parameters (Gao et al., 2017; Liu et al., 2017; Wang et al., 2017). Then, the phylogenetic tree of bHLHs in *A. thaliana* and *D. huoshanense* was constructed using maximum-likelihood methods. The candidate protein sequence was aligned using MEGA(v 6.0). (Tamura et al., 2013). The phylogenetic tree was then created using the IQ-TREE program. The best-fit model was optimized as JTT+F+R7 to produce a more trustworthy WRKY classification. The running parameter was "iqtree.exe -s./bidui.fas -m JTT+F+R7 -bb 1000-alrt 1000-nt AUTO".

The chromosome location, conserved motif and gene structure analysis of *DhbHLHs*

The chromosomal location of *DhbHLH* genes was visualized using the TBtools software (Chen et al., 2020). The gene density was retrieved from the gff file. The chromosome mapping was built using TBtools. The motif pattern and gene structure were also visualized using TBtools. The conserved motifs were identified by searching the candidate proteins with the MEME suite tool (https://meme-suite.org/). The number of motifs was limited to 20, while the width of the motifs ranged from 6 to 200. In terms of gene structure analysis, the gff files were used to visualize the UTR, exon, and intron regions.

The *cis*-acting element analysis of *DhbHLHs*

To discover the *cis*-acting elements in the promoter, the 2000 bp upstream sequence of the promoter was extracted from the *D. huoshanense* genome (Carvalho et al., 2015; Song et al., 2021b). Initially, the promoters of all genes were extracted using the GFF/GTF sequence extractor tool from TBtools software. The *cis*-acting elements were then extracted using the rapid fasta extractor and filter tools. The PlantCARE service (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was performed to obtain these responsive elements.

Expression profile analysis of *DhbHLHs* under different MeJA treatments

The differential expressed profiling analysis was performed under the control group and seven MeJA treatment groups to determine which *bHLH* family members are strongly up- and down regulated by JA signaling. The fragments per kilobase of exon model per million mapped fragments (FPKM) value was used to calculate the TPM value. Following that, the Salmon tool was used to analysis expression levels based on our RNA-seq datasets (Patro et al., 2017). The expression level for replicates in the same group is the average of all samples. The FPKM of all unigenes was utilized to create an expression profile of differentially expressed genes (Wang et al., 2019). The normalized FPKM ($\log_2$ (FPKM)) of all unigenes was used for the hierarchical clustering and expression analyses of *DhbHLH* genes (Song et al., 2022c).

The *q*RT-PCR analysis of the main *DhbHLHs* genes

Total RNA was extracted from *Dendrobium* leaves of the blank and MeJA treatment groups using the RNAprep Pure Plant Kit (Takara, Japan) (CK and T1–T7). The RNA content and purity were determined using an ultra-micro spectrophotometer. The quantitative analysis was performed using the 7500 series real-time fluorescence quantitative PCR (Bio-RAD, America). Twelve candidates were obtained from the genome. The PrimeScript RT reagent kit was used to transform the total RNA from the above samples into cDNA (Takara, Japan). For *q*RT-PCR test, a 20 μl reaction system was used: 10 μl SYBR Premix Ex Taq II, 2 μl cDNA, 0.8 μl *DhbHLH*-RT-F and *DhbHLH*-RT-R, 0.4 μl ROX Reference Dye II, and 6 μl dH_2O. The PCR reaction program refers to the previous method (Song et al., 2021a). The $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression with *actin* gene as the reference (Livak and Schmittgen, 2001).

Results

The reference genome comparison, novel gene identification, and gene structure analysis

In this study, the BGISEQ-500 platform was used for next-generation transcriptome sequencing, and a total of 8 groups of 24 samples were used for the high-throughput sequencing. On the basis of the published *D. huoshanense* genome, the average alignment rate of the samples with the genome was 92.29%, while the average alignment rate of the gene set was 71.66% (Tables S1, S2). Through a comprehensive quality evaluation of the sequencing data and transcript coverage, the results showed that the reads were evenly distributed in various regions of the transcript, the relative position of genes ranged from 0.2 to 0.8 (Figure 1A). According to the genome comparison results, the transcript coverage per sample were counted. The results showed that more than 50% of transcripts had read coverage above 90%, while 20% of transcripts had read coverage between 80% and 90% (Figure 1B). By calculating the saturation curve of the detected genes, the sample had enough sequencing data (Figure 1C). The length of most transcripts ranges from 500 to 2000 nucleotides, but when the sequence length exceeds 3000 nucleotides, the number of transcripts increased dramatically (Figure 1D). A total of 8,969 new genes were predicted; a total of 27,264 genes were identified, of which 18,418 were known genes and 8,846 new genes; a total of 41,069 new transcripts were identified, of which 17,374 belonged to novel alternative splicing isoforms of protein-coding genes, with 8,969 transcripts belonging to novel protein-coding genes and the remaining 14,726 belonging to long non-coding RNA (Table S3). Through gene structure and variant analysis, the type and number of alternative splicing of all genes were determined (Figure 1F). The results showed that nearly 60% of the alternative splicing events belonged to skipped exon (SE), followed by alternative 3' splicing site (A3SS), and mutually exclusive exons (MXE) events accounted for less than 10% (Figure 1E; Table S4). Differential analysis of alternative splicing of different samples showed that Control-vs-Time1, Time6-vs-Time7 and Time3-vs-Time4 had more SEs. Except for the Time3-vs-Time4 group, the proportion of retained intron (RI) events was the same in each group. Interestingly, the Control-vs-Time1 group contained fewer A3SS and alternative 5' splicing site (A5SS) events.

Gene annotation based on public database alignment

The PlantTFDB and the PRGdb were used to annotate transcription factors and plant resistance genes, respectively

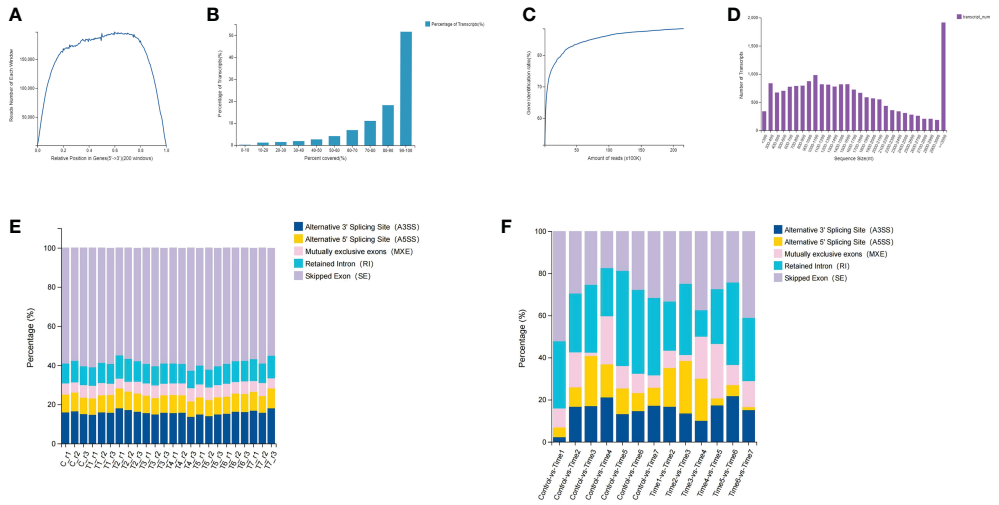


FIGURE 1
Data quality assessment and gene structure analysis. **(A)** Distribution of clean reads in transcripts; **(B)** Read coverage of all transcripts; **(C)** Sequencing reads saturation curve; **(D)** Length of all transcripts; **(E)** Alternative shearing of all transcripts **(F)** Differential alternative splicing analysis of transcripts.

(Sanseverino et al., 2009). A total of more than 40 transcription factors were identified, including several TF families with many members, such as MYB, AP2-EREBP, bHLH, NAC, *etc.* (Figure 2A). A total of 13 disease resistance genes were annotated, mainly including the coiled-coil nucleotide-binding leucine-rich repeat (CC-NB-LRR/CNL) domain, NBS domain (N), NBS and LRR (NL) domain, receptor serine-threonine kinase-like and extracellular leucine-rich repeat (RLP) domain, Toll-interleukin receptor-like (TIR) domain, and TIR-NB-LRR/TNL domain (Figure 2B). In addition, the GO and KEGG databases were used for gene annotation and functional enrichment analysis. The results of GO and KEGG annotations show that these genes were involved in genetic information processing, cellular processes, metabolic processes, membranes, protein binding, and catalytic activities (Figures 2C, D). The GO and KEGG functional enrichment showed that these genes mainly involved in ascorbate and aldarate metabolism, lysine degradation, isoquinoline alkaloid biosynthesis, ubiquitin-mediated proteolysis, glycoside synthesis and degradation, protein processing, and other metabolic pathways and molecular regulation processes (Figures 2E, F).

Gene quantification and time-series analysis

To confirm the correlation between samples, the Pearson correlation coefficient of the gene amounts were calculated between every two samples (Table S5). The gene expression similarity in the three biological replicates of each group is the highest, indicating that the reproducibility of the samples is good. The similarity between the two groups of T5 and T6 is relatively high, and the correlation coefficients are above 0.8%, followed by the two groups of T3 and T4 (Figure 3A). The principal component analysis shows that except for the T7 group, the differences among the other groups of samples are not significant (Figure 3B). However, some samples between groups, such as T2 and T3 groups, were not completely separated by PCA analysis. The expression density shows that most genes have a similar expression trend, and the expression ($\log_2$ (FPKM+1)) is mainly distributed between 0 and 2. When the gene abundance is between 0.2 and 1.5, the gene abundance is high (Figure 3C). The number of genes with FPKM values more than or equal to 10 is the greatest in each group, followed by those with expression levels between 1 and 10 (Figure 3D). Time-series analysis based on a loose clustering algorithm showed that these genes clustered into 12 expression patterns in eight different time periods (Figure 3E). The gene expression peak of cluster 1 is at T6, the gene expression peak of cluster 7 is at T3, the gene expression peak of cluster 4 is at T2, and the gene expression peak of cluster 3 is at T3. The gene expression of cluster 8 gradually increased with the MeJA treatment, and the general trend of the gene expression gradually decreased in cluster 8.

Hierarchical clustering, and functional enrichment analysis of DEGs

Firstly, the DEGs between the CK and the MeJA treatment group were analyzed. The number of DEGs in the Control-vs-Time6 group was at most 4151, including 1730 up-regulated genes and 2421 down-regulated genes. The next group is the Control-vs-Time7 group, with a total of 2427 DEGs. There were 1,504 differential genes in the Control-vs-Time5 group (Figure S1). The number of DEGs among the four groups from Control-vs-Time1 to Control-vs-Time4 was lower (Figures 4A–G). Heatmap clustering analysis showed that the DEGs between the CK and the MeJA group at different time points

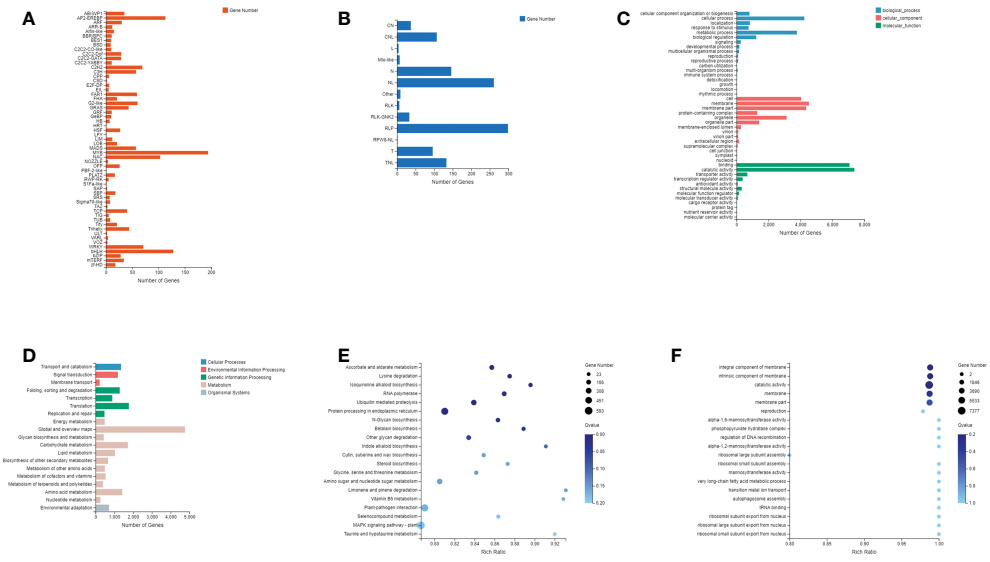


FIGURE 2
Annotation of identified genes and novel genes. **(A)** Family and number of all transcription factors identified; **(B)** Domain classification and number of plant resistance genes; **(C)** GO classification of all genes; **(D)** KEGG classification of all genes; **(E)** KEGG pathway enrichment analysis; **(F)** GO enrichment analysis.

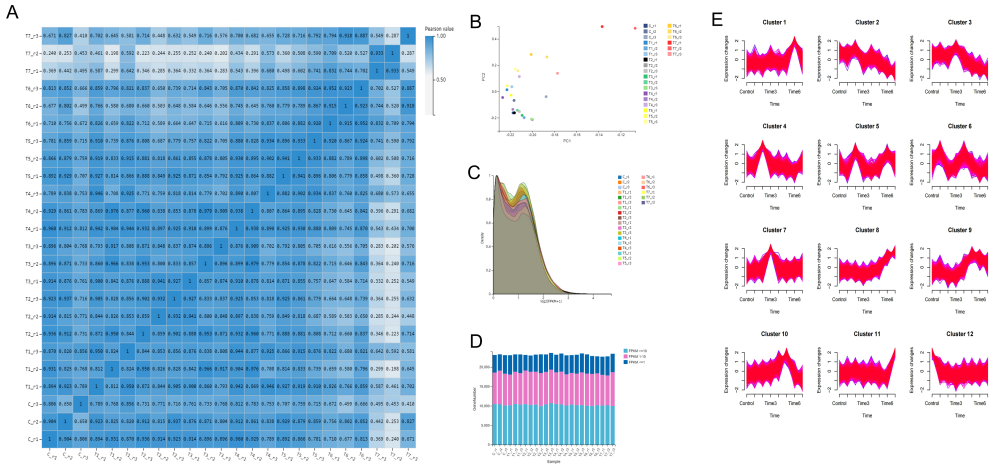


FIGURE 3
Gene quantification and gene expression analysis. **(A)** correlation analysis among all sample groups; **(B)** principal component analysis among samples; **(C)** gene expression density of all groups; **(D)** gene expression distribution of all samples; and **(E)** time series analysis of all genes.

showed various patterns. The expression of most genes showed an up-regulation trend under MeJA treatment (Figures 4H–N). Control-vs-Time1 and Control-vs-Time2 two groups contained relatively few DEGs; only 16 and 18 DEGs were highly expressed (FPKM > 8). However, most differential genes had moderate expression levels in the Control-vs-Time6 and Control-vs-Time7 groups. Furthermore, GO and KEGG pathway classification and functional enrichment were performed on these DEGs. GO classification showed that molecular functions involving molecular binding and catalytic activity contain the largest number of genes, followed by cellular components involving membrane structure and organelle composition and biological processes such as metabolic processes (Figures S2A–G). KEGG pathway classification mainly involves global maps and carbohydrate generation and involves protein folding, sorting, and degradation, as well as genetic information processing such as transcription and translation (Figures S2H–N). The GO functional enrichment results indicated that the enriched pathways and biological processes were quite different under different time treatments (Figures 5A–G). For example, the Control-vs-Time2 group mainly involves pattern binding, polysaccharide binding, and carbohydrate binding. The Control-vs-Time4 group mainly involves acylphosphatase activity and tetrahydrofolate regulator activity. The Control-vs-Time5 group mainly involves microtubule severing and the katanin complex. The Control-vs-Time6 mainly involves radiation, light stimulus responses and organelle assembly. The Control-vs-Time7 mainly involves cysteine biosynthesis and phosphorylase activity. KEGG pathway functional enrichment indicated that the enrichment of glycosphingolipid biosynthesis in the Control-vs-Time5 and Control-vs-Time7 groups was the most obvious (rich factor ratio close to 1). Followed by the Control-vs-Time6 group, the main pathways involved are monoterpene biosynthesis, sulfur metabolism, circadian rhythm, *etc.* Exceptionally, some secondary metabolic pathways involved in flavone and flavonol biosynthesis, vitamin B6 metabolism, brassinosteroid biosynthesis, glutathione metabolism, *etc.* were also significantly enriched in other groups (Figures 5H–N).

Acquisition of gene sequences and identification of the *bHLHs*

Through NCBI, EnsemblPlants, and TAIR databases, the genome sequences and annotation files of *D. huoshanense, A. thaliana, O. sativa, D. nobile*, and *D. chrysotoxum* were obtained. The established hidden Markov model was used to compare and predict the DhbHLH genes, and the identified bHLH genes were further retrieved and verified using the Pfam, Interpro, and Smart databases. A total of 83 bHLH genes were identified. Further, the physical characteristics of these bHLH candidates were analyzed (Table S6). The amino acid length of these bHLHs ranged from 85 to 1689, the relative molecular mass ranged from 9855 to 183707, and the protein isoelectric points ranged from 4.4 to 10.2. The grand average of hydropathicity (GRAVY) ranges from -0.798 to -0.076. The aliphatic index ranges from 57.3 to 103.1.

Collinearity analysis of the *bHLH* genes

To reveal the evolutionary relationship of *bHLH* genes among different species, all the synteny blocks of *D. huoshanense, A. thaliana, O. sativa, D. nobile*, and *D. chrysotoxum* were obtained by collinear analysis (Figure 6). The number of gene pairs of *D. huoshanense* and *D. nobile* was the highest, with a total of 92, followed by *D. chrysotoxum*, with 89 gene pairs. *D. huoshanense* had fewer syntenic blocks with *O. sativa* and *A. thaliana*, 31 and 6 blocks, respectively (Table S7). The above results are consistent with the genetic relationship between species. It is worth noting that there is synteny between some *DhbHLH* genes and multiple *bHLH* genes of *D. nobile*, and *D. chrysotoxum*. For example, *Dhu000000317* is collinear with *Maker76141, Maker76022, Maker79668*, and *Maker60892* of *D. chrysotoxum*. *Dhu000027551* is collinear with *Maker76141, Maker66639, Maker60892*, and *Maker86412* of *D. chrysotoxum*. *Dhu000009253* is collinear with *Dnobile06G01687.1,*

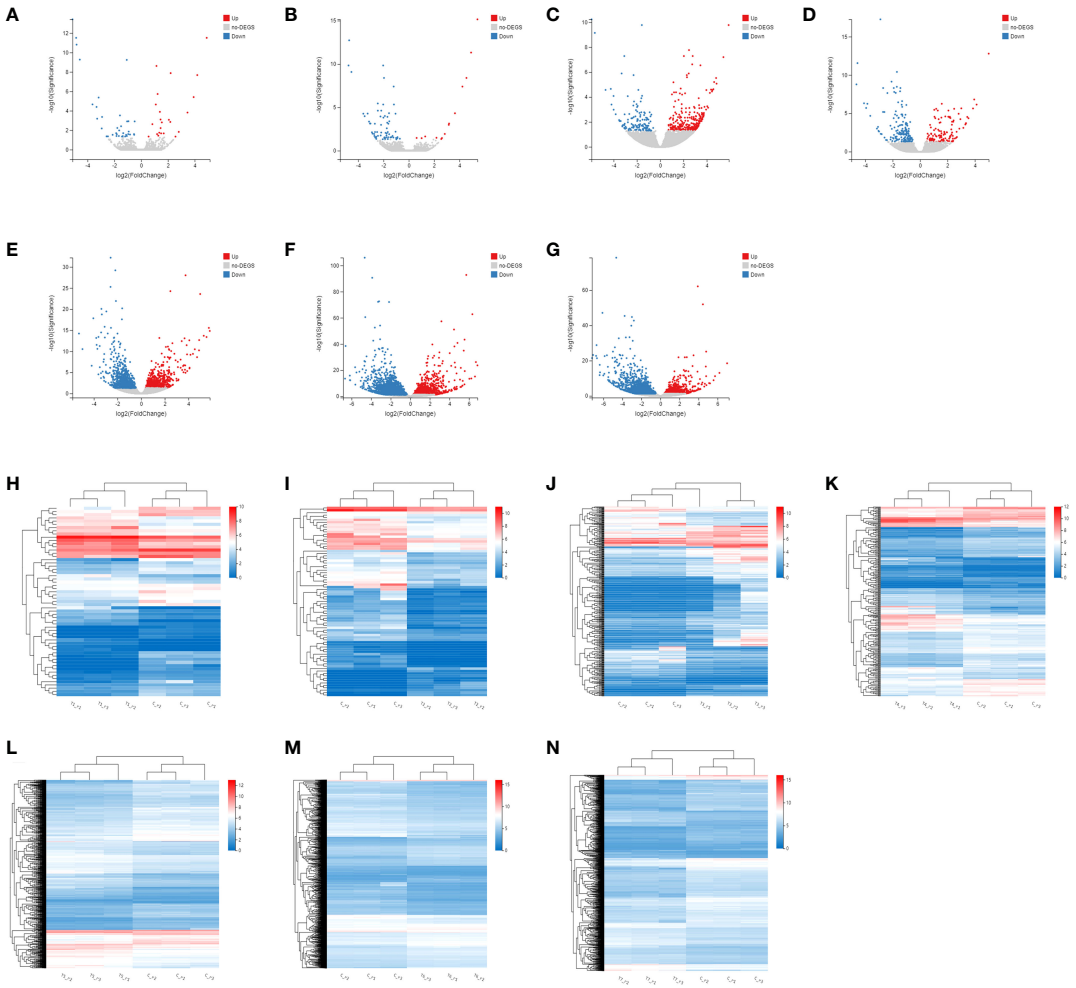


FIGURE 4
Screening and cluster analysis of DEGs. **(A)** DEGs between Control-vs-Time1; **(B)** DEGs between Control-vs-Time2; **(C)** DEGs between Control-vs-Time3; **(D)** DEGs between Control-vs-Time4; **(E)** DEGs between Control-vs-Time5; **(F)** DEGs between Control-vs-Time6; **(G)** DEGs between Control-vs-Time7; Hierarchical clustering of DEGs in **(H)** Control-vs-Time1; **(I)** Control-vs-Time2; **(J)** Control-vs-Time3; **(K)** Control-vs-Time4; **(L)** Control-vs-Time5; **(M)** Control-vs-Time6; **(N)** Control-vs-Time7.

Dnobile13G01347.1, *Dnobile13G01310.1*, and *Dnobile13G00716.1* of *D. nobile*. *Dhu000022245* is collinear with *Os03t0205300-00* and *Os07t0588400-00*. *Dhu000013850* has collinearity with *AT2G24260.1*, *AT4G30980.2*, and *AT5G58010.1*.

Phylogenetic analysis of the DhbHLHs and AtbHLHs

Based on the ML method, a phylogenetic tree for DhbHLHs and AtbHLHs was constructed. According to the *Arabidopsis* nomenclature and the known bHLH classification, the DhbHLH protein was divided into 12 subgroups (groups I–XII) (Figure 7). However, Group III can be further divided into six subclades: a, b, c, d, e, and f. Group IV can be divided into four subclades: a, b, c, and d. Group V can be divided into a and b subclades. Group VII can be divided into a and b subclades. Group VIII can be divided into three subclades: a, b, and c. From the ultrafast bootstrap (UFBoot) support and SH-aLRT test results, several DhWRKY proteins have high homology with AtWRKYs. For example, the UFBoot support of DhbHLH26 and AtbHLH71 was 99%, and the SH-aLRT test was 92.1%. The UFBoot support of DhbHLH44 and AtbHLH97 was 100%, and the SH-aLRT test was 97.2%.

Chromosomal location, conserved motif, and gene structure analysis of the *DhbHLH* gene

In order to determine the location of *bHLH* genes on the chromosome of *D. huoshanense*, the structural composition and motif composition of these genes were analyzed. The results showed that *DhbHLH* was unevenly distributed on 17 pseudochromosomes (Figure 8A). Chromosome 7 contains the most *bHLH* genes, with 11, followed by chromosome 13 and chromosome 15, with 8 and 7 *bHLH* genes, respectively. In addition, large-scale gene duplication and segmental duplication

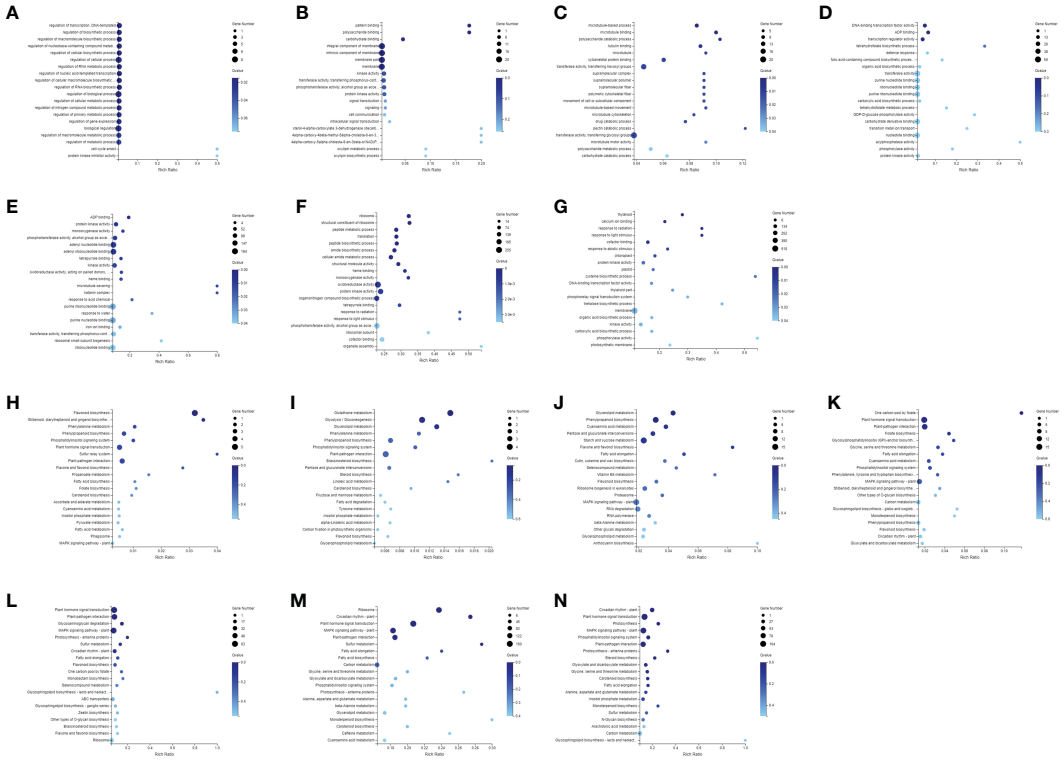


FIGURE 5
Annotation and functional enrichment of DEGs. GO enrichment analysis of DEGs of **(A)** Control-vs-Time1; **(B)** Control-vs-Time2; **(C)** Control-vs-Time3; **(D)** Control-vs-Time4; **(E)** Control-vs-Time5; **(F)** Control-vs-Time6; **(G)** Control-vs-Time7; KEGG pathway enrichment analysis of DEGs of **(H)** Control-vs-Time1; **(I)** Control-vs-Time2; **(J)** Control-vs-Time3; **(K)** Control-vs-Time4; **(L)** Control-vs-Time5; **(M)** Control-vs-Time6; **(N)** Control-vs-Time7.

events of *bHLH* genes existed on different chromosomes. For example, *DhbHLH80* and *DhbHLH81* are associated with *DhbHLH20/21* and *DhbHLH27/28*, respectively. For gene duplications on the same chromosome with a physical distance of less than 100 kb, these *DhbHLH* genes may be obtained through tandem duplication events, like *DhbHLH34* and *DhbHLH35*, *DhbHLH30* and *DhbHLH31*, *DhbHLH57* and *DhbHLH58*, *etc*. Conserved motif analysis showed that the *bHLH* members clustered in the same subgroup had similar motif compositions, and these motifs were regularly distributed among the *bHLH* genes of different subgroups according to the phylogenetic relationship (Figure 8B). A total of 20 motifs were identified, among which motifs 1 and 2 are conserved motifs shared by all genes, presumably the conserved domain of bHLH. DhbHLH67/68 from Group XII contains a unique motif 10, and DhbHLH30/31 from Group IX contains a unique motif 13. DhbHLH75/76 from Group VIIb has a unique motif 12. In group VIIIc, DhbHLH30/31 has a unique motif 18, while DhbHLH14/57/58 has a representative motif 17. In Group IIIe, DhbHLH20/21, DhbHLH27/28, and DhbHLH80/81 have a unique motif4, presumably the bHLH-MYC_N domain, but DhbHLH22/23/24/25 from Group IIId do not have this domain. Gene structure analysis showed that all *bHLH* proteins contained exons and introns, and more than half of the genes also contained a 3' or 5' UTR (Figure 8C). Most genes have multiple exons, but some genes contain only one exon.

Cis-acting elements analysis of the *DhbHLH* gene

To determine whether *bHLH* genes could be triggered by environmental stress, the regulatory elements within their promoters were examined. The results indicated that the promoters of the majority of genes had a dozen or more cis-acting elements (Figure 9A). Some tandemly replicated genes contain *cis*-acting elements of similar composition. Both *DhbHLH34* and *DhbHLH35* have ARE, AE-box, BOX-4, CGTCA-motif, *etc*. Both *DhbHLH30* and *DhbHLH31* contain ABRE, BOX-4, CGTCA-motif, GATA-motif, G-box, I-box, LTR, *etc*. Both *DhbHLH57* and *DhbHLH58* contain ARE, BOX-4, GATA-motif, G-box, GCN4-motif, MBS, P-box, TCT-motif, *etc*. Both *DhbHLH65* and *DhbHLH66* contain ABRE, ARE, BOX-4, G-box, GT1-motif, LTR, RY-element, TCT-motif, *etc*. Both *DhbHLH49* and *DhbHLH50* contain ABRE, ARE, CGTCA-motif, G-box, GT1-motif, LAMP-element, TCCC-motif, *etc*. A total of 52 *cis*-acting elements involved in the growth and stress responses were identified (Figure 9B). Box-4 is the most light-responsive component, accounting for 13% of all motifs. This is followed by the G-box motif, which is also associated with light stress. The third most common motif is ABRE, which is mainly related to the ABA-responsive gene. In addition to elements related to stress, some specific *cis*-acting elements were also identified. For example, the CAAAGATATC-motif is a element related to circadian rhythm

10.3389/fpls.2023.1169386

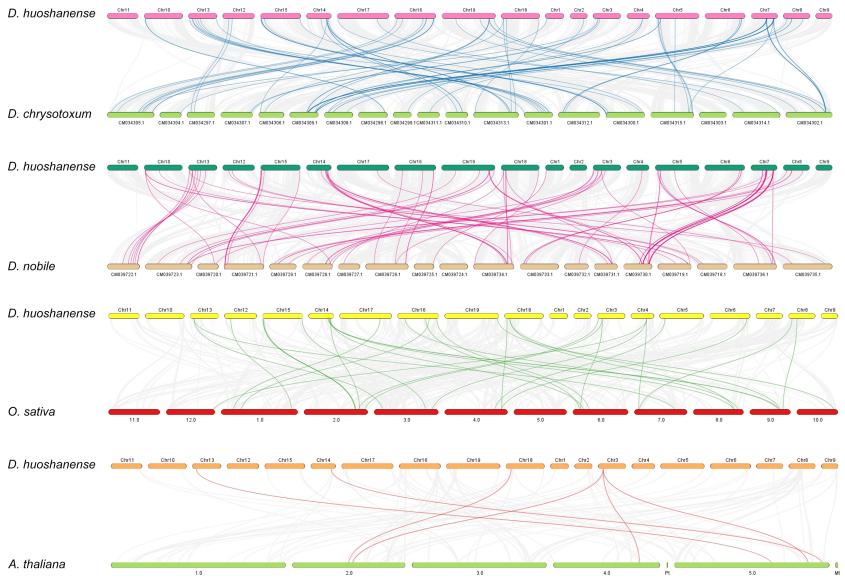


FIGURE 6
Collinearity analysis of DhbHLH and bHLHs from other species.

control, and these elements in the promoter regions of 19 *bHLH* genes were detected. MBSI elements in 6 *bHLH* genes were closely related to the biosynthesis of flavonoids. Functional classification also shows that the number of elements involved in light responsiveness is the largest, followed by MeJA responsiveness, ABA responsiveness, auxin-responsiveness, and other hormone-related regulatory elements (Figure 9C).

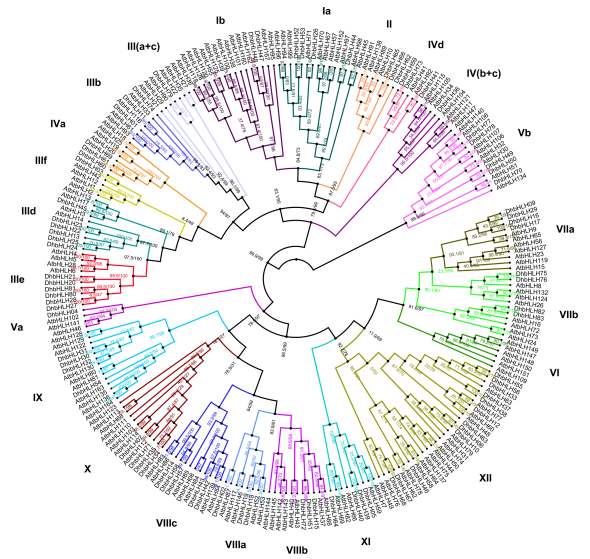


FIGURE 7
Phylogenetic analysis of the *DhbHLH* and *AtbHLH* genes.

Expression profile and correlation analysis of the *DhbHLH* gene under different MeJA treatments

Using transcriptome sequencing, a batch of DEGs were found in the early stages. To clarify which *DhbHLH* genes were induced or inhibited by MeJA, the heat map clustering analysis at different time points were performed (Figure 10; Table S8). More than half of the *bHLH* genes were low-expression genes (FPKM<1) throughout the treatment. But there were also some *bHLH* genes, such as *DhbHLH21, DhbHLH27, DhbHLH28, DhbHLH52, DhbHLH70*, and *DhbHL80*, which showed continuous expression throughout the time period. There are also individual genes, such as *DhbHLH46* and *DhbHLH48*, where the expression level of MeJA treatment first increased and then decreased. In addition, a correlation analysis on *bHLH* gene expression levels was performed to ensure which genes might be co-expressed. The blue modules in the figure represent positive correlations, while the red modules represent negative correlations (Figure 11). *DhbHLH16* and *DhbHLH17* have the highest correlation (R^2 = 0.98), followed by *DhbHLH16* and *DhbHLH17* (R^2 = 0.97). The correlation coefficient between *DhbHLH03* and *DhbHLH65* is 0.96. These significantly positively correlated genes may have the characteristics of co-expression. In addition to positively correlated expression patterns, there were also significant negatively correlated expression patterns of some *bHLH* genes, such as *DhbHLH23* and *DhbHLH80* (R^2 = -0.93), *DhbHLH23* and *DhbHLH48* (R^2 = -0.91).

10.3389/fpls.2023.1169386

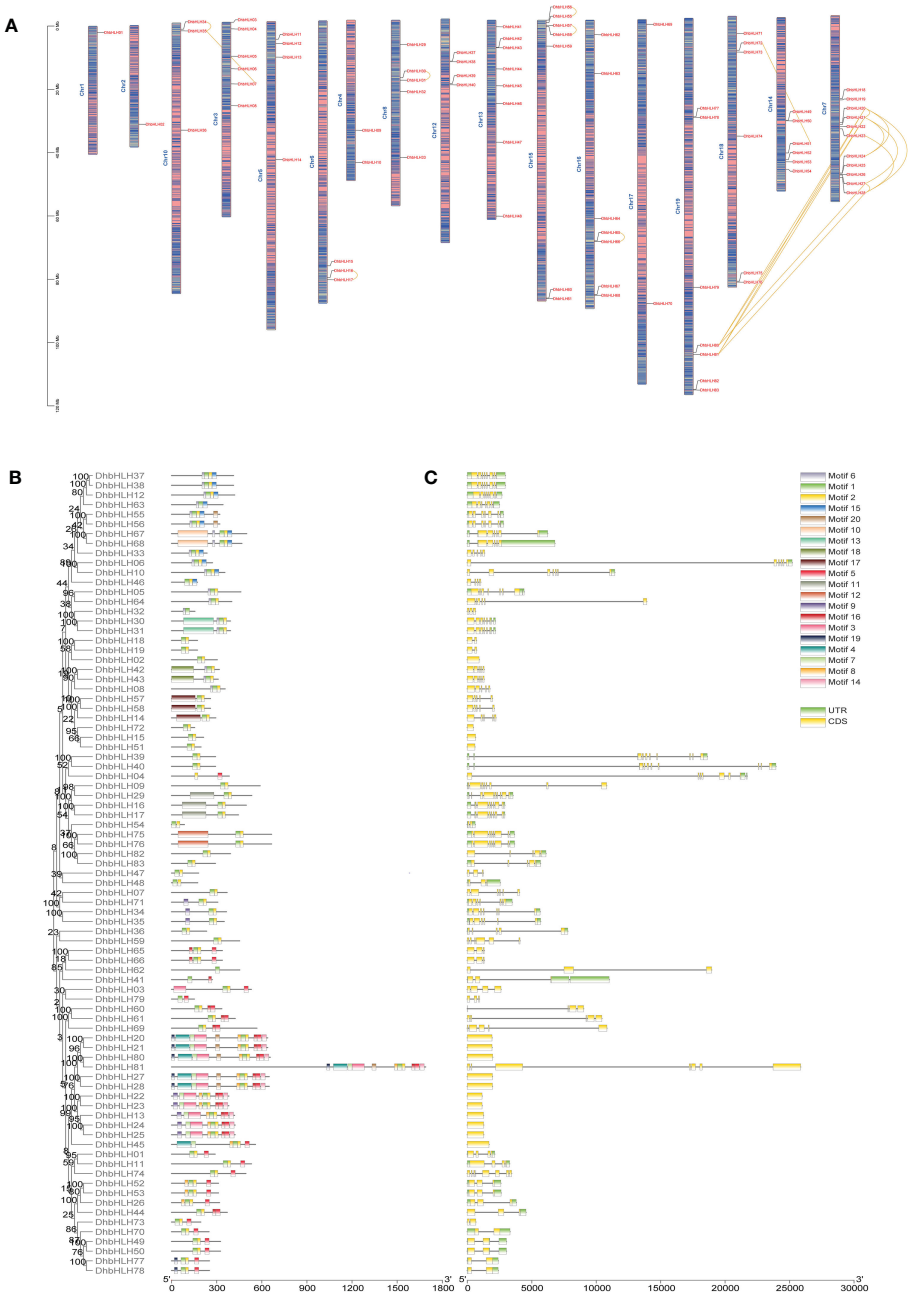


FIGURE 8
Chromosomal location, conserved motifs, and gene structure of the *DhbHLH* gene. **(A)** Chromosomal distribution of *DhbHLH*. **(B)** Distribution of the conserved motifs of *DhbHLH*. **(C)** The UTRs and exon-intron structures of *DhbHLH*. The chromosome number is represented at the top of each chromosome, and the left scale is in megabases (Mb). There is tandem duplication or segmental duplication between genes.

Conserved domain, phylogenetic tree, and *q*RT-PCR analysis of the *bHLH* subgroups IIId and IIIe

Multiple sequence alignment shows that the N-terminus of bHLH subgroups IIId/e has both the bHLH zipper conserved domain and a typical bHLH-MYC domain (Figure 12A). The bHLH-MYC_N domain contains JID and TAD, which are responsible for the binding of JAZ to MYC2 and the transcriptional activation of MYC-targeted genes, respectively. DhMYC2 contains GDG motifs, suggesting their ability to bind JAZ proteins (Song et al., 2022b). The presence of RA[K/L]QAQ motifs in DhMYC2 suggests that it is a canonical MYC2 gene. Interestingly, the 469th aa of DhbHLH20 and the 350th aa of AtbHLH28 are both M residues, while this specific aa in other species is a Q residue here. The SDXH motif is the site responsible for the phosphorylation modification of MYC2. The ACT-like domain is widely distributed in the bHLH subgroups IIId/e and forms the βαββαβ secondary structures. The ACT-like domain of DhMYC2 was not highly homologous to the MYC2 genes of several species. A phylogenetic tree was constructed with identified

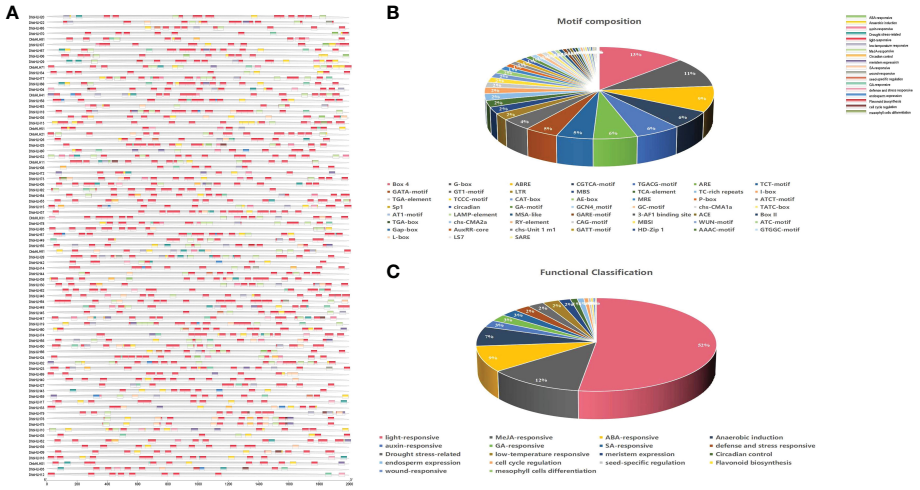


FIGURE 9
Cis-acting elements of the *DhbHLH* gene. **(A)** Distribution of all *cis*-acting elements in the upstream 2000 kb of the promoter; **(B)** motif composition and number; **(C)** functional classification and number.

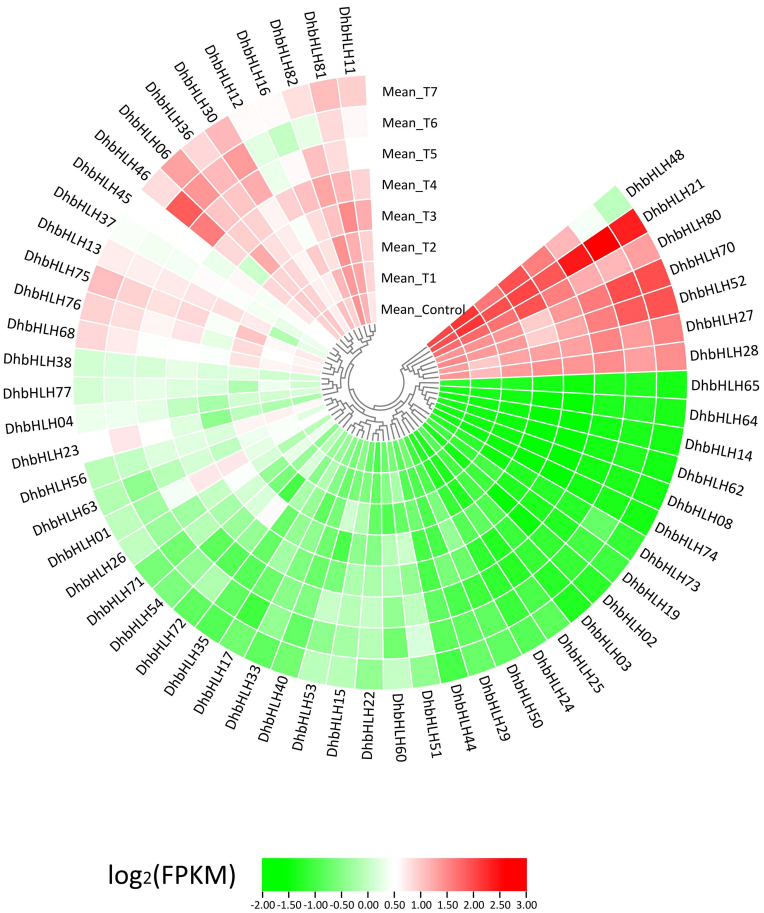


FIGURE 10
Expression profile of the *DhbHLH* genes under MeJA treatment. All FPKM values are the average of three replicates per group.

10.3389/fpls.2023.1169386

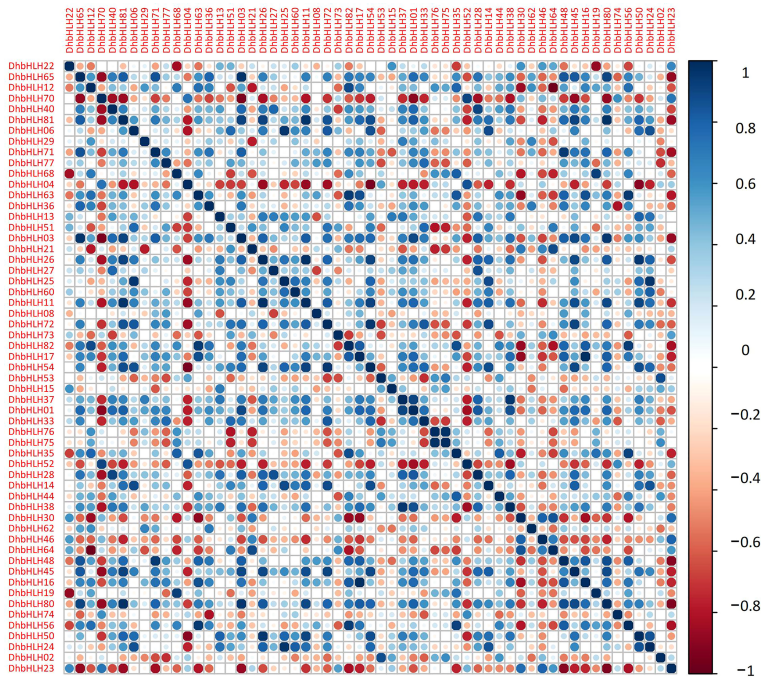


FIGURE 11
Correlation analysis of the expression levels of main *DhbHLHs*. Blue represents a positive correlation, and red represents a negative correlation.

MYC or MYC-like proteins from seven other species to determine which bHLH subgroups IIId/e have potential transcriptional regulatory functions (Stitz et al., 2011). The results indicated that bHLH subgroups IIId/e could be clustered into two major groups (Figure 12B). DhMYC2 subfamily members have higher homology with OsMYC2. DhbHLH subgroup IIId (DhbHLH13, DhbHLH23, DhbHLH24, and DhbHLH25) and AtbHLH14 clustered into one group, while DhbHLH subgroup IIId (DhbHLH45) clustered together with AtbHLH3. DhbHLH subgroup IIIf (DhbHLH03) is in a separate branch as an outgroup.

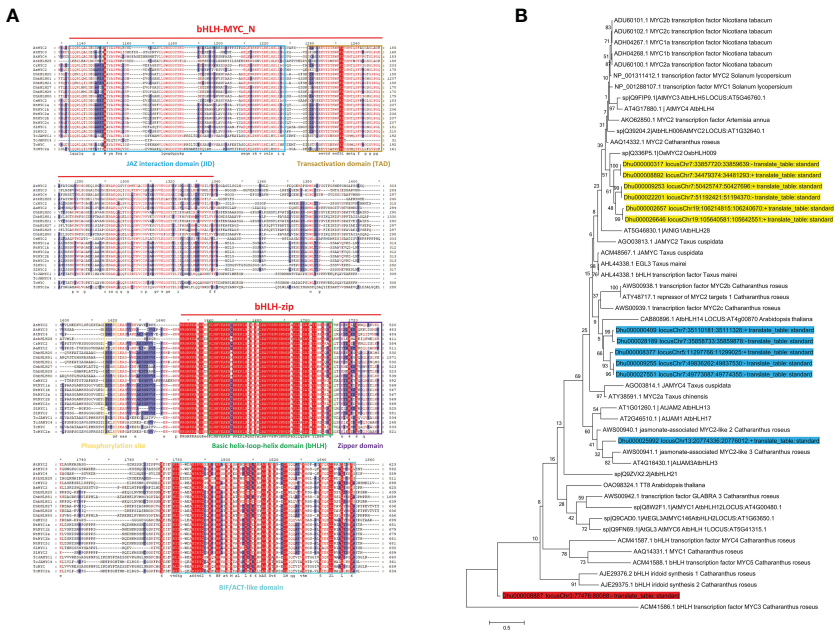


FIGURE 12
Amino acid sequence alignment and the phylogenetic tree of MYC2-related genes. **(A)** Multiple sequence alignment of DhMYC2 and MYC2-like genes of other species; **(B)** Phylogenetic tree of DhMYC2 and MYC2-like genes of other species. The bootstrap values are indicated in the nodes of the branches. Yellow indicates the MYC2 subclade, blue indicates the JAM subclade, and red is the outgroup. *represents intervals of every 20 amino acidresidues.

10.3389/fpls.2023.1169386

To figure out whether *DhbHLH* expression was induced or inhibited by MeJA, the expression pattern of *bHLH* subgroups IIId/e was investigated (Table S9). The results indicated that the six members of the *DhMYC2* subfamily showed differential spatiotemporal expression characteristics at different time periods (Figure 13A). The expression of *DhbHLH81* and *DhbHLH20* were the highest at 16 and 4 hours of treatment, which were more than 20 times those of the control group. Except for *DhbHLH81*, the other five *DhbHLHs* had the largest peak value during the treatment, and then their expression levels gradually decreased. Compared with the *DhMYC2* subfamily, the expression of *DhbHLH* subgroup IIId genes were generally weaker (Figure 13B). The expression of *DhbHLH23* and *DhbHLH13* peaked at the 4th hour and then gradually decreased. Interestingly, the expression of *DhbHLH22, DhbHLH24, DhbHLH25*, and *DhbHLH45* were lower than those of the control group at the 8th hour after treatment. In particular, the expression of *DhbHLH25* and *DhbHLH45* were strongly inhibited by MeJA.

Discussion

Comparative transcriptomics of MeJA treatment based on NGS

Comparative transcriptomics technology provides a powerful tool for the discovery of differential genes related to transcriptional regulation (Shoji, 2019). In this study, the time-series transcriptome analysis combined with genome data were integrated to construct the gene regulatory network of JA signaling in *D. huoshanense*. A spatiotemporal expression profile of genes was established under different MeJA treatments. 8,800 new genes were predicted from more than 27,000 identified genes. Nearly 60% of these genes had SE alternative splicing, and A3SS events were less than 20%. Different forms of RNA splicing provide kinetic energy for exon recombination at the transcriptional level (Chini et al., 2016). Through gene annotation and functional enrichment, these DEGs were mainly involved in transcription, translation, protein folding, and the biosynthesis of secondary metabolites (Wang et al., 2021).

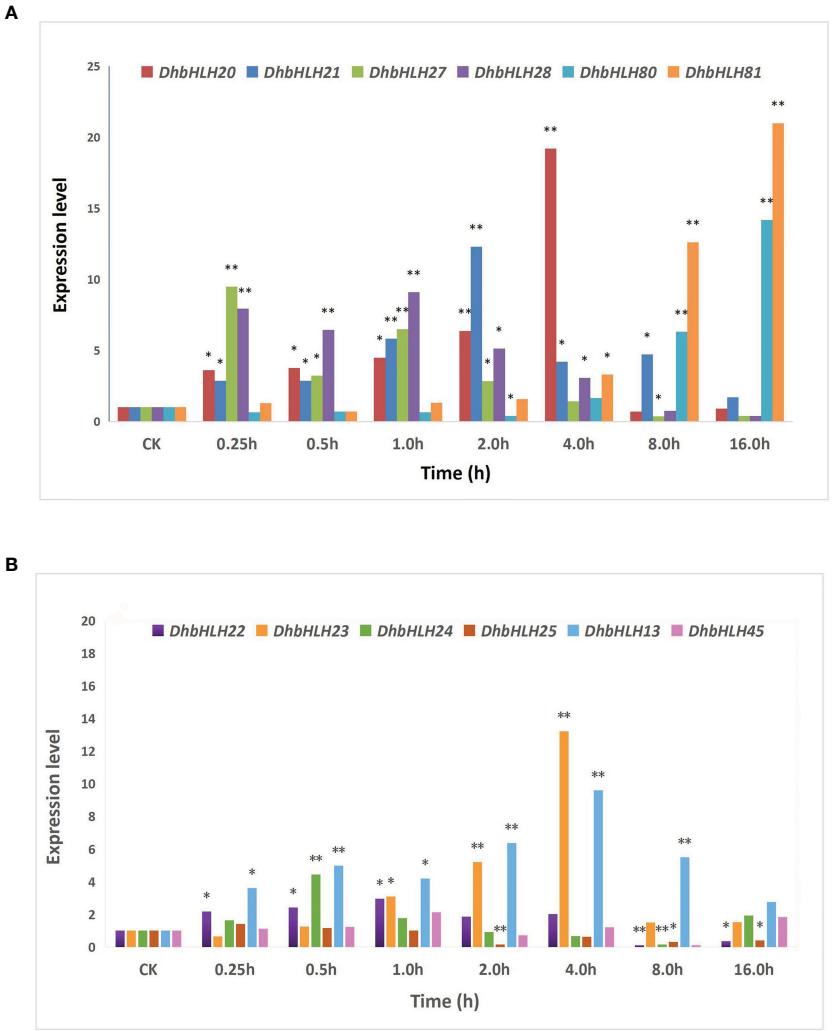


FIGURE 13
Expression analysis of *bHLH* subgroups IIId and IIIe. **(A)** the expression level of IIIe *DhbHLH* genes; **(B)** the expression level of IIId *DhbHLH* genes. '*'P<0.01, '**'P<0.001.

The JA-mediated gene regulatory network coordinates the spatiotemporal specific activation and co-expression of the genes involved in the aforementioned biological processes (Afrin et al., 2015; Yan et al., 2016). In the initiation stage (within 1 hour), JA preferentially regulated the expression of the JA-biosynthetic genes, primary metabolism-related genes, and some transcription factors. In the effect stage (1-2 hours), enzymes involved in the metabolism of amino acids, fatty acids, sugars, and other substances, as well as genes involved in the plant immune response, began to be expressed in large quantities. During the time-lapse stage (after 2 hours), genes involved in substance transport, hormone transport, and secondary metabolite biosynthesis started gradually activating (Hickman et al., 2017; Zander et al., 2020).

Identification of *DhbHLH* genes and gene duplication events

bHLH TFs have similar conserved domains of bHLH-MYC N and bHLH-zip, allowing orthologous genes of many species to preserve a high degree of homology. In this study, DhbHLHs maintained high homology with many members of AtbHLH (SH-aLRT ≥ 80% and UFBoot ≥ 95%), especially branches on IIIb and Ia subclades, such as AtbHLH116 and DhbHLH11, AtbHLH71 and DhbHLH26, AtbHLH97 and DhbHLH44, *etc.* (Guindon et al., 2010; Minh et al., 2013). The CDS sequence of the *DhbHLH* gene contains multiple exon and intron structures, and some genes also contain a 3' or 5' UTR. The UTR structure is generally considered to be related to gene transcription, but its sequence cannot be translated into protein (Thiel et al., 2021). *DhbHLH62* and *DhbHLH68* have a particularly long 3' UTR, and their role in gene transcription and expression needs further study. Several studies have shown that two WGDs occurred in the orchid lineage (Zhang et al., 2017; Song et al., 2022b). *Apostasia*, *Dendrobium*, and *Phalaenopsis* all experienced a WGD recently, which may have arisen around the *K/Pg* boundary. Putative peaks in the early *Ks* age distribution could mean that monocot ancestors went through more WGD events in the past (Cai et al., 2015; Zhang et al., 2017). Here, 17 chromosomes contained imbalanced duplicates of these genes. Collinearity analysis showed that some *bHLH* genes underwent WGD or segmental duplication. The *bHLH* genes with similar distances (< 100kb) on the same chromosome were doubled and expanded by paralogous genes through tandem duplication. It acts as a catalyst for the development of new genes.

Cis-acting elements and expression profile analysis of *DhbHLHs*

Cis-acting elements are motifs on the promoter region that are triggered by the external environment and typically have tissue- and time-specific transcriptional regulation (Zander et al., 2020). Here, a series of elements involved in hormone signal transduction, cell cycle regulation, circadian rhythm, abiotic stress, and secondary metabolism were identified. More than half of the elements associated with photoperiod control and ABA responsiveness. These elements include the Box 4 motif, the G-box motif, the AREB element, the CGTCA motif, the ARE element, *etc.* Using comparative transcriptomics, a collection of differentially expressed genes was identified to determine which *WRKY* genes can be activated by MeJA. A series of key genes involved in the JA signaling pathway were identified. Only a few *bHLH* genes, such as *DhbHLH21, DhbHLH70, DhbHLH52*, and *DhbHLH27*, were up-regulated in response to MeJA treatment, as shown by heatmap clustering and expression profiling. Previous studies showed that CrORCA3 regulates the JA-responsive genes expression in the biosynthesis of terpenoid indole alkaloids (Zhang et al., 2011). CrMYC2 has the capacity to bind to the promoter of *ORCA3*, hence modulating its mRNA expression (Paul et al., 2017). The MeJA-induced gene from *Solanum lycopersicum* (*SlJIG*) was dramatically induced by MeJA treatment (Cao et al., 2020). *LjbHLH7* regulates the production of cyanogenic glucosides by directly activating the expression level of the *CYP79D3* gene (Chen et al., 2022).

Expression pattern and function of the IIId/e subgroup *bHLH* genes

Studies have shown that bHLH subgroups IIId/e have unique functions in the JA signaling (Zhao et al., 2021). In this study, homologous sequence alignment and phylogenetic analysis were used to screen out *bHLH* proteins that may have transcriptional activation functions. DhbHLH subgroups IIId/e all contain a typical MYC2 conserved domain, namely bHLH-MYC_N (Liu et al., 2022). The phylogenetic tree indicated that the *bHLH* subgroups IIId and IIIe genes were further classified into two branches, namely the *JASMONATE-ASSOCIATED MYC2-LIKE* (*JAM*) subclade and the *MYC2* subclade. In *A. thaliana*, *bHLH* subgroup IIIe include *MYC2*, *MYC3*, *MYC4*, and *MYC5*, which mainly positively regulate the expression of JA-responsive genes and disease resistance responses (Fernández-Calvo et al., 2011; Niu et al., 2011; Qi et al., 2015b; Song et al., 2017). However, *bHLH* subgroup IIId like *bHLH17/JAM1*, *bHLH13/JAM2*, *bHLH3/JAM3*, and *bHLH14* function as repressors to antagonistically regulate JA responses (Nakata et al., 2013; Sasaki-Sekimoto et al., 2013; Song et al., 2013; Fonseca et al., 2014; Qi et al., 2015b). Six DhbHLH subgroup IIId genes with AtbHLH14, AtbHLH13, AtbHLH17, and AtbHLH3 clustered in one branch, suggesting they could play a negative regulatory role in JA signaling. The DhbHLH subgroup IIIe clustered in a branch with OsbHLH009 and CrMYC2, indicating that they may upregulate the expression of JA-responsive genes.

Conclusions

In this study, A total of 57 *DhbHLH* members were screened out using a large-scale genome identification. Comparative genomics

revealed multiple duplication events of the *bHLH* gene in the *D. huoshanense* genome, which partially explained why the *bHLH* gene's expansion was obtained through WGD and segmental duplication, as evidenced by collinearity analysis. Thousands of DEGs involved in the regulation of JA signaling were screened out using comparative transcriptomics. These *bHLH* genes showed different expression patterns, more than half of which were low-expression genes. Only a few genes were strongly induced by MeJA. Through the analysis of *cis*-acting elements, it was indicated that more than half of the elements are related to light signal, hormone signaling, and other abiotic stresses. There are also a small number of elements related to the cell cycle, circadian rhythm, and secondary metabolism. The *q*RT-PCR results showed that IIId and IIIe *DhbHLH* subgroups had distinct expression patterns. *DhbHLH20*, *DhbHLH32*, and *DhbHLH81* from the *DhbHLH* subgroup IIIe were strongly induced by JA, but the expression of *DhbHLH81* was somewhat delayed (peaking at 16 hours). The expression patterns of the IIIe *DhbHLH* subgroup were relatively consistent. The expression of the core JA-responsive genes, such as *DhbHLH13* and *DhbHLH23*, both peaked at the fourth hour. The research would provide scientific tools for the discovery of the *bHLH* genes in other *Dendrobium* species.

Data availability statement

The data presented in the study are deposited in the NCBI GenBank and NGDC GSArepository, accession number PRJNA597621 and CRA006607, respectively.

Author contributions

CS, YZ and XH discussed the writing plan. CS, XH, WZ and YW drafted the manuscript. CS and IS edited the manuscript. CS, FZ and JD conducted the experiment. CS, CJ, GL, and LL analyzed the data. CS, XH and CC acquired the funding. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Demonstration Experiment Training Center of Anhui Provincial Department of Education (2022sysx033, wxxy2022191), Anhui Province Key R&D Project (202104h04020008), Anhui Province Science and Technology Major Project (202003c08020004), Anhui Province Natural Science Key Research Project (KJ2020A0986), High-level Talents Research Initiation Fund of West Anhui University (WGKQ2022025) and National Innovation Training Program for College Students (202110376082).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1169386/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Number of DEGs among all sample groups

SUPPLEMENTARY FIGURE 2
Classification of DEGs. GO classification of DEGs of (A) Control-vs-Time1; (B) Control-vs-Time2; (C) Control-vs-Time3; (D) Control-vs-Time4 (E) Control-vs-Time5; (F) Control-vs-Time6; (G) Control-vs-Time7; KEGG pathway classification of DEGs of (H) Control-vs-Time1; (I) Control-vs-Time2; (J) Control-vs-Time3; (K) Control-vs-Time4; (L) Control-vs-Time5; (M) Control-vs-Time6; (N) Control-vs-Time7

SUPPLEMENTARY TABLE 1
Reference genome alignment ratios for each group.

SUPPLEMENTARY TABLE 2
Reference gene alignment ratios for each group.

SUPPLEMENTARY TABLE 3
Number of known and novel genes.

SUPPLEMENTARY TABLE 4
Alternative splicing classes and gene numbers for each group.

SUPPLEMENTARY TABLE 5
Correlation coefficients among different sample groups.

SUPPLEMENTARY TABLE 6
The physical characteristics of identified *bHLH* genes from the *D. huoshanense* genome.

SUPPLEMENTARY TABLE 7
The gene pairs between *DhbHLH* and *bHLHs* of other four species.

SUPPLEMENTARY TABLE 8
The FPKM values of *DhbHLH* used for expression profiling.

SUPPLEMENTARY TABLE 9
The primers of the *q*RT-PCR of the *DhbHLH* genes.

References

Afrin, S., Huang, J. J., and Luo, Z. Y. (2015). JA-mediated transcriptional regulation of secondary metabolism in medicinal plants. *Sci. Bull.* 60, 1062–1072. doi: 10.1007/s11434-015-0813-0

Altmann, M., Altmann, S., Rodriguez, P. A., Weller, B., Vergara, L. E., Palme, J., et al. (2020). Publisher correction: extensive signal integration by the phytohormone protein network (Nature, (2020), 583, 7815, (271-276), 10.1038/s41586-020-2460-0). *Nature* 584, E34. doi: 10.1038/s41586-020-2585-1

An, C., Li, L., Zhai, Q., You, Y., Deng, L., Wu, F., et al. (2017). Mediator subunit MED25 links the jasmonate receptor totranscriptionallyactive chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 114, E8930–E8939. doi: 10.1073/pnas.1710885114

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bjornson, M., Pimprikar, P., Nürnberger, T., and Zipfel, C. (2021). The transcriptional landscape of arabidopsis thaliana pattern-triggered immunity. *Nat. Plants* 7, 579–586. doi: 10.1038/s41477-021-00874-5.The

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanin in biosynthesis in arabidopsis thaliana. *Plant J.* 39, 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cai, J., Liu, X., Vanneste, K., Proost, S., and Tsai, W.-C. (2015). The genome sequence of the orchid phalaenopsis equestris. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149

Cao, Y., Li, K., Li, Y., Zhao, X., and Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biology (Basel)* 9, 1–16. doi: 10.3390/biology9030061

Cao, Y., Liu, L., Ma, K., Wang, W., Lv, H., Gao, M., et al. (2022). The jasmonate-induced bHLH gene SlJIG functions in terpene biosynthesis and resistance to insects and fungus. *J. Integr. Plant Biol.* 64, 1102–1115. doi: 10.1111/jipb.13248

Carvalho, S. G., Guerra-Sá, R., de, C., and Merschmann, L. H. (2015). The impact of sequence length and number of sequences on promoter prediction performance. *BMC Bioinf.* 16, S5. doi: 10.1186/1471-2105-16-S19-S5

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 8, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, C., Liu, F., Zhang, K., Niu, X., Zhao, H., Liu, Q., et al. (2022). MeJA-responsive bHLH transcription factor LjbHLH7 regulates cyanogenic glucoside biosynthesis in lotus japonicus. *J. Exp. Bot.* 73, 2650–2665. doi: 10.1093/jxb/erac026

Chen, Y., Wang, Y., Huang, J., Zheng, C., Cai, C., Wang, Q., et al. (2017). Salt and methyl jasmonate aggravate growth inhibition and senescence in arabidopsis seedlings *via* the JA signaling pathway. *Plant Sci.* 261, 1–9. doi: 10.1016/j.plantsci.2017.05.005

Chini, A., Gimenez-Ibanez, S., Goossens, A., and Solano, R. (2016). Redundancy and specificity in jasmonate signalling. *Curr. Opin. Plant Biol.* 33, 147–156. doi: 10.1016/j.pbi.2016.07.005

Du, M., Zhao, J., Tzeng, D. T. W., Liu, Y., Deng, L., Yang, T., et al. (2017). MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. *Plant Cell* 29, 1883–1906. doi: 10.1105/tpc.16.00953

Fan, Y., Yan, J., Lai, D., Yang, H., Xue, G., He, A., et al. (2021). Genome-wide identification and expression analysis of the bHLH transcription factor family and its response to abiotic stress in sorghum [Sorghum bicolor (L.) moench]. *BMC Genomics* 22, 1–18. doi: 10.1186/s12864-021-07848-z

Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66, 94–116. doi: 10.1111/j.1365-313X.2010.04459.x

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J. M., Gimenez-Ibanez, S., Geerinck, J., et al. (2011). The arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* 23, 701–715. doi: 10.1105/tpc.110.080788

Fonseca, S., Fernández-Calvo, P., Fernández, G. M., Díez-Díaz, M., Gimenez-Ibanez, S., López-Vidriero, I., et al. (2014). bHLH003, bHLH013 and bHLH017 are new targets of JAZ repressors negatively regulating JA responses. *PloS One* 9, 1–12. doi: 10.1371/journal.pone.0086182

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563

Gao, J., Yang, S., Cheng, W., Fu, Y., Leng, J., Yuan, X., et al. (2017). GmILPA1, encoding anAPC8-like protein, controls leaf petiole angle in soybean. *Plant Physiol.* 174, 1167–1176. doi: 10.1104/pp.16.00074

Goossens, J., Mertens, J., and Goossens, A. (2017). Role and functioning of bHLH transcription factors in jasmonate signalling. *J. Exp. Bot.* 68, 1333–1347. doi: 10.1093/jxb/erw440

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Han, B., Jing, Y., Dai, J., Zheng, T., Gu, F., Zhao, Q., et al. (2020). A chromosome-level genome assembly of dendrobium huoshanense using long reads and Hi-c data. *Genome Biol. Evol.* 12, 2486–2490. doi: 10.1093/gbe/evaa215

Hao, Y., Zong, X., Ren, P., Qian, Y., and Fu, A. (2021). Basic helix-loop-helix (Bhlh) transcription factors regulate a wide range of functions in arabidopsis. *Int. J. Mol. Sci.* 22, 1–20. doi: 10.3390/ijms22137152

Hickman, R., Van Verk, M. C., Van Dijken, A. J. H., Mendes, M. P., Vroegop-Vos, I. A., Caarls, L., et al. (2017). Architecture and dynamics of the jasmonic acid gene regulatory network. *Plant Cell* 29, 2086–2105. doi: 10.1105/tpc.16.00958

Howe, G. A., Major, I. T., and Koo, A. J. (2018). Modularity in jasmonate signaling for multistress resilience. *Annu. Rev. Plant Biol.* 69, 387–415. doi: 10.1146/annurev-arplant-042817-040047

Jiang, Y. Q., and Deyholos, M. K. (2009). Functional characterization of arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors inabiotic stresses. *Plant Mol. Biol.* 69, 91–105. doi: 10.1007/s11103-008- 9408-3

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kumar, L., and Futschik, M. E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2, 5–7. doi: 10.6026/97320630002005

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Liu, H., Wu, M., Zhu, D., Pan, F., Wang, Y., Wang, Y., et al. (2017). Genome-wide analysis of the AAAP gene family in moso bamboo (Phyllostachys edulis). *BMC Plant Biol.* 17, 1–18. doi: 10.1186/s12870-017-0980-z

Liu, J. P., Pineros, M. A., and Kochian, L. V. (2014). The role of aluminum sensing and signaling in plant aluminumresistance. *J. Integr. Plant Biol.* 56, 221–230. doi: 10.1111/jipb.12162

Liu, Z. J. (2015). The genome sequence of the orchid phalaenopsis equestris. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149

Liu, Y., Du, M., Deng, L., Shen, J., Fang, M., Chen, Q., et al. (2019). Myc2 regulates the termination of jasmonate signaling *via* an autoregulatory negative feedback loop [open]. *Plant Cell* 31, 106–127. doi: 10.1105/tpc.18.00405

Liu, R., Song, J., Liu, S., Chen, C., Zhang, S., Wang, J., et al. (2021a). Genome-wide identification of the capsicum bHLH transcription factor family: discovery of a candidate regulator involved in the regulation of species-specific bioactive metabolites. *BMC Plant Biol.* 21, 1–18. doi: 10.1186/s12870-021-03004-7

Liu, S., Wang, Y., Shi, M., Maoz, I., Gao, X., Sun, M., et al. (2022). SmbHLH60 and SmMYC2 antagonistically regulate phenolic acids and anthocyanins biosynthesis in salvia miltiorrhiza. *J. Adv. Res.* 42, 205–219. doi: 10.1016/j.jare.2022.02.005

Liu, R., Wang, Y., Tang, S., Cai, J., Liu, S., Zheng, P., et al. (2021b). Genome-wide identification of the tea plant bHLH transcription factor family and discovery of candidate regulators of trichome formation. *Sci. Rep.* 11, 1–13. doi: 10.1038/s41598-021-90205-7

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8

Major, I. T., Yoshida, Y., Campos, M. L., Kapali, G., Xin, X. F., Sugimoto, K., et al. (2017). Regulation of growth–defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module. *New Phytol.* 215, 1533–1547. doi: 10.1111/nph.14638

Minh, B. Q., Nguyen, M. A. T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024

Nakata, M., Mitsuda, N., Herde, M., Koo, A. J. K., Moreno, J. E., Suzuki, K., et al. (2013). A bHLH-type transcription factor, ABA-INDUCIBLE BHLH-TYPE TRANSCRIPTION FACTOR/JA-ASSOCIATED MYC2-LIKE1, acts as a repressor to negatively regulate jasmonate signaling in arabidopsis. *Plant Cell* 25, 1641–1656. doi: 10.1105/tpc.113.111112

Niu, Y., Figueroa, P., and Browse, J. (2011). Characterization of JAZ-interacting bHLH transcription factors that regulate jasmonate responses in arabidopsis. *J. Exp. Bot.* 62, 2143–2154. doi: 10.1093/jxb/erq408

Pan, H., Chen, Y., Zhao, J., Huang, J., Shu, N., Deng, H., et al. (2023). In-depth analysis of large-scale screening of WRKY members based on genome-wide identification. *Front. Genet.* 13. doi: 10.3389/fgene.2022.1104968

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in catharanthus roseus. *New Phytol.* 213, 1107–1123. doi: 10.1111/nph.14252

Peñuelas, M., Monte, I., Schweizer, F., Vallat, A., Reymond, P., García-Casado, G., et al. (2019). Jasmonate-related MYC transcription factors are functionally conserved in marchantia polymorpha. *Plant Cell* 31, 2491–2509. doi: 10.1105/tpc.18.00974

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Qi, T., Huang, H., Song, S., and Xie, D. (2015a). Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in arabidopsis. *Plant Cell* 27, 1620–1633. doi: 10.1105/tpc.15.00116

Qi, T., Wang, J., Huang, H., Liu, B., Gao, H., Liu, Y., et al. (2015b). Regulation of jasmonate-induced leaf senescence by antagonism between bHLH subgroup IIIe and IIId factors in arabidopsis. *Plant Cell* 27, 1634–1649. doi: 10.1105/tpc.15.00110

Sanseverino, W., Roma, G., De Simone, M., Faino, L., Melito, S., Stupka, E., et al. (2009). PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res.* 38, 814–821. doi: 10.1093/nar/gkp978

Sasaki-Sekimoto, Y., Jikumaru, Y., Obayashi, T., Saito, H., Masuda, S., Kamiya, Y., et al. (2013). Basic helix-loop-helix transcription factors JASMONATE-ASSOCIATED MYC2-LIKE1 (JAM1), JAM2, and JAM3 are negative regulators of jasmonate responses in arabidopsis. *Plant Physiol.* 163, 291–304. doi: 10.1104/pp.113.220129

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5593–E5601. doi: 10.1073/pnas.1419161111

Shoji, T. (2019). The recruitment model of metabolic evolution: jasmonate-responsive transcription factors and a conceptual model for the evolution of metabolic pathways. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00560

Song, C., Cao, Y., Dai, J., Li, G., Manzoor, M. A., Chen, C., et al. (2022a). The multifaceted roles of MYC2 in plants: toward transcriptional reprogramming and stress tolerance by jasmonate signaling. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.868874

Song, S., Huang, H., Wang, J., Liu, B., Qi, T., and Xie, D. (2017). MYC5 is involved in jasmonate-regulated plant growth, leaf senescence and defense responses. *Plant Cell Physiol.* 58, 1752–1763. doi: 10.1093/pcp/pcx112

Song, C., Jiao, C., Jin, Q., Chen, C., Cai, Y., and Lin, Y. (2020). Metabolomics analysis of nitrogen-containing metabolites between two dendrobium plants. *Physiol. Mol. Biol. Plants* 26, 1425–1435. doi: 10.1007/s12298-020-00822-1

Song, C., Li, G., Dai, J., and Deng, H. (2021a). Genome-wide analysis of PEBP genes in dendrobium huoshanense: unveiling the antagonistic functions of FT/TFL1 in flowering time. *Front. Genet.* 12. doi: 10.3389/fgene.2021.687689

Song, S., Qi, T., Fan, M., Zhang, X., Gao, H., Huang, H., et al. (2013). The bHLH subgroup IIId factors negatively regulate jasmonate-mediated plant defense and development. *PloS Genet.* 9, 1–20. doi: 10.1371/journal.pgen.1003653

Song, C., Wang, Y., Manzoor, M. A., Mao, D., Wei, P., Cao, Y., et al. (2022b). In-depth analysis of genomes and functional genomics of orchid using cutting-edge high-throughput sequencing. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1018029

Song, M., Wang, H., Wang, Z., Huang, H., Chen, S., and Ma, H. (2021b). Genome-wide characterization and analysis of bHLH transcription factors related to anthocyanin biosynthesis in fig (Ficus carica l.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.730692

Song, C., Zhang, Y., Manzoor, M. A., and Li, G. (2022c). Identification of alkaloids and related intermediates of dendrobium officinale by solid-phase extraction coupled with high-performance liquid chromatography tandem mass spectrometry. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.952051

Stitz, M., Gase, K., Baldwin, I. T., and Gaquerel, E. (2011). Ectopic expression of at JMT in nicotiana attenuata: creating a metabolic sink has tissue-specific consequences for the jasmonate metabolic network and silences downstream gene expression. *Plant Physiol.* 157, 341–354. doi: 10.1104/pp.111.178582

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Thiel, J., Koppolu, R., Trautewig, C., Hertig, C., Kale, S. M., Erbe, S., et al. (2021). Transcriptional landscapes of floral meristems in barley. *Sci. Adv.* 7, 1–15. doi: 10.1126/sciadv.abf0832

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Wang, P., Jin, S., Chen, X., Wu, L., Zheng, Y., Yue, C., et al. (2021). Chromatin accessibility and translational landscapes of tea plants under chilling stress. *Hortic. Res.* 8, 1–15. doi: 10.1038/s41438-021-00529-8

Wang, Y., Liu, H., Zhu, D., Gao, Y., Yan, H., and Xiang, Y. (2017). Genome-wide analysis of VQ motif-containing proteins in moso bamboo (Phyllostachys edulis). *Planta* 246, 165–181. doi: 10.1007/s00425-017-2693-9

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, 1–14. doi: 10.1093/nar/gkr1293

Wang, K., Wang, D., Zheng, X., Qin, A., Zhou, J., Guo, B., et al. (2019). Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. *Nat. Commun.* 10, 1–15. doi: 10.1038/s41467-019-12575-x

Wu, M., Ding, X., Fu, X., and Lozano-Duran, R. (2019). Transcriptional reprogramming caused by the geminivirus tomato yellow leaf curl virus in local or systemic infections in nicotiana benthamiana. *BMC Genomics* 20, 1–17. doi: 10.1186/s12864-019-5842-7

Yan, J., Li, S., Gu, M., Yao, R., Li, Y., Chen, J., et al. (2016). Endogenous bioactive jasmonate is composed of a set of (+)-7-iso-JA-amino acid conjugates. *Plant Physiol.* 172, 2154–2164. doi: 10.1104/pp.16.00906

Zander, M., Lewsey, M. G., Clark, N. M., Yin, L., Bartlett, A., Saldierna Guzmán, J. P., et al. (2020). Integrated multi-omics framework of the plant response to jasmonic acid. *Nat. Plants* 6, 290–302. doi: 10.1038/s41477-020-0605-7

Zhang, Z., Chen, J., Liang, C., Liu, F., Hou, X., and Zou, X. (2020). Genome-wide identification and characterization of the bHLH transcription factor family in pepper (Capsicum annuum l.). *Front. Genet.* 11. doi: 10.3389/fgene.2020.570156

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in catharanthus roseus. *Plant J.* 67, 61–71. doi: 10.1111/j.1365-313X.2011.04575.x

Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R., Hsiao, Y. Y., Niu, S. C., et al. (2017). The apostasia genome and the evolution of orchids. *Nature* 549, 379–383. doi: 10.1038/nature23897

Zhao, W., Liu, Y., Li, L., Meng, H., Yang, Y., Dong, Z., et al. (2021). Genome-wide identification and characterization of bHLH transcription factors related to anthocyanin biosynthesis in red walnut (Juglans regia l.). *Front. Genet.* 12. doi: 10.3389/fgene.2021.632509

Zheng, Y., Lan, Y., Shi, T., and Zhu, Z. (2017). Diverse contributions of MYC2 and EIN3 in the regulation of arabidopsis jasmonate-responsive gene expression. *Plant Direct* 1, 1–8. doi: 10.1002/pld3.15

Zhou, X., Liao, Y., Kim, S. U., Chen, Z., Nie, G., Cheng, S., et al. (2020). Genome-wide identification and characterization of bHLH family genes from ginkgo biloba. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-69305-3

Zhu, Y., Meng, C., Zhu, L., Li, D., Jin, Q., Song, C., et al. (2017). Cloning and characterization of DoMYC2 from dendrobium officinale. *Plant Cell. Tissue Organ Cult.* 129, 533–541. doi: 10.1007/s11240-017-1198-3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fpls.2023.1196176

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
CAS, China

REVIEWED BY
Aiping Song,
Nanjing Agricultural University, China
Zhitao Niu,
Nanjing Normal University, China
Li Feng,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Rui-Sen Lu
lurs@cnbg.net

RECEIVED 29 March 2023
ACCEPTED 26 April 2023
PUBLISHED 06 June 2023

CITATION
Hu K, Sun X-Q, Chen M and Lu R-S (2023) Low-coverage whole genome sequencing of eleven species/subspecies in *Dioscorea* sect. *Stenophora* (Dioscoreaceae): comparative plastome analyses, molecular markers development and phylogenetic inference.
Front. Plant Sci. 14:1196176.
doi: 10.3389/fpls.2023.1196176

COPYRIGHT
© 2023 Hu, Sun, Chen and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Low-coverage whole genome sequencing of eleven species/subspecies in *Dioscorea* sect. *Stenophora* (Dioscoreaceae): comparative plastome analyses, molecular markers development and phylogenetic inference

Ke Hu[1,2,3], Xiao-Qin Sun[1,2,3], Min Chen[1,2,3] and Rui-Sen Lu[1,2,3]*

[1]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, [2]Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Nanjing, China, [3]Jiangsu Provincial Science and Technology Resources Coordination Platform (Agricultural Germplasm Resources) Germplasm Resources Nursery of Medicinal Plants, Nanjing, China

Dioscorea sect. *Stenophora* (Dioscoreaceae) comprises about 30 species that are distributed in the temperate and subtropical regions of the Northern Hemisphere. Despite being evolutionarily "primitive" and medically valuable, genomic resources and molecular studies of this section are still scarce. Here, we conducted low-coverage whole genome sequencing of 11 *Stenophora* species/subspecies to retrieve their plastome information (whole plastome characteristics, plastome-divergent hotspots, plastome-derived SSRs, etc.) and polymorphic nuclear SSRs, as well as performed comparative plastome and phylogenetic analyses within this section. The plastomes of *Stenophora* species/subspecies ranged from 153,691 bp (*D. zingiberensis*) to 154,149 bp (*D. biformifolia*) in length, and they all contained the same 114 unique genes. All these plastomes were highly conserved in gene structure, gene order and GC content, although variations at the IR/SC borders contributed to the whole length differences among them. The number of plastome-derived SSRs among *Stenophora* species/subspecies varied from 74 (*D. futschauensis*) to 93 (*D. zingiberensis*), with A/T found to be the most frequent one. Seven highly variable regions and 12 polymorphic nuclear SSRs were identified in this section, thereby providing important information for further taxonomical, phylogenetic and population genetic studies. Phylogenomic analyses based on whole plastome sequences and 80 common protein coding genes strongly supported *D. biformifolia* and *D. banzhuana* constituted the successive sister species to the remaining sampled species, which could be furtherly divided into three clades. Overall, this study provided a new perspective for plastome

evolution of *Stenophora*, and proved the role of plastome phylogenomic in improving the phylogenetic resolution in this section. These results also provided an important reference for the protection and utilization of this economically important section.

KEYWORDS

***Dioscorea* sect. *Stenophora*, plastome, comparative analyses, hotspots, phylogeny**

1 Introduction

Dioscorea is the largest genus in the family Dioscoreaceae with over 600 species, which contains about ten major clades: *Stenophora*, New World I, New World II, African, Mediterranean, New World III, Compound Leaved, Malagasy, *Shannicorea* and *Enantiophyllum* (Viruel et al., 2016; Couto et al., 2018; Viruel et al., 2018). Among them, *Stenophora*, the subject of our study, coincides with the section *Stenophora* Uline circumscribed by Burkill (1960), and contains about 30 species disjunctively distributed in the Northern Hemisphere (Gao et al., 2008; Vinogradova et al., 2022). The sect. *Stenophora* likely originated in Himalayas-Hengduan Mountains, China, and is the most basal clade of *Dioscorea*, differing from the rest of this genus by having rhizomes, monosulcate pollen and a diploid chromosome number (x = 10) (Wilkin et al., 2005; Gao et al., 2008; Hsu et al., 2013; Viruel et al., 2016; Couto et al., 2018; Noda et al., 2020). Plant species in this section are reported to have great medicinal values. In particular, the rhizomes of *D. nipponica* and *D. zingiberensis* are extensively used to extract diosgenin, which is an important precursor for the synthesis of steroid drugs in the pharmaceutical industry (Gong et al., 2011; Cheng et al., 2021). The immense evolutionary and medicinal value of *Stenophora* species has also brought new challenges to their conservation and sustainable use. One major concern is that the increasing demands for naturally growing plants has threaten their wild populations and genetic variations (Das et al., 2013; Sun et al., 2017). For example, *D. nipponica* has been listed as a secondary-level endangered plant species in China, as its wild resource is facing extinction (Fu, 1992; Chen et al., 2007). Another concern is the misidentification and misuse of *Stenophora* species, as they are similar in morphological characteristics and local names. Therefore, accurate and rapid identification of *Stenophora* species (e.g., molecular markers) is urgently required.

Previous studies of *Stenophora* have concentrated on external morphology, cytology, pollen morphology and phytochemistry (Pei et al., 1979; Huang et al., 2010). Morphological and embryological features have been shown to be important for systematics and species identification of *Stenophora*, and could divide this section into several subclades (e.g., Titova and Torshilova, 2015; Vinogradova et al., 2022), however, it is still difficult to find clear gaps of morphological variations among closely related species (Noda et al., 2020). Moreover, although previous molecular-based studies have accelerated species identification and phylogenetic inference of *Stenophora*, the plastid loci used (e.g., *atpB*, *matK*, *rbcL*) always showed low discriminatory power (Gao et al., 2008; Noda et al., 2020). For example, Gao et al. (2008) revealed that *D. nipponica* was sister to *D. althaeoides*, and *D. nipponica* ssp. *rosthornii* was not related to these two species, but their interspecific relationships receive weak bootstrap support. Evidently more effective molecular markers are needed to solve the remaining phylogenetic dilemma.

Plastomes of land plants generally have a quadripartite circular structure, with a pair of inverted repeats (IRs) separated by a large single-copy (LSC) region and a small single-copy (SSC) region, ranging from 100 to 200 kb in length (Raubeson and Jansen, 2005; Jansen and Ruhlman, 2012; Olejniczak et al., 2016; Lu et al., 2023). Due to many advantages such as highly conserved structure, usually uniparental inheritance, absence of recombination, and large copy numbers, plastome sequences have been widely used for accurate species identification and phylogenetic inferences, especially at low taxonomic levels (Gitzendanner et al., 2018; Yang et al., 2022). Furthermore, comparative plastome genomics could provide essential information for plastome evolution, such as gene loss and IR boundary shifts, and can develop mutational hotspots, which may contribute to species discrimination, phylogenetic, and population genetic studies (Lu et al., 2021; Yang et al., 2022). In sect. *Stenophora*, although some representative plastomes have been sporadically released, previous studies mainly focused on the plastome characteristics of single species (e.g., Wu et al., 2016; Zhou et al., 2016), or performed comparative and phylogenetic analyses only based on a small number of plastomes (e.g., Zhao et al., 2018; Xia et al., 2019). Thus, it is necessary to provide more genomic resources for further understanding the plastome evolution and phylogeny of *Stenophora*.

With the rapid development of next generation sequencing (NGS) technologies, it is cheap and fast to obtain low-coverage (~0.1–10×) of the whole genome sequencing data (or called genome skimming data), which could provide sufficient data for complete plastome assemblies (Straub et al., 2012; Twyford and Ness, 2017; Jin et al., 2020). Besides, the assembled nuclear scaffolds from low-coverage whole genome sequencing data could be used for mining polymorphic nuclear SSRs (nSSRs) (e.g., Liu et al., 2018; Xia et al., 2018; Lu et al., 2022). Here, we performed low-coverage whole genome shotgun sequencing for 11 *Stenophora* species/subspecies (i.e., *D. banzhuana* Pei & Ting, *D. biformifolia* Pei & Ting, *D.*

collettii Hook.f., *D. deltoidea* Wall., *D. futschauensis* Uline ex R.Knuth, *D. gracillima* Miq., *D. nipponica* Makino, *D. nipponica* subsp. *rosthorni* (Prain & Burkill) C.T.Ting, *D. spongiosa* J.Q.Xi, M.Mizuno & W.L.Zhao, *D. tokoro* Makino, *D. zingiberensis* C.H.Wright). Using this data, we aimed to i) present the complete and annotated plastome sequences of these 11 *Stenophora* species/subspecies, and assess plastome structural evolution of them; ii) identify plastomic SSRs and mutational hotspot regions (plastome-derived markers); iii) develop polymorphic nSSRs based on assembled nuclear scaffolds of *Stenophora* species/subspecies; and iv) conduct phylogenetic analyses of these species/subspecies using plastome data. Overall, this study will not only provide a valuable resource for species identification and phylogenetic studies of *Stenophora*, but also be useful for conservation and utilization of this economically important section.

2 Materials and methods

2.1 Plant materials, DNA extraction and genomic sequencing

Fresh leaves of 11 *Stenophora* species/subspecies, i.e., *D. banzhuana*, *D. biformifolia*, *D. collettii*, *D. deltoidea*, *D. futschauensis*, *D. gracillima*, *D. nipponica*, *D. nipponica* subsp. *rosthorni*, *D. spongiosa*, *D. tokoro*, *D. zingiberensis*, were field-collected and dried with silica-gel. The voucher specimens were deposited at Herbarium of Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (NAS) [details about sampling information can be found in Gao et al. (2008)]. For each species/subspecies, genomic DNA was extracted from silica gel-dried leaves using DNAsecure Plant Kit (Tiangen Biotech, Beijing, China), following the manufacturer's protocol. DNA concentration and integrity were measured by Agilent 2100 BioAnalyzer and agarose gel electrophoresis. Paired-end library with insert size of 350 bp was constructed for each species/subspecies by using the Genomic DNA Sample Prep, and then sequenced on the Illumina HiSeqTM 4000 platform (Illumina, San Diego, California, USA) according to the paired-end 2 × 150 bp protocol. Library construction, genome sequencing and raw data filtering were conducted by Novogene Bioinformatics Technology Co., Ltd., Beijing, China.

2.2 Plastome assembly and annotation

After removing library barcodes and filtering low-quality data, the clean reads (about 4 Gb per sample) were used for *de novo* assembly of whole plastome sequences using GetOrganelle v.1.7.6 (Jin et al., 2020), with the default parameters as suggested by its authors. All the assembly graphs were subsequently visually inspected using Bandage v.0.8.1 (Wick et al., 2015). Initial annotations of all newly assembled plastomes were performed with MAFFT v.7 plugin (Katoh and Standley, 2013) in Geneious Prime® 2022.0.1 (https://www.geneious.com) by aligning them to two closely related and previously published plastomes, i.e., *D. aspersa* (NC_039807) and *D. collettii* (NC_037717), and transferring reference annotations to these newly assembled plastomes. Then, the initial annotations were checked and adjusted manually to confirm the accuracy of exon/intron boundaries and start/stop codon locations. All newly generated plastome sequences were deposited in GenBank (accession numbers: OQ525992-OQ526002). High-resolution circular plastome maps of these 11 *Stenophora* species/subspecies were generated using the web-based tool OrganellarGenomeDRAW (OGDRAW) v.1.3.1 (Greiner et al., 2019).

2.3 Whole plastome comparison within sect. *Stenophora*

To visualize sequence similarity of plastomes within sect. *Stenophora*, all 11 newly sequenced plastomes and one plastome of *D. villosa* (NC_034686) were aligned using the global alignment program LAGAN (Brudno et al., 2003), and visualized in VISTA browser (Frazer et al., 2004), taking annotations of *D. villosa* plastome as reference. To further illustrate the IR expansions and contractions among *Stenophora* plastomes, the four junctions between two invert repeats (IRs) and large/small single copy (LSC/SSC) regions were identified and compared by Repeat Finder plugin as implemented in Geneious Prime® 2022.0.1 (https://www.geneious.com/plugins/repeat-finder/).

2.4 Identification of mutational hotspots and plastome-derived SSRs

To identify mutational hotspot regions for PCR-based identification of *Stenophora* species and subspecies, a total of 12 plastomes (one plastome per species/subspecies, Table 1) were first aligned using the MAFFT v.7 plugin (Katoh and Standley, 2013) in Geneious Prime® 2022.0.1. Then, protein coding sequences (CDS), intergenic spacer regions (IGS), introns and tRNAs, with aligned length > 200 bp and the total number of mutations > 0 were extracted from the alignment matrix of these 12 plastome sequences. Finally, the nucleotide diversity (π) of these regions was calculated in DnaSP v.6.12.03 (Rozas et al., 2017). In addition, the MISA-web application (Beier et al., 2017) was employed to screen SSRs across the 12 *Stenophora* plastomes, with thresholds (minimum numbers) of 10, 5, 4, 3, 3, and 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide SSRs, respectively.

2.5 Development of polymorphic nuclear SSRs

To develop polymorphic nuclear SSRs in sect. *Stenophora*, low-coverage whole genome sequence reads of these 11 species/subspecies were aligned to the reference genome sequence of D. *zingiberensis* (Cheng et al., 2021) to remove mitochondria and plastome reads, using BWA-MEM v.0.7.17 (Li, 2013). Aligned files were then sorted using SAMtools v.1.9 (Li et al., 2009). The

TABLE 1 The basic features of 11 *Stenophora* plastomes newly generated in this study.

Species/Subspecies	Length (bp)					GC content (%)	Number of genes				
	Total	LSC	SSC	IR	PCGs		Total	CDS	tRNA	rRNA	Duplicated
Dioscorea banzhuana	153,989	84,004	18,959	25,513	79,071	37.20%	133	87	38	8	19
Dioscorea biformifolia	154,149	84,145	18,856	25,574	78,729	37.20%	133	87	38	8	19
Dioscorea collettii	153,746	83,903	18,657	25,593	78,867	37.20%	133	87	38	8	19
Dioscorea deltoidea	153,947	83,969	18,920	25,529	78,771	37.20%	133	87	38	8	19
Dioscorea futschauensis	153,948	83,981	18,909	25,529	78,759	37.20%	133	87	38	8	19
Dioscorea gracillima	153,996	83,970	18,908	25,559	78,777	37.20%	133	87	38	8	19
Dioscorea nipponica	153,885	83,950	18,919	25,508	78,780	37.20%	133	87	38	8	19
Dioscorea nipponica subsp. *rosthornii*	153,916	83,981	18,919	25,508	78,780	37.20%	133	87	38	8	19
Dioscorea spongiosa	153,947	83,969	18,920	25,529	78,771	37.20%	133	87	38	8	19
Dioscorea tokoro	153,946	83,968	18,920	25,529	78,765	37.20%	133	87	38	8	19
Dioscorea villosa	153,974	83,920	18,902	25,576	78,720	37.20%	133	87	38	8	19
Dioscorea zingiberensis	153,691	83,129	18,918	25,822	78,867	37.20%	133	87	38	8	19

resultant Binary Alignment/Map (BAM) data (only containing nuclear reads) were *de novo* assembled into scaffolds using a de Bruijn graph-based assembly program, SOAPdenovo v.1.0.4 (Xie et al., 2014). Based on these nuclear scaffolds, the potential polymorphic nuclear SSRs were identified using CandiSSR (Xia et al., 2016), with default parameters.

2.6 Phylogenetic analyses

Phylogenetic relationships among the 12 species/subspecies of sect. *Stenophora* (Table 1) were inferred based on two datasets: (1) whole plastome sequences and (2) 80 shared protein coding genes, taking *D. aspersa* (NC_039807) and *D. alata* (NC_039707) as outgroups. For the latter dataset, three partitioning scenarios: (1) unpartitioned scenarios; (2) partitioned by each gene and intergenic region; and (3) partitioned by codon position were employed. Both whole plastome sequences and protein coding sequences were aligned using the MAFFT v.7 plugin (Katoh and Standley, 2013) in Geneious Prime® 2022.0.1. The best nucleotide substitution model was determined by the Akaike Information Criterion (AIC) in jModelTest v2.1.4 (Darriba et al., 2012), and the GTR + G substitution model was selected for both datasets. Maximum likelihood (ML) analyses were performed using RAxML v.8.2.12 (Stamatakis, 2014) available in the CIPRES Science Gateway v.3.3 (http://www.phylo.org/portal2/). Clade support values were estimated by 1000 bootstrap replicates. Bayesian inference (BI) analyses were conducted on MrBayes v.3.2.7 (Ronquist et al., 2012), which consists of two independent runs of 1×10^6 generations, with four independent Markov chain Monte Carlo (MCMC) chains (i.e., one cold and three heated) each, and a sampling frequency of 1000 generations. The first 200 trees were discarded as 'burn-in', and the remaining trees were used to construct a majority-rule consensus tree and estimate posterior probabilities (PPs).

3 Results and discussion

3.1 Plastome characteristics

The whole length of these *Stenophora* plastomes ranged from 153,691 bp (*D. zingiberensis*) to 154,149 bp (*D. biformifolia*) (Figure 1; Table 1). All these plastomes shared the typical quadripartite structure of angiosperm plastomes, with a pair of IR regions (25,508–25,822 bp) separated by the LSC (83,129–84,145 bp) and SSC (18,657–18,959 bp) regions. The length variation of *Dioscorea* plastomes is a very common phenomenon (Zhao et al., 2018; Xia et al., 2019), which is often caused by the expansion and contraction of the IR regions (see details below). These *Stenophora* plastomes have the same overall GC content (37.20%), higher than that in LSC (35.00–35.10%) and SSC (31.2%) regions, but lower than that in IR regions (43.29–43.0%) (Table 1), possibly influenced by the high GC content (55.3%) of the four ribosomal RNA (rRNA) sequences.

All these *Stenophora* plastomes encoded the same 114 unique genes, including 80 protein-coding genes (PCGs), 30 transfer RNA (tRNA) genes, and four rRNA genes. Nineteen unique genes (seven PCGs, eight tRNA genes, and all four rRNA) were duplicated in the IRs, giving a total of 133 genes (Figure 1; Table S1). Among these unique genes, nine PCGs (i.e., *atpF*, *petB*, *petD*, *ndhA*, *ndhB*, *rpoC1*, *rpl2*, *rpl16*, and *rps16*) and six tRNAs (*trnK*-UUU, *trnG*-UCC, *trnL*-UAA, *trnV*-UAC, *trnI*-GAU and *trnA*-UGC) possessed a single intron, while three PCGs (*ycf3*, *rps12* and *clpP*) contained two introns (Figure 1; Table S1). The *rps12* gene consists of three exons that were trans-spliced together: exon 1 was located in the LSC

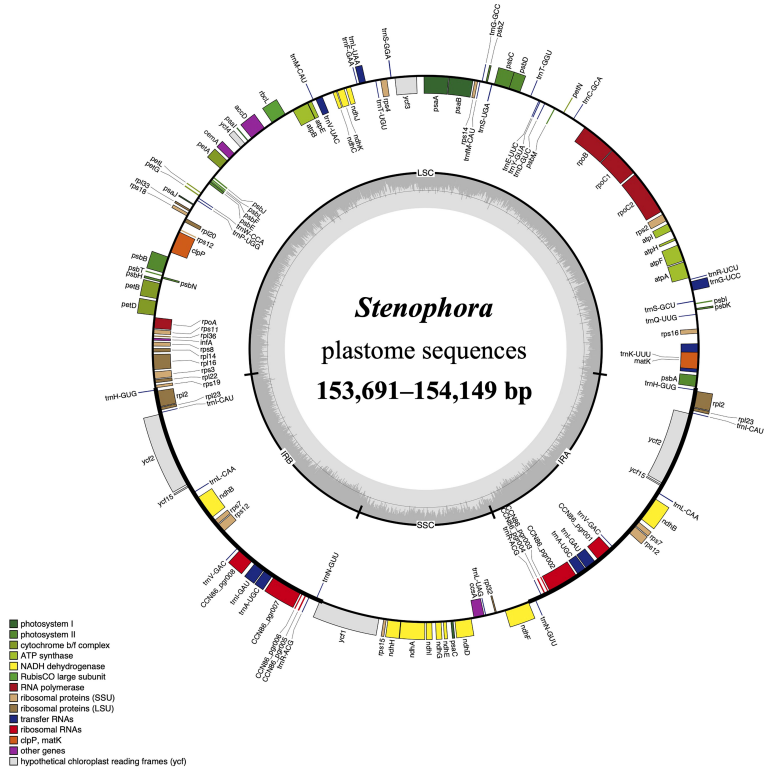


FIGURE 1
The plastome map of *Stenophora* species/subspecies. Thick lines on the outer complete circle identify the inverted repeat regions (IRa and IRb). Genes shown on the outside of the circle are transcribed clockwise, while genes inside are transcribed counter-clockwise. Genes are color coded according to their functional groups. GC/AT content is displayed by darker/lighter grey bars in the inner ring.

region, while exons 2 and 3 were proximal and located in the IR regions (Figure 1; Table S1). Furthermore, all *Stenophora* plastomes reported in this study harbored the complete *rps16* gene, contrary to previous studies indicating the entire loss of *rps16* gene in several clades of *Dioscorea* (Jansen et al., 2007; Lu et al., 2023). To further improve our understanding of *rps16* gene evolution in *Dioscorea*, a ML phylogenetic tree was reconstructed (with the same method above) based on whole plastome sequences of 42 *Dioscorea* species, using *Trichopus zeylanicus* and *Tacca leontopetaloides* as outgroups. Phylogenetic result showed that *rps16* gene was lost in all other *Dioscorea* clades except *Stenophora*, suggesting a single loss of this gene within *Dioscorea* (Figure S1). Since *Stenophora* and the rest of the genus diverged about 48.3 (47.6–49.1) million years ago (Mya) (Viruel et al., 2016), the gene loss mentioned above may have occurred in sync with this divergence event, implying that the loss of *rps16* gene may have occurred about 48.3 Mya.

3.2 Comparative plastome analyses of *Stenophora*

Comprehensive comparison of 12 *Stenophora* plastomes revealed a high degree of overall sequence similarity and collinearity within this section (Figure 2). Similar to previous monocot plastome studies (e.g., Asaf et al., 2017; Lu et al., 2017; Lu et al., 2021; Lu et al., 2022), our mVISTA analysis demonstrated that IRs exhibited a lower level of sequence divergence compared with LSC and SSC regions (Figure 2). This could be attributed to copy correction between IR sequences by gene conversion, and the abundance of conserved rRNA genes in the IRs (Khakhlova and Bock, 2006). In addition, the protein-coding regions were found to be more conserved than non-coding regions (including intergenic spacers and introns), which were likely to be subject to natural selection (Shaw et al., 2007; Lu et al., 2022).

Despite the similarity of plastome sequences, and the conservation of gene content and linear order of genes, the 12 *Stenophora* plastomes exhibited obvious differences at the IR/SC borders (Figure 3). For example, the *ndhF* gene crossed the SSC/IRa border in *D. collettii* and *D. zingiberensis*, while it was completely included in the SSC region in the other 10 species (Figure 3). The IRb region extended 238 bp into the *rps19* gene in *D. zingiberensis*, much deeper than those extended into all other species (2–22 bp), and further extended 280–296 bp into the *ycf1* gene. IR expansion into *rps19* gene has also been observed in other sections in the genus *Dioscorea*, e.g., *Opsophyton*, *Testudinaria*, *Enantiophyllum* (Zhao et al., 2018; Lu et al., 2023), suggesting that this phenomenon may be an ancestral symplesiomorphy of the genus *Dioscorea*. In addition, the *trnH* gene was totally located within the IR region and duplicated in all these species, 142–378 bp away from its proximal IR/SC border.

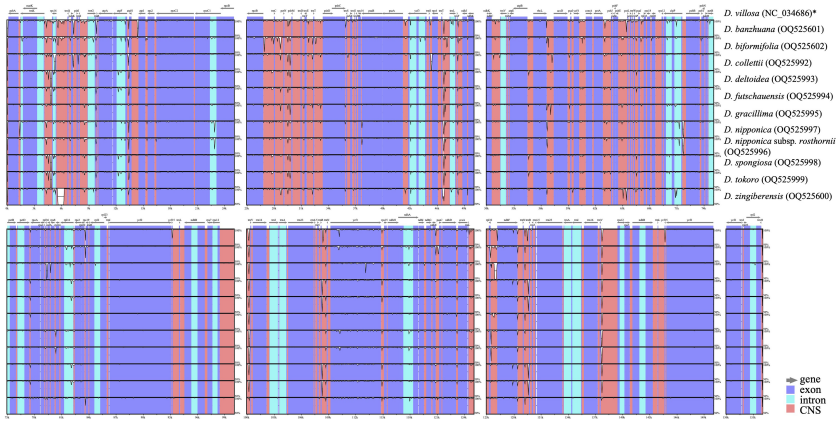


FIGURE 2
Sequence identity plots among *Stenophora* plastomes, with *D. villosa* (NC_034686) as a reference. Annotated genes are shown along the top. Gray arrows above the alignment indicates genes with their orientation. The vertical scale represents the percent identity between 50% and 100%. Genome regions are color coded as exon, intron, and conserved non-coding sequences (CNS). * previously published plastome sequence.

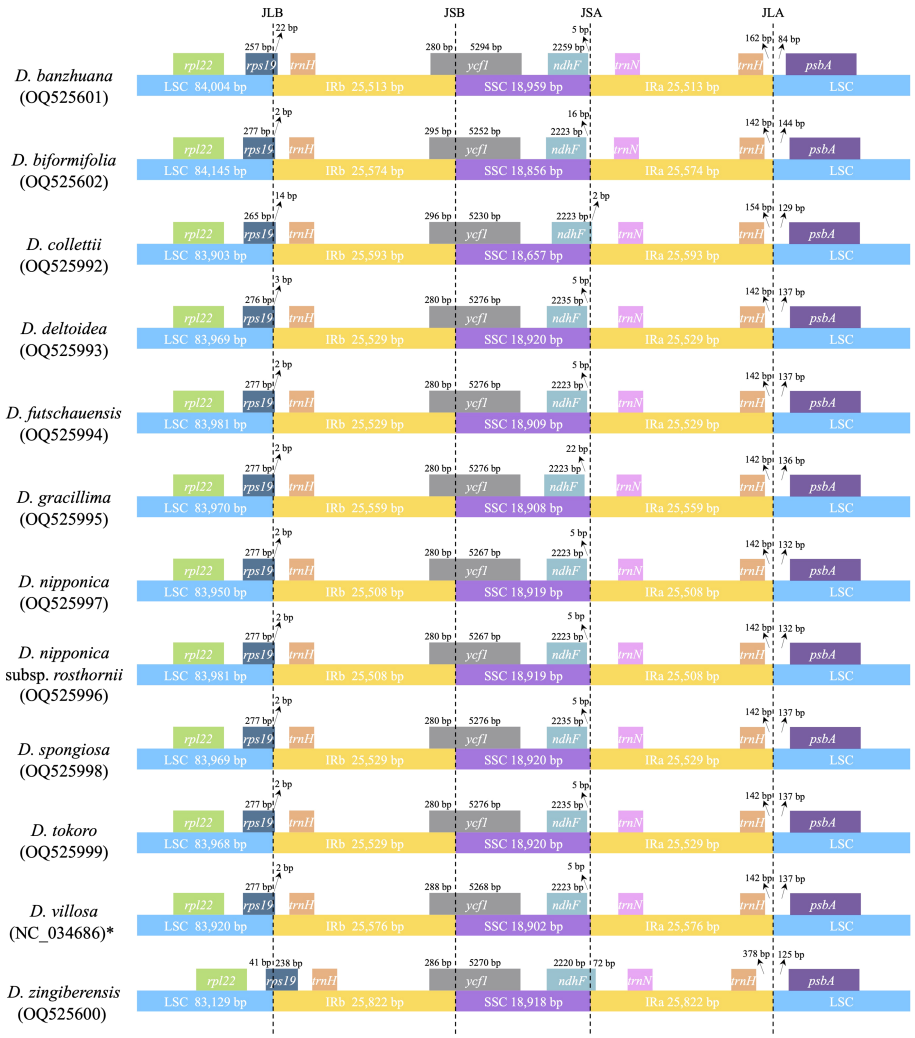


FIGURE 3
Comparison of IR/SC junctions among 12 *Stenophora* plastomes. * previously published plastome sequence.

3.3 Plastome-derived hotspot regions and SSRs for *Stenophora*

Morphology-based species identification in sect. *Stenophora* has always been difficult, because it is challenging to find clear gaps of morphological variations among closely related species (Kawabe et al., 1997; Wilkin et al., 2005; Gao et al., 2008). In this case, barcoding has been performed for this section using nuclear gene phosphoglucose isomerase (PGI) and plastid DNA (*matK*, *rbcL* and *trnL-F*) regions (Kawabe et al., 1997; Gao et al., 2008). However, these markers are today considered intermediately variable regions (Shaw et al., 2014), and always showed low species discriminatory power and poor phylogenetic resolution (Gao et al., 2008; Noda et al., 2020). Therefore, we here used these plastome sequences to develop novel genetic markers (hypervariable regions) for taxonomic and phylogenetic analysis of *Stenophora*. A total of 130 regions (58 CDS, 53 IGS, 13 introns, five tRNAs and one rRNA) was eventually extracted to calculate the nucleotide diversity, and the π values ranged from 0.01% (*rrn16*) to 3.36% (*ndhD–ccsA*) (Figure 4). Six IGS regions (i.e., *ndhD-ccsA*, *petA-psbJ*, *trnL-rpl32*, *psbZ-trnG*, *trnD-trnY* and *rpl32-ndhF*), and *rps16* intron sequence were the top seven highly variable regions (π > 1.00%) (Figure 4), which could be served as section-specific molecular markers for future identification, conservation and utilization of *Stenophora* species.

Plastome-derived SSRs (chloroplast SSRs, cpSSRs) are scattered in the plastomes across different plant taxa, and have been widely used in population genetic studies and breeding programs (Jiménez, 2010; Chmielewski et al., 2015; Hazra et al., 2021; Ping et al., 2021). In this study, the MISA analysis identified a total of 960 SSRs across the 12 *Stenophora* plastomes. The number of SSRs for each plastome ranged from 74 (*D. futschauensis*) to 93 (*D. zingiberensis*). Mononucleotide repeats were predominant, with numbers ranging from 35 (*D. collettii*) to 50 (*D. biformifolia*), followed by dinucleotides ranging from 14 (*D. banzhuana*) to 18 (*D. villosa*), and tetranucleotides (10 in *D. collettii* and 9 in the other 11 plastomes), while trinucleotides (4–6 per plastome), pentanucleotides (0–5 per plastome) and hexanucleotides (3 per plastome) were relatively few in *Stenophora* plastomes (Figure 5; Table S2). The most common motifs were A/T and AT/TA for mono- and dinucleotides, accounting for 46.75%–68.76% and 17.20%–22.97% of the total SSRs in *Stenophora* plastomes, respectively, which may lead to the AT richness of the *Stenophora* plastomes (Figure 5; Table S2). These results were also consistent with the previous findings that plastome-based SSRs are largely composed of short polyadenine (polyA) and polythymine (polyT) repeats, while rarely contained tandem guanine (G) and cytosine (C) repeats (Kuang et al., 2011; Lu et al., 2022). In addition, several potential species-specific SSRs were identified in the present study. For example, AGC/CTG and AAGTAT/ACTTAT were only observed in *D. collettii* and *D. biformifolia*, respectively, but not appeared in the other 10 species/subspecies. Both AATAG/ATTCT and AATAT/ATATT were only presented in *D. nipponica* and its subspecies, while absence in other 10 species (Figure 5; Table S2). Clearly, these SSRs could be developed as effective molecular markers for species identification.

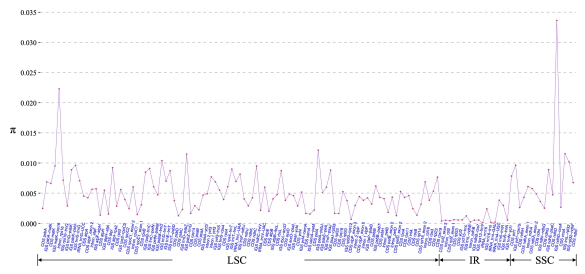


FIGURE 4
Nucleotide variability (π) values of 130 regions (58 CDS, 53 IGS, 13 introns, five tRNAs and one rRNA) extracted from the alignment matrix of 12 *Stenophora* plastome sequences.

3.4 Polymorphic nuclear SSRs of *Stenophora*

Unlike plastome-derived SSRs with a certain degree of conservatism and usually uniparental inheritance, nuclear SSR markers are co-dominant and generally highly polymorphic, thus can complement plastome-derived SSR analysis in plants (Aecyo et al., 2021). In this study, based on the multiple assembled nuclear scaffolds of 11 newly sequenced *Stenophora* species/subspecies, a total of 12 polymorphic nSSRs (including six dinucleotides and six trinucleotides) were determined within this section by using CandiSSR (Table S3). Among these polymorphic nSSRs, nSSR_7 could divide these *Stenophora* species/subspecies into five groups, while four nSSRs (i.e., nSSR_1, nSSR_2, nSSR_5, nSSR_6) and seven nSSRs (i.e., nSSR_3, nSSR_4, nSSR_8, nSSR_9, nSSR_10, nSSR_11, nSSR_12) could divide them into four and three groups, respectively (Table S3). Apparently, these polymorphic nSSRs would be useful for species identification and conservation of this section, especially in the population genetic context.

3.5 Phylogenetic relationships within *Stenophora*

Previous phylogenetic studies have laid an important foundation for the phylogeny and classification of *Stenophora* species, however the selected loci (e.g., *atpB*, *matK*, *rbcL* and *trnL-F*) unfortunately could not provide sufficient information for elucidating the phylogenetic and evolutionary relationships among them (Gao et al., 2008; Viruel et al., 2016; Noda et al., 2020). Recently, plastome sequences have been extensively used for phylogenetic analyses, especially in addressing unresolved relationship at low taxonomic levels (Carbonell-Caballero et al., 2015; Li et al., 2017). In this study, two datasets including the complete plastome sequences and 80 commonly present protein-coding genes of 12 *Stenophora* species/subspecies were used to perform phylogenetic analyses, with *D. aspersa* and *D. alata* as outgroups. Both ML and BI analyses of these two datasets (including different partitioning scenarios on protein-coding genes) yielded the same topology, with moderate to high bootstrap support values (BS =

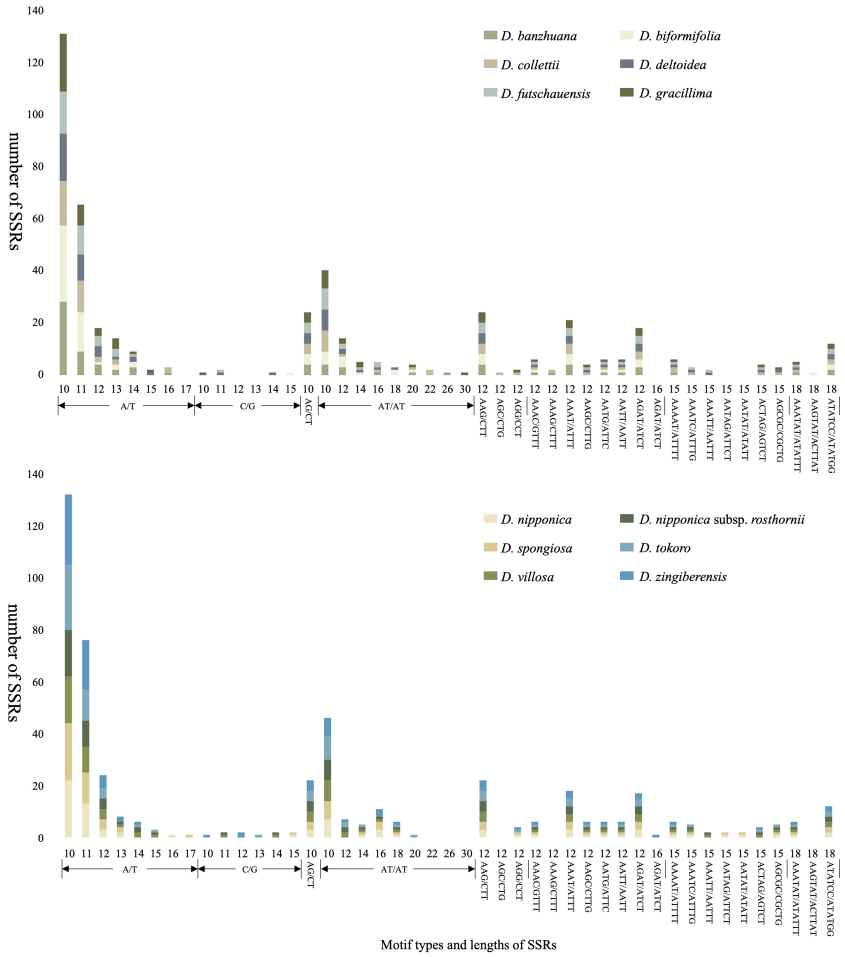


FIGURE 5
Plastome-derived SSRs in the 12 *Stenophora* species/subspecies.

65–100) and maximal posterior probability support values (PP = 1.0) at each node (Figure 6). The topology strongly supported *D. biformifolia* and *D. banzhuana* constituted the successive sister species to the rest. The remaining 10 sampled species/subspecies within this section could be further divided into three clades. Clade I contained four species, in which *D. futschauensis* was sister to the clade of *D. tokoro* + (*D. deltoidea* + *D. spongiosa*). Clade II, i.e., (*D. nipponica* + *D. nipponica* subsp. *rosthornii*) + *D. zingiberensis*, and clade III, i.e., (*D. gracillima* + *D. villosa*) + *D. collettii* were sister to each other, and jointly sister to Clade I (Figure 6). Contrary to previous hypothesis that *D. nipponica* ssp. *rosthornii* was not related to *D. nipponica* (Gao et al., 2008), our study strongly supported the monophyly of *D. nipponica* and *D. nipponica* ssp. *rosthornii*, which was consistent to the taxonomic treatments of this species in Flora of China (Ting et al., 2000).

Although our analyses have demonstrated the power of plastome phylogenetics to improve the resolutions of phylogenetic relationships in sect. *Stenophora*, this study was conducted based on insufficient taxa sampling, thus could not establish a complete picture of phylogenetic relationships within this section. Also considering that hybridization and polyploidization has been reported within this section (Qin and Zhang, 1985), plastome

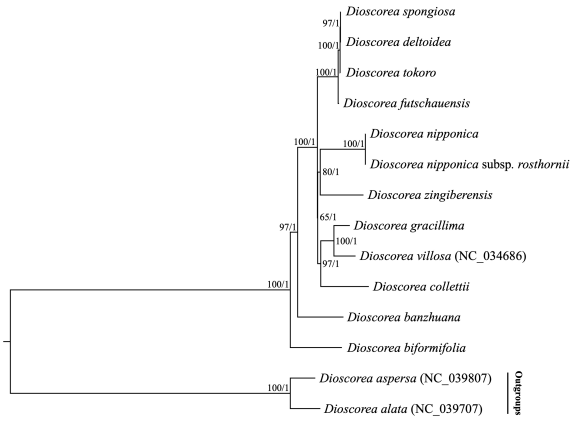


FIGURE 6
Phylogenetic relationships of 12 *Stenophora* species/subspecies inferred from Maximum likelihood (ML) and Bayesian inference (BI) methods, according to complete plastome sequences. Numbers above the lines represent ML bootstrap values and BI posterior probabilities. Phylogenetic trees based on 80 commonly present protein-coding genes with different partitioning scenarios are completely consistent with this topology.

data could not accurately capture hybridization and polyploidization events, as plastome is usually uniparentally inherited, and acts as a linked single locus (Birky, 1995; Stull et al., 2015). Thus, moving beyond the plastomes and analyzing multilocus nuclear DNA sequence data with more extensive sampling is necessary in the future, to explore the phylogenetic and biogeographic hypotheses of sect. *Stenophora*.

4 Conclusions

In this study, we first assembled and annotated the complete plastomes of 11 *D.* sect. *Stenophora* species/subspecies, based on low-coverage whole genome sequencing data. Together with previously published plastome sequence of *D. villosa*, we then provided comparative plastome analyses within this section. All sampled *Stenophora* plastomes (153,691–154,149 bp) shared the same gene content, gene order and GC content. The *rps16* gene was lost in all other *Dioscorea* clades except *Stenophora*, which may have occurred about 48.3 Mya. A total of 960 plastome-derived SSRs and seven plastomic mutational hotspots were identified in *Stenophora*. Besides, we also successfully developed 12 polymorphic nuclear SSRs within this section, based on multiple assembled nuclear scaffolds. Phylogenetic analyses strongly supported that *D. biformifolia* and *D. banzhuana* constituted the successive sister species to the rest, which can be furtherly divided into three clades. Overall, the data obtained here will not only contribute to our understanding of plastome evolution of *Stenophora*, but also aid in the conservation and utilization of their genetic resources.

Data availability statement

All data generated in this study has been publicly available.

Author contributions

R-SL conceived the ideas. KH, X-QS and MC contributed to the sampling. X-QS and MC performed the experiment and analyzed the data. The manuscript was written by KH and R-SL. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (32200192), the Jiangsu Key Laboratory for the Research and Utilization of Plant Resources (JSPKLB202206), the Jiangsu Forestry Science and Technology Innovation and Promotion Project (LYKJ[2020]08), the Jiangsu Institute of Botany Talent Fund (JIBTF202102), and the Jiangsu Innovative and Enterpreneurial Talent Programme (JSSCBS20211311).

Acknowledgments

The authors gratefully acknowledge helpful comments from reviewers on earlier versions of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1196176/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Phylogenetic relationships among *Dioscorea* based on complete plastome sequences. The loss of *rps16* gene is indicated on the branches of the tree.

References

Aecyo, P., Marques, A., Huettel, B., Silva, A., Esposito, T., Ribeiro, E., et al. (2021). Plastome evolution in the caesalpinia group (Leguminosae) and its application in phylogenomics and populations genetics. *Planta* 254, 1–19. doi: 10.1007/s00425-021-03655-8

Asaf, S., Waqas, M., Khan, A. L., Khan, M. A., Kang, S. M., Imran, Q. M., et al. (2017). The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00304

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Birky, J. C.W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci.* 92, 11331–11338. doi: 10.1073/pnas.92.25.11331

Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Green, E. D., et al. (2003). LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* 13, 721–731. doi: 10.1101/gr.926603

Burkill, I. H. (1960). The organography and the evolution of dioscoreaceae, the family of the yams. *J. Linn. Soc London Bot.* 56, 319–412. doi: 10.1111/j.1095-8339.1960.tb02508.x

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol. Biol. Evol.* 32, 2015–2035. doi: 10.1093/molbev/msv082

Chen, F. Q., Fu, Y., Wang, D. L., Gao, X., and Wang, L. (2007). The effect of plant growth regulators and sucrose on the micropropagation and microtuberization of *Dioscorea nipponica* makino. *J. Plant Growth Regul.* 26, 38–45. doi: 10.1007/s00344-005-0147-2

Cheng, J., Chen, J., Liu, X., Li, X., Zhang, W., Dai, Z., et al. (2021). The origin and evolution of the diosgenin biosynthetic pathway in yam. *Plant Commun.* 2, 100079. doi: 10.1016/j.xplc.2020.100079

Chmielewski, M., Meyza, K., Chybicki, I. J., Dzialuk, A., Litkowiec, M., and Burczyk, J. (2015). Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. *iforest-Biogeosciences Forestry* 8, 765–771. doi: 10.3832/ifor1597-008

Couto, R. S., Martins, A. C., Bolson, M., Lopes, R. C., Smidt, E. C., and Braga, J. M. A. (2018). Time calibrated tree of *Dioscorea* (Dioscoreaceae) indicates four origins of yams in the neotropics since the Eocene. *Bot. J. Linn. Soc* 188, 144–160. doi: 10.1093/botlinnean/boy052

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109

Das, S., Choudhury, M. D., and Mazumdar, P. B. (2013). Micropropagation of *Dioscorea alata* l. through nodal segments. *Afr. J. Biotechnol.* 12, 6611–6617. doi: 10.5897/AJB2013.12191

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458

Fu, L. (1992). *Red data book about Chinese plants* (Beijing: Science press), 278–279.

Gao, X., Zhu, Y. P., Wu, B. C., Zhao, Y. M., Chen, J. Q., and Hang, Y. Y. (2008). Phylogeny of *Dioscorea* sect. *Stenophora* based on chloroplast *matK*, *rbcL* and *trnL-F* sequences. *J. Syst. Evol.* 46, 315–321. doi: 10.3724/sp.J.1002.2008.08007

Gitzendanner, M. A., Soltis, P. S., Yi, T. S., Li, D. Z., and Soltis, D. E. (2018). "Plastome phylogenetics: 30 years of inferences into plant evolution," in *Advances in botanical research* (New York: Elsevier), 293–313.

Gong, G., Qin, Y., and Huang, W. (2011). Anti-thrombosis effect of diosgenin extract from *Dioscorea zingiberensis* C.H. Wright *in vitro* and *in vivo*. *Phytomedicine* 18, 458–463. doi: 10.1016/j.phymed.2010.08.015

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Hazra, A., Das, S., Bhattacharya, S., Sur, S., Sengupta, C., and Das, S. (2021). Phylogenetic inference of ericales based on plastid genomes and implication of cp-SSRs. *BioTechnologia. J. Biotechnol. Comput. Biol. Bionanotechnology* 102, 277–283. doi: 10.5114/bta.2021.108723

Hsu, K. M., Tsai, J. L., Chen, M. Y., Ku, H. M., and Liu, S. C. (2013). Molecular phylogeny of *Dioscorea* (Dioscoreaceae) in East and southeast Asia. *Blumea - Biodiversity Evol. Biogeography Plants* 58, 21–27. doi: 10.3767/000651913x669022

Huang, H. P., Gao, S. L., Huang, P., and Zhou, L. (2010). Resources of medicinal classification and geographical distribution on plants of sect. *Stenophora* uline (*Dioscorea* l.) in China. *Strait Pharm. J.* 22, 64–66.

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* 104, 19369–19374. doi: 10.1073/pnas.0709121104

Jansen, R. K., and Ruhlman, T. A. (2012). Plastid genomes of seed plants. *Genomics chloroplasts mitochondria* 35, 103–126. doi: 10.1007/978-94-007-2920-9_5

Jiménez, A. (2010). "Microsatellites: a powerful genetic marker for fern research," in *Working with ferns: issues and applications* (New York: Springer), 207–220.

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 1–31. doi: 10.1186/s13059-020-02154-5

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kawabe, A., Miyashita, N. T., and Terauchi, R. (1997). Phylogenetic relationship among the section *Stenophora* in the genus *Dioscorea* based on the analysis of nucleotide sequence variation in the phosphoglucose isomerase (Pgi) locus. *Genes Genet. Sys.* 72, 253–262. doi: 10.1266/ggs.72.253

Khakhlova, O., and Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85–94. doi: 10.1111/j.1365-313X.2006.02673.x

Kuang, D. Y., Wu, H., Wang, Y. L., Gao, L. M., Zhang, S. Z., and Lu, L. (2011). Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54, 663–673. doi: 10.1139/g11-026

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (*arXiv preprint*). doi: 10.48550/arXiv.1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, P., Lu, R. S., Xu, W. Q., Ohi-Toma, T., Cai, M. Q., Qiu, Y. X., et al. (2017). Comparative genomics and phylogenomics of East Asian tulips (*Amana*, liliaceae). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00451

Liu, L., Wang, Y., He, P., Li, P., Lee, J., Soltis, D. E., et al. (2018). Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics* 19, 1–17. doi: 10.1186/s12864-018-4633-x

Lu, R., Chen, M., Feng, Y., Yuan, N., Zhang, Y., Cao, M., et al. (2022). Comparative plastome analyses and genomic resource development in wild rice (*Zizania* spp., poaceae) using genome skimming data. *Ind. Crops Prod.* 186, 115224. doi: 10.1016/j.indcrop.2022.115244

Lu, R. S., Hu, K., Zhang, F. J., Sun, X. Q., Chen, M., and Zhang, Y. M. (2023). Panplastome of greater yam (*Dioscorea alata*) in China: intraspecific genetic variation, comparative genomics, and phylogenetic analyses. *Int. J. Mol. Sci.* 24, 3341. doi: 10.3390/ijms24043341

Lu, R. S., Li, P., and Qiu, Y. X. (2017). The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.02054

Lu, R. S., Yang, T., Chen, Y., Wang, S. Y., Cai, M. Q., Cameron, K. M., et al. (2021). Comparative plastome genomics and phylogenetic analyses of liliaceae. *Bot. J. Linn. Soc* 196, 279–293. doi: 10.1093/botlinnean/boaa109

Noda, H., Yamashita, J., Fuse, S., Pooma, R., Poopath, M., Tobe, H., et al. (2020). A large-scale phylogenetic analysis of *Dioscorea* (Dioscoreaceae), with reference to character evolution and subgeneric recognition. *Acta Phytotax. Geobot.* 71, 103–128. doi: 10.18942/apg.201923

Olejniczak, S. A., Łojewska, E., Kowalczyk, T., and Sakowicz, T. (2016). Chloroplasts: state of research and practical applications of plastome sequencing. *Planta* 244, 517–527. doi: 10.1007/s00425-016-2551-1

Pei, C., Ting, C. T., Chin, H. C., Su, P., Tang, S. Y., and Chang, H. C. (1979). A preliminary systematic study of *Dioscorea* l. sect. *Stenophora* uline. *Acta Phytotaxon. Sin.* 17, 61–72.

Ping, J., Feng, P., Li, J., Zhang, R., Su, Y., and Wang, T. (2021). Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecol. Evol.* 11, 4786–4802. doi: 10.1002/ece3.7381

Qin, H. Z., and Zhang, M. Z. (1985). A study on the classification of *Dioscorea* cells in China: chromosome number, the origin and evolution of the genus. *Acta Phytotaxon. Sin.* 23, 11–18.

Raubeson, L. A., and Jansen, R. K. (2005). "Chloroplast genomes of plants," in *Plant diversity and evolution: genotypic and phenotypic variation in higher plants* (Wallingford: CABI publishing), 45–68.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275

Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., and Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am. J. Bot.* 101, 1987–2004. doi: 10.3732/ajb.1400398

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 99, 349–364. doi: 10.3732/ajb.1100335

Stull, G. W., Duno de Stefano, R., Soltis, D. E., and Soltis, P. S. (2015). Resolving basal lamiid phylogeny and the circumscription of icacinaceae with a plastome-scale data set. *Am. J. Bot.* 102, 1794–1813. doi: 10.3732/ajb.1500298

Sun, W., Wang, B., Yang, J., Wang, W., Liu, A., Leng, L., et al. (2017). Weighted gene co-expression network analysis of the dioscin rich medicinal plant *Dioscorea nipponica*. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00789

Ting, C., Michael, G., Wu, Z., and Raven, P. (2000). "Dioscorea linnaeus," in *Flora in china*. Eds. Z. Y. Wu and P. H. Raven (St. Louis: Missouri Botanical Garden Press and Beijing: Science Press).

Titova, G. E., and Torshilova, A. A. (2015). To the evolution of seed morphological type in the *Dioscorea* section *Stenophora* (Dioscoreaceae). *Botanicheskiĭ Zhurnal* 100, 761–786.

Twyford, A. D., and Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* 17, 858–868. doi: 10.1111/1755-0998.12626

Vinogradova, G., Torshilova, A., and Machs, E. (2022). Flower morphology and phylogenetic analysis of some *Dioscorea* species of the section *Stenophora* (Dioscoreaceae). *Plant Syst. Evol.* 308, 42. doi: 10.1007/s00606-022-01834-y

Viruel, J., Forest, F., Paun, O., Chase, M. W., Devey, D., Couto, R. S., et al. (2018). A nuclear *Xdh* phylogenetic analysis of yams (*Dioscorea*: dioscoreaceae) congruent with plastid trees reveals a new Neotropical lineage. *Bot. J. Linn. Soc* 187, 232–246. doi: 10.1093/botlinnean/boy013

Viruel, J., Segarra-Moragues, J. G., Raz, L., Forest, F., Wilkin, P., Sanmartín, I., et al. (2016). Late Cretaceous–early Eocene origin of yams (*Dioscorea*, dioscoreaceae) in the laurasian palaearctic and their subsequent oligocene–Miocene diversification. *J. Biogeogr.* 43, 750–762. doi: 10.1111/jbi.12678

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Wilkin, P., Schols, P., Chase, M. W., Chayamarit, K., Furness, C. A., Huysmans, S., et al. (2005). A plastid gene phylogeny of the yam genus, *Dioscorea*: roots, fruits and Madagascar. *Syst. Bot.* 30, 736–749. doi: 10.1600/036364405775097879

Wu, L., Wang, B., Yang, J., Song, C., Wang, P., Chen, S., et al. (2016). The chloroplast genome sequence of an important medicinal plant *Dioscorea nipponica*. *Mitochondrial DNA Part A* 27, 2559–2560. doi: 10.3109/19401736.2015.1038803

Xia, Y., Luo, W., Yuan, S., Zheng, Y., and Zeng, X. (2018). Microsatellite development from genome skimming and transcriptome sequencing: comparison of strategies and lessons from frog species. *BMC Genom.* 19, 1–13. doi: 10.1186/s12864-018-5329-y

Xia, E. H., Yao, Q. Y., Zhang, H. B., Jiang, J. J., Zhang, L. P., and Gao, L. Z. (2016). CandiSSR: an efficient pipeline used for identifying candidate polymorphic SSRs based on multiple assembled sequences. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01171

Xia, W., Zhang, B., Xing, D., Li, Y., Wu, W., Xiao, Y., et al. (2019). Development of high-resolution DNA barcodes for *Dioscorea* species discrimination and phylogenetic analysis. *Ecol. Evol.* 9, 10843–10853. doi: 10.1002/ece3.5605

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., et al. (2014). SOAPdenovo-trans: *de novo* transcriptome assembly with short RNA-seq reads. *Bioinformatics* 30, 1660–1666. doi: 10.1093/bioinformatics/btu077

Yang, Z., Ma, W., Yang, X., Wang, L., Zhao, T., Liang, L., et al. (2022). Plastome phylogenomics provide new perspective into the phylogeny and evolution of betulaceae (Fagales). *BMC Plant Biol.* 22, 611. doi: 10.1186/s12870-022-03991-1

Zhao, Z., Wang, X., Yu, Y., Yuan, S., Jiang, D., Zhang, Y., et al. (2018). Complete chloroplast genome sequences of *Dioscorea*: characterization, genomic resources, and phylogenetic analyses. *PeerJ* 6, e6032. doi: 10.7717/peerj.6032

Zhou, W., Chen, C., Hua, W. P., and Wang, Z. Z. (2016). The complete chloroplast genome sequence of *Dioscorea zingiberensis* (Dioscoreceae). *Mitochondrial DNA Part A* 27, 2730–2731. doi: 10.3109/19401736.2015.1046168

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 03 July 2023
DOI 10.3389/fpls.2023.1190768

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Wenkai Hui,
Sichuan Agricultural University, China
Juncheng Li,
Guangdong Academy of Agricultural Sciences, China

*CORRESPONDENCE
Xiaoling Wang
nxy170026@163.com
Zhu Gao
44647268@qq.com

†These authors have contributed equally to this work

RECEIVED 21 March 2023
ACCEPTED 04 May 2023
PUBLISHED 03 July 2023

CITATION
Ning X, Lin M, Huang G, Mao J, Gao Z and Wang X (2023) Research progress on iron absorption, transport, and molecular regulation strategy in plants. *Front. Plant Sci.* 14:1190768. doi: 10.3389/fpls.2023.1190768

COPYRIGHT
© 2023 Ning, Lin, Huang, Mao, Gao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Research progress on iron absorption, transport, and molecular regulation strategy in plants

Xinyi Ning[1,2,3†], Mengfei Lin[1,3†], Guohua Huang[1,2,3], Jipeng Mao[1,3], Zhu Gao[1,3,4*] and Xiaoling Wang[1,3*]

[1]Institute of Biological Resources, Jiangxi Academy of Sciences, Nanchang, China, [2]College of Environmental And Chemical Engineering, Nanchang Hangkong University, Nanchang, China, [3]Kiwifruit Engineering Research Center of Jiangxi Province, Nanchang, China, [4]JInstitute of Biotechnology, Jiangxi Academy of Sciences, Ji'an, Jiangxi, China

Iron is a trace element essential for normal plant life activities and is involved in various metabolic pathways such as chlorophyll synthesis, photosynthesis, and respiration. Although iron is highly abundant in the earth's crust, the amount that can be absorbed and utilized by plants is very low. Therefore, plants have developed a series of systems for absorption, transport, and utilization in the course of long-term evolution. This review focuses on the findings of current studies of the Fe^{2+} absorption mechanism I, Fe^{3+} chelate absorption mechanism II and plant-microbial interaction iron absorption mechanism, particularly effective measures for artificially regulating plant iron absorption and transportation to promote plant growth and development. According to the available literature, the beneficial effects of using microbial fertilizers as iron fertilizers are promising but further evidence of the interaction mechanism between microorganisms and plants is required.

KEYWORDS

plant, iron, absorption, transport, molecular regulation

1 Introduction

Iron is the most abundant essential trace element in human body (Obi et al., 2022). Fe deficiency is considered to be the first of the three "hidden hunger" in the world, with approximately one in five people worldwide suffering from iron deficiencyanemia (Lowe N.M., 2021). As the fundamental food source for animals and human beings, plants can directly absorb iron nutrition from the soil environment and are direct carriers of iron received from the soil (Krishna et al., 2023). Iron is the first identified essential plant

element that plays an important role in the whole plant growth cycle, and it is involved in the process of life activities such as protein synthesis, DNA replication, and respiration in the plant (Takanori et al., 2018), as well as in the composition of chlorophyll and plays an important role in biological nitrogen fixation. Nearly 40% of arable plants worldwide show varying degrees of iron deficiency yellowing, especially in calcareous soils in arid and semi-arid regions, where iron deficiency in plants is widespread. Worldwide, the North American continent, the Mediterranean coast, and parts of South America are severely iron deficient. In China, iron deficiency occurs in areas from Sichuan Basin in the south to the Inner Mongolian Plateau in the north, Huaibei Plain in the east to the Loess Plateau in the west, Gansu, Qinghai, Xinjiang and other provinces (Satoshi, 1999). In general, Fe exists in the soil mainly in two valence states, Fe2+ and Fe3+, of which Fe2+ can be absorbed and utilized by plants. While Fe_{3+} is difficult to be absorbed by plant roots due to its low solubility and poor effectiveness. The active Fe content in soil is easily affected by factors such as soil pH, salts, and some ions. For example, under acidic and flooded conditions, it exists as dissolved Fe_{2+}; when the soil pH is greater than 6.5, the content of active Fe starts to decline and exists in an insoluble or insoluble state; when the pH exceeds 8.5, it exists mainly in insoluble iron carbonate form, causing the appearance of yellowing in plants (Huei-hsuan and Wolfgang, 2020). Iron in alkaline soils often combines with phosphate or hydroxyl ions to form precipitates, resulting in the inability to be absorbed by plants (Fu et al., 2017). Plants have evolved a series of molecular systems for Fe uptake and translocation both to obtain sufficient Fe from the soil and to avoid Fe toxicity due to excessive uptake. Depending on the uptake mechanism, they can be divided into mechanism I and mechanism II. Mechanism I mainly absorb Fe_{2+}, which consists of three systems: the H+-ATPase pump system, the Fe_{3+} reduction system, and the translocation system of Fe2+, which are in turn influenced by ILR3 (Tissot et al., 2019), YABBY (Sun et al., 2021), the FEP1, IMA3 (Okada et al., 2022), MNB1 (Song et al., 2022), and bHLH1b (Liang, 2022); Mechanism II is achieved through Fe3+ chelate uptake, which is synthesized by the lysergic acid-like plant iron carrier (phytosiderophore, PS). The genes associated with this mechanism are GRX (Kobayashi et al., 2022a), ACOs (Senoura et al., 2020), HRZ (Liang, 2022) and so on. There is also another reported mechanism: the mechanism of interaction between microorganisms and plants. This mechanism is regulated by genes such as NRAMP (Cailliatte et al., 2010) and TonB (Bruce et al., 2021), and plants can take up iron by interacting with microorganisms. Since the iron is difficult to be absorbed and utilized by plants in the soil environment, plants are facing the problem of iron deficiency stress. To explore ways to solve this difficult problem, this paper provides a comprehensive elucidation of the adaptive and non-adaptive mechanisms of Fe uptake by plants from plant morphology, physiological and biochemical, molecular regulatory mechanisms, and metabolic pathways and analyzes the regulatory networks and mechanisms of action associated with each type of mechanism, proposing insights into the study of Fe uptake and utilization by plants, to provide references for carrying out Fe recycling and utilization in the plant body.

1.1 Effect of iron on plant morphology and physiology

1.1.1 Effect of iron on plant morphological characters

Iron is not a constituent of chlorophyll, but it is involved in the process of chlorophyll synthesis (Norman and Javier, 1986). When plants are deficient in iron, they first show symptoms such as vein chlorosis, shrunken leaves, and plant dwarfing, which in turn affect the normal growth and development of the plant, leading to damage to fruit quality, serious yield loss of cash crops, even worse, it can cause the entire plant to die. The symptoms of iron deficiency in plants vary from species to species. In mild iron deficiency, young leaves lose green between veins, while the veins remain green, and the leaves become smaller and thinner; in severe iron deficiency, the veins also begin to fade to green, stem growth is hindered, and the terminal buds wither; in more severe cases, the leaves all lose green or even turn white, the leaf flesh is necrotic, the leaf edges turn red, the leaf tips and leaf edges are necrotic and scorched, and the stems begin to stop growing, such as apple and peach (Sun et al., 1987; Victoria et al., 2008); it also causes leaf loss of green color, and the length of new internodes, leaf weight, and leaf thickness is significantly reduced, such as grapes and pears (Victoria et al., 2008; José et al., 2016).

The root system, as the main organ for nutrient uptake by plants, adapts to iron deficiency through corresponding changes in root morphology and physiology when the iron content in the soil is reduced (Maribela et al., 2012). Iron deficiency causes plants to appear enlarged and thickened near the root tip, with an increase in root hairs, and to cope with iron deficiency adversity by secreting large amounts of H^+ and enhancing the activity of trivalent ferric ion chelating reductase, thereby increasing the effectiveness of iron in plant roots, as in cucumber and tomato; it also results in rapid growth and thickening of the root tip and the continuous accumulation of more and more phenolic compounds in the epidermis and cortex, as a way to reduce insoluble ferric ion compounds, thus slowing down the damage caused by Fe deficiency, such as soybean (Yi et al., 1998); it also leads to main root growth, reduced root number, increased root hairs, increased root dry weight, and increased root surface area, such as wheat and pear (Fan et al., 2018; Zeng et al., 2019).

1.1.2 Effect of iron on plant physiology

1.1.2.1 Chlorophyll synthesis

The chloroplasts in most plant leaves contain about 80% of iron (Giovanni et al., 2015), of which 60% is immobilized in the cystoid membrane, 20% in the chloroplast stroma, and only a very small fraction of iron is in other organs. Therefore, the impeded chlorophyll synthesis due to iron deficiency is the main cause of leaf greening and yellowing.

In iron-deficient environments, ultrastructural abnormalities were observed in plant chloroplasts and chloroplasts, and the number of cystoid membranes in the lamellar structure of plant chloroplasts was reduced (Zhou et al., 2008). Iron deficiency led to a decrease in cis-aconitase activity, which is inextricably linked to the

synthesis of chlorophyll precursors. Iron deficiency chlorophyll synthesis, chlorophyll content, and concentration in leaves decrease with increasing iron deficiency yellowing, as in balsam pear (Zeng et al., 2019) and kiwifruit (Wang et al., 2019); iron deficiency leads to a significant reduction of chloroplast basidia and stromal cysts in faded green leaves, with disrupted ultrastructure and significantly reduced chlorophyll concentration, common in iron deficiency-sensitive tangerine (Ding et al., 2016); More seriously, not only the leaf chloroplast structure is severely affected, but also the mesophyll cells, palisade cells and parenchyma cells of the main vein (Zhang et al., 2019), as in Mexican Lime (Ranferi et al., 2006).

1.1.2.2 Photosynthesis

PS-I (photosystem-I) is the photosynthetic apparatus with the highest Fe content in the entire electron transport chain, and studies in a variety of plants have shown that Fe deficiency leads to severe disruption of the structure and activity of PS-I. Current studies have reported that plant photosynthesis is affected by Fe in three ways: first of all, Fe affects the production of cystoid in plant chloroplasts, and Fe deficiency stress disrupts the formation of photosynthetic elements and has a direct effect on photosynthetic basic substances and sites of action (Yoshida et al., 2021); afterwards, many Fe-oxygen complexes intervene in the process of photochemical reactions in plants, such as cytochrome C oxidase complex, Fe oxygen reductase, heme, and bean heme, and in Fe deficiency adversity, the proportion of the content of these substances shrinks and the physiological activities of some Fe-related enzymes are abrogated (Agarwala et al., 1965; Yadavalli et al., 2012); finally, Fe is an influential factor in the electron transfer process in photosynthesis, and in Fe deficiency, interrupts the electron transfer chain and hinders the photolysis of water, thus blocking the process of photosynthesis, and can this leads to a decrease in the capacity of ROS (reactive oxygen species) detoxification enzymes, which ultimately leads to a negative impact on the photosynthetic rate of plants (Singh et al., 2005).

The factors that lead to the decrease of photosynthetic rate include stomatal limitation and non-stomatal limitation. Stomatal limitation means that the weakening of photosynthesis is caused by the diffusion of CO_2 to the carboxylation site caused by the decrease of stomatal opening, rather than the stomatal limitation that makes the CO_2 concentration in the intercellular space very high, but photosynthesis is still limited very weakly. This can be judged by the magnitude and trend of Ci (Intercellular CO_2 concentration) values, where poplar exhibits higher stomatal restriction with higher stomatal conductance and more CO_2 lost by photosynthesis (Jurca et al., 2022). In iron-deficient chlorosis 'Dangshan crisp pear', the Gs (stomatal conductance) of leaves decreased and Ci increased, which means that the decrease of Pn (Net-photosynthetic-rate) accompanied by the increase of Ci, and the decrease of the photosynthetic rate due to yellowing belonged to non-stomatal limitation.

1.1.2.3 Respiration

Iron is associated with the composition of several respiration-related enzymes and is one of the basic components of these enzymes. And such enzymes are involved in plant respiration, such as cytochrome, CAT (catalase), and POD (peroxidase). Iron deficiency weakens the activity of these enzymes, causing a sudden decrease in the effectiveness of a series of redox reactions in plants, hindering the normal electron transfer and reducing ATP synthesis, thus making respiration affected (Sandmann, 1985). Iron is a cofactor of cis-aconitase, the rate-limiting enzyme of the entire respiration of the TCA cycle, and when plants are in an iron-deficient environment, the activity of this enzyme is inhibited and the process of regulating respiration is affected. Electron transport also requires the assistance of iron-sulfur proteins, and the complexes in the respiratory electron transport chain process all contain iron-sulfur proteins, and iron deficiency interferes with the activity of these protein complexes, further affecting the rate of transport of the respiratory electron transport chain in plants and ultimately hindering plant growth and metabolism (Philip and Leonid, 2005).

1.1.2.4 Metabolism of nitrogen element and other products

Iron plays a non-negligible role in the microbial nitrogen cycle, it is involved in the metabolic processes of nitrogen and is an important component of some essential enzymes in nitrogen fixation reactions, such as nitrogen-fixing enzymes. In the presence of iron deficiency, nitrate reductase activity is reduced, while glutamine synthetase and glutamate synthetase levels are increased (Borlotti et al., 2012). Apples supply nitrate for uptake by using nitrate reductase and nitrite reductase to reduce nitrate to nitrite, which is further reduced to ammonia and finally synthesizes amino acids, proteins, and other substances (Sun et al., 2021b). It has also been shown that nitrogen nutritional status is a key determinant of iron reactivation in plants and regulates the transfer of iron from fully developed senescing leaves wanting to grow sites (Shamima et al., 2018).

Iron deficiency also causes the development of plant diseases, reduced quality, and lower yields. It has been found that iron inhibits the growth of tomato cyanobacteria and alleviates disease development (Truchon et al., 2022). Iron affects fruit quality and yield by influencing plant enzyme activity and metabolism. In Fe-deficiency, apple induces increased enzyme activities of POD, superoxide SOD (dismutase), and root iron ion chelating reductase (FRO) under adversity (Li, 2003). Espen et al. (2000) used 31P-NMR to analyze the metabolic responses of cucumber roots under Fe-deficiency conditions and found that Fe-deficiency induced activation of metabolism and depletion of stored carbohydrates. Iron deficiency significantly reduced the iron content in citrus rootstocks and the sugar content in leaves (Mary-Rus et al., 2013); strawberry fruit quality and yield were affected, and in severe cases plants were unable to hang fruit (Zhang, 2020); the application of organic fertilizers containing iron too early for grapefruit increased the content of reducing sugars, total sugars, soluble solids, and vitamin C in the fruit.

2 Mechanism of iron uptake and transport in plants

The transport process of trace elements such as iron in plants first reaches the extra plastic body of root tip cells by diffusion, then

the absorbed iron is transported from the extracellular to the intracellular by the iron carrier on the cell membrane, into the cellular cytoplasm, then into the xylem thin-walled tissue in the root system, during which divalent iron is oxidized to trivalent iron into the ducts of the roots, transported through the ducts to the leaf flesh cells of mature leaves, and then it is then redistributed in the leaves and finally reaches the growth cells through the bast. The main systems for long-distance iron transport in plants are the xylem and the phloem.

The transport of iron ions in the plant body: Root apoplast → root plastid → thin-walled cell tissue → ducts → leaf flesh cells → bast → developmental cells (Qu, 2005).

However, in the vast majority of soil environments, iron levels are often not sufficient for normal plant growth. Higher plant inter-roots have continuously evolved a series of mechanisms of iron deficiency stress tolerance to adapt to iron deficiency adversity. For example, adaptive mechanisms regulated by iron nutrition in plants, including Mechanism I and Mechanism II; mechanisms of plant-microbial interactions for iron uptake and transport.

2.1 Mechanism I of absorption by Fe^{2+}

The Fe uptake in mechanism I is accomplished by three protease systems: 1) H^+-ATPase protease system, plant roots reduce the pH value of the surrounding soil by releasing hydrogen ions, and increase the dissolution of iron in rhizosphere soil. With the enhancement of H^+-ATPase activity, the roots secrete coumarin and riboflavin to increase the mobility of Fe^{3+}(Robe et al., 2020a; Robe et al., 2020b). Among the numerous H^+-ATPase (HA) genes, the expression of some HA genes is induced by iron deficiency stress, among which AHA2 is a direct commitment protein for hydrogen ion secretion, and ABA-activated BAK1 phosphorylates AHA2 at its C-terminal Ser-944 and activates AHA2, leading to rapid hydrogen ion efflux, cytoplasmic alkalinization, and ROS accumulation (Pei et al., 2022), and *CsHA1* in cucumber plays a similar role (Santi et al., 2003). Meanwhile, the plasma membrane-localized H^+-ATPase HA6 regulates hydrogen ion efflux and the expression of this gene is expressed up-regulated in plants grown on soils with low Fe content, while MYB308 activates *HA6* to promote inter-root hydrogen ion efflux and Fe uptake (Fan et al., 2022). 2) Fe^{3+} reduction system, consisting of trivalent iron chelating reductase (FRO) and reductive coenzyme II (NADPH) dehydrogenase. Fe^{3+} released by the H^+-ATPase protease system enters the cell after forming chelates with chelators in the plant, and the reduction system converts Fe^{3+} chelates into Fe^{2+} chelates released for plant uptake and utilization. The iron deficiency-induced pea Fe^{3+} reductase gene, *PsFRO1*, was expressed in plant roots, mycorrhizae, stems, and leaves, and the expression was relatively higher in root epidermal cells with Fe^{3+} reduction system and Fe^{2+} transporter protein system; and the expression of *PsFRO1* in the nitrogen fixation zone of root nodules may play a role in plant nitrogen fixation (Waters et al., 2002); natural allelic variation of *FRO2* regulates *Arabidopsis* root growth under iron deficiency conditions (Satbhai et al., 2017). This step is particularly important as it is the limiting step in the mechanism I process. 3) Fe^{2+} transporter protein (IRT) system, which brings Fe^{2+} chelates reduced by Fe^{3+} chelating reductase into the cell by transmembrane transport and then transported by other transporter proteins to various organelles and organs of the plant to supply plant growth and development. The *Arabidopsis AtIRT1* gene was the first *IRT* gene obtained by heterologous expression in a transporter double mutant yeast that eventually complemented iron deficiency symptoms (Eide et al., 1996). AtERF4 and AtERF72 are both negative regulators of the iron deficiency response and maintain iron homeostasis by directly binding to the *IRT1* promoter (Liu et al., 2017), AtERF4 through yeast single hetero ERF95 transcription factor regulates iron accumulation in *Arabidopsis* seeds through an *EIN3*-ERF95-*FER1*-dependent signaling pathway (Sun et al., 2020); ERF96 loses its function under iron deficiency stress and then increases the expression of iron uptake genes through ethylene and growth hormone signaling pathways and reduced expression of chlorophyll-degrading genes and enhanced iron and chlorophyll accumulation in *Arabidopsis* (Yao et al., 2022); NRANP is involved in the uptake of small amounts of iron from the soil by relying on its specific affinity for iron and its response to environmental signals (Eide et al., 1996); and INO reduces iron accumulation in developing seeds by inhibiting *NRANP1* expression to reduce iron loading in developing seeds. Phenolic compounds in *Arabidopsis* root secretions were also found to convert insoluble iron in the plant growth medium into soluble iron, but ultimately only absorbed from it through the IRT1/FRO2 high-affinity iron transport system (Vert et al., 2021). f6′H1 (Feruloyl coenzyme A 60-hydroxylase 1) is a key enzyme for Fe-deficiency-induced coumarin synthesis. In iron deficient environments, the biosynthesis of coumarin is of great significance for plant growth, promoting the absorption of iron and inducing plant immunity. Iron deficiency induces up-regulation of F6′H1 expression and shifts the lignin synthesis pathway to coumarin synthesis. The F6′H1 homolog S8H (scopoletin 8-hydroxylase), also induced by iron deficiency, was found to generate Fe^{3+}-reducing fraxetin (7,8-dihydroxy-6-methoxy coumarin) by hydroxylating coumarin scopoletin, which can dissolve and bind iron to form stable complexes at neutral to alkaline pH (Zhang et al., 2021) (Figure 1) (Rajniak et al., 2018). Other genes such as *ILR3* (Tissot et al., 2019), *YABBY* (Sun et al., 2021a), *FEP1*, *IMA3* (Okada et al., 2022), *MNB1* (Song et al., 2022), and *bHLH1b* (Liang, 2022) are involved in this mechanism.

Soil pH and HCO_3^- have a strong influence on this mechanism, and bicarbonate stress induces plant roots to secrete relevant substances or protons into the inter-root soil to acidify it and inducing root iron reductase gene and iron transporter gene expression thereby enhancing iron reductase activity and iron uptake (Wang, et al., 2022). Plants that conform to the mechanism I for iron uptake mainly contain dicotyledons and non-grass monocotyledons, such as soybean, peanut, sunflower, and cucumber.

This strategy involves a protein complex consisting of HA, FRO, IRT and PDR. HA secreted protons to acidify the rhizosphere soil and mobilize Fe^{3+}, and BAK and MYB assisted in the activation of HA. After being reduced by FRO, Fe^{2+} enters

the plant roots with the help of IRT, and ERF negatively feeds back to regulate IRT. Coumarin also contributes to the reduction strategy, possibly by converting Fe^{3+} to Fe^{2+}. Sideretin is the major coumarin at acidic pH, while fraxetin is the major coumarin at neutral or alkaline pH. PRD mediates coumarin efflux from roots.

2.2 Mechanism II of absorption as Fe^{3+} chelate

In plants, Fe must be bound to other compounds for transport, and substances such as citric acid, phenolics, NA (nicotinamide), and MAs (myristic acid) are involved in the transport of Fe in plants. FDR3 (citrate transporter protein) and OsFDRL1 in rice are involved in the transport of Fe by citrate loading in the xylem; NA and DMA (2'-deoxymyristic acid) are present in the secretions of the best and are involved in iron transport in the bast (Takanori et al., 2018). YSL has its presence in all parts of the plant, OsYSL15 is a Fe(III)-DMA transporter protein located on the plasma membrane and is mainly responsible for iron uptake from the inter-root (Inoue et al., 2009). OsYSL2 is a Fe(II)-NA transporter protein responsible for Fe transport in the bast (Ishimaru et al., 2010); OsYSL16 is present in the plasma membrane of periplasmic cells and is involved in Fe partitioning through the vascular bundle (Kakei et al., 2012); OsYSL18 is involved in Fe transport in reproductive organs and critical bast (Aoyama et al., 2009). OsYSL18 is related to Fe transport in reproductive organs and critical sclerotia (Aoyama et al., 2009).

MAs are a class of low molecular weight non-protein amino acids that have a strong affinity for trivalent Fe (with six functional groups that chelate Fe) and can form stable, octahedral trivalent chelates (Fe (III)-MAs) that are then taken up by plants. The trivalent chelates (Fe (III)-MAs) are stable and octahedral, which are then absorbed and used by plants. The adaptation of plants to Fe-deficient environments through mechanism II uptake occurs in two major steps: 1) the plant root system synthesizes the myristica acid-like PS (phytosiderophore) *via* SAM (S-adenosyl methionine) through a series of enzymatic reactions and then actively secretes it into the root environment *via* carrier action (Higuchi et al., 1999); 2) The phytosiderophore chelates with trivalent iron ions in the inter-root environment, forming Fe (III)-PS, a chelate of phytosiderophore and trivalent iron ions, and migrates to the root plasma membrane, where it is transported to the plant by the specific transporter proteins YS1 (yellow stripe 1), YSL (yellow stripe-like) and so on (Takahashi et al., 2012; Yamagata et al., 2022). Nicotinamide (NA), 2'-deoxycholic acid (DMA) and maltogenic acid (MA) are chelating agents required for the uptake and transport of iron by plants. While nicotinic amide synthetase (NAS), nicotinic amide aminotransferase (NAAT), 2'-deoxygenate synthase (double deoxygenate synthetase (DMAs), MAs transporter protein (TOM), and NA efflux transporter protein (ENA) are involved in iron uptake and transport in mechanism II plants (Figure 2). To adapt to the iron-deficient environment, DMA accumulates in the xylem sap of rice, the release of NA and DMA/MAs is regulated by different efflux transport proteins, and the expression of DMA efflux transport proteins OsTOM1, OsTOM2, and OsTOM3 is induced in roots, OsTOM1 is involved in the secretion of MAs to the root interiors, while OsTOM2 and OsTOM3 in some tissues exhibited specific expression patterns out of iron transport-related (Nozoye et al., 2015). Meanwhile, expression of the homolog of OsTOM1, vesicular creatine transporter (OsVMT), was induced in roots (Che et al., 2019). The histone or heme-associated protein (HAP) transcription factor (TF) HAP5A is required for the response to iron deficiency in *Arabidopsis*, and the expression of plant genes encoding nicotinamide synthase is greatly reduced in the presence of hap5a mutations (Xiao et al., 2020). Transfer of genes that are capable of efficient synthesis and secretion of myristic acid-like iron carriers in barley, such as *HvNAS*, *HvNAAT*, and *HvIDS3*, into rice significantly improved rice myristic acid synthesis and secretion (Motofumi et al., 2008); combination of carbohydrate-binding module family 11 (CBM11) and iron-binding peptide (IBP) into a CBM-IBP fusion peptide, secreted into the cell walls of *Arabidopsis* and rice, transformed *Arabidopsis* and rice plants showed significantly increased iron accumulation and biotransformation (Yang et al., 2016); based on the natural phycocyanin 2'-deoxymalginate (DMA), a novel iron chelator proline-2'- deoxymalginate (PDMA) solubilizes insoluble iron and upregulates the expression of the yellow streak family gene *AhYSL1* to improve iron nutrition in peanut plants (Wang et al., 2022). Other genes related to this mechanism include *ACOs* (Senoura et al., 2020) and *HRZ* (Liang, 2022). It has been shown that the combined introduction of *OsHRZ* knockdown and *OsIRT1* promoter-Refre1/372 can further improve iron deficiency tolerance without affecting the knockdown effect on iron accumulation (Kobayashi et al., 2022b).

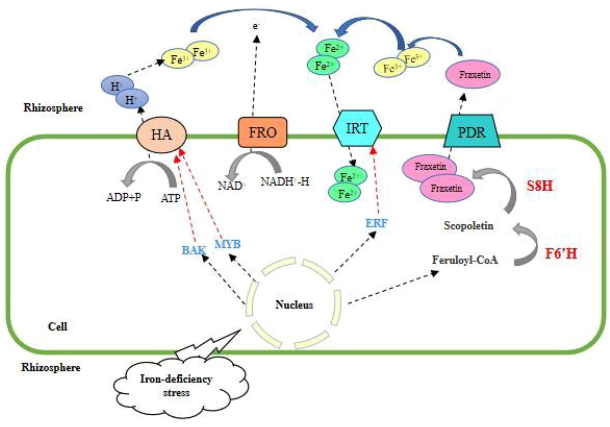


FIGURE 1
Mechanism I of plant root uptake and response to Fe deficiency stress.

Plants that take up iron by mechanism II are mainly monocotyledons, such as rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), and maize (*Zea mays* L.). Plants have two distinctive features when they take up iron by mechanism II: carrier release and high affinity of iron carriers on root cell membranes. Soil pH had little effect on this mechanistic physiological process, so mechanism II plants in medium/alkaline soils showed no significant iron deficiency symptoms.

As a typical monocotyledonous plant, rice is not a typical mechanism II plant, which not only synthesizes DMA in the

roots to chelate Fe^{3+} but also acquires Fe^{2+} through the transporter proteins OsIRT1 and OsIRT2. Three consecutive enzymatic reactions catalyzed by NAS, NAAT, and DMAS are also present in the synthesis of DMA. In addition, various transcription factors such as IDEF1, IDEF2, and bHLH were shown to regulate iron uptake and transport (Li et al., 2019).

In this mechanism, PEZ is responsible for secreting PCA/CA, reducing $Fe^{3+}$to Fe^{2+}, DMA excretes PS through TOM to chelate Fe^{3+}, and Fe^{3+}- PS is transferred to the root through YS1/YSL.

2.3 Microbial mechanism of plant iron uptake

The mechanism of iron uptake and transport by plants cannot be fully explained by Mechanism I and Mechanism II alone. As an indispensable nutrient element for living organisms, the mechanism of its regulation in different organisms varies greatly. When the fungus-iron ion-plant triad interacts, it not only helps to enhance a new understanding of plant and microbial life activities but also facilitates a breakthrough in the field of plant disease resistance mechanism research (Qiu and Liu, 2022). Iron plays a role in plant-pathogen protection of plant host cells from bacterial or fungal infection. In recent years, it has been found that under low Fe conditions, microorganisms produce Fe carriers (Siderophores), and microbial Fe carriers can provide Fe nutrition to plants for their growth and development. There are two possible mechanisms for plant-microbe interactions: 1) microbial Fe carriers with high redox potential can be reduced to provide Fe(II) to the plant transport system (Konrad, 1994); 2) Microorganisms synthesized siderophores and diffused around the rhizosphere after iron deficiency stress, and siderophores chelated with Fe^{3+} to form Fe^{3+}-Siderophores chelate. Part of the Fe^{3+}-Siderophores chelate is absorbed by the microorganism itself and reduced to Fe^{2+} after entering the cell, which is used for the growth and development of the microorganism; the other part is absorbed and utilized by the plant (Gu et al., 2020).

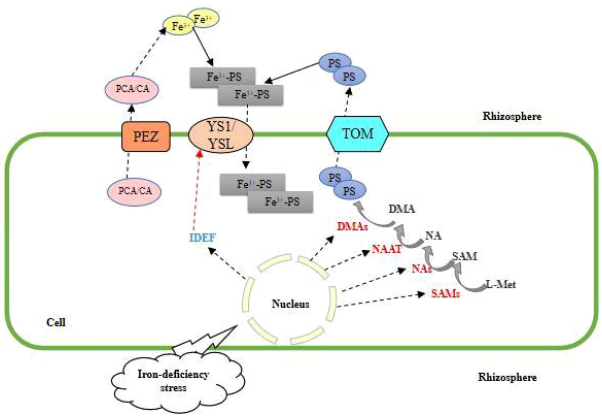


FIGURE 2
Mechanism II of plant root uptake and response to Fe deficiency stress.

As a transcriptional repressor of iron carrier synthesis, Fur is the key factor, in dimensioning cellular iron homeostasis. When iron is abundant, Fur binds divalent iron to form a complex that binds to the promoters of iron carrier synthesis genes or regulatory genes to block their transcription; when cells are exposed to iron deficiency, Fur derepresses and induces iron carrier synthesis (Hosni and Bryan, 2013). The expression of the *TonB* gene depends on the availability of surrounding iron and the orientation from a specific gene *fur*, and expressed *TonB* is associated with *ExbBH* and *ExbD* interact to form TonB complexes that act through conformational changes to modulate outer membrane receptors, which in turn activate iron transport across the membrane (Figure 3) (Bruce et al., 2021). The product of the *NRAMP* gene helps prevent the synthesis of active defense enzymes by phagocytosed microorganisms by extracting metals from the phagocytic vesicles of the phagocytosed microorganisms, which contain iron or other metals as cofactors. In such mechanisms, ferritin attached to the surface of the plasma membrane may be invaginated by phagocytosis, and iron is digested by proteases secreted by the endoplasmic reticulum around the endocytic vesicles and separated from the protein, which then absorbs the free iron into the cytoplasm *via* the NRAMP transporter (Satoshi, 1999). In soybean plants, iron deficiency causes abnormal regulation of the *NRAMP* gene, and *GmNRAMP* expression was found in soybean root nodules and showed that this gene is capable of translocating iron or other metal ions in soybean root nodules (Qin et al., 2017). Maize and sunflower grew better in a non-sterile environment than in a sterile environment and had severe iron deficiency under sterile conditions (Masalha et al., 2000). The endophytic *Streptomyces* sp. GMKU3100 produced iron carriers in Thai rice and promoted rice and mung bean growth and development (Rungin et al., 2012). In Fe-deficient leaves, fungal infection progressed rapidly through biotrophic growth to necrotrophic growth development, whereas adequate Fe-nutrient status inhibited the formation of gram-derived anthracnose structures at the early biotrophic growth stage (Ye et al., 2014).

In this mechanism, the expressed TonB interacts with ExbBH and ExbD to form a TonB complex that acts through conformational change, modulates the outer membrane receptor, secretes ferriferous carrier, chelates Fe^{3+} in the environment, enters the plant root cell *via* YS1/YSL, and is reduced to Fe^{2+}.

3 Techniques and methods for correcting iron deficiency in plants

3.1 Application of plant iron fertilizer

The application of iron fertilizer is the most direct and easy way to address the symptoms produced by iron deficiency in plants. Such as foliar spraying, soil application, and trunk injection. Leaf spraying mainly involves spraying yellowing leaves with iron salt solutions (iron citrate, EDHA-Fe (ethylene diamine tetraacetic acid-Fe), ferrous sulfate and so on), which can improve the yellowing disorder of plants, such as kiwifruit under the treatment of iron citrate and iron amino acid complex, which can

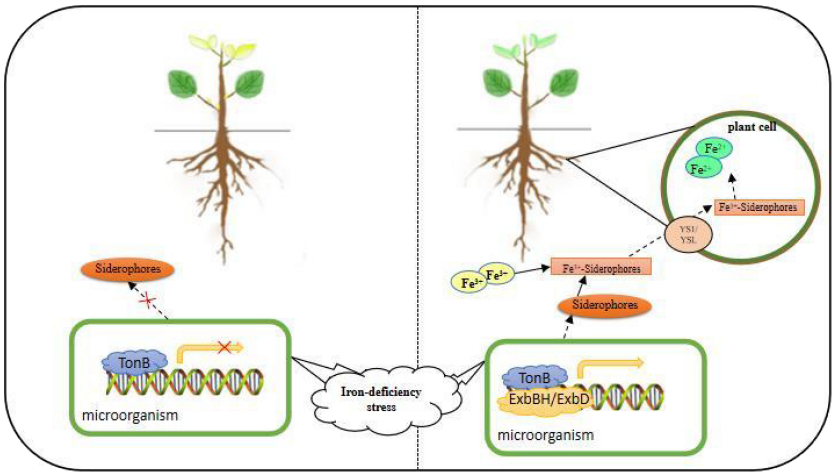


FIGURE 3
Microbial mechanisms of plant root uptake and response to Fe deficiency stress.

effectively improve the chlorophyll content and the content of vitamin C, soluble solids and whole iron in fruits and improve fruit quality (Wang et al., 2011); the foliar spray of ferric citrate balanced the sugar-acid ratio of grape berries and increased the contents of anthocyanins, flavanols, and flavonols in grape skins (Zhang et al., 2022); after the application of iron fertilizer complex containing iron citrate Fe and EDDHA-Fe, it effectively increased the content of active Fe in peanuts at the flowering hypocotyl and ripening stages and also significantly increased the content of whole Fe in leaves and kernels (Liu et al., 2016); in using different combinations of ferrous sulfate with amino acids (AA), urea (Urea), and EDTA-2Na (disodium ethylene diamine tetraacetic acid) chelated Fe for yellowing apple fruit trees. In the buried bottle and foliar spray experiments, it was found that buried bottle with foliar spray of Fe-EDTA-Urea significantly corrected apple yellowing by increasing the relative chlorophyll content, improving photosynthetic performance, and increasing iron utilization (Guo et al., 2018). Among others, inorganic iron fertilizer and ferrous sulfate can also be applied to the soil or iron fertilizer can be injected into the trunk to promote plant growth, such as using trunk injection to apply additional iron fertilizer during the budding, flowering, and fruit growth periods of gray jujube, which can improve the growth and quality of gray jujube to some extent (Song et al., 2013); using trunk injection to supplement tree iron nutrition during the fertility period of fruits can promote the growth and expansion of fruit, improve the yield of fruit trees and promote fruit quality (Ahmadi et al., 2023). Although the duration of iron fertilization application is short and the cost is relatively high, it is fast and effective.

3.2 Application of soil organic iron fertilizer/soil conditioner

Long-term application of inorganic iron fertilizer will not have an obvious effect on the treatment of plant yellowing disease, and ferrous sulfate is easy to form insoluble high iron substances in the soil, which eventually leads to the deterioration of the plant root environment. In contrast, soil organic fertilizers not only have high iron content, stable performance, and are mostly water-soluble, but also can increase the organic matter content in the soil, promote the release of soil nutrients, maintain the effectiveness of soil nutrients, improve the iron transfer capacity in plants and promote iron uptake by plants (Mohammad et al., 2018). For example, EDTA, DTPA (diacetyltriaminepentaacetic acid), HEDTA (hydroxyethyl ethylenediaminetetraacetic acid), and poultry manure can effectively increase the effective iron content in the soil, improve the symptoms of iron deficiency in plants, increase fruit weight and yield, and improve fruit quality, with obvious results in balsam pear (Ana et al., 2004).

The application of soil amendments to the soil can also improve the nutrient structure of the soil, such as applying acidifiers to the soil to lower the soil pH, thus increasing the effective iron content in the soil and thus improving plant yellowing disease (Dvořáčková et al., 2022); Si can increase the concentration of light and pigments, reduce lipid peroxidation, improve the efficiency of iron transport and utilization, and greater dry weight accumulation, and thus can be applied to reduce plant Fe deficiency symptoms by Si application (Gelza et al., 2020).

3.3 Application of microbial fertilizer

Direct application of inorganic iron fertilizer and chelate iron fertilizer has been widely used in production, but there are some problems such as high cost and easy degradation after application in soil (Nosheen et al., 2021). Microbial fertilizers are promising alternatives to chemical fertilizers and are becoming increasingly important in achieving sustainable agriculture. Biofertilization, also known as biofertilization and fungal fertilizer, can increase the

supply of plant nutrients, promote plant growth, and improve agricultural products and agro-ecological environment through the life activities of microorganisms contained in it (Nosheen et al., 2021). Biofertilizer are microbial preparations composed of PGPR (plant growth-promoting rhizobacteria) that can directly or indirectly promote plant growth by dissolving soil nutrients and producing plant growth stimulating hormones and ferritic metabolites (Aloo et al., 2022). When applied in combination with Bacillus or Bacillus like species, PGPR plays an important role in improving the growth, yield and quality of maize (Ahmad et al., 2023). Other studies have shown that plant growth promoting microorganisms produce various biomolecules, enhance the content of all the macronutrient synthesis micronutrients available in EFB (empty fruit bunches) biomass, transform them from the form that cannot be obtained by plants to the form that can be absorbed, and regulate the availability of iron in plants (Mahmud and Chong, 2021). The ground balloon mold can enhance the activity of trivalent iron reductase in the root system of Hovenia solids seedlings under alkaline conditions and reduce the production of its iron deficiency yellowing disease Fe-SPBs (siderophore producing bacterial strains) and soil fertilizer, can promote growth of peanut total iron absorption potential (Wang and Xia, 2009; Sarwar et al., 2022). Kumar et al. (2021) revealed the promise of rhizobium and plant growth-promoting rhizosphere bacterial aggregates for increasing biological yield and iron content of lentil seeds.

3.4 Crop intercropping

Intercropping is a common agricultural practice improving acquisition of nutrients including iron (Fe) (Noushin et al., 2020). There are generally two ways to change the iron nutrition of crops through intercropping. The first is to change the soil pH. For example, peanut-sesame intercropping can reduce the rhizosphere pH of peanut and increase the available iron content in soil, thus increasing the active iron content in various organs above ground of peanut and effectively improving the iron deficiency yellowing phenomenon of peanut (Wang et al., 2019). For leguminous plants, the improvement of iron nutrition can improve the formation of nodules and enhance the nitrogen fixation ability of roots (Ding et et al., 2009). The second is to secrete chelates to increase the absorption of iron in the soil. For example, grass can absorb iron from calcareous soil more effectively than dicotyledons, Iron deficiency in olives can be alleviated when they are intercropping with purple false brome and barley (Cañasveras et al., 2014). In maize-peanut intercropping, the insoluble iron in soil was chelated by wheat root acids secreted by maize roots, which improved the absorption and utilization of iron by peanuts and alleviated its yellowing (Zuo and Zhang, 2009; Dai et al., 2019).

4 Prospection

Iron plays an important role not only in human life but also in the growth and development of plants. Research on iron nutrition in plants has made great progress, and the molecular mechanisms of iron regulation in plants have also made great progress. In previous studies, the mechanisms of iron uptake in plants were divided into 2 types of iron uptake mechanisms dicotyledons and non-grass plants (Mechanism I) and grass plants (Mechanism II). However, plants do not live independently; therefore, the following studies should be enhanced in subsequent research.

Research on the genetic mechanism of plant tolerance to iron deficiency stress, continue to screen relevant regulatory genes, enrich the study of different genes on plant adaptation to iron deficiency, improve the genetic framework of plant response to iron deficiency, and actively apply biological factors such as plant ferritin and nicotinamide synthase to enhance plant iron adaptation to the environment.

Strengthen research on non-adaptive mechanisms. Plants absorb iron from soil by interacting with microorganisms. When the content of reducing iron in soil is insufficient, plants will sense the iron deficiency signal and transmit it to rhizosphere microorganisms, inducing them to secrete iron carriers for plant utilization and absorption. With the gradual deepening of research on microbial nutrition, the role of microorganisms in fertilizers has received increasing attention. The use of biological iron fertilizers can not only improve their fertilizer utilization rate, but also achieve the goal of environmental protection, ensuring the normal growth of plants while reducing fertilizer input, and protecting the environment. Currently, the production and application of microbial iron fertilizer is still in its infancy, and most studies on the amount of microbial fertilizer focus on whether it will have an impact on the environment. Subsequent research can focus on screening suitable microorganisms, such as Bacillus subtilis and soybean rhizobia, to explore the synthesis or degradation of specific molecules involved in plant iron absorption physiology (such as cellulose, hemicellulose, humic acid, and plant hormones), and to improve the mechanism by which plants absorb iron through the mechanism of microbial plant interaction through changes in microbial genes that affect plant traits, Provide a theoretical basis for production and application.

Strengthen the research application of iron deficiency mechanism in fruit arboriculture. As far as the existing research finds, the research on iron deficiency mechanism is mainly concentrated in legumes, rice, and other crops, and relatively little research is done in fruit trees. Therefore, it is important to strengthen the research on fruit trees to reduce the yield decline caused by soil iron deficiency, improve fruit quality, and enhance economic efficiency.

Continued screening of genotypic varieties against iron deficiency stress and screening of genotypic plants with high resistance to iron deficiency has become a fundamental way to overcome plant tolerance to iron deficiency stress.

Author contributions

XN and ML are the main author of the review, completing the search, collection, analysis, and writing of the first draft of relevant literature; XW and ZG are the creators and principals of the article; GH and JM participated in the writing and revision of the article. All the authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32101486), the Science and Technology Plan Project of Jiangxi Academy of Sciences (2020-GDRC-1) and Scientific research project of Jiangxi Academy of Sciences (2022YRCS003, 2022YRCC001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwala, S. C., Sharma, C. P., and Farooq, S. (1965). Effect of iron supply on growth, chlorophyll, tissue iron and activity of certain enzymes in maize and radish. *Plant Physiol.* 40 (3), 493–499. doi: 10.1104/pp.40.3.493

Ahmad, M., Hussain, A., Dar, A., Luqman, M., Ditta, A., Iqbal, Z., et al. (2023). Combating iron and zinc malnutrition through mineral biofortification in maize through plant growth promoting *Bacillus* and *Paenibacillus* species. *Front. Plant Science.* 46 (5), 782–800. doi: 10.3389/FPLS.2022.1094551

Ahmadi, H., Motesharezadeh, B., and Dadrasnia, A. (2023). Iron chlorosis in fruit stone trees with emphasis on chlorosis correction mechanisms in orchards: a review. *J. Plant Nutr.* 46 (5), 782–800. doi: 10.1080/01904167.2022.2087088

Aloo, B. N., Tripathi, V., Makumba, B. A., and Mbega, E. R. (2022). Plant growth-promoting rhizobacterial biofertilizers for crop production: the past, present, and future. *Front. Plant Science.* 13, 1002448. doi: 10.3389/FPLS.2022.1002448

Ana, Á.F., Sonia, G. M., and Juan, J. L. (2004). Evaluation of synthetic iron(III)-chelates (EDDHA/Fe^{3+}, EDDHMA/Fe^{3+} and the novel EDDHSA/Fe^{3+}) to correct iron chlorosis. *Eur. J. Agronomy.* 22 (2), 119–130. doi: 10.1016/j.eja.2004.02.001

Aoyama, T., Kobayashi, T., Takahashi, M., Nagasaka, S., Usuda, K., Kakei, Y., et al. (2009). OsYSL18 is a rice iron (III)-deoxymugineic acid transporter specifically expressed in reproductive organs and phloem of lamina joints. *Plant Mol. Biol.* 70 (6), 681–692. doi: 10.1007/s11103-009-9500-3

Borlotti, A., Vigani, G., and Zocchi, G. (2012). Iron deficiency affects nitrogen metabolism in cucumber (*Cucumis sativus l.*) plants. *BMC Plant Biol.* 12 (1), 189. doi: 10.1186/1471-2229-12-189

Bruce, T. J., Ma, J., Sudheesh, P. S., and Cain, K. D. (2021). Quantification and comparison of gene expression associated with iron regulation and metabolism in a virulent and attenuated strain of *Flavobacterium* psychrophilum. *J. fish diseases.* 44 (7), 949–960. doi: 10.1111/jfd.13354

Cailliatte, R., Schikora, A., Briat, J. F., Mari, S., and Curie, C. (2010). High-affinity manganese uptake by the metal transporter NRAMP1 is essential for *Arabidopsis* growth in low manganese conditions. *Plant Cell.* 22 (3), 904–917. doi: 10.1105/tpc.109.073023

Cañasveras, J. C., Campillo, M. C., Barrón, V., and Torrent, J. (2014). Intercropping with grasses helps to reduce iron chlorosis in olive. *J. Soil Sci. Plant Nutr.* 14 (3), 554–564. doi: 10.4067/s0718-95162014005000044

Che, J., Yokosho, K., Yamaji, N., and Ma, J. F. (2019). A vacuolar phytosiderophore transporter alters iron and zinc accumulation in polished rice grains. *Plant Physiol.* 181 (1), 276–288. doi: 10.1104/pp.19.00598

Dai, J., Qiu, W., Wang, N. Q., and Zuo, Y. M. (2019). From leguminosae/Gramineae intercropping systems to see benefits of intercropping on iron nutrition. *Front. Plant Science.* 10, 605. doi: 10.3389/fpls.2019.00605

Ding, D. K., Chai, L. J., Fu, L. N., Peng, S., and Pan, Z. Y. (2016). A novel citrus rootstock tolerant to iron deficiency in calcareous soil. *J. Am. Soc. Hortic. Science.* 141 (2), 112–118. doi: 10.21273/JASHS.141.2.112

Ding, H., Duan, L. H., Li, J., Yan, H. F., Zhao, M., Zhang, F. S., et al. (2009). Cloning and functional analysis of the peanut iron transporter AhIRT1 during iron deficiency stress and intercropping with maize. *J. Plant Physiol.* 167 (12), 996–1002. doi: 10.1016/j.jplph.2009.12.019

Dvořáčková, H., Dvořáček, J., Hueso, G. P., and Vlček, V. (2022). Effect of different soil amendments on soil buffering capacity. *PloS One* 17 (9), e0274858. doi: 10.1371/JOURNAL.PONE.0263456

Eide, D., Broderius, M., Fett, J., and Guerinot, M. L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc. Natl. Acad. Sci. United States America.* 93 (11), 5624–5628. doi: 10.1073/pnas.93.11.5624

Espen, L., Dell'Orto, M., and De Nisi, P. (2000). Metabolic responses in cucumber (*Cucumis sativus* l.) roots under fe-deficiency: a 31 p-nuclear magnetic resonance *in-vivo* study. *Planta.* 210, 985–992. doi: 10.1007/s004250050707

Fan, Q. Q., Sui, X. X., Huang, C. Y., Cui, D. Z., Li, Y. B., and Chu, X. S. (2018). Effect of iron deficiency stress on seedling-related traits in wheat. *Shandong Agric. Science.* 50 (05), 28–33. doi: 10.14083/j.issn.1001-4942.2018.05.006

Fan, Z. Y., Wu, Y. F., Zhao, L. Y., Fu, L. N., Deng, L. L., Deng, J. R., et al. (2022). MYB308-mediated transcriptional activation of plasma membrane h^+-ATPase 6 promotes iron uptake in citrus. *Horticulture Res.* 9, uhac088. doi: 10.1093/hr/uhac088

Fu, L., Zhu, Q. Q., Sun, Y. Y., Du, Y., Pan, Z. Y., and Peng, S. (2017). Physiological and transcriptional changes of three citrus rootstock seedlings under iron deficiency. *Front. Plant Science.* 8, 1104. doi: 10.3389/fpls.2017.01104

Gelza, C. M. T., Renato-de, M. P., Kamilla, S. O., Victor, D. A. D., and Gilmar-da, S. S. J. (2020). Silicon increases leaf chlorophyll content and iron nutritional efficiency and reduces iron deficiency in sorghum plants. *J. Soil Sci. Plant Nutr.* 20 (3), 1311–1320. doi: 10.1007/s42729-020-00214-0

Giovanni, F., Dimitris, P., Martino, T., Serena, F., Emeline, S., Valeria, V., et al. (2015). Ions channels/transporters and chloroplast regulation. *Cell Calcium.* 58 (1), 86–97. doi: 10.1016/j.ceca.2014.10.002

Gu, S. H., Wei, Z., Shao, Z. Y., and Jousset, A. (2020). Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nat. Microbiol.* 5 (8), 1002–1010. doi: 10.1038/s41564-020-0719-8

Guo, A. X., Wang, Y. X., Tang, J., Chen, B. H., Hu, Y., Zhu, Y. F., et al. (2018). Effect of different compound iron fertilizers on the correction of apple yellowing disease in calcareous soil. *J. Gansu Agric. University.* 53 (03), 76–82 + 89. doi: 10.13432/j.cnki.jgsau.2018.03.012

Higuchi, K., Suzuki, K., Nakanishi, H., Yamaguchi, H., Nishizawa, N. K., and Mori, S. (1999). Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. *Plant Physiol.* 119 (2), 471–480. doi: 10.1104/pp.119.2.471

Hosni, M. H., and Bryan, E. T. (2013). Transcriptional regulation by ferric uptake regulator (Fur) in pathogenic bacteria. *Front. Cell. Infection Microbiol.* 3, 59. doi: 10.3389/fcimb.2013.00059

Huei-hsuan, t., and Wolfgang, S. (2020). Biotechnology - genomics; new findings on genomics from institute of plant and microbial biology summarized (pH-dependent transcriptional profile changes in iron-deficient *Arabidopsis* roots). *Biotech. Week.* 21, 694. doi: 10.1186/s12864-020-07116-6

Inoue, H., Kobayashi, T., Nozoye, T., Takahashi, M., Kakei, Y., Suzuki, K., et al. (2009). Rice OsYSL15 is an iron-regulated iron (III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in the early growth of the seedlings. *J. Biol. Chem.* 284 (6), 3470–3479. doi: 10.1074/jbc.m806042200

Ishimaru, Y., Masuda, H., Bashir, K., Inoue, H., Tsukamoto, T., Takahashi, M., et al. (2010). A rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. *Plant journal: Cell Mol. Biol.* 62 (3), 379–390. doi: 10.1111/j.1365-313x.2010.04158.x

José, I. C., Cristián, R., Silvia, D., and Claudio, P. (2016). Contrasting physiological responses to iron deficiency in *Cabernet sauvignon* grapevines grafted on two rootstocks. *Scientia Horticulturae.* 199, 1–8. doi: 10.1016/j.scienta.2015.12.013

Jurca, M., Sjölander, J., Ibáñez, C., Matrosova, A., Johansson, M., Kozarewa, I., et al. (2022). ZEITLUPE promotes ABA-induced stomatal closure in *Arabidopsis* and *Populus*. *Front. Plant Science.* doi: 10.3389/FPLS.2022.829121

Kakei, Y., Ishimaru, Y., Kobayashi, T., Yamakawa, T., Nakanishi, H., Nishizawa, N. K., et al. (2012). OsYSL16 plays a role in the allocation of iron. *Plant Mol. Biol.* 79 (6), 583–594. doi: 10.1007/s11103-012-9930-1

Kobayashi, T., Maeda, K., Suzuki, Y., and Nishizawa, N. K. (2022b). Simultaneous enhancement of iron deficiency tolerance and iron accumulation in rice by combining the knockdown of OsHRZ ubiquitin ligases with the introduction of engineered ferric-chelate reductase. *Rice.* 15 (1), 54. doi: 10.1186/s12284-022-00598-w

Kobayashi, T., Shinkawa, H., Nagano, A. J., and Nishizawa, N. K. (2022a). The basic leucine zipper transcription factor OsbZIP83 and the glutaredoxins OsGRX6 and OsGRX9 facilitate rice iron utilization under the control of OsHRZ ubiquitin ligases. *Plant journal: Cell Mol. Biol.* 110 (6), 1731–1750. doi: 10.1111/TPJ.15767

Konrad, M. (1994). Iron availability in plant tissues – iron chlorosis on calcareous soils. *Plant Soil.* 165 (2), 275–283. doi: 10.1007/bf00008070

Krishna, T. P. A., Ceasar, S. A., and Maharajan, T. (2023). Biofortification of crops to fight anemia: role of vacuolar iron transporters. *J. Agric. Food Chem.* 71 (8), 3583–3598. doi: 10.1021/ACS.JAFC.2C07727

Kumar, A., Jha, M. N., Singh, D., Pathak, D., Rajawat, M., and Vikram, S. (2021). Prospecting catabolic diversity of microbial strains for developing microbial consortia and their synergistic effect on lentil (*Lens esculenta*) growth, yield and iron biofortification. *Arch. Microbiol.* 203 (8), 4913–4928. doi: 10.1007/S00203-021-02446-9

Li, Z. X. (2003). Effect of iron stress on physiological and biochemical characteristics of apple rootstocks (Hebei Agricultural University).

Li, Q., Chen, L., and Yang, A. (2019). The molecular mechanisms underlying iron deficiency responses in rice. *Int. J. Mol. Sci.* 21 (1), 43. doi: 10.3390/ijms21010043

Liang, G. (2022). Iron uptake, signaling, and sensing in plants. *Plant Commun.* 3 (5), 100349. doi: 10.1016/j.xplc.2022.100349

Liu, W. L., Chen, H. N., Chen, H. K., Li, Y. P., and Zhang, L. (2016). Study on the effect of compound iron fertilizer to correct iron deficiency yellowing in peanuts. *Phosphate fertilizer and compound fertilizers* (04), 46–48.

Liu, W., Karemera, N. J. U., Wu, T., Yang, Y. F., Zhang, X. Z., Xu, X. F., et al. (2017). The ethylene response factor AtERF4 negatively regulates the iron deficiency response in *Arabidopsis* thaliana. *PloS One* 12 (10), e0186580. doi: 10.1371/journal.pone.0186580

Lowe, N. M. (2021). The global challenge of hidden hunger: perspectives from the field. *Proc. Nutr. Society.* 80 (3), 11–14. doi: 10.1017/s0029665121000902

Mahmud, M. S., and Chong, K. P. (2021). Formulation of biofertilizers from oil palm empty fruit bunches and plant growth-promoting microbes: a comprehensive and novel approach towards plant health. *J. King Saud Univ. - Science.* 33 (8), 101647. doi: 10.1016/J.JKSUS.2021.101647

Maribela, P., Pedro, J. C., Teresa, S., Florinda, G., Anunciación, A., and Amarilis, D. V. (2012). Development and recovery of iron deficiency by iron resupply to roots or leaves of strawberry plants. *Plant Physiol. Biochem.* 53, 1–5. doi: 10.1016/j.plaphy.2012.01.001

Mary-Rus, M.-C. M., Ángeles, F.-G., Iglesias., D. J., and Francisco, L. (2013). Strategy I responses to fe-deficiency of two citrus rootstocks differing in their tolerance to iron chlorosis. *Scientia Horticulturae.* 153, 56–63. doi: 10.1016/j.scienta.2013.01.009

Masalha, J., Kosegarten, H., Elmaci, Ö., and Mengel, K. (2000). The central role of microbial activity for iron acquisition in maize and sunflower. *Biol. Fertility Soils.* 30 (5-6), 433–439. doi: 10.1007/s003740050021

Mohammad, K. S., Moin, N., and Maryam, A. (2018). Effect of fe-glycine aminochelate on pod quality and iron concentrations of bean (*Phaseolus vulgaris l.*) under lime soil conditions. *Commun. Soil Sci. Plant Analysis.* 49 (2), 215–224. doi: 10.1080/00103624.2017.1421655

Motofumi, S., Kendi, C. M., Hiromi, N., Michiko, T., Masahiko, S., Satoshi, M., et al. (2008). Transgenic rice lines that include barley genes have increased tolerance to low iron availability in calcareous paddy soil. *Soil Sci. Plant Nutr.* 54 (1), 77–85. doi: 10.1111/j.1747-0765.2007.00205.x

Norman, T., and Javier, A. (1986). Function of iron in chloroplasts. *J. Plant Nutr.* 9 (3-7), 609–646. doi: 10.1080/01904168609363470

Nosheen, S., Ajmal, I., and Song, Y. (2021). Microbes as biofertilizers, a potential approach for sustainable crop production. *Sustainability.* 13 (4), 1868. doi: 10.3390/SU13041868

Noushin, S., Roghieh, H., Narges, M., and Charlotte, P. (2020). Growth enhancement of *Brassica napus* under both deficient and adequate iron supply by intercropping with *Hordeum vulgare*: a hydroponic study. *Plant Biosyst. - Int. J. Dealing all Aspects Plant Biol.* 155 (3), 632–646. doi: 10.1080/11263504.2020.1769215

Nozoye, T., Nagasaka, S., Kobayashi, T., Sato, Y., Uozumi, N., Nakanishi, H., et al. (2015). The phytosiderophore efflux transporter TOM2 is involved in metal transport in rice. *J. Biol. Chem.* 290 (46), 27688–27699. doi: 10.1074/jbc.m114.635193

Obi, C. D., Bhuiyan, T., Dailey, H. A., and Medlock, A. E. (2022). Ferrochelatase: mapping the intersection of iron and porphyrin metabolism in the mitochondria. *Front. Cell Dev. Biol.* 10, 894591. doi: 10.3389/fcell.2022.894591

Okada, S., Lei, G. J., Yama, J. N., Huang, S., Ma, J. F., Mochida, K., et al. (2022). FE UPTAKE INDUCING PEPTIDE1 maintains fe translocation by controlling fe deficiency response genes in the vascular tissue of *Arabidopsis*. *Plant Cell environment.* 45 (11), 3322–3337. doi: 10.1111/pce.14424

Pei, D., Hua, D. P., Deng, J. J., Wang, Z. F., Song, C. P., Wang, Y., et al. (2022). Phosphorylation of the plasma membrane h^+-ATPase AHA2 by BAK1 is required for ABA-induced stomatal closure in *Arabidopsis*. *Plant Cell.* 34 (7), 2708–2729. doi: 10.1093/plcell/koac106

Philip, H., and Leonid, A. S. (2005). Organization of iron-sulfur clusters in respiratory complex I. *Science*. 309 (5735), 771–774. doi: 10.1126/science.1113988

Qin, L., Han, P. P., Chen, L. Y., Walk, T. C., Li, Y. S., Hu, X. J., et al. (2017). Genome-wide identification and expression analysis of *NRAMP* family genes in soybean (*Glycine max* l.). *Front. Plant Science.* 8, 1436. doi: 10.3389/fpls.2017.01436

Qiu, T., and Liu, S. M. (2022). Progress in the functional mechanism of iron ion signaling in plant pathogenic fungi. *Mol. Plant Breed.* 1–15.

Qu, S. C. (2005). Identification of trans-tomato iron carrier protein gene (LeIRT2) in begonia macrocarpa. [master's thesis]. *Nanjing Agric. University.*

Rajniak, J., Giehl, R. F. H., Chang, E., Murgia, I., von-Wirén, N., and Sattely, E. S. (2018). Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nat. Chem. Biol.* 14 (5), 442–450. doi: 10.1038/s41589-018-0019-2

Ranferi, M. T., Jorge, D. E.-B., Gabriel, A.-G., and Maria, T. C.-L. (2006). Morphological changes in leaves of mexican lime affected by iron chlorosis. *J. Plant Nutr.* 29 (4), 615–628. doi: 10.1080/01904160600564337

Robe, K., Conejero, G., Gao, F., LefebvreLegendre, L., SylvestreGonon, E., Rofidal, V., et al. (2020a). Coumarin accumulation and trafficking in *Arabidopsis* thaliana: a complex and dynamic process. *New phytologist.* 229 (4), 2062–2079. doi: 10.1111/nph.17090

Robe, K., Izquierdo, E., Vignols, F., Rouached, H., and Dubos, C. (2020b). The coumarins: secondary metabolites playing a primary role in plant nutrition and health. *Trends Plant Science.* 26 (3), 248–259. doi: 10.1016/J.TPLANTS.2020.10.008

Rungin, S., Indananda, C., Suttiviriya, P., Kruasuwan, W., Jaemsaeng, R., and Thamchaipenet, A. (2012). Plant growth-enhancing effects by a siderophore-producing endophytic streptomycete isolated from a Thai jasmine rice plant (*Oryza sativa* l. cv. KDML105). *Antonie van Leeuwenhoek.* 102 (3), 463–472. doi: 10.1007/s10482-012-9778-z

Sandmann, G. (1985). Consequences of iron deficiency on photosynthetic and respiratory electron transport in blue-green algae. *Photosynthesis Res.* 6 (3), 261–271. doi: 10.1007/BF00049282

Santi, S., Locci, G., Monte, R., Pinton, R., and Varanini, Z. (2003). Induction of nitrate uptake in maize roots: expression of a putative high-affinity nitrate transporter and plasma membrane h^+-ATPase isoforms. *J. Exp. botany.* 54 (389), 1851–1864. doi: 10.1093/jxb/erg208

Sarwar, S., Khaliq, A., Yousra, M., and Sultan, T. (2022). Iron biofortification potential of siderophore producing rhizobacterial strains for improving growth, yield and iron contents of groundnut. *J. Plant Nutr.* 45 (15), 2332–2347. doi: 10.1080/01904167.2022.2063733

Satbhai, S. B., Setzer, C., Freynschlag, F., Slovak, R., Kerdaffrec, E., and Busch, W. (2017). Natural allelic variation of FRO2 modulates *Arabidopsis* root growth under iron deficiency. *Nat. Commun.* 8 (1), 15603. doi: 10.1038/ncomms15603

Satoshi, M. (1999). Iron acquisition by plants. *Curr. Opin. Plant Biol.* 2 (3), 250–253. doi: 10.1016/S1369-5266(99)80043-0

Senoura, T., Kobayashi, T., An, G., Nakanishi, H., and Nishizawa, N. K. (2020). Defects in the rice aconitase-encoding *OsACO1* gene alter iron homeostasis. *Plant Mol. Biol.* 104, 629–645. doi: 10.1007/s11103-020-01065-0

Shamima, P., Ranjeet, K. R., Anjali, A., and Bhupinder, S. (2018). Combined deficiency of nitrogen and iron increases senescence-induced remobilization of plant immobile iron in wheat. *Acta Physiologiae Plantarum.* 40 (12), 211. doi: 10.1007/s11738-018-2782-9

Singh, A. K., Li, H., Bono, L., and Sherman, L. A. (2005). Novel adaptive responses revealed by transcription profiling of a *Synechocystis* sp. PCC 6803 delta-isiA mutant in the presence and absence of hydrogen peroxide. *Photosynthesis Res.* 84 (1-3), 65–70. doi: 10.1007/s11120-004-6429-x

Song, H., Chen, F., Wu, X., Hu, M., Geng, Q. L., Ye, M., et al. (2022). *MNB1* gene is involved in regulating the iron-deficiency stress response in *Arabidopsis* thaliana. *BMC Plant Biol.* 22 (1), 151. doi: 10.1186/s12870-022-03553-5

Song, F. H., Zhang, P., and Shi, Y. J. (2013). Effect of trunk injection of iron fertilizer on yield and quality of xinjiang gray jujube. *Northwest J. Agriculture.* 06), 102–107.

Sun, Y., Li, J. Q., Yan, J. Y., Yuan, J. J., Li, G. X., Wu, Y. L., et al. (2020). Ethylene promotes seed iron storage during *Arabidopsis* seed maturation *via* the ERF95 transcription factor. *J. Integr. Plant Biol.* 62 (08), 1193–1212. doi: 10.1111/jipb.12986

Sun, X. P., Wang, S. Y., Tong, Y. A., and Faust, M. (1987). Metabolic changes in iron-deficient apple seedlings. *J. Plant Nutr.* 10 (9–16), 1021–1030. doi: 10.1080/01904168709363630

Sun, L., Wei, Y. Q., Wu, K. H., Yan, J. Y., Xu, J. N., Wu, Y. R., et al. (2021a). Restriction of iron loading into developing seeds by a YABBY transcription factor safeguards successful reproduction in *Arabidopsis*. *Mol. Plant* 14 (10), 1624–1639. doi: 10.1016/j.molp.2021.06.005

Sun, W. J., Zhang, J. C., Ji, X. L., Feng, Z. Q., Wang, X., Huang, W. J., et al. (2021b). Low nitrate alleviates iron deficiency by regulating iron homeostasis in apple. *Plant Cell environment.* 44 (6), 1869–1884. doi: 10.1111/PCE.13998

Takahashi, K., Hayashi, K. I., and Kinoshita, T. (2012). Auxin activates the plasma membrane h+-ATPase by phosphorylation during hypocotyl elongation in *Arabidopsis*. *Plant Physiol.* 159 (2), 632–641. doi: 10.1104/pp.112.196428

Takanori, K., Tomoko, N., and Naoko, K. N. (2018). Iron transport and its regulation in plants. *Free Radical Biol. Med.* 133, 11–20. doi: 10.1016/j.freeradbiomed.2018.10.439

Tissot, N., Robe, K., Gao, F., Grant-Grant, S., Boucherez, J., Bellegarde, F., et al. (2019). Transcriptional integration of the responses to iron availability in *Arabidopsis* by the bHLH factor ILR3. *New phytologist.* 223 (3), 1433–1446. doi: 10.1111/nph.15753

Truchon, A. N., Hendrich, C. G., Bigott, A. F., Dalsing, B. L., and Allen, C. (2022). NorA, HmpX, and NorB cooperate to reduce NO toxicity during denitrification and plant pathogenesis in ralstonia solanacearum. *Microbiol. spectrum.*, e0026422. doi: 10.1128/SPECTRUM.00264-22

Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M. L., Briat, J. F., et al. (2021). CORRECTION: IRT1, an *Arabidopsis* transporter essential for iron uptake from the soil and plant growth. *Plant Cell.* 33 (2), 439–440. doi: 10.1093/plcell/koaa033

Victoria, F., Thomas, E., Víctor, D. R., and Javier, A. (2008). Leaf structural changes associated with iron deficiency chlorosis in field-grown pear and peach: physiological implications. *Plant Soil.* 311 (1/2), 161–172. doi: 10.1007/s11104-008-9667-4

Wang, G. Z., Han, H. T., Che, J. X., Zhang, X. D., and Zai, B. N. (2011). Effect of different iron preparations on the correction of iron deficiency yellowing in kiwifruit under calcareous soil conditions. *J. Fruit Science.* 01), 61–65. doi: 10.13925/j.cnki.gsxb.2011.01.009

Wang, N. N., Jiao, X. Y., Guo, T. L., Li, C. L., Liu, Z. D., and Ma, F. W. (2019). Time course of physiological responses in kiwifruit induced by bicarbonate. *Trees.* 33 (6), 1711–1722. doi: 10.1007/s00468-019-01892-9

Wang, F., Sun, Z. G., Jiao, N. Y., Xu, Y. H., Chen, Y., Ping, F., et al. (2019). Effects of intercropping between sesame and peanut on photosynthetic fluorescence characteristics of sesame functional leaves. *Chin. J. Appl. Ecology.* 30 (11), 3787–3794. doi: 10.13287/j.1001-9332.201911.014

Wang, B. L., Wei, H. F., Chen, Z., Li, Y. T., and Zhang, W. H. (2022). Carbonate-induced chemical reductants are responsible for iron acquisition in strategy I wild herbaceous plants native to calcareous grasslands. *Plant Cell Physiol.* 63 (6), 770–784. doi: 10.1093/pcp/pcac038

Wang, M. Y., and Xia, R. X. (2009). Effects of mycorrhizal fungi on the growth and iron uptake of *Citrus aurantium* at different pH values. *Acta Microbiologica Sinica.* 49 (10), 1374–1379. doi: 10.13343/j.cnki.wsxb.2009.10.011

Waters, B. M., Blevins, D. G., and Eide, D. J. (2002). Characterization of FRO1, a pea ferric-chelate reductase involved in root iron acquisition. *Plant Physiol.* 129 (1), 85–94. doi: 10.1104/pp.010829

Xiao, F. Z., Qi, W., Yu, T. M., Ye, T., and Ren, F. S. (2020). AtHAP5A regulates iron translocation in iron-deficient *Arabidopsis* thaliana. *J. Integr. Plant Biol.* 26 (12), 1910–1924. doi: 10.1111/jipb.12984

Yadavalli, V., Jolley, C. C., Malleda, C., Thangaraj, B., Fromme, P., and Subramanyam, R. (2012). Alteration of proteins and pigments influences the function of photosystem I under iron deficiency from chlamydomonas reinhardtii. *PloS One* 7 (4), e35084. doi: 10.1371/journal.pone.0035084

Yamagata, A., Murata, Y., Namba, K., Terada, T., Fukai, S., and Shirouzu, M. (2022). Uptake mechanism of iron-phytosiderophore from the soil based on the structure of yellow stripe transporter. *Nat. Commun.* 13 (1), 7180. doi: 10.1038/s41467-022-34930-1

Yang, H. B., Wei, H., Ma, G. J., Antunes, M. S., Vogt, S., Cox, J., et al. (2016). Cell wall targeted in planta iron accumulation enhances biomass conversion and seed iron concentration in *Arabidopsis* and rice. *Plant Biotechnol. J.* 14 (10), 1998–2009. doi: 10.1111/pbi.12557

Yao, Z. C., Hao, W. T., Wang, Y. J., Chen, Z. P., Cao, S. Q., and Jiang, L. (2022). Loss-of-function mutations in the ERF96 gene enhance iron-deficient tolerance in *Arabidopsis*. *Plant Physiol. Biochem.* 175, 1–11. doi: 10.1016/j.plaphy.2022.02.002

Ye, F., Albarouki, E., Lingam, B., Deising, H. B., and von-Wirén, N. (2014). An adequate fe nutritional status of maize suppresses infection and biotrophic growth of colletotrichum graminicola. *Physiologia Plantarum.* 151 (3), 280–292. doi: 10.1111/ppl.12166

Yi, C. L., Wang, H., Zhang, F. S., and Li, C. J. (1998). Differences in adaptive responses of cucumber, tomato, and soybean to iron deficiency stress. *Chin. J. Plant Ecology.* 06), 80–86.

Yoshida, K., Seger, A., Corkill, M., Heil, P., Karsh, K., McMinn, A., et al. (2021). Low fe availability for photosynthesis of Sea-ice algae: ex situ incubation of the ice diatom fragilariopsis cylindrus in low-fe sea ice using an ice tank. *Front. Mar. Science.* 8. doi: 10.3389/FMARS.2021.632087

Zeng, T., Li, C. L., Song, J. L., Zhou, A. T., Wu, Y. X., and He, T. M. (2019). Effect of iron deficiency treatment on leaf growth and photosynthetic characteristics of different pear lines. *Shandong Agric. Sci.* 51 (10), 91–95. doi: 10.14083/j.issn.1001-4942.2019.10.017

Zhang, J. L. (2020). Study on the effects of iron stress on the growth and outcome of two hydroponic strawberry species. [master's thesis] (Southwest University). doi: 10.27684/d.cnki.gxndx.2020.000235

Zhang, S., Chen, H., Gao, M., Gu, C. F., and Wang, R. (2022). Effects of different iron treatments on wine grape berry quality and peel flavonoid contents. *Food Sci. Nutr.* 10 (11), 3598–3607. doi: 10.1002/fsn3.2957

Zhang, X. Y., Jia, X. M., Zhang, R., and Wang, Y. X. (2019). Metabolic analysis in malus halliana leaves in response to iron deficiency. *Scientia Hortic.* 258, 108792. doi: 10.1016/j.scienta.2019.108792

Zhang, L. L., Liu, X. X., Zhu, Y. X., and Jin, C. W. (2021). Research progress on iron uptake, translocation, and their signaling pathways in strategy I plant. *J. Plant Nutr. Fertilizers.* 07), 1258–1272.

Zhou, H. J., Korcak, R. F., Wergin, W. P., Fan, F., and Faust, M. (2008). Cellular ultrastructure and net photosynthesis of apple seedlings under iron stress. *J. Plant Nutr.* 7 (6), 911–928. doi: 10.1080/01904168409363254

Zuo, Y., and Zhang, F. (2009). Iron and zinc biofortification strategies in dicot plants by intercropping with gramineous species. a review. *Agron. Sustain. Dev.* 29 (1), 63–71. doi: 10.1051/agro:2008055

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 July 2023
DOI 10.3389/fpls.2023.1223782

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Aamir Manzoor,
School of Agriculture and Biology,
Shanghai Jiao Tong University, China

REVIEWED BY
Fasih Ullah Haider,
South China Botanical Garden (CAS), China
Irfan Ullah Khan,
Jiangsu University, China
Zoia Arshad Awan,
Teagasc Food Research Center, Ireland

*CORRESPONDENCE
Shimei Wang
wangshimei0551@163.com
Qian Zhu
zhuqianwalk@163.com

†These authors have contributed equally to this work

RECEIVED 16 May 2023
ACCEPTED 04 July 2023
PUBLISHED 25 July 2023

CITATION
Yi Y, Hassan MA, Cheng X, Li Y, Liu H, Fang W, Zhu Q and Wang S (2023) QTL mapping and analysis for drought tolerance in rice by genome-wide association study. *Front. Plant Sci.* 14:1223782. doi: 10.3389/fpls.2023.1223782

COPYRIGHT
© 2023 Yi, Hassan, Cheng, Li, Liu, Fang, Zhu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

QTL mapping and analysis for drought tolerance in rice by genome-wide association study

Yueming Yi[1,2†], Muhammad A. Hassan[1,3†], Xinxin Cheng[2], Yiru Li[1,2], Huan Liu[1,3], Wuyun Fang[1,3], Qian Zhu[1,3*] and Shimei Wang[1,3*]

[1]Rice Research Institute, Anhui Academy of Agricultural Sciences, Hefei, China, [2]College of Agriculture, Anhui Science and Technology University, Fengyang, China, [3]Key Laboratory of Rice Genetics and Breeding in Anhui Province, Hefei, China

Rice drought resistance is a complicated quantitative feature involving a range of biological and agronomic variables, but little is known about the underlying genetics and regulatory mechanisms that regulate drought tolerance. This study used 120 recombinant inbred lines (RILs), derived from a cross between drought tolerant Lvhan 1 and susceptible Aixian 1. The RILs were subjected to drought stress at the first ear stage, and phenotypic data of 16 agronomic and physiological traits under varying conditions were investigated. Genome-wide association study (GWAS) on the drought resistance index of traits was carried out. A total of 9 quantitative trait loci (QTLs) associated with drought-related traits were identified on chromosomes 2, 6, 7, 8, 9, and 10, which includes QTLs for plant height (PH) *qPH10.1*, effective panicles number (EPN) *qEPN6.1*, panicle length (PL) *qPL9.1*, thousand-grain weight (TGW) *qTGW2.1*, *qTGW6.1*, *qTGW8.1*, leaf length (LL) *qLL7.1*, leaf width (LW) *qLW7.1*, and leaf area (LA) *qLA7.1*. The fraction of phenotypic variation explained by individual QTL varied from 10.6% to 13.9%. Except for days to flowering (DTF), the mean values of all traits under normal water management conditions were considerably higher than those under drought conditions. Except for the DTF, the drought resistance index of all rice traits was less than 1, indicating that drought treatment reduced the EPN, FGPP, SSR, PH, and LA, which affected the growth and development of rice. The drought resistance index of DTF was 1.02, indicating that drought prolonged the heading time of rice and diminish the yield parameters. Along with identifying QTLs, the results also predicted ten candidate genes, which are directly or indirectly involved in various metabolic functioning related to drought stress. The identification of these genomic sites or QTLs that effectively respond to water scarcity will aid in the quest of understanding the drought tolerance mechanisms. This study will facilitate the marker-assisted rice breeding and handy in the breeding of drought-tolerant rice varieties.

KEYWORDS

rice, drought tolerance, agronomic yield traits, genome-wide association studies, quantitative trait loci (QTL)

1 Introduction

Rice (*Oryza sativa* L.) is an important grain crop, and water is crucial for maintaining optimal growth and development. It has been estimated that about 3000 liters of irrigation is needed to produce 1 kg of rice grains (Oladosu et al., 2019). However, in recent years, global warming, water scarcity, and frequent seasonal drought spells seriously restrict the development of China's agriculture (Hassan et al., 2021). Rice's development, growth, and physiological processes were severely hampered by drought (Kumar et al., 2017; Ikmal et al., 2019; Yang et al., 2021). According to Dixit et al. (2017) it is imperative to investigate the potential to improve rice productivity in facing the challenge of limited water resources. It is the need of the hour to develop rice cultivars that can withstand drought stress without compromising their yield potential (Wang & Qin, 2017). Multiple genes control rice drought tolerance. Combining rice drought tolerance gene mining with molecular marker-assisted breeding technology is beneficial for developing rice varieties resistant to drought, increasing rice yield, conserving water, and preserving national food security.

The ability of rice to withstand drought is a complicated quantitative feature regulated by numerous quantitative trait loci (QTLs). Although researchers have discovered thousands of QTLS, few have been successfully combined with breeding (Liang et al., 2021; Wang et al., 2021). Traditional QTL mapping techniques are incapable of locating genes properly and efficiently. Genome-wide association study (GWAS) has become a popular tool for breeding rice with the advancement of biotechnology. The basis of GWAS is the population's linkage disequilibrium (LD) and single nucleotide polymorphisms (SNPs) in the genome. An analytical method that combines population structure, genome-wide LD level, and phenotypic data to identify the relationship between target traits and genetic markers/candidate genes within a population. Its advantages include high accuracy, rapid processing, and no construction of population-based mapping. GWAS provides an efficient method and approach for studying the genetic mechanism of rice drought resistance and mining potential drought resistance genes (Li et al., 2017; Bhandari et al., 2020). Wang et al. (2020) performed GWAS investigation on 272 indica materials, analyzed source-sink relationships and yield-related variables, and identified 70 QTLs influencing 11 related traits. Guo et al. (2018) used 507 diverse rice varieties to conduct a genome-wide association analysis on 51 image qualities and traditional parameters like green holding and yield, and they discovered 470 loci related to drought resistance. Additionally, employing RIL populations for GWAS and linkage analysis, 69 image trait association loci were also found. It has also been proved that some image traits and related genes can be used for drought resistance improvement in the field. Ma et al. (2016) used GWAS to find 29 QTLS related to plant height, yield, and drought resistance index in 270 cultivars. He also found a candidate gene for drought resistance, *OsRLK5*, vital in increasing rice productivity under drought stress. The GWAS based on deep sequencing is useful for detecting genetic variation in rice drought resistance enhancement. Several genes for rice drought resistance have been cloned and examined recently, including *OsMYB6* (Tang et al., 2019), *DROT1* (Sun et al., 2022), and *OsRINGzf1* (Chen et al., 2022), which have demonstrated positive benefits in controlling rice drought tolerance. However, they have not been utilized in developing new, drought-resistant rice cultivars.

In brief, GWAS has been employed to research rice attributes linked to drought resistance and to examine the genetic basis of rice drought resistance. This study constructed a recombinant inbred line population of indica rice line Lvhan 1 and drought-sensitive japonica strain Aixian 1, which had been verified in production. The population's agronomic, physiological, and other relevant traits were examined under various water stress conditions, and GWAS was used to identify the main effect of QTLs associated with drought resistance. The results of this study provided a foundation for breeding and enhancing high-yielding and high-quality rice cultivars that are drought-resistant.

2 Material and methods

2.1 Test materials and location

The experiment was conducted in 2021 at the Lujiang Base of the Rice Research Institute of Anhui Academy of Agricultural Sciences, using the recombinant inbred line population of drought-tolerant rice Lvhan 1 and drought-sensitive rice Aixian 1, including 120 lines with significant differences in drought resistance of F_{10}.

2.2 Test methods

2.2.1 Experiment design

Conventional and drought water treatments were designed for the experiment and planted in a mobile greenhouse and an open-air field. The experiment was conducted using a randomized complete block design (RCBD). The rice seeds were sown on May 30, 2021, and transplanted on June 22, 2021. Each plot had three rows with ten plants each, with a plant and row spacing of 20 cm and 26 cm, respectively. Each water treatment was replicated three times. The open-air field is managed according to daily field management; the drought management field is 1 m higher than the ground and covered with portable greenhouses, which are managed in the open air on sunny and rainy days. The soil water potential is maintained at -15 kPa ~ 0 kPa. When the soil water potential is lower than -15 kPa, immediately supplementary irrigation is applied, and irrigation is stopped at the first ear stage of rice and subjected to drought stress. When the leaf wilting reached 50%, or the soil water potential dropped below -50 kPa for more than five days, irrigation was resumed until harvesting maturity. Table 1 shows the weather conditions during the course of the experiment.

2.2.2 Investigation of agronomic traits

The following agronomic traits were investigated:

TABLE 1 Weather conditions during the cropping season 2020-2021.

Months	Temperature			Relative humidity	Total rainfall	Total sunshine	Mean sunshine
	Max.	Min.	Mean				
	°C	°C	°C	%	mm	h	h
May	27.1	18.7	22.9	58.0	173.5	150.4	4.9
June	30.5	22.9	26.7	59.3	100.6	146.3	4.9
July	31.5	25.3	28.4	66.3	213.1	130.1	4.2
August	30.8	24.3	27.6	68.0	354.5	131.1	4.3
September	31.1	22.3	26.7	56.9	91.2	182.9	6.1
October	21.8	13.9	17.9	59.2	73.8	125.0	4.0

Source: Meteorological station of Lujiang county, Anhui, China.
Here, mm, milli meter; d, days; h, hours.

(1) Tiller number (TN): Five plants were chosen as observation points in every treatment plot. The tiller dynamics were investigated, and the effective tiller number was recorded at the heading stage.

(2) Days to flowering (DTF): The number of days from sowing to flowering was counted.

(3) Agronomic traits of leaves: After one week of drought treatment, each plot's middle row was chosen to determine chlorophyll content (CC). The CC values in the upper, middle, and base parts of the uppermost three main stem leaves were measured by handheld chlorophyll SPAD apparatus, and the average value was calculated. Leaf length (LL), leaf width (LW), leaf area (LA), and leaf aspect ratio (LAR) were measured by the YMJ-A leaf area measuring instrument.

(4) Yield traits: Five plants were selected from each plot in each treatment at harvesting maturity to examine traits associated with yield, such as plant height (PH), effective panicles number (EPN), panicle length (PL), total grains per panicle (GPP), filled grains per panicle (FGPP), seed setting rate (SSR), thousand-grain weight (TGW), grain yield plant (GYP), and aboveground biomass (AB) which was measured by drying individual plants at 80°C till constant weight. Supplementary File 1 shows the raw data of all above mentioned phenotypic traits.

2.2.3 Calculation of drought resistance index

The drought-resistance index (DRI) was calculated using the following formula.

DRI = [measured value under drought stress/measured value under normal irrigation] ×100%.

The drought-resistance index of all traits was calculated for association analysis.

2.2.4 Genotype identification

The population genotype identification was carried out by Huazhi biotechnology limited company. For this, a random leaf sample from one plant was collected for each RIL to extract DNA. Repeated PCR amplifications were performed using the Huazhi 1K rice SNP chip. The amplified fragments were obtained and constructed with a sequencing kit from Beijing Genomics Institution (BGI) and sequenced in the BGI MGISEQ-T7 sequencing machine. All experiments were conducted following the standard procedures of the sequencing kit. Supplementary File 2 shows the raw data of genotypes in each RIL.

2.3 Data analysis

Excel 2018 and SPSS 26 were used for trait description and correlation analysis. The sequencing data were compared with the reference genome of Nipponbare to screen SNPs with polymorphism among parents. The correlation analysis combined the SNP marker obtained and the drought resistance index of drought resistance-related traits. The genotype data was filtered by TASSEL (v5.2.24) software first, removing the heterozygote sites, the deletion rate was less than 10%, and the genotype frequency was set to be 0.05-1.0. A total of 3550 high-quality SNP markers were selected from 5429 SNPS, and association analysis was performed using a mixed linear model(MLM)combined with Scaled IBS Kinship, genotype, and phenotypic data. The corresponding value of the observed log10p is used as the P value to plot the Q-Q scatter plot and the Manhattan plot. For the interval with a p-value less than 0.001, QTL was considered to exist in the region. The phenotypic contribution rate (R^2) was used to evaluate the overall contribution rate of associated QTL to the drought resistance index of phenotypic traits. Refer to the reference for QTL naming principles (Yonemaru et al., 2010).

3 Results and analysis

3.1 Genetic analysis of population traits

The drought-resistance index of each agronomic character of Lvhan 1 and Aixian 1 was significantly different. The drought resistance index of each character of Aixian 1 was mostly lower than that of Lvhan 1 (Figures 1, 2). Sixteen drought-resistant

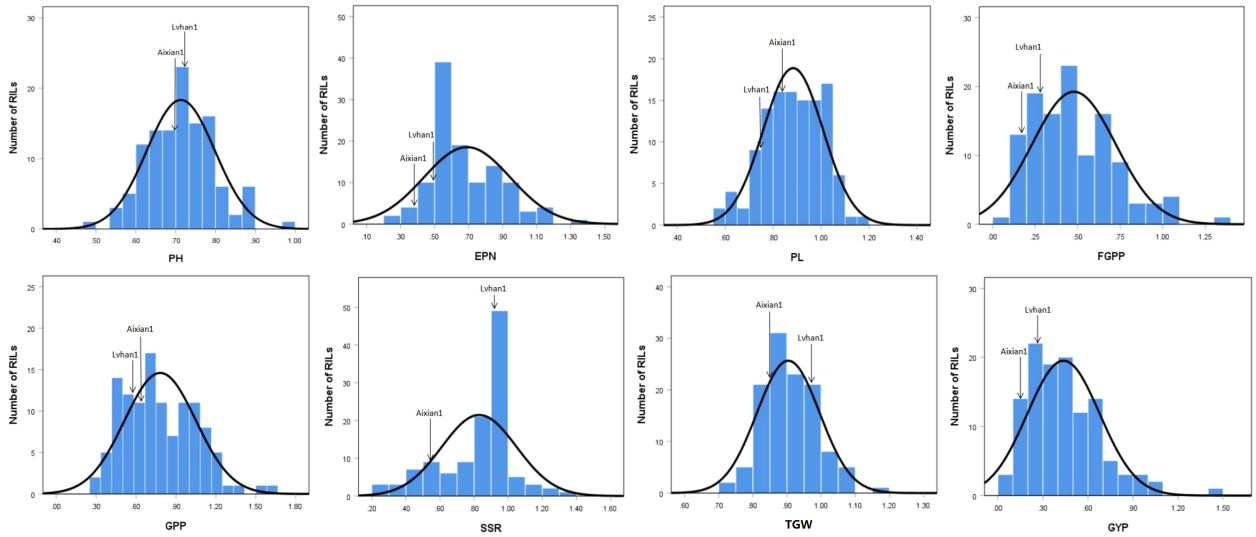


FIGURE 1
Histogram of the drought-resistance index of drought-related traits in the RIL population. PH=plant height, EPN=effective panicles number, PL=panicle length, FGPP=filled grains per panicle, GPP=grains per panicle, SSR=seed setting rate, TGW= thousand-grain weight, GYP=grain yield plant.

agronomic traits, including EPN, FGPP, SSR, PH, and LA of the RIL population, showed normal distribution, and most of the traits showed right skew and peak state. The PL is left skew, low peak state. The SSR, TGW, and CC are left skewed and in a peak state, and the data are relatively centralized. The characters with high coefficient of variation were GYP (55.18), FGPP (51.18), LA (44.17), TN (38.4), LL (37.32) and EPN (36.76). The phenotypic values of the traits showed a large range and bidirectional transgressive segregation (Table 2), indicating that these drought-related traits were quantitative traits controlled by multiple genes and suitable for QTL mapping. Except for the DTF, the drought resistance index of all rice traits was less than 1, indicating that drought treatment reduced the EPN, FGPP, SSR, PH, and LA, which affected the growth and development of rice. The drought resistance index of DTF was 1.02, indicating that drought prolonged the heading time of rice.

3.2 Correlation analysis

By analyzing the correlation of the drought resistance index of various agronomic and yield traits of rice, it was found that PL, FGPP, GPP, SSR, TGW, AB, CC, and PH showed a significant

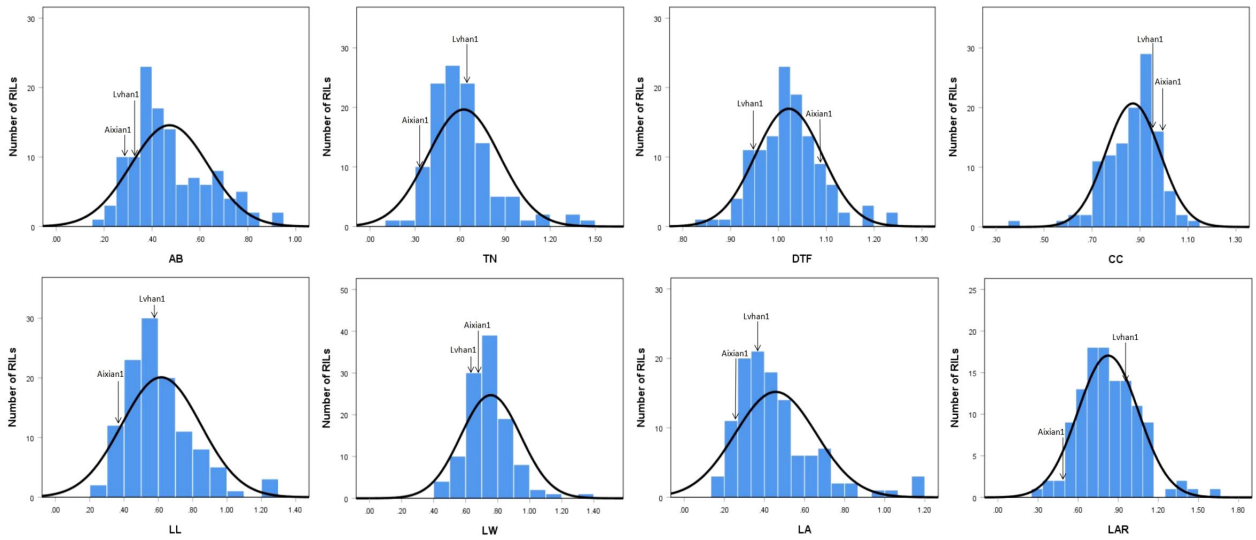


FIGURE 2
Histogram of the drought-resistance index of drought-related traits in the RIL population AB, aboveground biomass; TN, tiller number; DTF, days to flowering; CC, chlorophyll content; LL, leaf length; LW, leaf width; LA, leaf area; LAR, leaf aspect ratio.

10.3389/fpls.2023.1223782

TABLE 2 Statistical analysis of the drought-resistance index of measured traits in the RIL population.

Trait	Mean value ± standard deviation	Range	Skewness	Kurtosis	Coefficient variation/%
PH	0.71 ± 0.09	0.49-0.99	0.20	0.29	12.01
EPN	0.69 ± 0.25	0.22-1.83	1.73	5.08	36.76
PL	0.88 ± 0.12	0.59-1.17	-0.25	-0.43	14.12
FGPP	0.48 ± 0.24	0.08-1.38	0.80	0.76	51.18
GPP	0.78 ± 0.27	0.31-1.6	0.50	-0.02	34.35
SSR	0.83 ± 0.22	0.21-1.31	-0.92	0.67	26.43
TGW	0.90 ± 0.09	0.42-1.2	-0.85	6.02	10.15
GYP	0.44 ± 0.24	0.07-1.5	1.14	2.33	55.18
AB	0.47 ± 0.16	0.16-0.92	0.73	-0.08	34.15
TN	0.62 ± 0.24	0.17-1.8	1.93	5.84	38.40
DTF	1.02 ± 0.07	0.85-1.24	0.47	0.90	6.85
CC	0.87 ± 0.11	0.36-1.12	-1.03	2.72	12.95
LL	0.62 ± 0.23	0.25-1.93	2.15	8.69	37.32
LW	0.76 ± 0.19	0.04-1.76	2.86	13.16	24.75
LA	0.46 ± 0.2	0.16-1.17	1.49	2.78	44.17
LAR	0.82 ± 0.23	0.29-1.61	0.56	0.91	27.43

PH, plant height; EPN, effective panicles number; PL, panicle length; FGPP, filled grains per panicle; GPP, grains per panicle; SSR, seed setting rate; TGW, thousand-grain weight; GYP, grain yield plant; AB, aboveground biomass; TN, tiller number; DTF, days to flowering; CC, chlorophyll content; LL, leaf length; LW, leaf width; LA, leaf area; LAR, leaf aspect ratio.

positive correlation (Table 3), indicating that various traits of rice plant affected drought resistance. A highly significant correlation existed between the number of FGPP and PH, EPN, PL, GPP, SSR, GYP, AB, and CC. The correlation coefficient between FGPP and GYP was the highest (0.981), followed by AB (0.795). The correlation between GPP and TGW achieved significant levels; the correlation coefficient was 0.211. There were highly significant correlations between GYP and yield-related traits. CC was significantly correlated with PH, PL, FGPP, GPP and SSR, and AB and DTF. There was a negative correlation between the DTF and EPN. LL was positively correlated with LW, leaf area, and LAR.

3.3 Statistical analysis of SNP markers

The 5429 SNP loci were obtained by sequencing, and 3550 SNP loci were obtained by screening according to the deletion rate of less than 10% for GWAS analysis. SNP markers were evenly distributed on rice chromosomes, and the total length of the genome was 364.86 Mb, among which chromosome 1 was the longest (42.3 Mb) and chromosome 9 was the shortest (21.82 Mb). The distribution information of SNP markers on chromosomes of the whole genome is shown in Table 4.

3.4 Genome-wide association study

The TASSEL software was used to conduct a genome-wide association study on the drought-resistance index of 16 traits in the RIL population and the selected marker loci. As can be seen from the Q-Q diagram in Figure 3, when the abscissa was greater than 1.5, there was a significant difference between the P value of the GWAS result and the theoretical P value, indicating that there was indeed a significant correlation between phenotype and genotype. The Q-Q diagram illustrates that the model observations are close to the anticipated values.

The SNPs whose Manhattan map peak value exceeded the horizontal threshold of 3 may be the gene loci significantly associated with the measured phenotypic traits. There were 9 loci associated with rice drought resistance index located on multiple chromosomes (Figures 4, 5). One leaf area-related QTL was located on chromosome 7, with a P value of 3.719, and the explanation rate of phenotypic variation was 13.9%. One QTL related to leaf length was located on chromosome 7, with a P value of 3.568, and the explanation rate of phenotypic variation was 13.2%. One QTL related to ear length was located on chromosome 9 with a P value of 3.217, and the explanation rate of phenotypic variation was 11.5%. A QTL related to effective panicle number was identified on chromosome 6, with a P value of 3.199, indicating an explanatory rate of 11.4% of phenotypic variation. One QTL related to leaf width was located on chromosome 7, with a P value of 3.124, and the explanation rate of phenotypic variation was 11.5%. A QTL related to plant height was identified on chromosome 10, with a P value of 3.121, and the explanation rate of phenotypic variation was 11.1%. Three QTLS related to centroid weight were located on chromosomes 2, 6, and 8, with P values of 3.028, 3.052, and 3.219, respectively, among which *qTGW8.1* contributed the most to the phenotype (11.6%). Detailed data are shown in Table 5.

10.3389/fpls.2023.1223782

TABLE 3 The correlation of drought resistance index of measured traits.

Trait	PH(cm)	EPN	PL(cm)	FGPP	GPP	SSR	TGW(g)	GYP(g)	AB(g)	TN	DTF(d)	CC (SPAD)	LL(cm)	LW(cm)	LA (cm^2)
EPN	-0.184														
PL(cm)	0.475**	-0.176													
FGPP	0.410**	0.382**	0.585**												
GPP	0.510**	-0.187*	0.837**	0.633**											
SSR	0.422**	-0.180	0.436**	0.565**	0.405**										
TGW(g)	0.099	0.048	0.130	0.211*	0.061	0.353**									
GYP(g)	0.413**	0.369**	0.560**	0.981**	0.599**	0.574**	0.370**								
AB(g)	0.248**	0.597**	0.364**	0.795**	0.457**	0.192*	0.253**	0.804**							
TN	0.075	0.211*	-0.022	0.077	-0.064	-0.080	0.051	0.086	0.145						
DTF(d)	0.197*	0.299**	0.181	0.166	0.205*	.424**	0.081	0.167	-0.065	-0.061					
CC(SPAD)	0.319**	-0.141	0.440**	0.299**	0.416**	0.288**	0.025	0.281**	0.196*	-0.102	0.221*				
LL(cm)	-0.040	-0.112	-0.128	-0.096	-0.065	-0.009	-0.059	-0.098	-0.080	-0.068	0.157	-0.051			
LW(cm)	-0.141	-0.020	-0.216*	-0.179	-0.185	-0.058	0.169	-0.134	-0.102	-0.176	0.108	-0.133	0.523**		
LA(cm^2)	-0.113	-0.075	-0.150	-0.138	-0.111	-0.052	0.011	-0.125	-0.077	-0.071	0.185*	-0.048	0.853**	0.687**	
LAR	0.102	-0.139	0.071	0.052	0.114	0.082	-0.177	0.015	-0.029	0.013	0.080	0.078	0.771**	-0.021	0.390**

PH, plant height; EPN, effective panicles number; PL, panicle length; FGPP, filled grains per panicle; GPP, grains per panicle; SSR, seed setting rate; TGW, thousand-grain weight; GYP, grain yield plant; AB, aboveground biomass; TN, tiller number; DTF, days to flowering; CC, chlorophyll content; LL, leaf length; LW, leaf width; LA, leaf area; LAR, leaf aspect ratio.
* significance at $P < 0.05$; ** significance at $P < 0.01$.

3.5 Candidate gene association analysis

A total of 10 candidate gene loci were predicted on 12 rice chromosomes (Table 6). There were 23 candidate genes in the interval of 29.42 to 29.56Mb on chromosome 2. *LOC_Os02g0712000* encoding a serine endopeptidase. There were 36 candidate genes between 6.7 and 10.4Mb on chromosome 6. *LOC_Os06g0229800* was a cloned *ALK*, a key gene controlling rice's gelatinization temperature and encoding soluble starch synthase II (Raza et al., 2020). *LOC_Os06g0286500* encodes *NBS-LRR* disease resistance protein

TABLE 4 Distribution of SNP markers on chromosomes.

Chr	Start(bp)	End(bp)	Size(Mb)	Count	Spacing
1	222246	42526927	42.3	423	106026.77
2	1880	35926165	35.92	354	105659.66
3	398768	36102121	35.7	348	107217.28
4	332258	35126387	34.79	305	135914.57
5	170536	29536776	29.37	333	91199.50
6	122412	31198224	31.08	392	82868.83
7	16195	29564900	29.55	288	109035.81
8	17661	27999202	27.98	233	123812.13
9	465587	22289880	21.82	262	84590.28
10	46485	23155278	23.11	233	103626.87
11	1739463	28013085	26.27	332	80841.91
12	343927	27299780	26.96	238	121422.76
Total			364.86	3741	

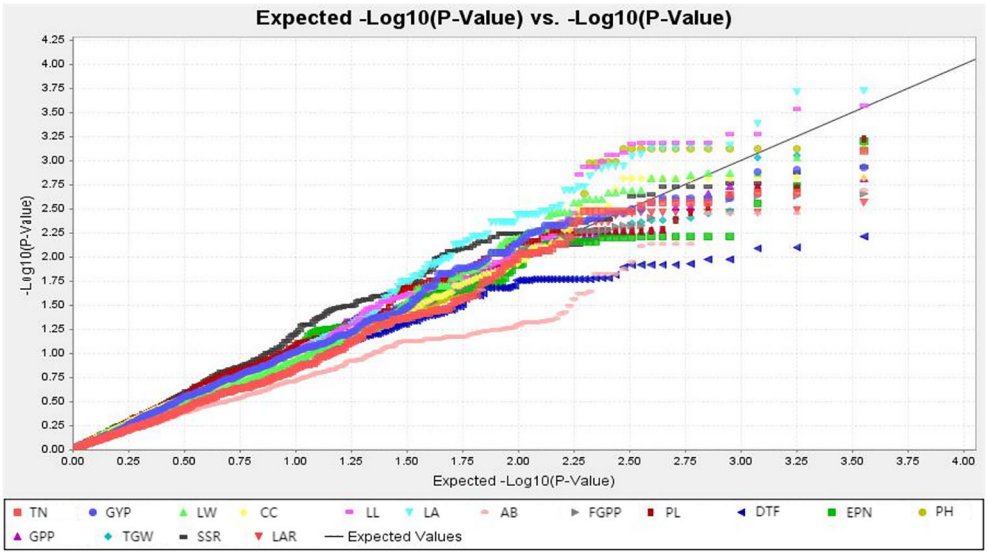


FIGURE 3
Q-Q plots illustration of drought tolerance index of measured trait. The abscissa of the Manhattan plot is 12 chromosomes of rice, the ordinate is -log10 (p) of SNPs, and the dashed horizontal line is the threshold of genome-wide significance.

homologous; There were 261 candidate genes between 22.2-24.2MB on chromosome 7, among which *LOC_Os07g05868* encoded a triacylglycerol lipase. *LOC_Os07g0586700* encodes the transcriptional suppressant *HOTR*. *LOC_Os07g0571800* is similar to the *YABBY4* gene, which regulates plant growth and development by regulating the gibberellin pathway (Yang et al., 2016). *LOC_Os07g0558500* encodes a fibro-alcohol tyrosine phosphatase antibody protein closely related to the *NYC4* gene, which is involved in the degradation of chlorophyll-protein complexes during leaf aging (Yamatani et al., 2013). *LOC_Os07g0558400* encodes chlorophyll a-b binding protein *CP29.1*, and *LOC_Os07g0591100* encodes *DUF620* family

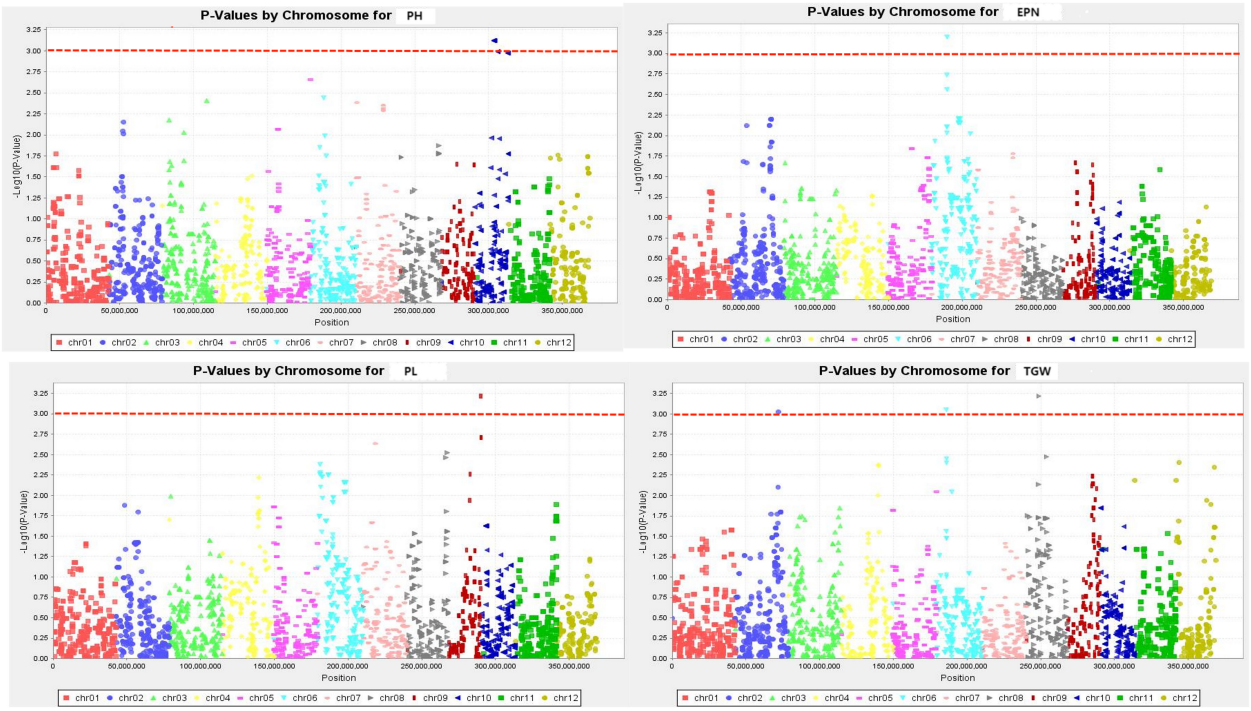


FIGURE 4
Manhattan plots exhibiting the drought tolerance index of measured traits. The abscissa of the Manhattan plot is 12 chromosomes of rice, the ordinate is -log10 (p) of SNPs, & the dashed horizontal line is the threshold of genome-wide significance.

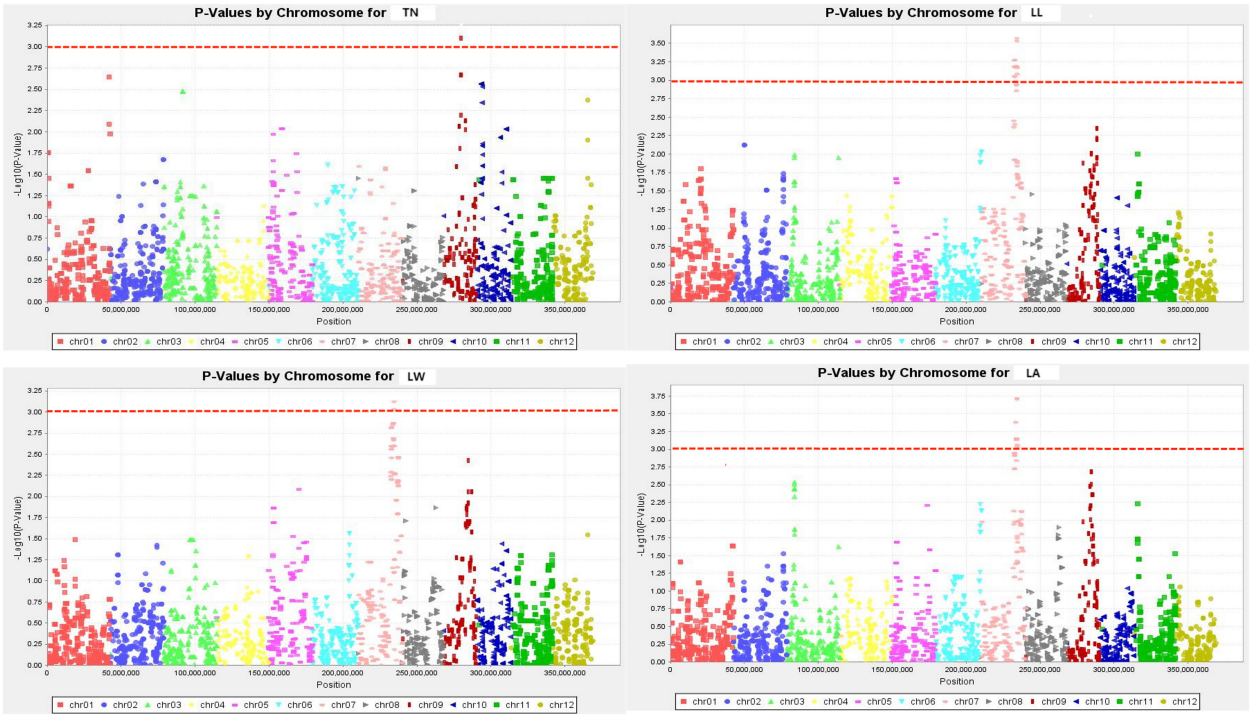


FIGURE 5
Manhattan plots exhibiting the drought tolerance index of measured traits. The abscissa of the Manhattan plot is 12 chromosomes of rice, the ordinate is -log10 (p) of SNPs, & the dashed horizontal line is the threshold of genome-wide significance.

proteins. Seven candidate genes were in the 8.7-8.8MB region on chromosome 8, *LOC_Os08g0243500* encoding NADPH oxidoreductases. *LOC_Os09g0360400* encodes a holoenzyme synthase, and *LOC_Os09g0555800* encodes a protein-containing AMP binding domain. There were 29 candidate genes in the 11.7-22.2MB range of chromosome 9. Five candidate genes encoding an oxidoreductase were between 13.6 and 13.7Mb of chromosome 10, *LOC_Os09g0403400*.

4 Discussion and conclusion

4.1 Heritability and correlation analysis of drought-resistant traits

Except for the DTF, the drought resistance index of all rice traits was less than 1. These results indicated that the phenotypic values of each trait under water stress were lower than those under normal

TABLE 5 QTL for drought-related traits detected by association analysis.

Trait	QTL	SNP Marker	Chromosome	Position	Internal	Degrees of freedom	P-value	Phenotypic contribution rate R^2(%)
PH	*qPH10.1*	GPSOS4495	chr10	13697942	13697787-13727984	12.026	3.121	11.1
EPN	*qEPN6.1*	GPSOS2698	chr06	10377542	10377542	12.394	3.199	11.4
PL	*qPL9.1*	GPSOS4239	chr09	22081469	22081469	12.488	3.217	11.5
TGW	*qTGW2.1*	GPSOS0981	chr02	29489095	29489095	11.575	3.028	10.6
TGW	*qTGW6.1*	GPSOS2563	chr06	6751340	6751340	11.693	3.052	10.7
TGW	*qTGW8.1*	GPSOS3648	chr08	8762649	8762649	12.505	3.219	11.6
LL	*qLL7.1*	GPSOS3462	chr07	23850719	22318310-24092102	14.193	3.568	13.2
LW	*qLW7.1*	GPSOS3467	chr07	24092102	24092079-24092102	12.059	3.124	11.5
LA	*qLA7.1*	GPSOS3462	chr07	23850719	23087748-24092102	14.935	3.719	13.9

TABLE 6 Candidate genes of important QTLs.

QTL	Candidate gene	Gene description
qPH10.1	*LOC_Os10g0403700*	oxido-reductase
qEPN6.1	*LOC_Os06g0286500*	*NBS-LRR* disease-resistance protein homologs
qPL9.1	*LOC_Os09g0555800*	Amp-binding domains containing proteins
qTGW2.1	*LOC_Os02g0712000*	Serine endopeptidase
qTGW6.1	*LOC_Os06g0229800*	*ALK* gene
qTGW8.1	*LOC_Os08g0243500*	NADPH oxido-reductase
qLL7.1	*LOC_Os07g0558400*	Chlorophyll a-b binding protein
qLA7.1	*LOC_Os07g0571800*	*YABBY4* gene
qLA7.1	*LOC_Os07g0586700*	transcription repressor *HOTR*
qLL7.1/qLA7.1	*LOC_Os07g0581300*	
qLL7.1/qLW7.1/qLL7.1	*LOC_Os07g0591100*	*DUF620* family protein

water conditions. Drought stress had serious effects on rice morphology, yield, and physiology. Drought stress lengthened the heading time while decreased growth period and seed setting rate; this resulted in the significant reduction in the rice grain yield (Shamsudin et al., 2016; Sandhu & Kumar, 2017). The variation coefficients of FGPP and GYP were the highest under flood and drought conditions, indicating that drought had the most serious effect on grain number and yield. There was a significant positive correlation between yield-related and physiological traits in rice. The GYP was positively correlated with EPN, GPP, FGPP, and SSR, which was consistent with the analysis of Cu et al. (2021). There was a significant positive correlation among leaf traits.

4.2 GWAS location analysis

A total of 9 QTL loci related to drought resistance traits were identified by GWAS, which was distributed on chromosomes 2, 6, 7, 8, 9, and 10, which may provide important genetic resources for future breeding. In order of contribution rate value, they are *qLA7.1*, *qLL7.1*, *qTGW8.1*, *qPL9.1*, *qEPN6.1*, *qLW7.1*, *qPH10.1*, *qTGW6.1*, *qTGW2.1*. The site with the largest contribution rate is *qLA7.1*, which is related to leaf area. Studies have concluded that leaves are one of the most sensitive organs in plants to drought stress, and leaf area is negatively correlated with drought tolerance. Under mild drought stress, biomass accumulation was inhibited, plant area decreased, and leaf number decreased. Under severe drought stress, plant leaves will age, die and fall off rapidly (Chen et al., 2023).

The QTL related to plant height was located on chromosome 10. The QTL related to effective panicles number was located on chromosome 6. The physical location of the loci related to panicle length located by Zheng et al. (2022) on chromosome 9 was similar to the *qPL9.1* loci region located in this study. QTLS related to 1000-grain weight were located on chromosomes 2, 6, and 8. The QTL related to the tiller number was located on chromosome 9. QTLS related to leaf length, leaf width, and leaf area were located on chromosome 7, and the loci of *qLL7.1* and *qLA7.1* overlap. The two loci are on the same chromosome, and there may be gene linkage. The phenotypic contribution rate of QTL loci ranged from 13.2% to 13.9%. Zhu and Xiong (2013) studied 1016 global rice core germplasm resources using GWAS analysis and identified 12 QTLS affecting the length of the rice blade, 7 QTLS affecting the width of the rice blade, and 6 QTLS affecting the area of the rice blade. The loci related to yield and leaf traits of each plant in this study were not similar to those of Zhu and Xiong (2013) so they may be new loci.

Leaf length, leaf width, and leaf area were controlled by *LOC_Os07g0591100* closely linked gene, and leaf length and leaf area were controlled by *LOC_Os07g0581300* closely linked gene, which suggested that there was a polytropic between the two genes. *LOC_Os02g0712000*, *LOC_Os06g0229800*, and *LOC_Os08g0243500* tightly linked genes control the 1000-grain weight. *LOC_Os02g0712000* encodes a serine endopeptidase. *LOC_Os06g0229800* is an *ALK* gene. *LOC_Os08g0243500* encodes a NADPH oxido-reductase. *Loc_os08g0243500* is a crucial gene of rice gelatinization temperature, promoting starch and sucrose metabolism. The panicle length was mainly controlled by *LOC_Os09g0555800* and *LOC_Os09g0555850*, closely linked genes. *LOC_Os09g0555800* encodes a protein-containing AMP binding domain, and *LOC_Os09g0555850* does not predict a known function. Effective panicle number was controlled by *LOC_Os06g0286500* closely linked gene *LOC_Os06g0286500* encoding *NBS-LRR* resistance protein homolog; Tiller number was mainly controlled by *LOC_Os09g0360400* closely linked gene, encoding a holoenzyme synthase. Plant height was controlled by LOC_Os10g0403400 and *LOC_Os10g0403700* closely linked genes. *LOC_Os09g0403400* encodes an oxidoreductase, and *LOC_Os10g0403700* has no predictive function.

4.3 Conclusion

Breeding for abiotic conditions like drought remains challenging due to the complex nature of the genetic system and the risk of unexpected developments (Debnath et al., 2023). To find the genes that control characteristics of drought tolerance, a modern molecular technique called QTL analysis was employed. In this study, nine QTL

loci related to drought resistance were identified and located on chromosomes 2, 6, 7, 8, 9, and 10, which includes *qPH10.1*, *qEPN6.1*, *qPL9.1*, *qTGW2.1*, *qTGW6.1*, *qTGW8.1 and qLL7.1*, *qLW7.1* and *qLA7.1*, respectively. The higher contribution rate sites were *qLA7.1* related to leaf area, and *qLL7.1*, related to leaf length. Ten candidate genes were also predicted, that includes *LOC_Os10g0403700* (oxido-reductase), *LOC_Os06g0286500* (*NBS-LRR* disease-resistance protein homologs), *LOC_Os09g0555800* (Amp-binding domains containing proteins), *LOC_Os02g0712000* (Serine endopeptidase), *LOC_Os06g0229800* (*ALK* gene), *LOC_Os08g0243500* (*NADPH* oxido-reductase), *LOC_Os07g0558400* (Chlorophyll a-b binding protein), *LOC_Os07g0571800* (*YABBY4* gene), *LOC_Os07g0586700* (transcription repressor *HOTR*), and *LOC_Os07g0591100* (*DUF620* family protein). It is necessary to continue verifying and fine-locating these QTLs and candidate genes through the conduction of genetic transformation and functional verification of candidate genes. This research study laid a foundation for a better understanding of rice drought tolerance genetic basis and facilitated the rice breeding program with respect to drought tolerance.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Files, further inquiries can be directed to the corresponding author/s.

Author contributions

YY and MH designed the experiment and wrote the manuscript. XC, SW and QZ contributed to the data interpretation and writing. YL and HL designed the figures and formulated tables. WF, QZ, MH, and YY participated in the processing of the experimental material. All authors read and approved the final manuscript.

Funding

The work was supported by a grant from the Anhui Natural Science Foundation Project (2208085QC82; 1908085MC93), the Young Talent Project of Anhui Academy of Agricultural Sciences, Asian Cooperation Fund Project (125001012000210004), and Anhui Provincial Key Research and Development Program (202104f06020017).

Acknowledgments

We thank Changrong Ye from Higentec Co. for the data analysis, thank Jingde Wu, Yonggeng Chen, Cunli Gong, and Yanyan Ma who helped with data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1223782/full#supplementary-material

References

Bhandari, A., Sandhu, N., Bartholome, J., Cao-Hamadoun, T. V., Ahmadi, N., Kumari, N., et al. (2020). Genome-Wide Association Study for Yield and Yield Related Traits under Reproductive Stage Drought in a Diverse indica-aus Rice Panel. *Rice* 13 (1), 1–22. doi: 10.1186/S12284-020-00406-3/FIGURES/9

Chen, L., Ma, J., Ma, X. D., Cui, D., Han, B., Sun, J. C., et al. (2023). QTL analysis of drought tolerance traits in rice during the vegetative growth period. *Euphytica* 219 (3), 1–16. doi: 10.1007/S10681-022-03151-4/METRICS

Chen, S., Xu, K., Kong, D., Wu, L., Chen, Q., Ma, X., et al. (2022). Ubiquitin ligase OsRINGzf1 regulates drought resistance by controlling the turnover of OsPIP2;1. *Plant Biotechnol. J.* 20 (9), 1743–1755. doi: 10.1111/PBI.13857

Cu, S. T., Warnock, N. I., Pasuquin, J., Dingkuhn, M., and Stangoulis, J. (20212021). A high-resolution genome-wide association study of the grain ionome and agronomic traits in rice Oryza sativa subsp. indica. *Sci. Rep.* 11, 1. doi: 10.1038/s41598-021-98573-w

Debnath, S., Kant, A., Bhowmick, P., Malakar, A., Purkaystha, S., Jena, B. K., et al. (2023). The enhanced affinity of WRKY reinforces drought tolerance in Solanum lycopersicum L.: An innovative bioinformatics study Plants. *Plants (Basel)* 12, 4, 762. doi: 10.3390/plants12040762

Dixit, S., Singh, A., Sandhu, N., Bhandari, A., Vikram, P., and Kumar, A. (2017). Combining drought and submergence tolerance in rice: marker-assisted breeding and QTL combination effects. *Mol. Breeding : New Strategies Plant Improvement* 37 (12), 1–12. doi: 10.1007/S11032-017-0737-2

Guo, Z., Yang, W., Chang, Y., Ma, X., Tu, H., Xiong, F., et al. (2018). Genome-Wide Association Studies of Image Traits Reveal Genetic Architecture of Drought Resistance in Rice. *Mol. Plant* 11 (6), 789–805. doi: 10.1016/J.MOLP.2018.03.018

Hassan, M. A., Xiang, C., Farooq, M., Muhammad, N., Yan, Z., Hui, X., et al. (2021). Cold Stress in Wheat: Plant Acclimation Responses and Management Strategies. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.676884

Ikmal, A. M., Nurasyikin, Z., Riana, T. A. T. N. A., Ellina, Z. P. D., Wickneswari, R., and Noraziyah, A. A. S. (2019). Drought Yield QTL (qDTY) with Consistent Effects on Morphological and Agronomical Traits of Two Populations of New Rice (Oryza sativa) Lines. *Plants (Basel Switzerland)* 8 (6), 1–16. doi: 10.3390/PLANTS8060186

Kumar, A., Basu, S., Ramegowda, V., and Pereira, A. (2017). "Mechanisms of drought tolerance in rice," *in Achieving sustainable cultivation of rice, Chapter 8.* 1, 131–163. doi: 10.19103/AS.2106.0003.08

Li, X., Guo, Z., Lv, Y., Cen, X., Ding, X., Wu, H., et al. (2017). Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *PLoS Genet.* 13 (7), 1–24. doi: 10.1371/JOURNAL.PGEN.1006889

Liang, Y., Liu, H. J., Yan, J., and Tian, F. (2021). Natural Variation in Crops: Realized Understanding, Continuing Promise. *Annu. Rev. Plant Biol.* 72, 357–385. doi: 10.1146/ANNUREV-ARPLANT-080720-090632

Ma, X., Feng, F., Wei, H., Mei, H., Xu, K., Chen, S., et al. (2016). Genome-Wide Association Study for Plant Height and Grain Yield in Rice under Contrasting Moisture Regimes. *Front. Plant Sci.* 7 (NOVEMBER2016). doi: 10.3389/FPLS.2016.01801

Oladosu, Y., Rafii, M. Y., Samuel, C., Fatai, A., Magaji, U., Kareem, I., et al. (2019). Drought Resistance in Rice from Conventional to Molecular Breeding: A Review. *Int. J. Mol. Sci.* 20 (14), 1–21. doi: 10.3390/IJMS20143519

Raza, Q., Riaz, A., Bashir, K., and Sabar, M. (2020). Reproductive tissues-specific meta-QTLs and candidate genes for development of heat-tolerant rice cultivars. *Plant Mol. Biol.* 104 (1–2), 97–112. doi: 10.1007/S11103-020-01027-6

Sandhu, N., and Kumar, A. (2017). Bridging the Rice Yield Gaps under Drought: QTLs, Genes, and their Use in Breeding Programs. *Agronomy* 7 (2), 27. doi: 10.3390/AGRONOMY7020027

Shamsudin, N. A. A., Swamy, B. P. M., Ratnam, W., Sta. Cruz, M. T., Sandhu, N., Raman, A. K., et al. (2016). Pyramiding of drought yield QTLs into a high quality Malaysian rice cultivar MRQ74 improves yield under reproductive stage drought. *Rice (New York N.Y.)* 9 (1), 21–21. doi: 10.1186/S12284-016-0093-6

Sun, X., Xiong, H., Jiang, C., Zhang, D., Yang, Z., Huang, Y., et al. (20222022). Natural variation of DROT1 confers drought adaptation in upland rice. *Nat. Commun.* 13, 1. doi: 10.1038/s41467-022-31844-w

Tang, Y., Bao, X., Zhi, Y., Wu, Q., Guo, Y., Yin, X., et al. (2019). Overexpression of a MYB Family Gene, OsMYB6, Increases Drought and Salinity Stress Tolerance in Transgenic Rice. *Front. Plant Sci.* 10. doi: 10.3389/FPLS.2019.00168

Wang, Y., Pang, Y., Chen, K., Zhai, L., Shen, C., Wang, S., et al. (2020). Genetic bases of source-, sink-, and yield-related traits revealed by genome-wide association study in Xian rice. *Crop J.* 8 (1), 119–131. doi: 10.1016/J.CJ.2019.05.001

Wang, H., and Qin, F. (2017). Genome-Wide Association Study Reveals Natural Variations Contributing to Drought Resistance in Crops. *Front. Plant Sci.* 8. doi: 10.3389/FPLS.2017.01110

Wang, C., Song, B., Yu, S., Xiao, W., and Huang, M. (2021). Construction of a genetic map of rice RILs based on whole genome sequencing. *J. South China Agric. Univ* 42 (2), 44–50. doi: 10.7671/J.ISSN.1001-411X.202006039

Yamatani, H., Sato, Y., Masuda, Y., Kato, Y., Morita, R., Fukunaga, K., et al. (2013). NYC4, the rice ortholog of Arabidopsis THF1, is involved in the degradation of chlorophyll - Protein complexes during leaf senescence. *Plant J.* 74 (4), 652–662. doi: 10.1111/TPJ.12154

Yang, X., Lu, M., Wang, Y., Wang, Y., Liu, Z., and Chen, S. (2021). Response Mechanism of Plants to Drought Stress. *Horticulturae* 7 (3), 50. doi: 10.3390/HORTICULTURAE7030050

Yang, C., Ma, Y., and Li, J. (2016). The rice YABBY4 gene regulates plant growth and development through modulating the gibberellin pathway. *J. Exp. Bot.* 67 (18), 5545–5556. doi: 10.1093/JXB/ERW319

Yonemaru, J., Yamamoto, T., Fukuoka, S., Uga, Y., Hori, K., and Yano, M. (2010). Q-TARO: QTL annotation rice online database. *Rice* 3 (2–3), 194–203. doi: 10.1007/S12284-010-9041-Z/FIGURES/4

Zheng, H., Sun, S., Bai, L., Jiang, S., Ding, G., Wang, T., et al. (2022). Identification of candidate genes for panicle length in Oryza sativa L. ssp. japonica *via* genome-wide association study and linkage mapping. *Euphytica* 218 (2), 1–12. doi: 10.1007/S10681-022-02972-7/METRICS

Zhu, X., and Xiong, L. (2013). Putative megaenzyme DWA1 plays essential roles in drought resistance by regulating stress-induced wax deposition in rice. *Proc. Natl. Acad. Sci. U.S.A.* 110 (44), 17790–17795. doi: 10.1073/PNAS.1316412110/-/DCSUPPLEMENTAL/PNAS.201316412SI.PDF

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fpls.2023.1205683

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Aamir Manzoor,
Shanghai Jiao Tong University, China

REVIEWED BY
Jingjian Li,
South China Agricultural University, China
Muhammad Anwar,
Hainan University, China

*CORRESPONDENCE
Chen Feng
fengc@lsbg.cn

†These authors have contributed equally to this work

RECEIVED 14 April 2023
ACCEPTED 04 July 2023
PUBLISHED 28 July 2023

CITATION
Wang S, Gao J, Li Z, Chen K, Pu W and Feng C (2023) Phylotranscriptomics supports numerous polyploidization events and phylogenetic relationships in *Nicotiana*.
Front. Plant Sci. 14:1205683.
doi: 10.3389/fpls.2023.1205683

COPYRIGHT
© 2023 Wang, Gao, Li, Chen, Pu and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Phylotranscriptomics supports numerous polyploidization events and phylogenetic relationships in *Nicotiana*

Shuaibin Wang[1†], Junping Gao[1†], Zhaowu Li[2], Kai Chen[1], Wenxuan Pu[1] and Chen Feng[3*]

[1]Tobacco Research Institute of Technology Centre, China Tobacco Hunan Industrial Corporation, Changsha, China, [2]Puai Medical College, Shaoyang University, Shaoyang, China, [3]Jiangxi Provincial Key Laboratory of ex-situ Plant Conservation and Utilization, Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang, China

Introduction: *Nicotiana* L. (Solanaceae) is of great scientific and economic importance, and polyploidization has been pivotal in shaping this genus. Despite many previous studies on the *Nicotiana* phylogenetic relationship and hybridization, evidence from whole genome data is still lacking.

Methods: In this study, we obtained 995 low-copy genes and plastid transcript fragments from the transcriptome datasets of 26 *Nicotiana* species, including all sections. We reconstructed the phylogenetic relationship and phylogenetic network of diploid species.

Results: The incongruence among gene trees showed that the formation of *N. sylvestris* involved incomplete lineage sorting. The nuclear–plastid discordance and nuclear introgression absence indicated that organelle capture from section *Trigonophyllae* was involved in forming section *Petunioides*. Furthermore, we analyzed the evolutionary origin of polyploid species and dated the time of hybridization events based on the analysis of PhyloNet, sequence similarity search, and phylogeny of subgenome approaches. Our results highly evidenced the hybrid origins of five polyploid sections, including sections *Nicotiana, Repandae, Rusticae, Polydicliae*, and *Suaveolentes*. Notably, we provide novel insights into the hybridization event of section *Polydicliae* and *Suaveolentes*. The section *Polydicliae* formed from a single hybridization event between maternal progenitor *N. attenuata* and paternal progenitor *N. undulata*; the *N. sylvestris* (paternal progenitor) and the *N. glauca* (maternal progenitor) were involved in the formation of section *Suaveolentes*.

Discussion: This study represents the first exploration of *Nicotiana* polyploidization events and phylogenetic relationships using the high-throughput RNA-seq approach. It will provide guidance for further studies in molecular systematics, population genetics, and ecological adaption studies in *Nicotiana* and other related species.

KEYWORDS

***Nicotiana* L., transcriptomic analysis, polyploid formation, evolutionary relationship, divergence time**

1 Introduction

Nicotiana L. (Solanaceae) is of great scientific and economic importance, containing the cultivated tobaccos (*N. tabacum* and *N. rustica*), the model plant (*N. benthamiana*), as well as some essential ornamentals (e.g., *N. alata* and *N. sylvestris*) (Wang and Bennetzen, 2015). The genus *Nicotiana* comprises about 87 species, including one recently reported new Australian species in the section *Suaveolentes* (*N. paulineana*) (Bally et al., 2021), nearly half of which are allotetraploids (Knapp et al., 2004). The classification of *Nicotiana* mainly relied on geographical distribution, morphological characters, and cytological investigations, which were first reported by Goodspeed (1956). It was subsequently updated by Knapp et al. (2004) based on the phylogenetic analysis and morphological description. The current classification of *Nicotiana* comprises three subgenera (*Rustica*, *Tabacum*, and *Petunioides*) and 13 sections, five of which contain polyploids formed by interspecific hybridization (seven diploid sections: *Alatae*, *Noctiflorae*, *Petunioides*, *Undulatae*, *Paniculatae*, *Trigonophyllae*, *Tomentosae*, and five polyploidy sections: *Suaveolentes*, *Repandae*, *Nicotiana*, *Polydicliae*, *Rusticae*) (Knapp et al., 2004; Leitch et al., 2008). The *Nicotiana* species are distributed across tropical and temperate regions and are primarily endemic to South America, North America, and Australia (Knapp et al., 2004), of which *N. tabacum* is one of the most widely cultivated non-food crops, having been spread worldwide by humans. About 75% of *Nicotiana* species occur naturally in America and 25% in Australia (Aoki and Ito, 2000; Clarkson et al., 2004). Interestingly, all native Australian *Nicotiana* species belong to section *Suaveolentes* (Bally et al., 2021). In addition, *Nicotiana* species exhibit a spectacular range of floral morphology and color, genome size, and karyotypic diversity (Leitch et al., 2008; Marks et al., 2011; Renny-Byfield et al., 2013; McCarthy et al., 2015; McCarthy et al., 2016) and the polyploids of *Nicotiana* formed at different stages of evolutionary divergence (Leitch et al., 2008). The genus *Nicotiana* is, therefore, an excellent system in which to take advantage of recent advances in the research of speciation, biodiversity, and phytogeography (Aoki and Ito, 2000; McCarthy et al., 2016).

Phylogenetic relationships have been the subject of study in this genus for around two decades based on the plastid markers (coding and noncoding) (Aoki and Ito, 2000; Clarkson et al., 2004), low-copy nuclear genes (Kitamura et al., 2001; Clarkson et al., 2017), nuclear ribosomal internal transcribed spacer (ITS) (Chase et al., 2003) and random amplified polymorphic DNA (RAPD) analysis (Khan and Narayan, 2007). The previous phylogenetic studies have provided new insights into interspecific relationships (Clarkson et al., 2004) and also led to a modification in the traditional classification of the genus *Nicotiana* (Knapp et al., 2004). However, these phylogenetic analyses are based on only several molecular markers or short sequences (glutamine synthetase gene, leafy/floricaula gene, ITS, *trnL-F*, *trnS-G*, *ndhF*, and *matK*) and provide limited resolution of relationships among *Nicotiana* species. And the deep relationships of *Nicotiana* have usually neither been resolved nor well-supported. Recent analyses using the complete chloroplast genomes of *Nicotiana* in 11 sections have recovered a nearly fully resolved phylogenetic relationship and deduced a potential maternal progenitor of polyploid species (Wang et al., 2022). However, evidence of phylogenetic relationships and diploid progenitors of polyploid species from the nuclear are still lacking. Therefore, sufficient evidence from whole genome data is needed to deduce the deep phylogenetic relationships and demonstrate the phylogenetic discordance among this genus.

Polyploidization has been pivotal in shaping this genus (Leitch et al., 2008; McCarthy et al., 2016). Approximately half of the *Nicotiana* species were considered natural tetraploid species of different ages (Goodspeed, 1956; Knapp et al., 2004; Leitch et al., 2008). The majority of the *Nicotiana* species possess 12 or 24 chromosome pairs, except for several diploid species in section *Alatae* with 9 or 10 pairs and several polyploidy species in section *Suaveolentes* with 15, 16, 18, 19, 20, 21, or 22 pairs (Khan and Narayan, 2007; Marks et al., 2011). So far, the morphological, distributional, cytogenetic, and molecular evidence has been used to discover the diploid progenitors for each tetraploid species (Chase et al., 2003; Lim et al., 2005; Kelly et al., 2013; Schiavinato et al., 2020). One of the most studied polyploids in this genus was the tetraploid *N. tabacum* (2n = 4x = 48), with different lines of evidence suggesting *N. sylvestris* (2n = 2x = 24, maternal donor) and *N. tomentosiformis* (2n = 2x = 24, paternal donor) as candidate parents based on morphological observation (Goodspeed, 1956), plastid genome comparison (Yukawa et al., 2006), RAPD analysis (Khan and Narayan, 2007), the genome size (Leitch et al., 2008), and whole genome sequencing (Sierro et al., 2014; Sierro et al., 2018). In addition, the species in section *Repandae* were proposed to have a hybrid origin between *N. sylvestris* and *N. obtusifolia* based on a phylogenetic context (Leitch et al., 2008). The species in the section *Rusticae* were suggested as hybrids between the ancestral species of *N. paniculata* of section *Paniculatae* and *N. undulata* of section *Alatae* based on the comparison of karyotype and genome size (Khan and Narayan, 2007; Leitch et al., 2008). The ancestral species of section *Suaveolentes* possibly were related to sections *Acuminatae* (which should be called *Petunioides*, following Knapp et al., 2004), *Noctiflorae*, and *Alatae* based on external morphology (Khan and Narayan, 2007). Still, this hypothesis has never been formally tested (Leitch et al., 2008), and the ancestry of allopolyploid species in section *Polydicliae* is unresolved (Khan and Narayan, 2007; Leitch et al., 2008). Thus, a genome-wide perspective on the origin and evolution of allopolyploid species, including estimation of divergence dates, has been lacking.

The rise of high-throughput sequencing techniques has produced massive amounts of genomic or transcriptomic data, providing an unprecedented opportunity for systematic and evolutionary studies in great depth (Lemmon and Lemmon, 2013). Notably, because of the relatively low cost of transcriptome sequencing compared with genome sequencing and the fact that phylotranscriptomics is almost as reliable as phylogenomics (Cheon et al., 2020), phylotranscriptomic analysis has emerged as the preferred method for studying evolutionary biology (Cheon et al., 2021). Recent studies based on nuclear genes, especially phylotranscriptomics, have been successful in resolving relationships of various scales from the genus (Yang et al., 2018; Chen et al., 2021) to angiosperm-wide (Yang et al., 2020), even

gymnosperms (Liu et al., 2022) and ferns (Qi et al., 2018) plants. During the last decade, large amounts of transcriptomic data have been generated for this genus, which has provided new opportunities for studying the phylogenetic relationships and evolution of polyploids at the scale of whole genomes. The newer approaches based on transcriptomic studies applying to *Nicotiana* species will provide a more accurate evaluation of speciation and polyploid events of *Nicotiana* (Clarkson et al., 2017).

Nicotiana species have significant economic importance. Most notably, tobacco (*N. tabacum*) is a major cash crop widely used in the production of tobacco products (Wang and Bennetzen, 2015). Additionally, *Nicotiana* species have pharmaceutical and research value. They are extensively used in the preparation of medicines, including both traditional herbal remedies and modern pharmaceuticals (Wang and Bennetzen, 2015). Moreover, due to the rich genomic and genetic diversity of *Nicotiana* species, they serve as ideal model organisms for studying genetic engineering, molecular biology, and genetics (Wang and Bennetzen, 2015).

In this study, we obtained the transcriptome datasets from 26 *Nicotiana* species, including 17 diploid species in eight sections and nine allopolyploid species in five sections. Two outgroup species (*Petunia axillaris* and *Petunia inflata*) from the genus *Petunia* were used as the sister taxa in the family of Solanaceae. We performed the analysis of phylogeny and PhyloNet based on the low-copy nuclear genes and transcript fragments of plastid genomes, respectively. Our study aimed to address the following topics: (1) to re-examine the classification and phylogenetic relationships reported in previous studies of *Nicotiana* and provide a relationship strongly supported among the lineages of *Nicotiana*; (2) to assess the conflict between nuclear and plastid phylogenetic topology and the inconsistency of gene trees; (3) to investigate the potential parental origin for the species of five polyploid sections; (4) to estimate the divergence time of diploid species and the time of the interspecific hybridization events that gave rise to polyploid species. This study represents the first exploration of *Nicotiana* phylogeny and timing of diversification in allopolyploids utilizing the high-throughput RNA-seq approach. It will provide valuable insights for future research in molecular systematic, population genetic, and ecological adaptation studies in *Nicotiana* and other related species.

2 Materials and methods

2.1 Data source, reads trimming, and transcriptome assembly

Here the transcriptome datasets were obtained from 26 species of genus *Nicotiana* including 17 diploid *Nicotiana* species and nine *Nicotiana* allotetraploid species representing each of the 13 *Nicotiana* sections: *Alatae* (n = 12), *Nicotiana* (n = 24), *Noctiflorae* (n = 12), *Paniculatae* (n = 12), *Petunioides* (*Noctiflorae*) (n = 12), *Polydicliae* (n = 24), *Repandae* (n = 24), *Rusticae* (n = 24), *Suaveolentes* (n = 16-24), *Sylvestres* (n = 12), *Tomentosae* (n = 12), *Trigonophyllae* (n = 12), and *Undulatae* (n = 12) (Clarkson et al., 2004). These transcriptome sequencing reads were retrieved from the NCBI Sequence Read Archive (SRA) database using the fastq-dump software of the SRA Toolkit package (available from https://github.com/ncbi/sra-tools). Additionally, two species of the genus *Petunia* (*P. axillaris* and *P. inflata*) from the non-*Nicotiana* Solanaceae family were included as outgroups. The protein-coding sequences of *P. axillaris* and *P. inflata* (Bombarely et al., 2016) were retrieved from Solanaceae Genomics Network (available from https://solgenomics.net/).

For transcriptome analyses, the quality control of raw data was performed using the fastp v0.23.2 (Chen et al., 2018), and the adapter, short reads (min. read length 50 bp), reads containing N, and reads with low-quality score (min. quality 20) were removed. All subsequent analyses were based on the filtered clean data. The clean data was first aligned into the genome sequences of allotetraploid *N. tabacum* (version 4.5, available from https://solgenomics.net/ftp/genomes/Nicotiana_tabacum/edwards_et_al_2017/) (Edwards et al., 2017) using STAR v2.7.10a (Dobin et al., 2012) with the default parameters. The aligned BAM files from the same species were merged, sorted, and indexed using samtools v1.15.1 (Danecek et al., 2021). Then, the quality of BAM files were assessed using samtools with the 'flagstat' and 'depth' algorithms. The Trinity v2.14.0 (Grabherr et al., 2011) was used to perform genome-guided *de novo* transcriptome assembly based on the BAM file of each species with default parameters. The BUSCO v5.3.2 (Manni et al., 2021) was applied to evaluate the completeness of the assembly results based on the solanales homologous gene database (available from https://busco-data.ezlab.org/v5/data/lineages/). The TransDecoder v5.5.0 (available from http://transdecoder.github.io) was employed to predict the open reading frame (ORF) within the assemblies with a minimum protein length of 100 bp. A BLAST search against the Uniref90 proteins database (Suzek et al., 2007) using BLAST (Camacho et al., 2009) and an HMMER search against the Pfam protein domain database v35.0 to identify common protein domains using PfamScan (Mistry et al., 2020). TransDecoder leveraged the outputs generated above to ensure that those peptides with blast hits or domain hits were retained in the set of reported likely coding regions. Then, CD-HIT v4.8.1 (Huang et al., 2010) was applied to remove redundant sequences with a threshold value of 0.95 identities. The statistic of assembly result was evaluated using 'TrinityStats.pl' script in the Trinity utility (Grabherr et al., 2011).

2.2 Identification and filtering of orthogroups

The OrthoFinder v2.5.4 (Emms and Kelly, 2015) was employed to infer orthologous genes from the predicted non-redundant CDS of the *Nicotiana* species and outgroups with default settings. Its utility has been tested in previous phylogenomic studies of plants (Peng et al., 2020) and animals (Fukushima and Pollock, 2020). To further increase the robustness of phylogenetic analyses, we filtered the orthologous groups according to the following criteria: (1) the max gene number of each sample in any orthologous group < 5; (2) the max gene number of each diploid sample in any orthologous group < 3; (3) the sequences of coding regions with the length >300 bp; (4) coverage of at least 50% of the 26 samples in each

orthologous group. Thus, a total of 995 low-copy orthologous were generated to investigate *Nicotiana*'s phylogenetic relationships. The gene number and the presence/absence of orthologous groups in each species were shown as a heatmap using Python script.

A significant difficulty in reconstructing phylogeny was that polyploidization events involved many species across the genus *Nicotiana*, resulting in hidden paralogs (remaining single-copy genes after the loss of distinct paralogs in different taxa), which should be avoided in phylogenetic analysis. Thus, low-copy orthologous from only the diploid species were selected to build the phylogenetic backbone structure. The low-copy orthologous groups, including polyploid species, were used to construct the phylogenetic network and infer the hybrid process.

2.3 Diploid gene tree and species tree reconstruction

The polyploid samples were excluded from the filtered low-copy orthologous groups. Thus, 17 *Nicotiana* ploidy species and two *Petunia* species as outgroups with one-to-one single-copy orthologous were extracted for phylogenetic analyses. The concatenation- and coalescent-based methods were applied to reconstruct *Nicotiana*'s gene and species tree, respectively. For the concatenation-based method, individual single-copy orthologous were aligned using MAFFT v7.490 (Katoh and Standley, 2013) with the 'L-INS- I' algorithm, and poorly aligned regions were removed using Gblock v0.91b (Talavera and Castresana, 2007) with parameters '-b4 = 5 -b5=h -t=d'. All trimmed alignments were concatenated as a supermatrix and then performed summary statistics using AMAS (Borowiec, 2016). Two conventional approaches were used to construct a phylogenetic tree: (1) A maximum-likelihood (ML) analysis was performed using RAxML-NG v1.1.0 (Kozlov et al., 2019), with 1,000 bootstraps to find the best-scoring ML tree and the best substitution model GTR +I+G4 under the Akaike information criterion (AIC) calculated by ModelTest-NG v0.1.7 (Darriba et al., 2019). (2) A Bayesian-inference (BI) analysis was performed using MrBayes v3.2.7a (Ronquist et al., 2012), with the parameters: nst = 6 and rates = gamma. Four independent Markov chain Monte Carlo (MCMC) chains were run for 1,000,000 generations with random initial trees. Trees were sampled per 100 generations. The first 25% of trees were discarded as burn-in, with the remaining trees being used for generating the consensus tree. Tracer v1.7.1 (available from http://beast.community/tracer) was used to assess the quality of the MCMC simulations and the stability of runs. For the coalescent-based method, a massively parallel tool ParGenes v1.2.0 (Morel et al., 2018; Zhang et al., 2018; Kozlov et al., 2019) was used for model selection. And species tree inference based on the single-copy orthologous with the parameters applied as: 'pargenes.py -a coalescence -o pargenes -c 32 -d nt -m –use-astral -b 1000'. Finally, the concatenated and coalescent phylogenetic results were visualized using FigTree v1.4.4 (available from http://tree.bio.ed.ac.uk/software/figtree/).

2.4 Identification of conflict and concordance among gene trees

Phyparts v0.0.1 (Smith et al., 2015) was employed to summarize the conflict and concordance information by comparing the gene tree of each single-copy orthologous against the species tree described above. Specifically, the gene tree of each single-copy orthologous with bootstrap support (BS) values was exported from the outputs of ParGenes and optimally rooted with the outgroups (*P. axillaris* and *P. inflata*) using phyx v1.3 (Brown et al., 2017). Still, in cases where the outgroups were missing, the gene trees were rooted by Minimal Ancestor Deviation (MAD) using MADroot (Tria et al., 2017). The branches in each gene tree with less than 33% BS were considered uninformative and collapsed using Newick utilities (Junier and Zdobnov, 2010). Two hundred four remaining gene trees and the species tree were used as inputs for phyparts to summarize the conflict and concordance information. The results of support and conflict statistics between gene trees and species trees were visualized with phypartspiecharts.py (available from https://github.com/mossmatters/phyloscripts/).

Furthermore, to quantify branch support values for the species tree, Quartet Sampling v1.3.1 (QS) (Pease et al., 2018) analysis was conducted with 1,000 replicates, and the log-likelihood cutoff was 2. QS was a method to analyze molecular phylogeny by calculating branch support using repeated sampling of quartets. It provided four values in the outputs: QC (the Quartet Concordance score), QD (the Quartet Differential score), QI (the Quartet Informativeness score), and QF (the Quartet Fidelity score). The QS outputs were visualized as a figure with plot_QC_ggtree.R (available from https://github.com/ShuiyinLIU/QS_visualization). The results from phyparts and QS could provide alternative evidence for evaluating the discordance of gene trees.

2.5 Phylogenetic network analysis of diploid species

Nuclear-plastid discordance indicated a hybrid origin for section *Petunioides*, comprising *N. acuminata*, *N. miersii*, and *N. attenuata* (see Results). Moreover, the discordance of species and gene trees also suggested a hybridization or incomplete lineage sorting (ILS) for the section *Sylvestres* (see Results). We then used the maximum pseudolikelihood (MPL) approach implemented in PhyloNet v3.8.2 (Wen et al., 2018) to assess corroborative evidence supporting these conclusions. A total of 596 gene trees covering all 17 *Nicotiana* ploidy species and two outgroups were used to infer the phylogenetic network with the command 'InferNetwork_MPL' (Yu and Nakhleh, 2015). The section *Sylvestres* and section *Petunioides* were set as a potential hybrid, respectively. Uncertain nodes were bypassed in the gene trees by applying a bootstrap support threshold of 30 using the -b flag. The networks were visualized using Dendroscpoe v3.82 (Huson and Scornavacca, 2012).

In addition, the NeighborNet method implemented in SplitsTree v4.18.2 (Huson and Bryant, 2005) were used to reconstruct phylogenetic networks for the concatenated dataset. The K2P model was used for distance analysis, and support values at each node were estimated by running 1,000 bootstrap replicates.

2.6 Inference on the origin of allopolyploid species

Two strategies were used to infer the hybridization process within polyploid species of *Nicotiana*: (1) Inference based on the phylogenetic network. The phylogenetic network of each allopolyploid species was carried out using PhyloNet v.3.8.2 (Wen et al., 2018) with the command 'InferNetwork_MP_Allopp' under the MDC criterion (Yan et al., 2021). Network searches were performed using only nodes in the gene trees with BS support of at least 30%, allowing for one hybridization event. The networks were visualized using Dendroscpoe v3.82 (Huson and Scornavacca, 2012). (2) Inference based on the phylogenetic relationship of subgenomes of the allopolyploid. The first step was to split the CDS sequences of each allopolyploid species into two subsets based on the sequence similarity between polyploid and all diploid species. Specifically, each low-copy orthologous group was used to perform a sequence similarity search using BLAST+ (Camacho et al., 2009) against itself with the default parameters. The diploid species with the best hit of the CDS sequences in each allopolyploid species were recorded using the custom Python script from the blast tabular output. The total number of diploid species with the best hit in the CDS sequences of each polyploid species was counted and visualized as a heatmap. The diploid species most closely related to the progenitors of each allopolyploid would yield the highest number of best hits. Therefore, we sought to identify the diploid progenitors' representatives for each allopolyploid. Then, each CDS sequence of allopolyploid species was assigned to the maternal or paternal group based on the putative results from the heatmap and previous PhyloNet analysis. The CDS sequences with ambiguous classification were discarded. The second step was to reconstruct the phylogenetic relationship based on the subgenomes of the polyploid species. Specifically, the phylogenetic relationship was inferred based on the concatenated datasets of the CDS sequences of diploid species and the classified CDS sequences of polyploid species using the ML analysis method described above. In addition, the inferred hybridization among the section *Sylvestres* and section *Petunioides* was also evaluated based on the sequence similarity search in the second strategy described above.

2.7 Plastid transcript assembly and phylogenetic analysis

Plastid genomes inherit maternally in the genus *Nicotiana* (Svab and Maliga, 2007). Thus the plastid phylogenetic tree could be used to determine the maternal progenitor of tetraploid species. Concretely, the clean reads were mapped into a previously published plastid genome (*N. sylvestris*: NC_007500.1) (Yukawa et al., 2006) with only one copy of inverted repeat regions using STAR v2.7.10a (Dobin et al., 2012), following the unmapped reads were filtered using samtools v1.15.1 (Danecek et al., 2021) with the parameter: -F 12. Genome-guided *de novo* transcriptome assembly was performed based on the filtered BAM file of each species. CD-HIT v4.8.1 (Huang et al., 2010) was used to exclude similar sequences with the parameter: -c 0.95. The local collinear regions among the filtered assemblies were identified using Mugsy v1.2.3 (Angiuoli and Salzberg, 2011). Only the conserved collinear regions with a length > 100 bp and coverage of at least 50% of the samples were extracted from the output of Mugsy. The filtered collinear regions were aligned and filtered according to the above-mentioned method. Then the combined supermatrix was used to perform phylogeny analysis using RAxML-NG v1.1.0 (Kozlov et al., 2019) with 1,000 bootstraps.

2.8 Dating of the divergence time and interspecific hybridization event

The divergence times within the species tree were inferred using BEAST v.2.6.7 (Bouckaert et al., 2019) optimized for OpenGL graphics. The dating analysis of concatenated single-copy genes of diploid species and the classified genes of allopolyploid species was performed with a strict clock, HKY substitution model, gamma site heterogeneity model, estimated base frequencies, and an ML starting tree. A Calibrated Yule model was specified as the tree prior. As no reliable fossils were available to calibrate the internal nodes of the *Nicotiana*, one secondary calibration from a recently published dated phylogeny of the Solanaceae (Särkinen et al., 2013) was used to calibrate the crown age of Solanaceae at 30.4 (95% HPD 26.3-34.0) million years ago (Ma). Two independent MCMC analyses of 10 million generations with 10% burn-in and sampling every 1,000 generations were conducted to evaluate the credibility of posterior distributions of parameters. The log files from BEAST were analyzed with Tracer v.1.7.0 (available from http://beast.community/tracer) to evaluate and ensure convergence, effective sample size (ESS) values, density plots, and trace plots. A maximum clade credibility tree with median heights was generated with TreeAnnotator v.1.8.4 (Bouckaert et al., 2019). The final tree was visualized using FigTree v1.4.4.

3 Results

3.1 RNAseq data, transcriptomes assembly, and ortholog identification

Here, a total of ~300 Gb transcriptome data from 26 *Nicotiana* species were collected, covering all 13 *Nicotiana* sections (Table 1). After read trimming, the number of clean reads per species ranged from 29.4 million to 454.6 million, with an average of 139.2 million.

For the genome-guide transcript assembly, the clean reads of each species were mapped to the *N. tabacum* genome, which resulted in an average of 81.6% primary mapped reads, with the fewest in *N. otophora* (66.9%) and the most in *N. tabacum* (93.8%)

10.3389/fpls.2023.1205683

TABLE 1 Summary information of the 26 *Nicotiana* species and the genome-guided transcriptome assemblies.

Species	Section	Ploidy level	Number of clean reads	Rate of properly mapping	Total number of trinity transcripts	Total length of contigs (bp)	Number of unigenes	GC content (%)	N50 of unigenes (bp)	BUSCO score
N. acuminata	*Petunioides*	diploid	279,341,354	75.59%	96,809	59,036,752	32,407	43.17	975	55.40%
N. amplexicaulis	*Suaveolentes*	tetraploid	44,726,966	76.82%	124,846	79,200,669	34,582	43.07	978	52.60%
N. attenuata	*Petunioides*	diploid	311,000,849	77.39%	134,204	90,057,835	28,812	42.45	1,515	96.60%
N. benthamiana	*Suaveolentes*	tetraploid	122,804,829	85.04%	234,682	188,327,938	71,557	42.49	1,242	95.80%
N. bonariensis	*Alatae*	diploid	55,090,758	88.53%	186,220	116,471,091	42,432	42.83	963	60.20%
N. cavicola	*Suaveolentes*	tetraploid	51,118,833	80.36%	131,916	87,576,621	35,640	43.01	1,059	60.30%
N. clevelandii	*Polydicliae*	tetraploid	51,559,778	86.65%	323,807	185,462,084	63,141	42.85	927	64.90%
N. cordifolia	*Paniculatae*	diploid	103,447,274	83.90%	201,692	134,731,202	41,923	42.62	1,098	70.80%
N. glauca	*Noctiflorae*	diploid	99,004,138	87.97%	243,441	183,438,304	54,161	42.63	1,047	70.50%
N. knightiana	*Paniculatae*	diploid	197,180,374	84.75%	220,167	198,049,253	48,534	42.6	1,263	84.00%
N. miersii	*Petunioides*	diploid	190,154,848	74.46%	77,747	45,701,251	29,068	43.4	897	51.50%
N. noctiflora	*Noctiflorae*	diploid	217,788,050	83.63%	208,075	122,281,149	41,406	42.62	1,056	67.80%
N. obtusifolia	*Trigonophyllae*	diploid	210,487,655	75.06%	85,016	50,875,955	28,194	43.3	963	50.00%
N. otophora	*Tomentosae*	diploid	36,545,666	66.86%	85,009	56,383,761	26,388	43.19	1,155	58.10%
N. paniculata	*Paniculatae*	diploid	454,631,909	83.46%	216,763	148,555,716	38,348	42.68	1,227	77.80%
N. petunioides	*Noctiflorae*	diploid	42,827,560	81.29%	108,580	66,927,564	33,327	43.06	981	54.80%
N. plumbaginifolia	*Alatae*	diploid	56,375,891	85.87%	168,392	106,350,036	45,195	42.74	933	61.80%
N. raimondii	*Paniculatae*	diploid	46,982,714	76.73%	111,544	85,905,570	31,998	42.92	1,170	66.50%
N. rosulata	*Suaveolentes*	tetraploid	84,859,415	82.19%	157,340	111,804,498	38,078	42.88	1,107	67.90%
N. rustica	*Rusticae*	tetraploid	84,391,418	83.68%	166,515	97,606,255	39,238	43.06	996	59.30%
N. stocktonii	*Repandae*	tetraploid	29,444,062	76.11%	112,297	73,805,927	35,211	43.12	984	53.60%
N. sylvestris	*Sylvestres*	diploid	302,837,182	91.04%	242,196	337,818,150	35,627	43.07	1,254	71.10%
N. tabacum	*Nicotiana*	tetraploid	248,141,718	93.84%	319,020	283,849,802	58,908	42.47	1,179	81.70%
N. tomentosiformis	*Tomentosae*	diploid	151,687,921	76.97%	170,353	136,584,817	35,233	42.78	1,293	77.20%
N. undulata	*Undulatae*	diploid	121,444,311	81.36%	158,486	97,525,737	38,222	42.93	1,047	64.70%
N. velutina	*Suaveolentes*	tetraploid	52,835,465	77.55%	138,048	103,844,488	34,822	42.96	1,158	67.10%

(Table 1). These assemblies produced between 77,747 and 323,807 contigs (≥ 300 bp) for each species, with an N50 length ranging from 757 bp to 2,756 bp and a total length ranging from 45.7 Mb to 337.8 Mb (Table 1). A total of 26,388 to 71,557 unigenes were detected in each species, with N50 ranging from 897 to 1,515 bp and GC content from 42.4% to 43.4%. The assembly completeness evaluation results showed that all assembled unigenes had relatively high BUSCO scores, ranging from 50.0% (*N. obtusifolia*) to 96.6% (*N. attenuata*) among the 26 species (Table 1). These values, below a fully satisfactory BUSCO score, could be explained by the absence of tissue diversity and (or) enough data. Nevertheless, we considered that these genome-guided assembled gene sets would provide a reasonably good representation of the transcriptomes of *Nicotiana* species. For the plastid transcript assembly, the clean reads of each species were mapped to the plastid genome of *N. sylvestris*. The unmapped reads were filtered from the mapping results subsequently, which resulted in an average of 0.22 million mapped reads and an average of 85.6% reads coverage rate (>5x), with the fewest in *N. petunioides* (68.32%) and the most in *N. velutina* (99.7%) (Supplementary Table S1). After the transcript assemblies, ORF prediction, and de-redundancy, there were between 12 and 144 unigenes detected in each species, with N50 ranging from 873 to 19,600 bp and GC content ranging from 36.50% to 38.97% (Supplementary Table S1).

Based on the combined datasets from the unigenes of *Nicotiana* species and the outgroups, a total of 73,634 orthologous groups were identified, of which 995 low-copy genes passed our filtering criteria as mentioned above. Among these low-copy orthologous

groups, the number of genes in each species ranges from 752 to 1,185. A total of 776 orthologous groups contained at least one of the outgroups, 558 shared in all 26 *Nicotiana* species, and 510 shared in all 28 species (26 *Nicotiana* specie and two outgroups) (Supplementary Figure S1). All the filtered low-copy genes were used to perform the phylogeny analysis, detect the parental progenitor, and date the hybridization event.

3.2 Diploid phylogenetic inference

We obtained the sequences of 995 genes with at least 273 bp in length for the nuclear phylogenetic dataset of diploid species. The alignment length for these single-copy genes ranged from 249 to 5,192 bp, with a mean length of 813 bp. After concatenation, the aligned 995-gene super matrix reached 808,952 bp in length, with 138,331 (17.1%) variable sites, 85,084 (10.5%) parsimony informative sites, and 154,142 (19.1%) missing sites (gaps and undetermined characters) (Supplementary Table S2).

For the nuclear phylogenetic dataset, the phylogenetic relationships reconstructed by ML and BI methods had identical topologies that separated the genus into two major clades with strongly supported (MLBS=100, PP=1.0) (Figure 1A). The first of them was integrated by a clade of section *Tomentosae*, sister of section *Trigonophyllae*, and a clade where section *Paniculatae* was sister to the clade of section *Undulatae*. The second was integrated by a clade, where section *Alatae* and *Sylvestres* were recovered as successive sister species of the clade of section *Petunioides* plus section *Noctiflorae* (Figure 1A). The coalescent-based species tree inferred from the diploid nuclear dataset yielded a concordant phylogenetic relationship (ASTRAL LPP=1.00) with the concatenation analysis, except for the *N. sylvestris* clade with a slightly less confident (ASTRAL LPP=0.94) (Figure 1A).

For the plastid phylogenetic dataset, the concatenated transcript fragment from plastid genomes reached 45,728 bp in length, with 2,103 (4.6%) variable sites, 1,125 (2.5%) parsimony informative sites, and 4,448 (9.72%) missing sites (Supplementary Table S2). The diploid plastid genomic phylogeny showed that the genus *Nicotiana* was monophyletic. Unlike nuclear phylogeny, within this plastid phylogenetic topology, three main clades with strong support were recovered (Figure 1B). The first was section *Tomentosae* as the basal-most clade. A clade of section *Petunioides* integrated the second, sister of section *Trigonophyllae*, and a clade of section *Paniculatae*, sister of section *Undulatae*. The third was integrated by a clade, where section *Alatae* and *Sylvestres* were recovered as successive sister species of section *Noctiflorae*.

Phylogenetic discordance was observed between topologies inferred from the nuclear and plastid concatenated datasets (Figures 1A, B). In nuclear phylogeny, section *Petunioides* (*N. acuminata*, *N. miersii*, and *N. attenuata*) were strongly supported (MLBS = 100) to be sister to section *Noctiflorae* (*N. noctiflora*, *N. petunioides*, and *N. glauca*). Section *Petunioides* were placed as sister to section *Trigonophyllae* (*N. obtusifolia*) with strongly supported (MLBS = 92) in plastid phylogeny. This result showed the hybridization of section *Petunioides* between section *Noctiflorae* and *Trigonophyllae*, which might explain the discordance of nuclear and plastid phylogenetic topologies. In addition, the section *Tomentosae* (*N. otophora* and *N. tomentosiformis*) was recovered as a basal taxon in plastid phylogeny. In construction, two main groups were recovered within the nuclear phylogeny.

3.3 Gene tree concordance and conflict

Our species tree inferred from coalescent nuclear data suggested that two major clades were identified in the phylogeny of the genus *Nicotiana* with full support (node A). Likewise, the result of Quartet Sampling (QS) demonstrated that node A was confirmed with full support (1/-/1; i.e., all informative quartets support that lineage) and phyparts result with almost all the informative gene trees (750 out of 751) support this topology (Figures 1C, D). Although the nuclear phylogenetic tree confirmed *N. sylvestris* sister to species of the clade of section *Petunioides* plus section *Noctiflorae* with full support, this clade (node B) was supported by only 28% of informative quartets with a skewed frequency for alternative discordant topologies (QS score = 0.28/0/1). Likewise, the result of phyparts supported this clade with only 148 out of the 716 informative gene trees (20.7%). This result revealed that ILS or hybridization might explain this phylogenomic discordance.

All the conflict nodes between nuclear and plastid phylogenetic topology showed relatively low informative gene trees supported in the phyparts result (Figure 1C). For example, nodes C and D have alternative discordant topologies. The result of phyparts supported this clade with only 275 of the 739 informative gene trees (37.2%) and 253 of the 655 informative gene trees (38.6%), respectively. In contrast, the QS result demonstrated that all these two nodes related to three sections were confirmed with full support (1/-/1) (Figure 1D).

3.4 Diploid phylogenetic networks

The phylogenetic discordances of nuclear-plastid and gene trees showed a complex evolutionary history among *N. sylvestris* and the species of section *Petunioides*. Using the *N. sylvestris* as potential hybrid species, our PhyloNet analysis indicated that *N. sylvestris* was a potential hybrid species between the section *Alatae* and the common ancestor of section *Noctiflorae* and *Petunioides*, or at least introgressed with section *Alatae* (Figure 2A). When using the species of section *Petunioides* as potential hybrid species, our PhyloNet analysis did not support the hybrid origin of *N. sylvestris* and the species of section *Petunioides* (data not shown). The analysis of sequence similarity search in *N. sylvestris* showed that a total of 264 genes in section *Petunioides*, 119 in section *Noctiflorae*, and 249 in section *Alatae* were similar to the gene sequences in *N. sylvestris* with the best hit (Figure 2B). In the sequence similarity search of section *Petunioides*, the best hit mainly occurred in *N. sylvestris*, section *Noctiflorae*, and *Alatae*, not in the section *Trigonophyllae* (Figure 2B).

Phylogenetic networks reconstructed using NeighborNet revealed apparent clustering among the *Nicotiana* diploid sections. When rooted in the genus *Petunia*, the topology of

10.3389/fpls.2023.1205683

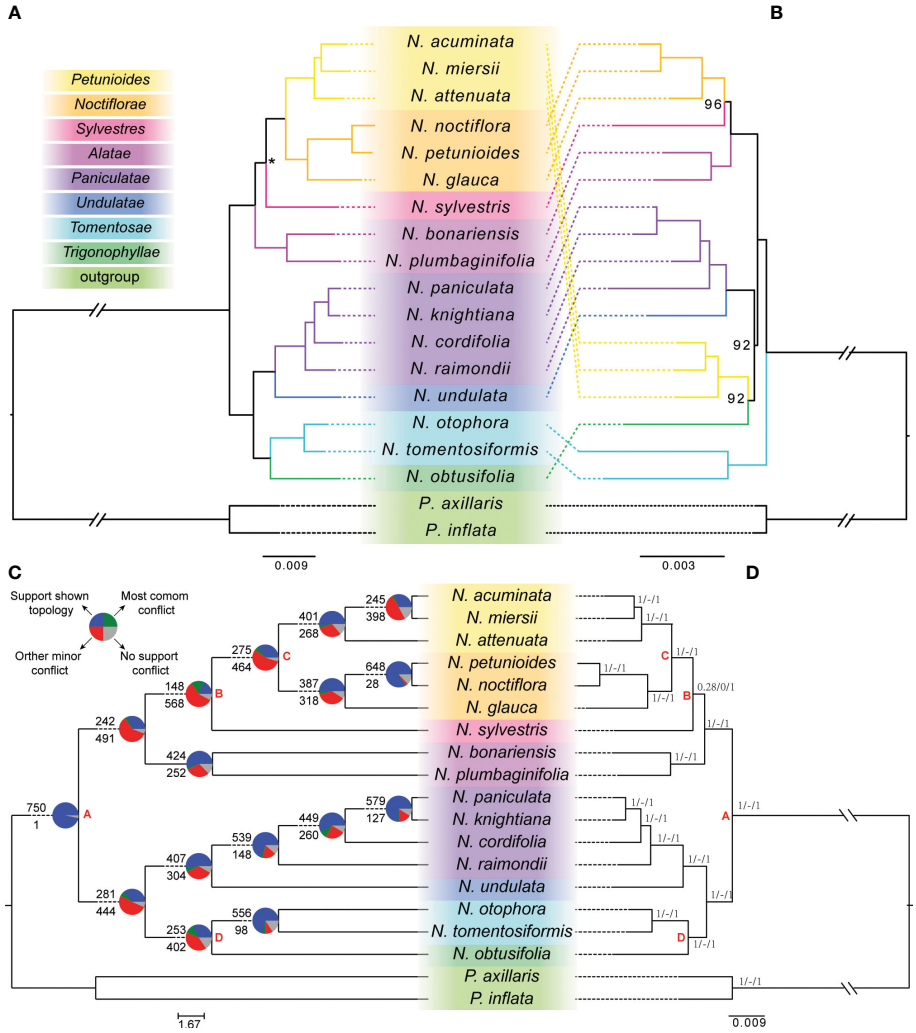


FIGURE 1
Tanglegram illustrating the nuclear-plastid discordance and the conflict signal among the gene trees in *Nicotiana* diploid species. **(A)** The diploid nuclear phylogeny recovered from concatenation- or coalescent-based methods based on the 995 single-copy genes (left). **(B)** The diploid plastid phylogeny recovered from the maximum-likelihood method based on the concatenated local collinear regions of transcript fragments (right). Bootstrap percentages were indicated beside the branches, and only values less than 100 were shown. **(C)** Patterns of gene-tree concordance and conflict of *Nicotiana* based on the phyparts analysis (left). The tree topology used was inferred by ASTRAL. The pie charts at each node show the proportion of genes in concordance (blue), conflict (green: a single dominant alternative; red: all other conflicting trees), and without enough information (gray). The numbers above and below each branch were the numbers of concordant and conflicting genes at each bipartition, respectively. **(D)** Information on *Nicotiana*'s gene-tree concordance and conflict based on the result from quartet sampling (right). Branch labels show quartet concordance (QC), quartet differential (QD), and quartet informativeness (QI), respectively, for each relationship. Corresponding clades in the nuclear and plastid phylogeny were colored. Asterisks on the branches of nuclear phylogeny (left) indicated local posterior probabilities of 0.94 in the coalescent-based species tree. Sections were classified according to Knapp et al. (2004) and labeled to the left.

phylogenetic networks (Figure 2C) was very similar to the ML and Bayes trees (Figure 1A). However, *N. sylvestris* was alternative splits connecting to the section *Alatae* or the common ancestor of section *Petunioides* and *Noctiflorae* (Figure 2C), which showed uncertainty regarding the phylogenetic placement of this clade.

3.5 Putative diploid progenitors of polyploid species

Based on the InferNetwork_MP_Allopp approach implemented in PhyloNet, we inferred the allopolyploid network from gene trees under the MDC criterion (Figures 3A–E). For the *N. tabacum* hybridization test, the PhyloNet result showed a hybridization event between the *N. sylvestris* and *N. tomentosiformis* (Figure 3A). For the *N. stocktonii*, the PhyloNet result showed a hybridization event between the *N. obtusifolia* and *N. sylvestris* (Figure 3B). For the *N. rustica*, the PhyloNet result showed a hybridization event between the *N. knightiana* and *N. undulata* (Figure 3C). For the *N. clevelandii*, the PhyloNet detected a hybridization between the *N. attenuata* and *N. undulata* into the clade *N. clevelandii* (Figure 3D). For the hybridization tests of the *N. velutina*, *N. cavicola*, *N. rosulata*, *N. benthamiana*, and *N. amplexicaulis*, the PhyloNet

10.3389/fpls.2023.1205683

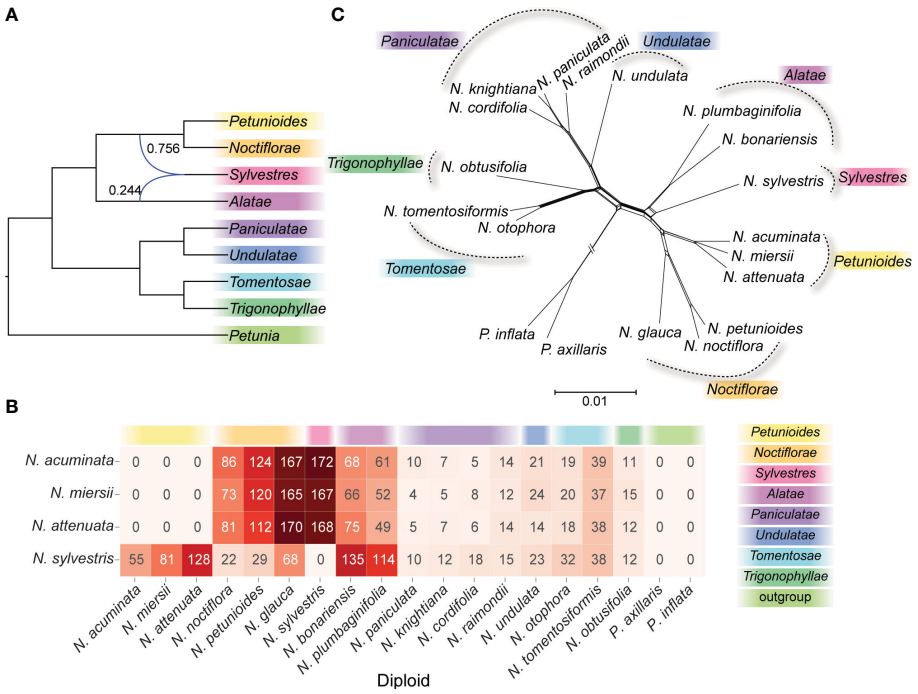


FIGURE 2
The phylogenetic network analysis of diploid species. **(A)** Phylogenetic networks were inferred by setting *N. sylvestris* as the hybrid species using the InferNetwork_ML method in PhyloNet. Blue branches indicated lineages involved in reticulated histories, and numerical values were the inheritance probabilities for each reticulation. **(B)** The heatmap showed the best hit number of the CDS sequences in *N. sylvestris* and section *Petunioides* against the other diploid species. **(C)** Phylogenetic networks generated by NeighborNet method in SplitTree4.

detected a hybridization between the *N. glauca* and *N. sylvestris* into the clade of section *Suaveolentes* (Figure 3E).

For each polyploid species, we counted the number of diploid species with the best hit of the CDS sequences (Figure 3F). In the CDS sequences of tetraploid *N. tabacum*, 341 and 456 CDS sequences had the best hit against the *N. sylvestris* and *N. tomentosiformis*, respectively, which suggested that *N. sylvestris* and *N. tomentosiformis* were the putative diploid progenitors of *N. tabacum*. Similarly, in tetraploid *N. Stocktonii*, a total of 217 and 241 CDS sequences against the *N. sylvestris* and *N. obtusifolia* with the best hit, respectively, suggested that *N. sylvestris* and *N. obtusifolia* were the putative diploid progenitors of *N. Stocktonii*. In the tetraploid *N. rustica*, 370 CDS sequences against the *N. sylvestris* with the best hit and 212 and 185 CDS sequences against the *N. paniculata* and *N. knightiana*, respectively, supported the hybrid simulation of *N. sylvestris* and *N. paniculata* and (or) *N. knightiana* into *N. rustica*. In the tetraploid *N. clevelandii*, 191 CDS sequences against the *N. attenuata* with the best hit. And 235 and 123 CDS sequences against the *N. undulata* and *N. paniculata*, respectively. This result supported the hybrid simulation between *N. attenuata* and *N. undulata* or *N. paniculata* into *N. clevelandii*. In the tetraploid *N. Velutina*, *N. cavicola*, *N. rosulata*, *N. benthamiana*, and *N. amplexicaulis*, 233–258 CDS sequences against the *N. sylvestris* with the best hit, and another parent could be origin from section *Petunioides*, *Noctiflorae* or *Alatae*.

Based on the classified sequences of polyploid species and transcript fragments of plastid, we construct the phylogenetic relationship of nuclear and plastid, respectively, including all diploid and polyploid species. In the plastid phylogenetic relationship (Supplementary Figure S2), the *N. tabacum* was sister to the *N. sylvestris*; the *N. rustica* was sister to the *N. knightiana*; the *N. clevelandii* was sister to the *N. obtusifolia*; both *N. stocktonii* and section *Suaveolentes* was sister to the section *Noctiflorae*. In the nuclear phylogenetic relationship (Figure 3G), the *N. tabacum* was sister to the *N. tomentosiformis* and *N. sylvestris*, respectively; the *N. clevelandii* was sister to the *N. undulata* and *N. attenuata*, respectively; the *N. stocktonii* was sister to the *N. obtusifolia* and *N. sylvestris*, respectively; and the *N. rustica* was sister to *N. undulata* and the common ancestor of *N. knightiana* and *N. paniculata*, respectively. The results of phylogenetic relationships provided clues about the parent origin of polyploid species. We concluded the hybridization event by combining the phylogenetic relationships of nuclear and plastid (Figures 3G, S2). Namely, the *N. tabacum* formed from a single hybridization event between extant relatives of maternal progenitor *N. sylvestris* and paternal progenitor *N. tomentosiformis*; the *N. stocktonii* formed from a single hybridization event between maternal progenitor *N. sylvestris* and paternal progenitor *N. obtusifolia*; the *N. rustica* formed from a single hybridization event between paternal progenitor *N. undulata* and maternal progenitor common ancestor of *N. knightiana* and *N. paniculata*; the *N. clevelandii* formed from a single hybridization event between maternal progenitor *N. attenuata* and paternal progenitor *N. undulata*; the section *Suaveolentes* could be formed from a single hybridization event between maternal progenitor *N. sylvestris* and paternal progenitor of *N. glauca*.

10.3389/fpls.2023.1205683

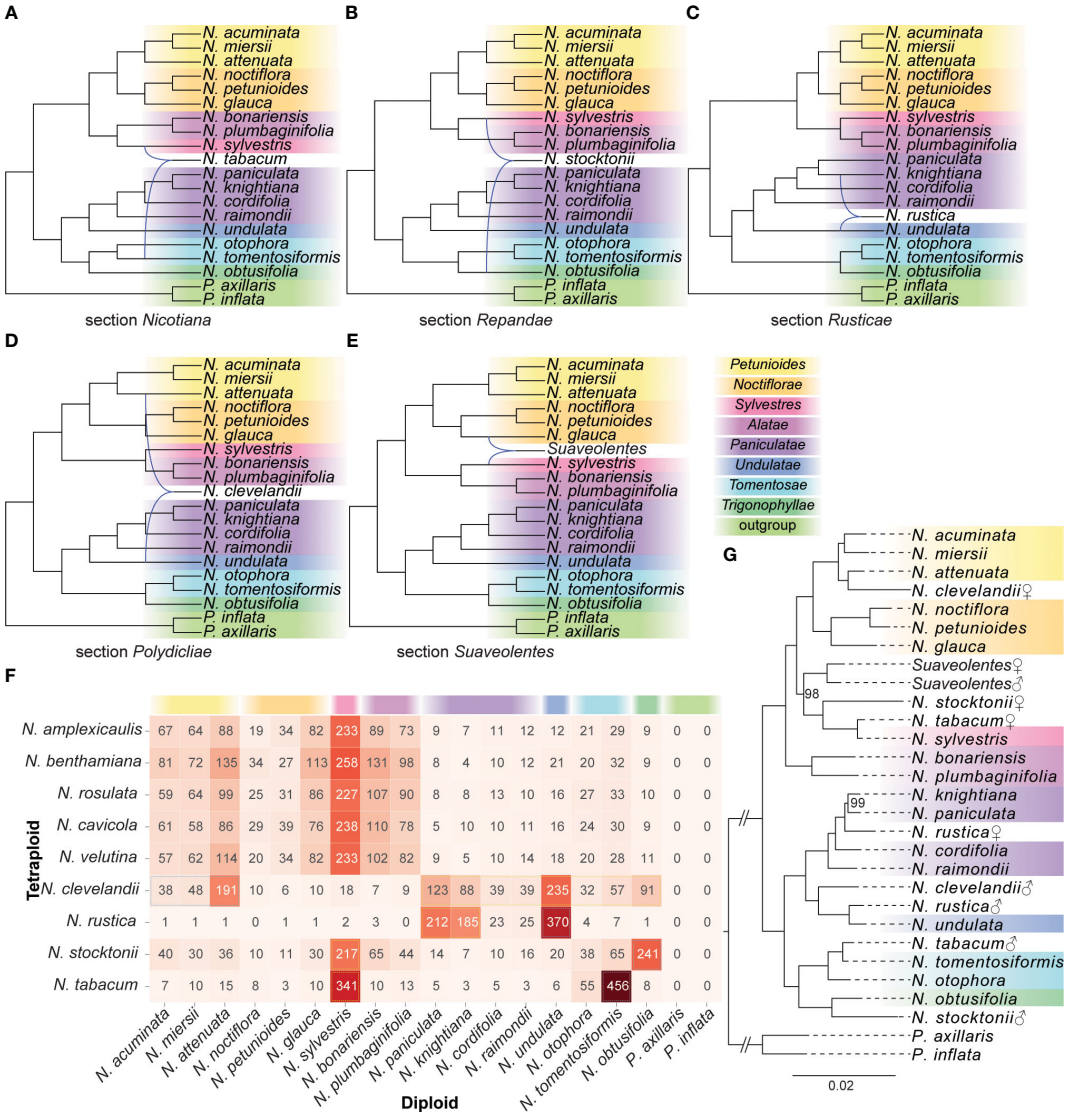


FIGURE 3
The diploid origin analysis of polyploid species. **(A–E)** Phylogenetic networks were inferred by setting each polyploid species as the hybrid species using the InferNetwork_MP_Allopp method in PhyloNet. Blue branches indicated lineages involved in reticulated histories. **(F)** The heatmap showed the best hit number of the CDS sequences in the polyploid species against the other diploid species. **(G)** The nuclear phylogeny recovered from the maximum-likelihood method based on the classified low-copy genes. Bootstrap percentages were indicated beside the branches, and only values less than 100 were shown. Sections were classified according to Knapp et al. (2004) and labeled to the right. ♀: maternal origin; ♂: paternal origin.

3.6 Dating the time of hybridization event among *Nicotiana* species

To date the time of hybridization events that had led to the formation of polyploid species, we then used the classified CDS based on the sequence similarity search to obtain the time distance between the hybrid and the parental species (a hybridization date) (Figure 4; Supplementary Table S3). The divergence between the genus *Nicotiana* and *Petunia* was dated to c. 30.2 Ma (95% Highest Posterior Density (HPD) = 25.6–35.2 Ma). The diversification of *Nicotiana* was inferred to begin at c. 9.24 Ma (95% HPD = 7.84–10.77 Ma). The *N. tabacum*, as the only species in section *Nicotiana*, was the youngest allotetraploid section with age estimates of c. 0.42 Ma from its maternal progenitor *N. sylvestris* (95% HPD = 0.35-0.50 Ma) and c. 0.58 Ma from its paternal progenitor *N. tomentosiformis* (95% HPD = 0.49–0.68 Ma). Section *Rusticae* was a monotypic section containing only *N. rustica* and yielded age estimates of c.1.52 Ma from its maternal progenitor, the common ancestor of *N. paniculata* and *N. knightiana* (95% HPD = 1.28-1.76 Ma) and c. 1.30 Ma from its paternal progenitor *N. undulata* (95% HPD = 1.09-1.52 Ma). Section *Polydicliae* consists of two species, of which *N. clevelandii* yielded age estimates of c. 3.71 Ma from its maternal progenitor, the common ancestor of section *Petunioides* (95% HPD = 3.11-4.30 Ma) and c. 3.73 Ma from its paternal progenitor *N. undulata* (95% HPD = 3.12-4.30 Ma). Section *Repandae* contained four species, of which *N. Stocktonii* yielded age estimates of c. 5.02 Ma from its maternal progenitor *N. sylvestris* (95% HPD = 4.25-5.84 Ma) and c. 3.39 Ma from its paternal

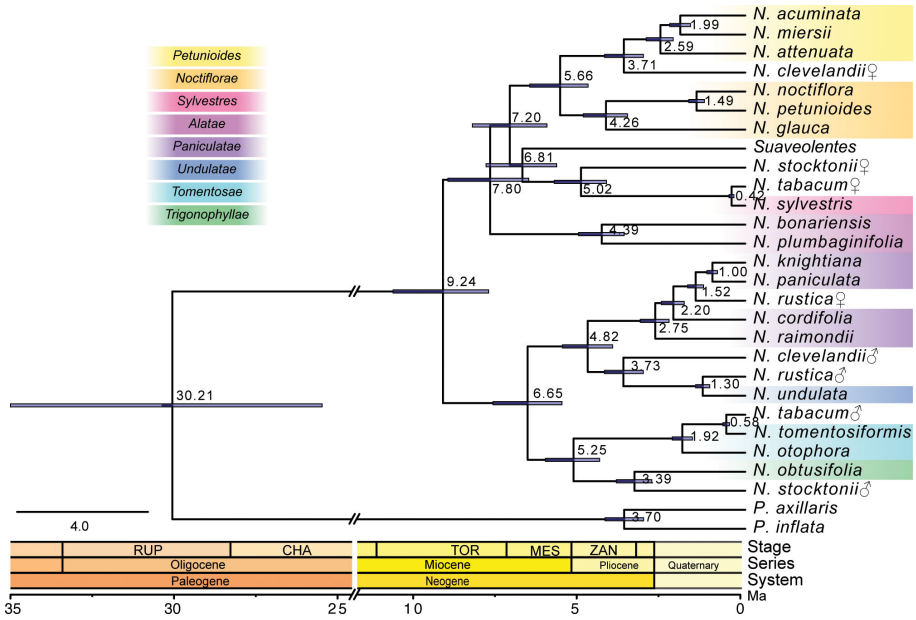


FIGURE 4
The phylogenetic tree showed the topology and divergence time for 26 *Nicotiana* species. Divergence times were indicated by light blue bars at the internodes; the range of these bars indicates 95% of the highest posterior density (HPD) interval of the divergence time. Numbers at the internodes indicate the mean divergence time. The geological timescale was illustrated at the bottom. ♀: maternal origin; ♂: paternal origin.

progenitor *N. obtusifolia* (95% HPD = 2.86-3.94 Ma). Section *Suaveolentes* was the oldest and most species-rich allotetraploid section with age estimates of c. 6.81 Ma from its maternal progenitor, the common ancestor of *N. sylvestris* (95% HPD = 5.76-7.92 Ma).

4 Discussion

4.1 Strongly supported diploid phylogeny and nuclear-plastid phylogenetic discordance

The Molecular Phylogeny of the *Nicotiana* genus has been researched for more than two decades (Aoki and Ito, 2000). The previous phylogenetic analyses using a combination of the internal transcribed spacer (ITS) and several plastid markers suggested that section *Tomentosae* was the base taxa and the section *Petunioides* was sister to the MCRA of section *Noctiflorae*, *Alatae*, and Sylvestres (Leitch et al., 2008). Recently, the nuclear phylogenetic tree based on GS and LFY genes was used to investigate the timing of diversification (Clarkson et al., 2017). However, these phylogeny relationships in the genus *Nicotiana* were inferred from several nuclear or plastid makers, which lack enough support and reliable results. Our recent analysis of whole plastid genomes provided a well-supported phylogenetic relationship of 11 sections in *Nicotiana* (Wang et al., 2022), which supported the section *Tomentosae* as the base clade of all others, and the section *Petunioides* was sister to the section *Trigonophyllae*.

Over the past ten years, high-throughput transcriptome sequencing has provided an unprecedented volume of available genetic data. The transcriptome data have been widely used for reconstructing the phylogenetic relationship of plants (Leebens-Mack et al., 2019) regardless of the tissue origin of the transcriptomes (Cheon et al., 2020). We used gene and species tree approaches to construct a diploid phylogeny of *Nicotiana* based on RNAseq data that includes representatives from all 13 sections recognized in the *Nicotiana* genus. Our phylogenetic analyses of nuclear and plastid datasets produced mostly harmonious and well-supported relationships (i.e., ≥92 BP) among major lineages in *Nicotiana*, including those not well resolved in previous studies. According to the nuclear phylogeny, the earliest divergence in *Nicotiana* involves two major clades (Figure 1A). The first major clade was formed by the most recent common ancestor (MRCA) of sections *Trigonophyllae*, *Tomentosae*, *Undulatae*, and *Paniculatae*. Section *Trigonophyllae* was sister to section *Tomentosae*, while section *Undulatae* was sister to section *Paniculatae*. The second major clade was formed by the MRCA of sections *Petunioides*, *Noctiflorae*, *Sylvestres*, and *Alatae*. All of these relationships are strongly supported, except for the placement of section *Sylvestres* (ASTRAL LPP = 0.94/ASTRAL BS = 90/concatenated BS > 99; 148/568 informative gene trees). According to the plastid phylogeny, section *Tomentosae* was sister to all remaining sections, with a grade formed by the MRCA of section *Trigonophyllae*, *Petunioides*, *Undulatae*, *Paniculatae* and the MRCA of section *Alatae*, *Sylvestres*, and *Noctiflorae*.

Our nuclear analyses revealed two major clades, which were not indicated by previous molecular or morphological analyses (McCarthy et al., 2016; Clarkson et al., 2017). The plastid phylogenetic tree in this study showed a consistent placement with our previous result based on the whole plastid genomes (Wang et al., 2022). Furthermore, the availability of genome-scale

data allowed us to examine the consistency of phylogenetic signals in the nuclear and plastid genomes for the first time, and several incongruent have been identified between nuclear and plastid phylogeny, especially for section *Petunioides* and *Sylvestres* (Figure 1). Incomplete lineage sorting and hybridization are the two main evolutionary processes that could lead to incongruent topologies between nuclear and organelle genomes (Willyard et al., 2009; Kao et al., 2022). It is often difficult to disentangle these processes. For section *Sylvestres*, incomplete lineage sorting should count for the incongruent topologies between gene trees. In addition, our analyses show an inconsistent placement of section *Petunioides* between nuclear and plastid phylogeny (Figures 1A, B), which is likely the result of an interspecific hybrid origin of this section between section *Noctiflorae* and *Trigonophyllae* in more early time. In contrast, the hybrid signal and nuclear introgression were lacking in section *Petunioides* based on the sequence similarity search and PhyloNet analysis, respectively (Figure 2B), which means that organelle capture can explain this observation. Using multiple gene trees enabled us to detect evidence for hybridization events between diploid species and resolved the phylogeny more robustly than in the previous studies.

4.2 Inferring polyploid parentage of *Nicotiana* species

As the genus *Nicotiana* contains several groups of tetraploids that formed at different times from different diploid progenitors (Kelly et al., 2013), it provides an ideal system for understanding polyploidization, a pervasive and powerful evolutionary force in plants. We identified the most likely diploid progenitors of five allopolyploid sections using the combined approaches of the phylogenetic network, sequence similarity search, and phylogenetic tree of subgenomes (Figure 3). Our results provide novel insights into parental species for the *Nicotiana* polyploids. Below we discussed the putative parental species of *Nicotiana* polyploids and proposed the results obtained in this study (Figure 5).

4.2.1 Section Nicotiana

N. tabacum (common tobacco), the only species in section *Nicotiana*, was most widely grown commercially for tobacco production. Its diploid ancestors (*N. sylvestris* as maternal progenitor and *N. tomentosiformis* as paternal progenitor) and the details of the hybridization have been well characterized (Sierro et al., 2013; Sierro et al., 2014). This study also provided three pieces of evidence to support that the diploid *N. sylvestris* and diploid *N. tomentosiformis* were the maternal and paternal progenitors of *N. tabacum*, respectively (Figure 5), which validated the reliability and accuracy of these strategies (Figures 3A, G).

4.2.2 Section Repandae

Section *Repandae* consists of four allopolyploid species (*N. Nudicaulis*, *N. repanda*, *N. Stocktonii*, and *N. Nesophila*; 2n = 4x = 48) (Knapp et al., 2004; Clarkson et al., 2005). It has been thought that section *Repandae* formed from a single hybridization event between extant relatives of maternal progenitor *N. sylvestris*

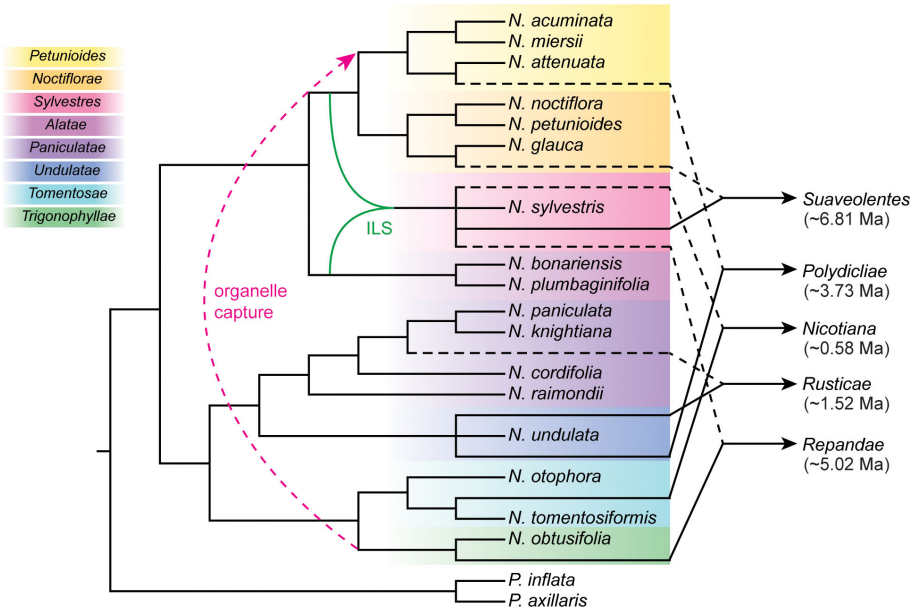


FIGURE 5
Cladogram summary of the polyploidization events and phylogenetic relationships in *Nicotiana*. Sections of allotetraploid origin were indicated by dashed black lines and solid black lines from their maternal lineages and paternal lineages, respectively. The section *Petunioides* involved organelle capture event was indicated by red dashed lines. The *N. sylvestris* involved incomplete lineage sorting (ILS) event was indicated by solid green lines. The time of hybridization events was noted under the allotetraploid sections. Sections were classified according to Knapp et al. (2004) and labeled to the left.

and paternal progenitor *N. obtusifolia*. Subsequently, four allopolyploid species were produced following speciation (Chase et al., 2003; Clarkson et al., 2010; Kelly et al., 2013; Dodsworth et al., 2017). Based on our strategies, the diploid *N. sylvestris* and *N. obtusifolia* were recognized as the maternal and paternal progenitors of section *Repandae*, respectively (Figure 5), consistent with the previous results (Dodsworth et al., 2017).

4.2.3 Section rustica

Like common tobacco, *N. rustica* (Aztec tobacco) was the only species in section *Rusticae* and was an allotetraploid native to South America formed through a recent hybridization event (Knapp et al., 2004). Based on morphology, karyotype analyses, and breeding experiments, Goodspeed (1956) proposed that the diploid parents of *N. rustica* were ancestors of *N. paniculata* and *N. undulata*. Based on the comparative nuclear genome analysis, Sierro et al. (2018) found that the tetraploid species *N. rustica* inherited about 41% of its genome from its paternal progenitor, *N. undulata*, and 59% from its maternal progenitor, the common ancestor of *N. paniculata* and *N. knightiana*. Whereas, the pieces of evidence from comparative plastid genes and genome analysis revealed that the maternal parent of the tetraploid *N. rustica* was from the species of section *Paniculatae*, and the diploid *N. knightiana* was genetically closer than *N. paniculata* to *N. rustica* (Clarkson et al., 2004; Mehmood et al., 2020; Wang et al., 2022). The result from sequence similarity analysis support that both *N. knightiana* and *N. paniculata* might have donated the maternal genome of *N. rustica* (Figure 3B). Although the plastid genome of *N. knightiana* appears to be closer than that of *N. paniculata* to the *N.rustica* chloroplast genome, our analysis of PhyloNet and the phylogeny of subgenomes still suggested that a common ancestor of both *N. knightiana* and *N. paniculata* served as the maternal donor to *N. rustica* (Figure 5).

4.2.4 Section Polydicliae

Section *Polydicliae* consists of two species, *N. quadrivalvis* and *N. clevelandii*, the only allopolyploid section found in western North America (Goodspeed, 1956). In early research, plastid-based analyses indicated that a diploid species from section *Trigonophyllae* was the maternal genome donor of section *Polydicliae* (Clarkson et al., 2004). Based on the analysis of genome size, Leitch et al. (2008) found that the genome size of *N. attenuata* was most closely related to the paternal genome donor of section *Polydicliae*. Clarkson et al. (2010) suggested that section *Polydiclieae* was the product of a cross between the ancestors of section *Trigonophyllae* (maternal) and *N. attenuata* (paternal). Based on the strict consensus trees from the ADH and LFY/FLO datasets, Kelly et al. (2013) proposed that section *Polydicliae* formed from a single hybridization event between extant relatives of maternal progenitor *N. obtusifolia* and paternal progenitor *N. attenuata*. Subsequently, the analysis of floral evolution in section *Polydicliae* was performed based on this hybridization model (McCarthy et al., 2015; Bombarely et al., 2016). Based on the phylogenetic tree of five plastid loci and two nuclear genes, Clarkson et al. (2017) suggested a similar result. However, we obtained a different conclusion. In the plastid phylogenetic relationship (Supplementary Figure S2), the *N. clevelandii* was sister to the *N. obtusifolia*, the plastid of which was transferred to section *Petunioides*. Thus, we now believe that section *Polydiclieae* is the product of a cross between the ancestors of *N. attenuata* (maternal) and *N. undulata* (paternal) based on our strategies (Figure 5).

4.2.5 Section Suaveolentes

Section *Suaveolentes* was an almost all-Australian clade (the exception being *N. africana* of Namibia) of allopolyploid species, including the vital plant model *N. benthamiana* (Schiavinato et al., 2020). Karyotypic variation within this section was very enrichment from n = 15 to n = 32 chromosomes (Marks et al., 2011). It seems likely that section *Suaveolentes* has explosive radiation of taxa occurred, primarily accompanied by diploid reductions probably due to fusions of chromosomes (Clarkson et al., 2004). As the oldest *Nicotiana* polyploids (Clarkson et al., 2017), the diploid progenitors of section *Suaveolentes* were poorly understood (Marks et al., 2011). Goodspeed (1956) proposed that several diploid sections of *Nicotiana*, namely sections *Alatae*, *Sylvestres*, *Noctiflorae*, and *Petunioides*, were involved in the formation of the allotetraploid section *Suaveolentes*. Kelly et al. (2013) reconstructed the evolutionary origin of sect *Suaveolentes* using four regions from the nuclear and plastid genome. They identified a likely scenario where a member of the *N. Sylvestres* acted as the paternal progenitor, and a member of either section *Petunioides* or *Noctiflorae* was the maternal progenitor. Recently, Schiavinato et al. (2020) showed that the maternal progenitor of *N. benthamiana* was a member of section *Noctiflorae* and confirmed a member of section *Sylvestres* as a paternal subgenome donor. Our analysis based on the PhyloNet approaches supported a scenario where the *N. sylvestris* acted as the paternal progenitor, and the *N. glauca* of section *Noctiflorae* acted as the maternal progenitor in the formation of section *Suaveolentes* (Figure 3E), in line with previous findings (Schiavinato et al., 2020). But the analysis from the sequence similarity search and the phylogeny of the subgenome only support that *N. sylvestris* could have been involved in its formation (Figure 3G). Lastly, we summarized that the *N. sylvestris* (paternal progenitor) and the *N. glauca* (maternal progenitor) were involved in the formation of section *Suaveolentes* (Figure 5).

5 Conclusion

In conclusion, this study sheds light on the genetic diversity, phylogenetic relationships, and evolutionary history of *Nicotiana* species. The findings provide valuable insights into the classification and phylogenetic relationships within the genus. The identification of parental origins and the estimation of divergence times of polyploid species contribute to our understanding of speciation and hybridization events. Furthermore, the application of high-throughput RNA-seq technology in this study demonstrates its efficacy in phylogenetic studies and paves the way for future molecular systematic and population genetic investigations. The comprehensive dataset and analytical approaches used in this study serve as a valuable resource for further research in *Nicotiana* and related species.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

SW and JG conceived and designed the study. SW and ZL conducted the bioinformatics analysis. ZL, WP, and KC assisted in data collection. SW, JG, and CF wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Special Project for Lushan Plants [2021ZWZX18] and the Key Funding of CNTC [No. 110201801030 (JY-07)] and CNTC [No. 110202101003 (JY-03)].

Conflict of interest

SW, JG, KC, and WP were employed by China Tobacco Hunan Industrial Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This study received funding from CNTC No. 110201801030 JY-07 and CNTC No. 110202101003 JY-03. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. All authors declare no other competing interests.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1205683/full#supplementary-material

References

Angiuoli, S. V., and Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27, 334–342. doi: 10.1093/bioinformatics/btq665

Aoki, S., and Ito, M. (2000). Molecular phylogeny of *Nicotiana* (Solanaceae) based on the nucleotide sequence of the *matK* gene. *Plant Biol.* 2, 316–324. doi: 10.1055/s-2000-3710

Bally, J., Marks, C. E., Jung, H., Jia, F., Roden, S., Cooper, T., et al. (2021). *Nicotiana paulineana*, a new Australian species in *Nicotiana* section *Suaveolentes*. *Aust. Syst. Bot.* 34, 477–484. doi: 10.1071/SB20025

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2, 16074. doi: 10.1038/nplants.2016.74

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660. doi: 10.7717/peerj.1660

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PloS Comput. Biol.* 15, e1006650. doi: 10.1371/journal.pcbi.1006650

Brown, J. W., Walker, J. F., and Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics* 33, 1886–1888. doi: 10.1093/bioinformatics/btx063

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421

Chase, M. W., Knapp, S., Cox, A. V., Clarkson, J. J., Butsko, Y., Joseph, J., et al. (2003). Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann. Bot.* 92, 107–127. doi: 10.1093/aob/mcg087

Chen, C.-L., Zhang, L., Li, J.-L., Mao, X.-X., Zhang, L.-S., Hu, Q.-J., et al. (2021). Phylotranscriptomics reveals extensive gene duplication in the subtribe Gentianinae (Gentianaceae). *J. Syst. Evol.* 59, 1198–1208. doi: 10.1111/jse.12651

Chen, S., Zhou, Y., Chen, Y., and Bioinformatics, J. G. J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cheon, S., Lee, S.-G., Hong, H.-H., Lee, H.-G., Kim, K. Y., and Park, C. (2021). A guide to phylotranscriptomic analysis for phycologists. *Algae* 36, 333–340. doi: 10.4490/algae.2021.36.12.7

Cheon, S., Zhang, J., and Park, C. (2020). Is phylotranscriptomics as reliable as phylogenomics? *Mol. Biol. Evol.* 37, 3672–3683. doi: 10.1093/molbev/msaa181

Clarkson, J. J., Dodsworth, S., and Chase, M. W. (2017). Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). *Plant Syst. Evol.* 303, 1001–1012. doi: 10.1007/s00606-017-1416-9

Clarkson, J. J., Kelly, L. J., Leitch, A. R., Knapp, S., and Chase, M. W. (2010). Nuclear glutamine synthetase evolution in *Nicotiana*: Phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. *Mol. Phylogenet. Evol.* 55, 99–112. doi: 10.1016/j.ympev.2009.10.003

Clarkson, J., Knapp, S., Garcia, V. F., Olmstead, R. G., Leitch, A., and Chase, M. (2004). Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Mol. Phylogenet. Evol.* 33, 75–90. doi: 10.1016/j.ympev.2004.05.002

Clarkson, J. J., Lim, K. Y., Kovarik, A., Chase, M. W., Knapp, S., and Leitch, A. R. (2005). Long-term genome diploidization in allopolyploid *Nicotiana* section Repandae (Solanaceae). *New Phytol.* 168, 241–252. doi: 10.1111/j.1469-8137.2005.01480.x

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2019). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dodsworth, S., Jang, T.-S., Struebig, M., Chase, M. W., Weiss-Schneeweiss, H., and Leitch, A. R. (2017). Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). *Plant Syst. Evol.* 303, 1013–1020. doi: 10.1007/s00606-016-1356-9

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18, 448. doi: 10.1186/s12864-017-3791-6

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2

Fukushima, K., and Pollock, D. D. (2020). Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. *Nat. Commun.* 11, 4459. doi: 10.1038/s41467-020-18090-8

Goodspeed, T. H. (1956). The genus *Nicotiana*. *J. Am. Pharm. Assoc.* 45, 193–193. doi: 10.1002/jps.3030450326

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–U130. doi: 10.1038/nbt.1883

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Huson, D. H., and Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062

Junier, T., and Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* 26, 1669–1670. doi: 10.1093/bioinformatics/btq243

Kao, T.-T., Wang, T.-H., and Ku, C. (2022). Rampant nuclear–mitochondrial–plastid phylogenomic discordance in globally distributed calcifying microalgae. *New Phytol.* 235, 1394–1408. doi: 10.1111/nph.18219

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kelly, L. J., Leitch, A. R., Clarkson, J. J., Knapp, S., and Chase, M. W. (2013). Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section *suaveolentes*). *Evolution* 67, 80–94. doi: 10.1111/j.1558-5646.2012.01748.x

Khan, M. Q., and Narayan, R. K. J. (2007). Phylogenetic diversity and relationships among species of genus *Nicotiana* using RAPDs analysis. *Afr. J. Biotechnol.* 6, 148–162. doi: 10.5897/AJB06.442

Kitamura, S., Inoue, M., Shikazono, N., and Tanaka, A. (2001). Relationships among *Nicotiana* species revealed by the 5S rDNA spacer sequence and fluorescence in *situ* hybridization. *Theor. Appl. Genet.* 103, 678–686. doi: 10.1007/s001220100643

Knapp, S., Chase, M. W., and Clarkson, J. J. (2004). Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon* 53, 73–82. doi: 10.2307/4135490

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Leitch, I. J., Hanson, L., Lim, K. Y., Kovarik, A., Chase, M. W., Clarkson, J. J., et al. (2008). The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.* 101, 805–814. doi: 10.1093/aob/mcm326

Lemmon, E. M., and Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44, 99–121. doi: 10.1146/annurev-ecolsys-110512-135822

Lim, K. Y., Matyasek, R., Kovarik, A., Fulnecek, J., and Leitch, A. R. (2005). Molecular cytogenetics and tandem repeat sequence evolution in the allopolyploid *Nicotiana rustica* compared with diploid progenitors *N. paniculata* and *N. undulata*. *Cytogenetic Genome Res.* 109, 298–309. doi: 10.1159/000082413

Liu, Y.-Y., Jin, W.-T., Wei, X.-X., and Wang, X.-Q. (2022). Phylotranscriptomics reveals the evolutionary history of subtropical East Asian white pines: further insights into gymnosperm diversification. *Mol. Phylogenet. Evol.* 168, 107403. doi: 10.1016/j.ympev.2022.107403

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

Marks, C. E., Ladiges, P. Y., and Newbigin, E. (2011). Karyotypic variation in *Nicotiana* section *Suaveolentes*. *Genet. Resour. Crop Evol.* 58, 797–803. doi: 10.1007/s10722-011-9724-3

McCarthy, E. W., Arnold, S. E. J., Chittka, L., Le Comber, S. C., Verity, R., Dodsworth, S., et al. (2015). The effect of polyploidy and hybridization on the evolution of floral colour in *Nicotiana* (Solanaceae). *Ann. Bot.* 115, 1117–1131. doi: 10.1093/aob/mcv048

McCarthy, E. W., Chase, M. W., Knapp, S., Litt, A., Leitch, A. R., and Le Comber, S. C. (2016). Transgressive phenotypes and generalist pollination in the floral evolution of *Nicotiana* polyploids. *Nat. Plants* 2, 16119. doi: 10.1038/nplants.2016.119

Mehmood, F., Abdullah, Ubaid, Z., Shahzadi, I., Ahmed, I., Waheed, M. T., et al. (2020). Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). *PeerJ* 8, e9552. doi: 10.1101/2020.01.13.905158

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Morel, B., Kozlov, A. M., and Stamatakis, A. (2018). ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics* 35, 1771–1773. doi: 10.1093/bioinformatics/bty839

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* 105, 385–403. doi: 10.1002/ajb2.1016

Peng, J.-C., Ma, X.-G., Wang, Y.-H., and Sun, H. (2020). New insights into the evolutionary history of Megacodon: Evidence from a newly discovered species. *Plant Diversity* 42, 198–208. doi: 10.1016/j.pld.2020.05.003

Qi, X., Kuo, L.-Y., Guo, C., Li, H., Li, Z., Qi, J., et al. (2018). A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Mol. Phylogenet. Evol.* 127, 961–977. doi: 10.1016/j.ympev.2018.06.043

Renny-Byfield, S., Kovarik, A., Kelly, L. J., Macas, J., Novak, P., Chase, M. W., et al. (2013). Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J.* 74, 829–839. doi: 10.1111/tpj.12168

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13, 214. doi: 10.1186/1471-2148-13-214

Schiavinato, M., Marcet-Houben, M., Dohm, J. C., Gabaldón, T., and Himmelbauer, H. (2020). Parental origin of the allotetraploid tobacco *Nicotiana benthamiana*. *Plant J.* 102, 541–554. doi: 10.1111/tpj.14648

Sierro, N., Battey, J. N. D., Bovet, L., Liedschulte, V., Ouadi, S., Thomas, J., et al. (2018). The impact of genome evolution on the allotetraploid *Nicotiana rustica* – an intriguing story of enhanced alkaloid production. *BMC Genomics* 19, 855. doi: 10.1186/s12864-018-5241-5

Sierro, N., Battey, J. N. D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 5, 3833. doi: 10.1038/ncomms4833

Sierro, N., Battey, J. N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., et al. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* 14, R60. doi: 10.1186/gb-2013-14-6-r60

Smith, S. A., Moore, M. J., Brown, J. W., and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.* 15, 150. doi: 10.1186/s12862-015-0423-0

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi: 10.1093/bioinformatics/btm098

Svab, Z., and Maliga, P. (2007). Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proc. Natl. Acad. Sci.* 104, 7003–7008. doi: 10.1073/pnas.0700063104

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *System. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tria, F. D. K., Landan, G., and Dagan, T. (2017). Phylogenetic rooting using minimal ancestor deviation. *Nat. Ecol. Evol.* 1, 193. doi: 10.1038/s41559-017-0193

Wang, X., and Bennetzen, J. (2015). Current status and prospects for the study of *Nicotiana* genomics, genetics, and nicotine biosynthesis genes. *Mol. Genet. Genomics* 290, 1–11. doi: 10.1007/s00438-015-0989-7

Wang, S., Gao, J., Chao, H., Li, Z., Pu, W., Wang, Y., et al. (2022). Comparative chloroplast genomes of *Nicotiana* species (Solanaceae): Insights into the genetic variation, phylogenetic relationship, and polyploid speciation. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.899252

Wen, D., Yu, Y., Zhu, J., and Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* 67, 735–740. doi: 10.1093/sysbio/syy015

Willyard, A., Cronn, R., and Liston, A. (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Mol. Phylogenet. Evol.* 52, 498–511. doi: 10.1016/j.ympev.2009.02.011

Yan, Z., Cao, Z., Liu, Y., Ogilvie, H. A., and Nakhleh, L. (2021). Maximum parsimony inference of phylogenetic networks in the presence of polyploid complexes. *Syst. Biol.* 71, 706–720. doi: 10.1093/sysbio/syab081

Yang, Y., Moore, M. J., Brockington, S. F., Mikenas, J., Olivieri, J., Walker, J. F., et al. (2018). Improved transcriptome sampling pinpoints 26 ancient and more recent

polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* 217, 855–870. doi: 10.1111/nph.14812

Yang, L., Su, D., Chang, X., Foster, C. S. P., Sun, L., Huang, C.-H., et al. (2020). Phylogenomic insights into deep phylogeny of Angiosperms based on broad nuclear gene sampling. *Plant Commun.* 1, 100027. doi: 10.1016/j.xplc.2020.100027

Yu, Y., and Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 16, S10. doi: 10.1186/1471-2164-16-S10-S10

Yukawa, M., Tsudzuki, T., and Sugiura, M. (2006). The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol. Genet. Genomics* 275, 367–373. doi: 10.1007/s00438-005-0092-6

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19, 153. doi: 10.1186/s12859-018-2129-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 August 2023
DOI 10.3389/fpls.2023.1238624

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Muhammad Anwar,
Hainan University, China
Xiangpeng Leng,
Qingdao Agricultural University, China

*CORRESPONDENCE
Ronggao Gong
gongronggao@sicau.edu.cn

RECEIVED 12 June 2023
ACCEPTED 26 July 2023
PUBLISHED 16 August 2023

CITATION
Zhang Y, Chen C, Cui Y, Du Q, Tang W, Yang W, Kou G, Tang W, Chen H and Gong R (2023) Potential regulatory genes of light induced anthocyanin accumulation in sweet cherry identified by combining transcriptome and metabolome analysis. *Front. Plant Sci.* 14:1238624. doi: 10.3389/fpls.2023.1238624

COPYRIGHT
© 2023 Zhang, Chen, Cui, Du, Tang, Yang, Kou, Tang, Chen and Gong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Potential regulatory genes of light induced anthocyanin accumulation in sweet cherry identified by combining transcriptome and metabolome analysis

Yao Zhang, Chaoqun Chen, Yiling Cui, Qinsong Du, Wenjing Tang, Wenlong Yang, Guanqiong Kou, Wanjia Tang, Hongxu Chen and Ronggao Gong*

College of Horticulture, Sichuan Agricultural University, Chengdu, China

Anthocyanins exist widely in various plant tissues and organs, and they play an important role in plant reproduction, disease resistance, stress resistance, and protection of human vision. Most fruit anthocyanins can be induced to accumulate by light. Here, we shaded the "Hong Deng" sweet cherry and performed an integrated analysis of its transcriptome and metabolome to explore the role of light in anthocyanin accumulation. The total anthocyanin content of the fruit and two of its anthocyanin components were significantly reduced after the shading. Transcriptome and metabolomics analysis revealed that *PAL, 4CL, HCT, ANS* and other structural genes of the anthocyanin pathway and cyanidin 3-O-glucoside, cyanidin 3-O-rutinoside, and other metabolites were significantly affected by shading. Weighted total gene network analysis and correlation analysis showed that the upstream and middle structural genes *4CL2, 4CL3*, and *HCT2* of anthocyanin biosynthesis may be the key genes affecting the anthocyanin content variations in fruits after light shading. Their expression levels may be regulated by transcription factors such as *LBD, ERF4, NAC2, NAC3, FKF1, LHY, RVE1*, and *RVE2*. This study revealed for the first time the possible role of *LBD, FKF1*, and other transcription factors in the light-induced anthocyanin accumulation of sweet cherry, thereby laying a preliminary foundation for further research on the role of light in anthocyanin accumulation of deep red fruit varieties and the genetic breeding of sweet cherry.

KEYWORDS

sweet cherry, anthocyanin, Shading, WGCNA, 4CL

1 Introduction

Various kinds of plants are growing on the earth, and they show rich colors because they are rich in different pigments. These colors not only meet the needs of the plant's own growth and reproduction, but also bring visual enjoyment to people. Among all the colors, red, blue, and purple are mainly affected by anthocyanin concentration (Rao et al., 2021). Anthocyanins are important secondary metabolites in higher plants. They belong to the flavonoid group and are water-soluble pigments (Zhou et al., 2022). Anthocyanins are usually synthesized in the cytoplasm and transported to vacuole storage through a complex transport system (Passeri et al., 2016). As one of the three major pigments of plants (Zhao et al., 2022b), anthocyanins promote pollination by insects to help plant propagation and help plants to resist low temperature, drought, salt, and other stresses to a certain extent (Alcalde-Eon et al., 2013; Zhang et al., 2019; Kim et al., 2022). With the deepening of research, increasing evidence from recent years indicates that anthocyanins also play an important role in maintaining physical health, such as anti-aging and vision protection properties and the reduction of the risk of diabetes and cancer (Gonçalves et al., 2021).

As the main environmental factors affecting plant growth and development, light, temperature, and water also affect the synthesis of plant anthocyanins (Li et al., 2019). In many studies, plants under different stress conditions will show varying anthocyanin contents compared with the counterparts under a normal environment. The anthocyanin content of red pears increased after white light exposure compared with double-layer paper bag coverage (Bai et al., 2019). After low temperature treatment at 16 °C, the upper leaves of the "Gala" apple showed obvious red spots (Song et al., 2019a). In the study of Jian-Ping An et al, apple seedlings after drought stress also showed an increase in anthocyanin accumulation (An et al., 2020b). Among these environmental factors, light is particularly important for anthocyanin synthesis in horticultural plants.

At present, the anthocyanin synthesis pathway has been examined in-depth in Arabidopsis, apple, strawberry, grape and other plants (Leong et al., 2018; Jiu et al., 2021; Liu et al., 2021; Nguyen et al., 2023). Anthocyanin is mainly synthesized by phenylalanine and malonyl-coenzyme A through a variety of enzymatic reactions involving structural genes such as *PAL*, *4CL*, *CHS*, *CHI*, *F3H*, *F3'H*, *DFR*, *ANS*, and *UFGT* (Jaakola, 2013). In addition to the influence of structural genes, the synthesis of anthocyanins is also controlled by many regulatory genes. The MBW ternary complex is the most widely studied regulatory factor in the anthocyanin synthesis pathway (Ma and Constabel, 2019). The said complex consists of the *R2R3 MYB*, basic helix-loop-helix *bHLH*, and *WD40* regulatory factors in plants and directly regulates various structural genes in anthocyanin biosynthesis. Anthocyanin synthesis is also regulated by other transcription factors such as alkaline leucine zipper *bZIP*, *WRKY*, and *NAC*, which function mainly through direct or indirect action on the MBW complex(An et al., 2018; Sun et al., 2019; Yue et al., 2022).

Similarly, the induction of light in plant anthocyanin biosynthesis is regulated by transcription factors. In red pear, *BBX16* can positively regulate the accumulation of photoinduced anthocyanins by activating *MYB10*, which significantly increases anthocyanin content in fruits after light restoration (Bai et al., 2019). *PybZIPa*, another important transcription factor, activates *PyUFGT* to participate in the light-induced anthocyanin accumulation in red pears by binding to the tandem G-box in the promoter (Liu et al., 2019a). Anthocyanin accumulation in apples is also induced by light (Yang et al., 2019). Studies have shown that *MdMYB1*, *MdBBX21*, *MdTCP46* and other regulatory factors are involved in anthocyanin accumulation (An et al., 2020a; Yang et al., 2021; Zhang et al., 2021), and the degree of anthocyanin accumulation in apples varies under different light intensities (Chen et al., 2019). In peaches, *PpHYH* activates the transcription of three *PpMYB10* gene clusters in the presence of cofactor *PpBBX4*, leading to anthocyanin accumulation in the peel under sunlight (Zhao et al., 2022a). Purple broccoli's anthocyanin biosynthesis genes such as *PAL*, *4CL*, and *CHI* were significantly downregulated after shading, and the anthocyanin content decreased (Liu et al., 2020). These results indicate that light plays an important role in regulating anthocyanin biosynthesis in plants.

Sweet cherry (*Prunus avium L.*) is a perennial light-loving horticulture crop of the genus *Prunus* in the Rosaceae family, originating in the region between northeastern Anatolia, the Caucasus, and the Caspian Sea (Davis, 1975). Today, the fruit tree is grown globally in more than 40 countries with mild climates (Ceccarelli et al., 2020). Given their attractive appearance, delicious taste and rich nutritional value, sweet cherries are widely favored by consumers and have considerable economic value (Wei et al., 2015). At present, however, we know very little about the role of light in sweet cherries, especially red sweet cherry varieties. Thus, this study preliminarily explored the effect of light on anthocyanin accumulation in red sweet cherry varieties and its mechanism, thereby laying a foundation for further research on the role of light in anthocyanin accumulation in deep red fruit varieties and the genetic breeding of sweet cherry.

2 Materials and methods

2.1 Plant materials and treatment

"Hong deng" sweet cherry grafted on Prunus tomentosa was used as the test material in this study and was obtained from the sweet cherry test site in Buwa Village, Wenchuan County, Aba Tibetan and Qiang Autonomous Prefecture, Sichuan Province. Six sweet cherry trees aged 12 years with the same growth conditions and growth period were randomly divided into two groups of three trees each (i.e., three replicates). The test materials were uniformly managed from the end of March 2022 (initial flowering stage). In the experiment, the 'Hong Deng' sweet cherry fruit after shading bagging (SHD) was used as the treatment, and the normal light fruit without bagging (LHD) was used as the control. The sweet cherry fruit in the treatment group was bagged on April 25, 2022 (23 days after anthesis [DAA], fruit expansion stage), 300 fruits per tree were used for experimental treatment. Then, we randomly removed

some of the fruit bags (RHD) the treatment group at the fruit maturity stage (55 DAA). Sampling began 23 DAA on the bagging day.15 sweet cherry fruits of the same size and which were free of diseases, pests, and mechanical damage were picked from the east, south, west, and north directions of the crown of each tree in each community according to a random sampling method, and were collected every 3 days until the fruits were ripe (63 DAA). A total of 45 fruits were collected per treatment in each period and a total of 11 period were collected. The samples collected each time were immediately placed in an ice box and brought back to the laboratory. After being photographed, the samples were quickly cut into uniform blocks, subjected to liquid nitrogen quick freezing, wrapped in foil, and stored in a refrigerator at -80 °C for use. Three biological replicates were set for each sample.

2.2 Determination of fruit color, anthocyanin content and composition

A total of 15 fruit samples with similar maturity and color at each stage were selected for each treatment. The skin color difference at three points on the equatorial line of fruits was measured by an automatic chromometer. The L*, a*, and b* values were also recorded.

Total anthocyanin content was determined by the hydrochloric methanol method described as follows (Zhang et al., 2018a). Weighed 0.2 g sample and add 5 mL HCl/methanol (1/99, v/v), shaked the mixture well, and stored at 4 °C away from light for more than 20 h, followed by ultrasonic extraction at 4 °C for 30 min. Then, the mixture was centrifuged at 8000 rpm and 4 °C for 10 min. Absorbance was measured at 530, 620, and 650 nm. Calculate A=(A530-A620)-0.1(A650-A620), and the result was expressed as the amount of nmol anthocyanins contained in each g of sample.

Anthocyanin components were determined by high performance liquid chromatography (HLPC) using the Agilent 1260 II liquid chromatograph on Comatex C18 column (250 mm × 4.6 mm, 5 μm) with a diode array detector (Chen et al., 2022). The detection wavelength was 520 nm, the column temperature was 30 °C, the sample size was 10 μm, and the flow rate was 1 mL/min. The mobile phase A was acetonitrile and the mobile phase B was 1.6% formic acid aqueous solution. The gradient elution procedure is as follows: 0–15 min 95%–85% B, 15–21 min 85%–72% B, 21–22 min 72%–60% B, 22–24 min 60%–40% B, 24–27 min 40%-95% B, and 27–30 min 95% B.

2.3 Metabolite assay

2.3.1 Metabolite extraction

A tissue sample of 100 mg liquid nitrogen grinding was taken and placed in an EP tube with 500 μL of 80% methanol aqueous solution. The sample was subjected to vortex shock, ice bath for 5 min, centrifugation at 15000 g and 4 °C for 20 min. A certain amount of the supernatant was taken, diluted with mass spectrometry water until the methanol content was 53%, and centrifuged at 15000 g and 4 °C for 20 min. The supernatant was then collected and injected into LC-MS for analysis.

2.3.2 Metabolite detection and analysis of chromatographic conditions

We used a HypesilGoldcolumn (C18, 100 × 2.1 mm, 1.9 μm) column to inject 2 μL of samples at a flow rate of 0.2 mL/min and column temperature of 40°C with an automatic injector set at 8°C. The positive and negative modes were adopted, with 0.1% formic acid as the positive mobile phase A and methanol as the mobile phase B. The negative mode mobile phase A is 5 mM ammonium acetate, pH 9.0, and the mobile phase B is methanol. The gradient elution procedure is as follows: 0–1.5 min, 98%–15% A, 2%–85% B; 1.5–3 min, 15%–0% A, 85%–100% B; 3–10 minutes, 0%–98% A, 100%–2% B; 10–10.1 minutes, 98% A, 2% B; and 11–12 minutes, 98% A, 2% B.

The mass spectrum conditions are as follows. The scanning range is 100–1500 m/z. The ESI source settings include the spray voltage of 3.5 kV; sheath gas flow rate of 35 psi; Aux gas flow rate of 10 L/min; ion transfer tube temperature (capillary temp) of 320°C; ion import RF level (S-lens RF level) of 60; Aux gas heater temp of 350°C; and polarities: positive, negative. The MS/MS secondary scans are data-dependent scans.

2.3.3 Metabolome data preprocessing and metabolite identification

The raw file was imported into the CD 3.1 library search software for processing, and the retention time, mass–charge ratio, and other parameters of each metabolite were screened. The retention time deviation of 0.2 min and mass deviation of 5 ppm were set for the peak alignment of different samples, followed by peak extraction and quantification of the peak area. Then, the target ions were integrated, and the molecular formula was predicted by molecular ion peak and fragment ion and compared with the mzCloud (https://www.mzcloud.org/), mzVault, and Masslist databases. After standardized processing, the relative peak area was obtained. Compounds with CV greater than 30% of the relative peak area were deleted from the QC samples. Finally, the identification and relative quantitative results of metabolites were obtained. The data processing involved a Linux operating system (CentOS version 6.6) and the R and Python software.

2.3.4 Statistical analysis of metabolome data

The KEGG database (https://www.genome.jp/kegg/pathway.html) was employed to identify the metabolites. In the multivariate statistical analysis, the metabolomics data processing software metaX was used to transform the data. Then, principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) were performed to obtain the VIP value of each metabolite. In the univariate analysis, the statistical significance (P value) of each metabolite between the two groups was calculated based on a t test, and the fold change (FC value) of the metabolite between the two groups was calculated. The default criteria for differential metabolite screening were VIP >1, P value <0.05, and FC ≥2 or ≤0.5. The cluster heatmap was drawn with the R-packet Pheatmap, and the metabolite data were normalized with a z-score.

The KEGG database was also utilized to examine the function and metabolic pathway of the metabolites. When x/n>y/n, the metabolic pathway was considered enriched; when the P value of the metabolic pathway was <0.05, the metabolic pathway was considered significantly enriched.

2.4 Transcriptome sequencing

2.4.1 RNA extraction and transcriptome sequencing library preparation

Total RNA was extracted using a total RNA kit (Beijing Tiangen Biotechnology Co., LTD., China), and the integrity and total RNA was accurately detected by the Agilent 2100 bioanalyzer. The starting RNA of the library was total RNA. mRNA with polyA tail was enriched by Oligo(dT) magnetic beads. The mRNA fragments were randomly interrupted by divalent cations in the fragmentation buffer. The first cDNA strand was synthesized in the M-MuLV reverse transcriptase system, then degraded by RNaseH. The second cDNA strand was synthesized in the DNA polymerase I system using dNTPs as raw material. The purified double-stranded cDNA was end-repaired, A-tail was added, and sequencing joints were connected. The cDNA of approximately 370~420 bp were screened with AMPureXP beads for PCR amplification, and AMPureXP beads were used again to purify the PCR products to obtain the A library.

After the library construction, the library quality inspection was performed. Once the library inspection was qualified, different libraries were employed for Illumina sequencing by pooling according to the requirements of effective concentration and target data volume. Consequently, a 150 bp paired end reading was generated. The basic principle of sequencing entails sequencing by synthesis described as follows. Add four kinds of fluorescently labeled dNTP, DNA polymerase and connector primers to the sequencing flow cell for amplification. When each sequencing cluster extends the complementary chain, each addition of a fluorescently labeled dNTP can release a corresponding fluorescence. The sequencer will capture the fluorescence signal and convert the light signal into a sequencing peak through a computer software, thereby obtaining the sequence information of the fragment to be tested.

2.4.2 Transcriptome data analysis

Image data measured by high-throughput sequencers were converted into sequence data (reads) by CASAVA base recognition in the fastq format. The raw data from the sequencing contained a small number of reads with sequencing connectors or of low sequencing quality. To ensure the quality and reliability of data analysis, the raw data required filtration. These raw data include reads with the adapter, reads containing N (which means that base information cannot be determined), and low-quality reads (those with Qphred <=20 base number accounting for more than 50% of the entire read length). The Q20, Q30, and GC contents of the clean data were also calculated. All subsequent analyses were based on high quality analysis conducted by clean data.

We downloaded the reference Genome and Gene Model annotation file directly from the genome website. Then, using 3M HISAT2 v2.0.5, an index of the reference genome was built and the paired end clean reads were matched to the reference genome. New gene prediction was performed with StringTie (1.3.3b) (Pertea et al., 2015). Gene expression level quantification featureCounts (1.5.0-p3) was employed to calculate the readings mapped to each gene. The FPKM of each gene was then calculated according to the length of the gene, and the reading mapped to that gene was calculated.

Differential expression analysis between the two comparison combinations was performed using the DESeq2 software (1.20.0). Benjamini and Hochberg's method was used to adjust the resulting P-values to control the error discovery rate. DESeq2 found that genes with adjusted P values <=0.05 were assigned as differentially expressed. The corrected P-value and |log2foldchange| were employed as thresholds for significant different-expression. The statistical enrichment of differentially expressed genes in the KEGG pathway was analyzed using the clusterProfiler (3.8.1) software. We utilized a local version GSEA analysis tool http://www.broadinstitute.org/gsea/index.jsp to analyze the GSEA KEGG data set for this species.

2.5 Combined transcriptome and metabolome analysis

All the obtained differentially expressed genes (DEGs) and differentially expressed genes metabolites (DEMs) were mapped simultaneously in the KEGG pathway database to obtain co-enriched KEGG pathway information. According to the enrichment pathways of the DEG, the top 10 metabolic pathways for the co-significant enrichment of the DEM and DEG in each comparison group were identified, and GraphPad 8.0.2 was used to generate a histogram. The anthocyanin biosynthesis pathway map was drawn with reference to the KEGG pathway database. The FPKM values of the DEGs and the relative quantitative values of the DEMs involved in the path were standardized and plotted into heat maps to the pathway map.

2.6 Weighted correlation network analysis

Weighted total gene network analysis (WGCNA) was conducted with the TBtools software (https://github.com/ShawnWx2019/WGCNA-shinyApp). A total of 4217 DEGs FPKM values were calculated from 24 transcriptome samples (4 time points, 3 replicates). The network type was unsigned and the correlation type was Pearson. R^2 > 0.85 was selected as the soft threshold standard and the soft threshold was set to 10, the merge cut height was 0.2, the min module size was 30. Correlation analysis was conducted between the content of the two anthocyanins and the obtained modules, and the criterion for selection of candidate modules was correlation coefficient >0.75. The hub gene was screened with gene significance (GS) >0.5 and module

membership (MM) >0.8 (MM>0.9 in the turquoise module) as the criteria (Song et al., 2019b).

2.7 Analysis of candidate gene correlation and promoter cis-acting elements

All candidate gene sequences were submitted to the PlantTFDB database (http://planttfdb.gao-lab.org/) to predict possible transcription factors. The correlation of the selected candidate genes and anthocyanin components in the co-expression network was analyzed with Cytoscape_V.3.7.1. Plant CARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict promoter cis-acting elements of upstream 2000 bp initiation codon of candidate genes. Then, the component prediction information was input into the TBtools (http://www.tbtools.org/) software for mapping.

2.8 qRT-PCR analysis

Total RNA from the plants was extracted using a total RNA extraction kit. The integrity of the total RNA was detected by 1.0% agarose gel electrophoresis with 3 ul of the obtained RNA products. Then, cDNA was synthesized by the HiScript® III RT SuperMix for qPCR (+gDNA wiper) kit. The reaction conditions of reverse transcription PCR were 25 °C, 10 min; 50 °C, 15 min; and 85 °C, 5 min. The configuration of the reaction solution needed to be performed on the ice, and after its reverse transcriptional synthesis, the first strand of cDNA was stored in the refrigerator at -20 °C for use. qRT-PCR was performed using the CFX96TM real-time system (Bio-Rad, California, USA) and 2 × TSINGKE® Master qPCR Mix (SYBR Green I) (TSINGKE, Beijing, China) reagents. The PCR primer (S1) was designed using Primer6.0. Using the cDNA as template and β-actin gene as the internal reference, real-time fluorescence quantitative PCR was employed to analyze the expression of related genes. Three biological repeats and technical repeats were observed in each reaction. The amplification procedure involved the following parameters: 95°C, 30 s predenaturation; 95° C, 10 s denaturation, 60°C, 30 s annealing and extension, 40 cycles. The dissolution curve used the instrument default acquisition procedure, using $2^{-\Delta\Delta Ct}$ to calculate the relative gene expression (Livak & Schmittgen, 2001).

3 Results and analysis

3.1 Changes of anthocyanin content in sweet cherry fruit after shading treatment

After the shading treatment, the color of sweet cherry fruits changed significantly. Compared with the fruits under normal light conditions, a lighter red color was observed on the sweet cherry fruits after the shading treatment (Figure 1A). No significant difference in the color of sweet cherry fruits was noted in the early stage of shading (23-31 DAA), the a*/b* value was negative, and all fruits were green. At approximately 35 DAA, the sweet cherry fruit entered the whitening stage and began to turn color, and the a*/b* value approached zero. Then, the color gradually turned to deep red, the a*/b* values of the fruits began to change,

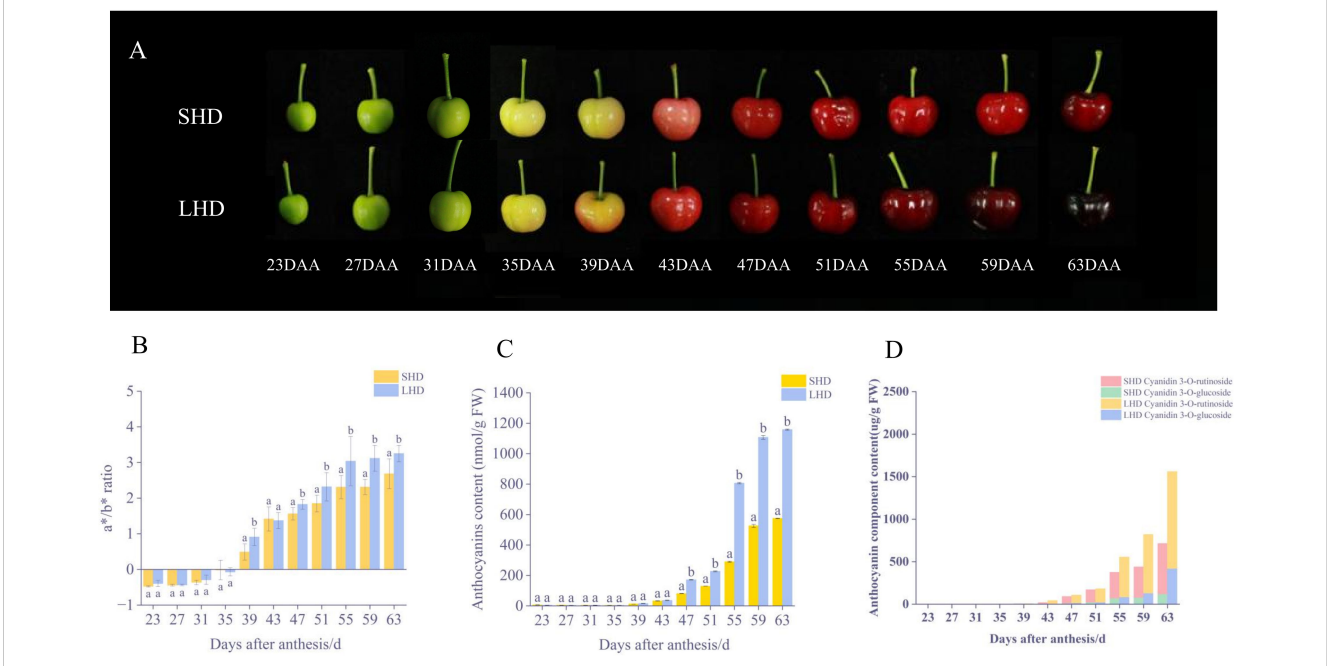


FIGURE 1
Fruit color changes after shading. **(A)** SHD means Shading "Hong Deng," LHD means "Hong Deng" under normal light, DAA means days after anthesis during fruit growth and development. **(B)** Changes in fruit color **(C)** Changes in total anthocyanin content **(D)** Changes in anthocyanin component content of fruit.

and the SHD was almost always lower than the LHD till fruit harvest (Figure 1B). Consistent with the changes of fruit phenotype and color, the anthocyanin content of the sweet cherry fruits after shading treatment was lower than that of fruits under normal light since 39 DAA. This difference has become significant since 55 DAA, and the difference level has been maintained (Figure 1C).

To investigate the effects of shading on anthocyanin components, the said components and contents were determined by HPLC. Two anthocyanins were obtained, cyanidin 3-O-glucoside and cyanidin 3-O-rutinoside (Figure 1D). Between them, cyanidin 3-O-rutinoside was the main anthocyanin component in "Hong Deng" sweet cherry, accounting for 65.34% of the total anthocyanin. The results showed that the contents of these two components decreased significantly under the influence of shading treatment, an outcome which was consistent with the changes of total anthocyanin and fruit phenotype.

3.2 Metabolome analysis

Figure 1 showed that after shading treatment, the color and anthocyanin content of sweet cherry fruit changed strongly at 39, 47, 55, and 63 DAA, with substantial differences. To explore the mechanism of influence of light shading on sweet cherry fruit color change, samples from the above four periods were selected for further metabolome and transcriptome analysis.

The metabolome PCA results were shown in Figure 2A. A total of 48 samples could be clearly distinguished by the first two principal components, which accounted for 31.2% and 16.8% of the total variability, respectively. Metabolites were divided into 8 groups in total. The repeated samples within each group were closely related, and the distance between the groups was relatively far, indicating the reliability of the metabolome data. Obvious differences were observed in the diagram of sweet cherry samples in the four periods. Fruit samples at 55 and 63 DAA were distributed in the negative end of PC1, and those at 39 and 47 DAA were distributed in the positive end of PC1. That is, fruit samples at the first two periods were significantly different from those at the last two periods. The close distance between SHD63 and LHD55 indicated the strong similarity of metabolites between the two groups. LHD63 was separately distributed in the negative end of PC1 and the positive end of PC2 and was far away from SHD63, indicating that the metabolites were significantly different between the two groups at 63 days after flowering. A total of 1065 metabolites were identified in the metabolome, including 341 positive ions and 724 negative ions. According to KEGG functional annotation, the identified metabolites can be roughly divided into three categories, of which the metabolism group had

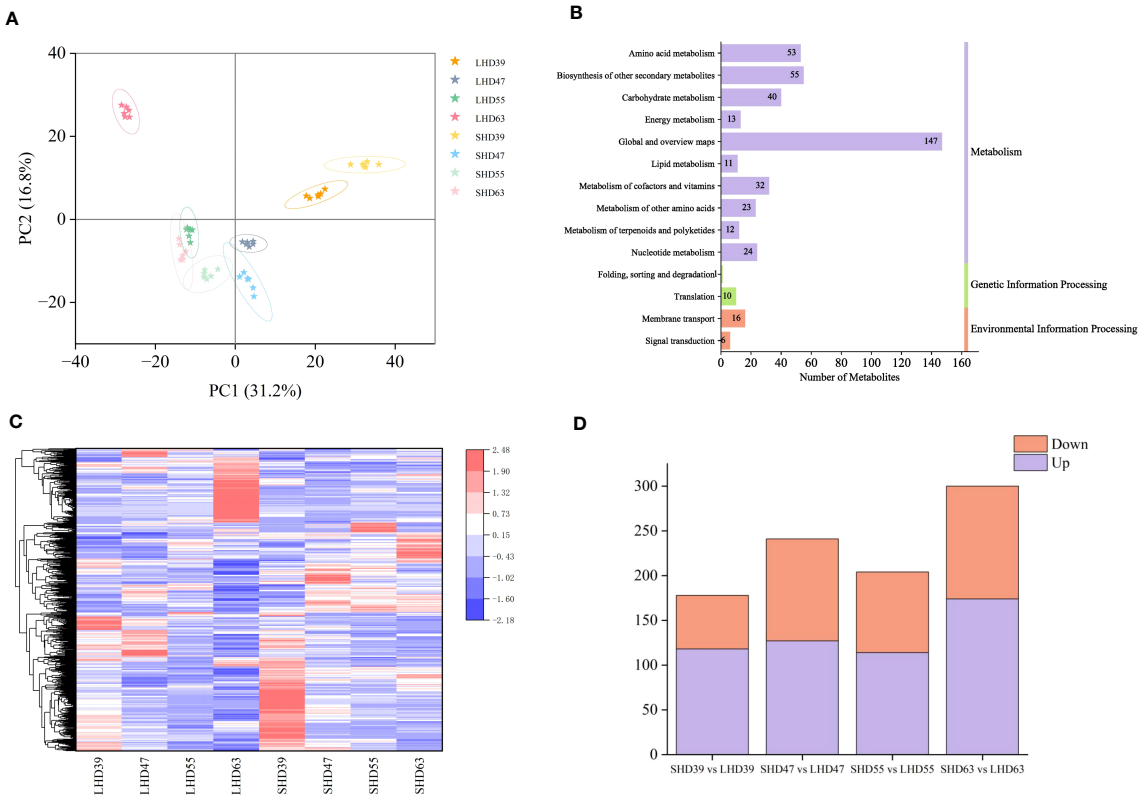


FIGURE 2
Metabolomic analysis of four periods of fruit after shading treatment. **(A)** PCA analysis of metabolites in different treatment groups at each stage; **(B)** KEGG pathway annotation of metabolites; **(C)** Cluster heat maps of all DEMs in the four periods, with relative levels of metabolites ranging from low (blue) to high (red); **(D)** Numbers of DEMs in the upper (purple) and lower (orange) tones during the four periods.

the largest number of metabolites, reaching 410, accounting for 92.55% (Figure 2B).

With threshold VIP > 1.0, difference multiple FC > 1.5 or FC < 0.667, and P-value < 0.05, 678 DEMs were selected. The expression patterns of the DEMs were significantly different among the groups. As shown in Figure 2C, the low expression levels of DEMs in the treatment group may be the main cause of the fruit color differences. Comparison of the LHD and SHD samples at different periods revealed that the combinations with the most differentiated metabolites were LHD63 vs. SHD63 (a total of 300 metabolites, 174 up-regulated and 126 down-regulated), and the least differentiated metabolites were LHD39 vs. SHD39. In general, more up-regulated metabolites were observed in the four periods (Figure 2D). In summary, shading exerted an effect on the metabolites of sweet cherry fruits in multiple post-flowering periods, and the effect was greatest at 63 days after flowering.

3.3 Transcriptome analysis

The transcriptomic data of sweet cherry fruit were analyzed by RNA-seq technique after light shading treatment. A total of 1,165,020,240 raw data were generated from the fruit samples over four periods, and 1,145,349,294 high-quality clean readings were obtained after filtering out junction sequences, uncertain readings, and low-quality readings. On average, 95.01% of the clean readings were located on the sweet cherry genome. A total of 45,071 transcripts were used for subsequent analysis (S2). PCA analysis of transcription sample expression was performed (Figure 3A). Each sample was clearly well distinguished on the score map, and the focus between each reset was tight, indicating differences in fruit transcripts after shading treatment. Unlike the metabolome results, LHD39, LHD47, SHD39, SHD47, and some LHD55 were on the negative end of PC1, and the rest were on the

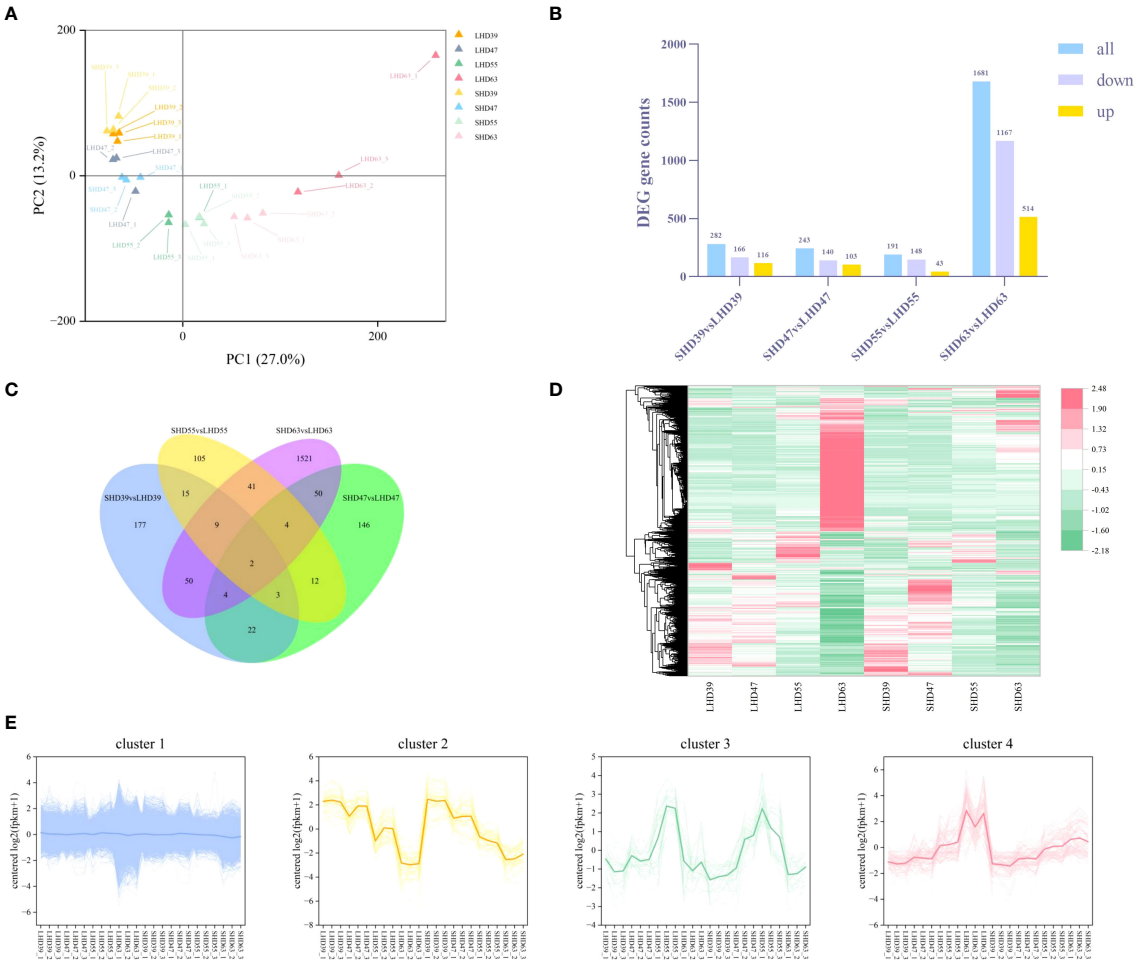


FIGURE 3
Transcriptome data analysis of fruit at four stages after shading treatment **(A)** PCA analysis of gene expression in different treatment groups at each stage; **(B)** Quantity statistics of the DEGs in the four periods, with yellow representing up-regulation, purple representing down-regulation, and blue representing total DEGs quantity; **(C)** Venn plots of common or unique expression of DEGs within and between comparison groups in each period; **(D)** Cluster heat maps of the expression of all differentially metabolized genes in the four periods, with red representing up-regulation and green representing down-regulation; **(E)** Expression patterns of all DEGs in the four periods, with different colors representing different expression trends, and dark thick lines representing the average expression profiles of all genes in each cluster.

10.3389/fpls.2023.1238624

positive end of PC1. Interestingly, the dispersion between LHD63 and SHD63 was greater in the score map than in the other three periods, suggesting that the shading treatment had a greater effect on the samples in the last period.

The threshold was set to |log2(FoldChange)| >= 1&padj <= 0.05. A total of 4217 DEGs were identified. The DEGs results for comparison among the treated samples in different periods were shown in Figure 3B. Similar to the results of the metabolic group, LHD63 vs. SHD63 had the most DEGs among the three groups of comparison, reaching 1681. The down-regulated DEGs outnumbered the up-regulated counterparts. The common or unique DEGs results among the three comparison groups were analyzed (Figure 3C). Only 2 genes were differentially expressed in all samples, and most differentially expressed genes were only in LHD63 vs. SHD63 (1521). The FPKM value of the DEGs was used to draw the hierarchical clustering heat map (Figure 3D). Significant differences occurred among the treated samples, an outcome which was similar to the results of the metabolic group. Moreover, the expression levels of numerous genes in SHD63 were significantly down-regulated. The H-cluster function was used to divide all DEGs expression patterns into 4 groups, and all samples revealed good repeatability (Figure 3E). Cluster 1 contained the largest number of DEGs at 4041, accounting for 96% of the total DEGs. The expression level of DEGs in Cluster 2 was down-regulated with growth and development. In Cluster 3, the DEGs were gently up-regulated to 55 DAA and then down-regulated. In Cluster 4, the expression level of DEGs was up-regulated with fruit growth and development, and the expression level of DEGs in SHD was generally lower than that in LHD. Similar to the variation trend of anthocyanin content, the DEGs in Cluster 4 may be related to the decrease of fruit anthocyanin content after shading.

3.4 Transcriptome metabolome combined analysis

3.4.1 KEGG enrichment analysis

To further determine the metabolic pathways of co-enrichment of DEGs and DEMs, we mapped the top 10 KEGG pathways of co-enrichment of DEGs and DEMs in each comparison group, showing 26 enriched pathways (Figure 4). Cysteine and methionine metabolism and carbon metabolism appeared most frequently (3 times). Of all the DEMs enriched KEGG pathways, the most significantly enriched metabolic pathways occurred at 63 DAA. These pathways included phenylpropanoid biosynthesis, glutathione metabolism, nitrogen metabolism, amino sugar, and nucleotide sugar metabolism. Among all the DEMs enriched KEGG pathways, an extremely apparent flavonoid biosynthesis occurred at 55 DAA. In addition, the circadian rhythm-plant pathway was present in the top 20 co-enriched metabolic pathways of both DEGs and DEMs, as detailed in S3. These results indicated a greater effect of shading treatment on structural genes and metabolites related to the upper and middle pathway of anthocyanin biosynthesis (phenylpropanoid biosynthesis and flavonoid biosynthesis in fruit).

3.4.2 Analysis of the anthocyanin biosynthesis pathway

To more clearly and visually demonstrate the changes of genes and metabolites in the anthocyanin biosynthesis pathway of fruit after shading treatment, we mapped the said pathway and thermologically mapped the DEGs and DEMs data (Figure 5). The pathway mainly included three paths: phenylpropanoid biosynthesis (upstream of anthocyanin biosynthesis), flavonoid biosynthesis (middle stream of anthocyanin biosynthesis), and anthocyanin metabolism (downstream

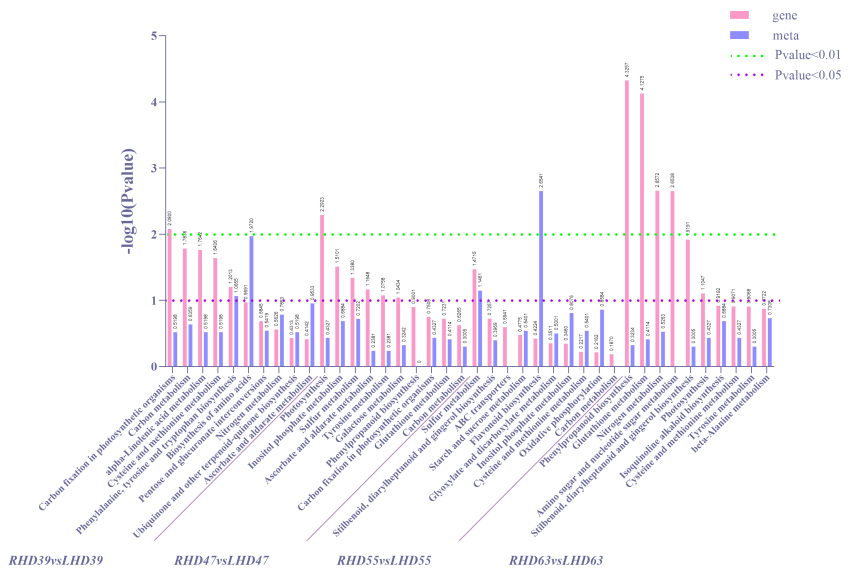


FIGURE 4
Statistical diagrams of the first 10 KEGG pathways co-enriched by DEMs and DEMs according to DEGs in each comparison group after shading treatment, in which differential gene P-value for enrichment analysis is ≤0.5.

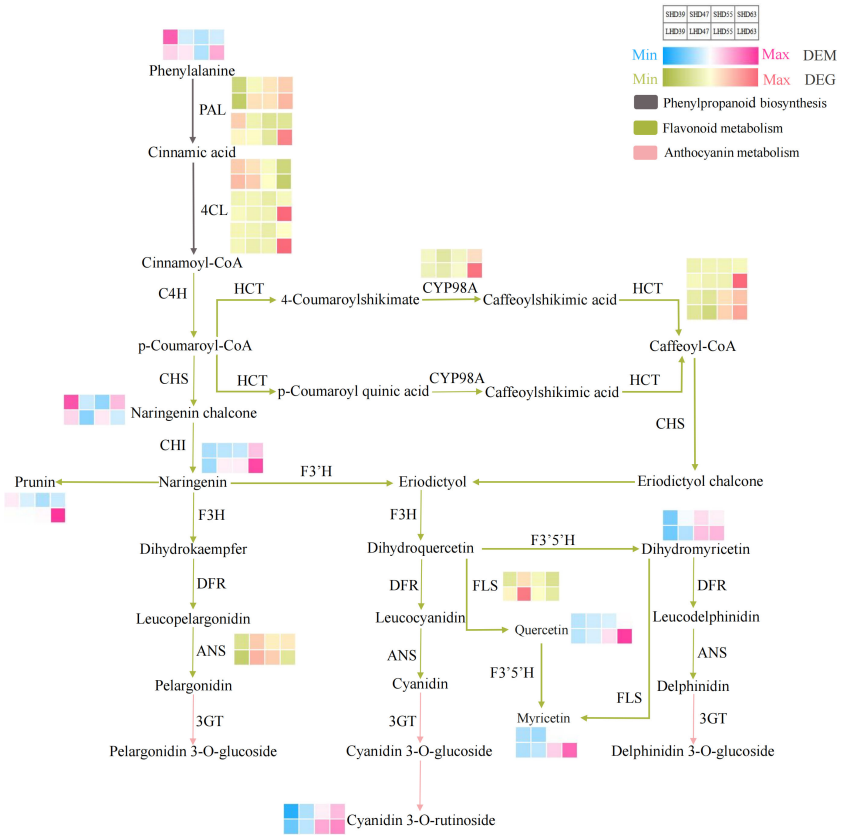


FIGURE 5
Anthocyanin biosynthesis pathway composed of phenylpropanoid biosynthesis, flavonoid biosynthesis, and anthocyanin metabolism. The red and blue blocks represent the DEMs. Red and green blocks represent DEGs.

of anthocyanin biosynthesis). The pathway involved 10 DEGs and 8 DEMs.

Phenylalanine was an important precursor substance in phenylpropanoid biosynthesis and the anthocyanin biosynthesis pathway. As the only DEM in the phenylpropanoid biosynthesis pathway, the phenylalanine content increased in the early period of blackout (39 DAA), and decreased in the following three periods compared with the counterpart with normal light. This phenylpropanoid biosynthesis pathway also involved 5 DEGs, 2 *PAL* and 3 *4CL*. The expression of *PAL* was slightly up-regulated at 39 DAA and was down-regulated at other periods. The expression levels of two *4CL* genes were significantly down-regulated at 63 DAA. However, the expression level of one *4CL* gene decreased with fruit development, slightly down-regulated in the early period of anthesis, and up-regulated in the later period (63 DAA).

Flavonoid biosynthesis and anthocyanin metabolism involved 5 DEGs and 7 DEMs. In the flavonoid biosynthesis pathway, the difference of *CYP98A* expression was most significant at 63 DAA. The expression trended of the two *HCT* vary, but both their expression levels were down-regulated after fruit shading. No significant difference occurred in the expression of *ANS* except for slight up-regulation in the fruit at 63 DAA. The expression level of *FLS* decreased after shading and was most obvious at 47 DAA. The DEMs in the flavonoid biosynthesis pathway included prunin, quercetin, naringenin chalcone, naringenin, dihydromyricetin, and myricetin. Except for naringenin chalcone, the contents of the other DEMs in this pathway were decreased by shading treatment at 55 DAA and postflowering 63 DAA, with the highest decrease at 63 DAA. One DEM (cyanidin 3-O-rutinoside) was detected in anthocyanin biosynthesis, and the content was reduced in the shading fruit entity. Which suggested that it may be the main component leading to the lightening of the shading fruit color and the decrease of anthocyanin content.

In general, most of the DEGs showed a down-regulated trend in the shading fruit during the entire biosynthesis pathway. The DEGs in the pathways were more abundant the DEMs, and most of the differences occurred at 63 days postflowering. This finding indicated that the shading treatment had the greatest effect on sweet cherry fruits at 63 days after flowering. The upstream pathway of anthocyanin biosynthesis DEGs, such as the *PAL* and *4CL* genes, changed obviously after sunshade treatment. Then, the contents of anthocyanin biosynthesis precursors such as naringenin decreased in the shading fruit during the three periods. We posited that shading may inhibit the expression of *PAL*, *4CL*, *CYP98A*, and other genes, which led to the reduction of the precursors of anthocyanin synthesis and ultimately causes the decrease of anthocyanin accumulation in the light-shading fruits.

3.5 WGCNA analysis of DEGs and anthocyanin components

To further determine the molecular mechanism of fruit color change caused by shading, 4217 DEGs were subjected to WGCNA analysis (Figure 6A). The 4217 DEGs were divided into 11 modules (S4), with the turquoise module containing the most DEGs (2080) and the gray module with the least (35).

Then, we used the contents of two anthocyanin components (cyanidin 3-O-glucoside and cyanidin 3-O-rutinoside) measured in Figure 1D as trait indicators and correlated them with 11 modules. The results were as shown in Figure 6B, in which the contents of the two anthocyanin components were highly positively correlated with the turquoise and pink modules and highly negatively correlated with the blue module. The correlation coefficients of the turquoise, pink, and blue modules with the content of cyanidin 3-O-rutinoside were 0.85 ($p = 1.4 \times 10^{-9-8}$), 0.81 ($p = 1.9 \times 10^{-7}$), and -0.88 ($p = 1.4 \times 10$), respectively. The corresponding correlation coefficients with cyanidin 3-O-glucoside were 0.91 ($p = 4.5 \times 10^{-11}$), 0.87 ($p = 3.3 \times 10^{-6-9}$), and 0.77 ($p = 1.3 \times 10$). Given the high correlation between the three modules and the components of anthocyanin ($r>0.75$), a total of 3036 DEGs (2080 turquoise, 55 pink, and 901 blue) were

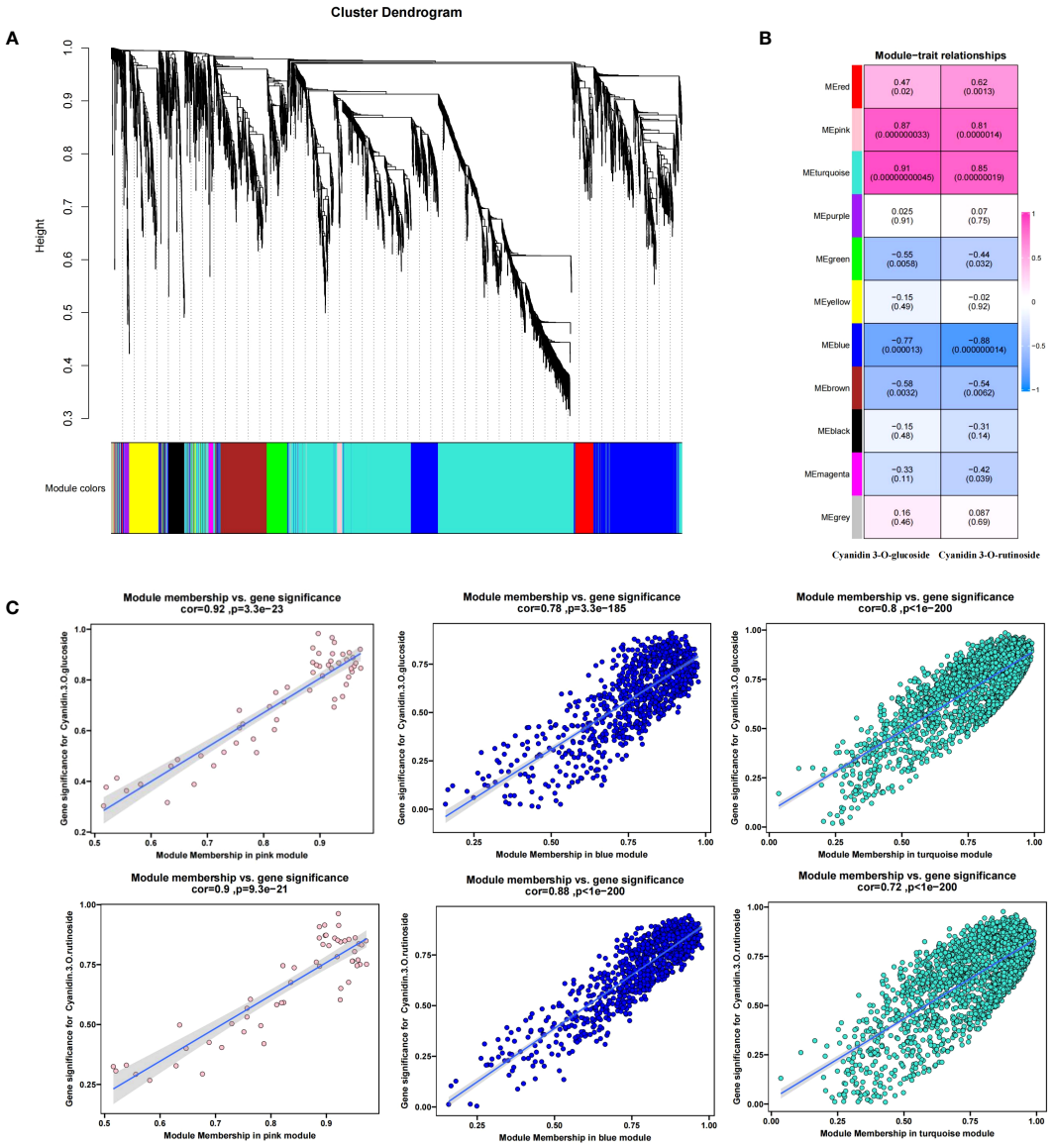


FIGURE 6
Weighted correlation network analysis modules of all DEGs established after the shading treatment. **(A)** Tree graphs of 4217 DEGs by hierarchical clustering of topological overlapping dissimilarities. **(B)** Heat maps of the correlation between modules and sample traits. Different colors represent different modules (11 in total); the number in the grid indicates the Pearson correlation between the module and the trait, with values ranging from -1 (blue) to 1 (red); and the value in parentheses represents the p-value, such that the smaller the value, the more significant the outcome. **(C)** Gene significance (GS) and module membership (MM) relationships between genes and phenotypic traits (anthocyanin components) in the three candidate modules.

identified from the three modules (S4). The relationship between MM and GS between genes and phenotypic traits (anthocyanin components) in each module was shown in Figure 6C (further details are presented in S5). To further narrow down the scope of candidate genes, we set the threshold value MM >0.8 (MM >0.9 in the turquoise module) and GS >0.5 on the basis of the WGCNA. A total of 57 transcription factors such as *bHLH*, *ERF*, *NAC*, *LBD*, and *WRKY*, 5 anthocyanin synthesis structural genes (2 *HCT* and 3 *4CL*) and 4 photoinducable proteins (*RVE1*, *RVE2*, *LHY*, and *FKF1*) were screened, which were chosen as our candidate genes (S6).

3.6 Analysis of candidate gene correlation and promoter cis-acting elements

To more clearly and intuitively show the relationship between candidate genes and anthocyanin components, we further performed correlation analysis on the obtained candidate genes and two anthocyanin components (S7). Further, the correlations >0.8 and <-0.8 were presented using the Cytoscape software (Figure 7A). Among the candidate genes, 4 of 9 genes were in the inner circle (*RVE2*, *LHY*, *HCT1*, and *4CL3*) were clustered in Cluster 4 (Figure 3E), an outcome which was consistent with the change trend of total anthocyanin and component contents. Meanwhile, all candidate genes in the figure were closely correlated with anthocyanin components, thereby indicating the credibility of the WGCNA screening results. The 57 transcription factors were divided into a 24-transcription factor family, among which the *bHLH*, *bZIP*, *ERF*, *G2*-like, *LBD*, *HSF*, *NAC*, and *WRKY* transcription factor family members were more closely related to structural genes and anthocyanin components. For example, *bHLH3*, *ERF4*, *LBD*, *NAC2*, *NAC3*, *HSF4*, *FAR11*, and other transcription factors had a correlation coefficient of more than 0.8 with all structural genes and anthocyanin components except for *4CL1*. In addition, transcription factors strongly correlated with structural gene *4CL1* were negatively correlated, but the *4CL2* and *4CL3* were positively correlated with the two components of anthocyanin and most transcription factors. This difference suggested that *4CL* may be the key structural gene that caused the decrease of anthocyanin accumulation in fruits after shading. In the network diagram, the *LHY* gene related to light response was negatively correlated, and most of the transcription factors related to it were highly positively correlated with the content of the two anthocyanin components. Similarly, the MYB-related proteins *RVE1* and *RVE2* associated with light response were significantly positively correlated with *4CL2*, *4CL3*, *HCT1*, *HCT2*, and the two anthocyanin components, suggesting an important role in the reduction of fruit anthocyanin content by shading.

To elucidate the possible regulatory mechanisms of conserved elements in the promoter region of candidate structural genes and photoinducible proteins, we performed cis-acting element analysis on the upstream 2000 bp of these gene sequences (Figure 7B). Moreover, 22 elements were screened to participate in light responsiveness, hormonal responsiveness, and abiotic stress responsiveness. Among the elements, the cis-acting regulatory element involved in light responsiveness was the most numerous

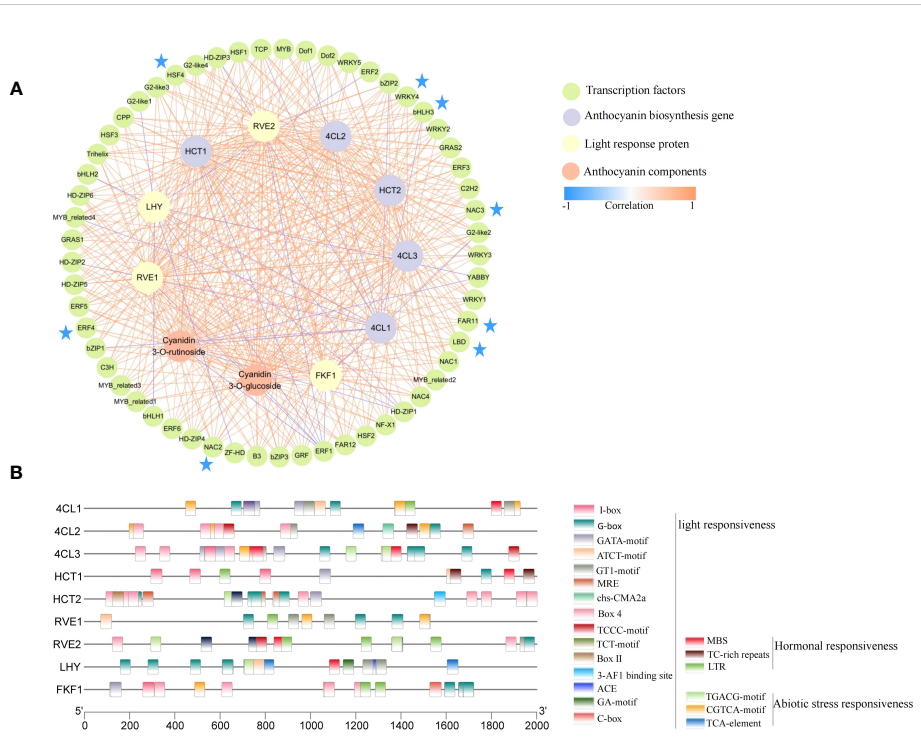


FIGURE 7
Correlation network and cis-acting elements of candidate genes. **(A)** Correlation network diagram constructed by candidate genes and two anthocyanin components. The colors of the lines represent Pearson correlations (correlations between transcription factors are not included in the figure), from blue (-1) to orange (1). **(B)** cis-acting elements of candidate gene promoters.

(15), suggesting that light may play an important role in the regulatory network of these genes.

3.7 Fruit phenotype and candidate gene analysis at 63 DAA after light restoration

To further narrow down the range of candidate genes that affect the reduction of anthocyanin content in shading fruits, we employed light restoration treatment by removing fruit bags at 55 DAA and harvested fruits at 63 DAA (See S2 for relevant transcriptome information and Figure S1 for transcriptome PCA). As can be seen from Figure 8A, the fruit with restored light exhibited deeper redness, but the fruit color did not return under normal light. The total anthocyanin content, the a*/b* value, and the contents of two anthocyanin components all increased, but were below counterparts for fruits under normal light (Figures 8B, C). That is, after shading, the anthocyanin content was between the normal light and shading fruits, and light restoration after shading could not completely restore the anthocyanin accumulation in fruits. Heat maps showed that the expression levels of most transcription factors did not change significantly after light restoration, but the expression levels of transcription factors *LBD*, *bHLH1*, *ERF4*, *FARF11*, *HSF2*, *HFS4*, *NAC2*, *NAC3*, *NF-X1*, and *WRKY4* were significantly down-regulated after shading and up-regulated after light restoration (Figure 8D). These transcription factors were consistent with the content of anthocyanin components and the expression levels of structural genes *4CL2*, *4CL3*, and *HCT2*.

3.8 qRT-PCR verification

Next, we randomly performed qRT-PCR analysis on some of the candidate genes to verify the validity of the transcriptome data Figure 9. Although the FPKM values, qPCR relative quantitative values, and specific difference multiples of the selected genes varied in the four stages of sweet cherry fruit, the expression trends of these genes detected by the two methods were the same at the four stages. According to the Pearson's correlation coefficient (S1), the correlation between the qPCR relative quantitative values and FPKM values ranged from 0.811(*RVE1*) to 0.992(*4CL3*, *EFR4*),

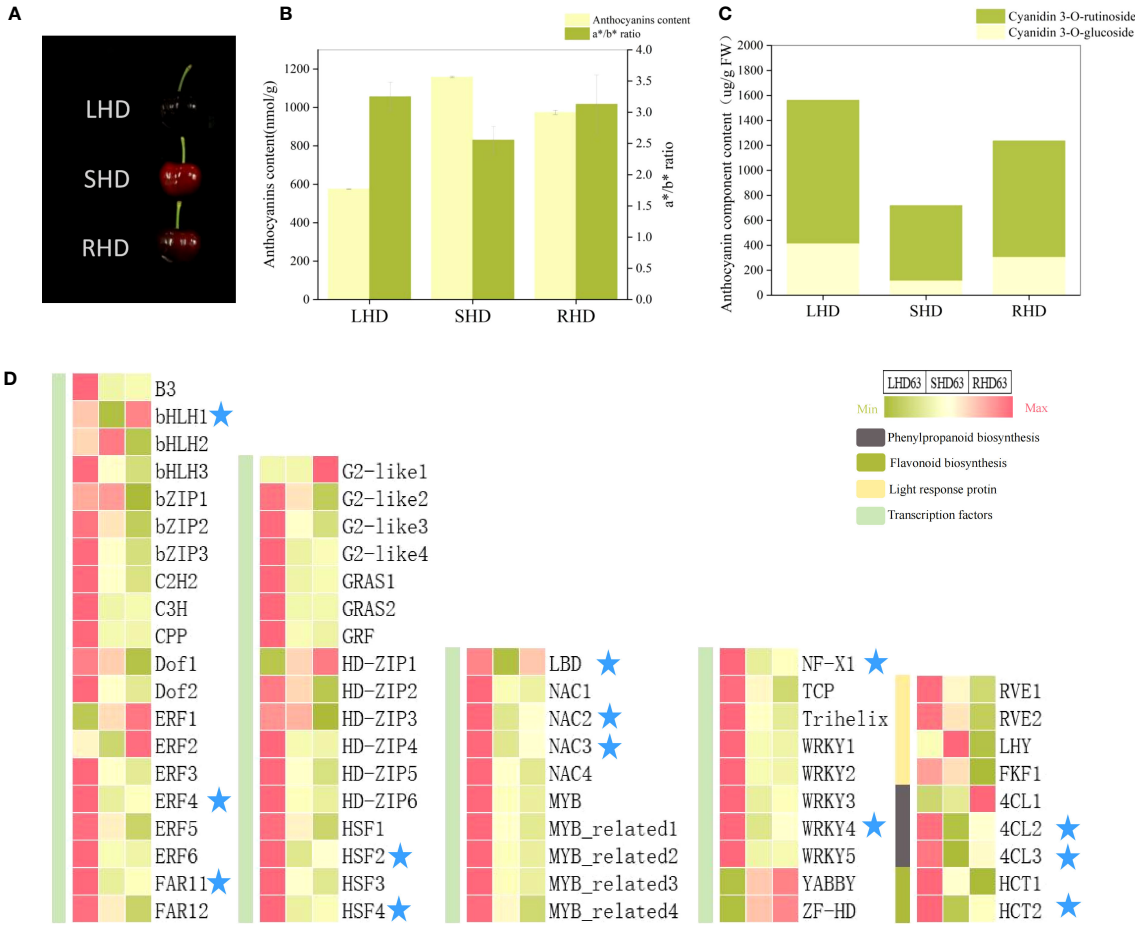


FIGURE 8
Changes of fruit at 63 DAA after restoring light by removing fruit bags **(A)** changes of fruit phenotype; **(B)** Total anthocyanin content and color changes of fruit; **(C)** Content map of fruit anthocyanin components; **(D)** Expression heat maps of candidate structural genes and regulatory factors in three treated fruits at 63 DAA.

10.3389/fpls.2023.1238624

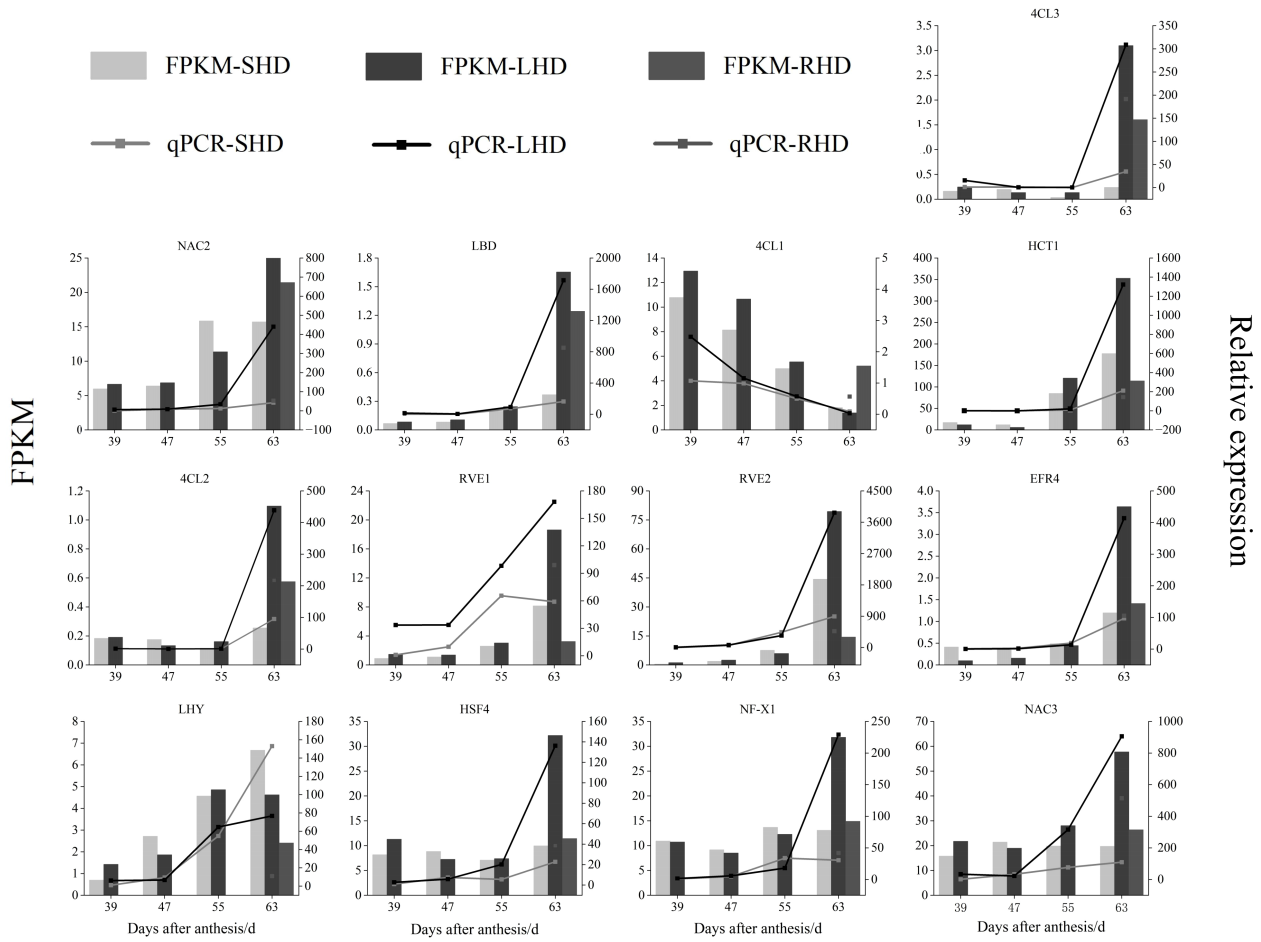


FIGURE 9
qRT-PCR verification of genes and transcription factors related to the content of anthocyanins in sweet cherry after shading treatment.

indicating the reliability of transcriptome data. The gene expression levels of *4CL2* and *4CL3* in sweet cherry fruits after shading were lower than those in the fruits under normal light, and the expression levels were up-regulated after light restoration. Thus, *4CL2* and *4CL3* may be the key structural genes affecting anthocyanin accumulation. We also verified the expression of *LBD*, *NAC2*, *NAC3*, *ERF4*, and other transcription factors, and their expression trends were consistent with the expression levels of structural genes *4CL2* and *4CL3* and the anthocyanin content. qRT-PCR further proved that structural genes (*4CL1*, *4CL2*, *4CL3*, and *HCT1*) and transcription factors related to anthocyanin biosynthesis showed significant differences among the treated groups, indicating that the reduction of anthocyanin content of "Hong Deng" sweet cherry was related to the upstream structural genes *4CL* and *HCT* and a variety of transcription factors in anthocyanin biosynthesis.

4 Discussion

Light is an indispensable environmental factor for plant growth, affecting multiple growth and development processes from seed germination and flowering to fruit coloring. Sweet cherry is a light-loving plant which is widely loved by consumers because of its bright fruit color and juicy taste. The "Hong Deng" sweet cherry is the main red sweet cherry variety in the market at present. In previous studies, considerable evidence indicates that light plays an important role in the coloring of apples, grapes, strawberries, and other fruits (Jiu et al., 2021; Liu et al., 2021; Nguyen et al., 2023). However, until now, few reports are available on the role of light in the coloring of red sweet cherry varieties and its mechanism.

The effect of light on anthocyanin accumulation is achieved mainly through two ways: light quality and light intensity. Currently, the research on the effects of light quality on anthocyanins mainly focus on red light and blue light. In a study applying different light-emitting diodes to strawberries, the anthocyanin content in strawberries increased rapidly after the application of blue light (Zhang et al., 2018b). Tao et al. applied a different light to the Hongzaoyu pear. After 72 h, blue light increased anthocyanin accumulation in pear, but red-light treatment had little effect (Tao et al., 2018). Jian et al. verified that after applying different light intensity to uncolored apples, anthocyanin accumulation increased significantly under medium

and high light intensity, and anthocyanin biosynthesis related genes were also upregulated (An et al., 2020a). Under low light intensity (15 μmol/m^2), Perilla showed a green color, but turned dark red under medium light intensity (180 μmol/m^2) (Xie et al., 2022).

In this study, the "Hong Deng" sweet cherry fruit was subjected to shading treatment, with the following results: the red color of the fruit became lighter, the anthocyanin content was significantly reduced, and the red color of the fruit was partially restored after the restoration of light. Thus, light played an important role in the color of the "Hong Deng" sweet cherry fruit. In addition to affecting the anthocyanin content of the fruit, the shading treatment also led to a reduction in the weight (Figure S2), longitudinal longitude (Figure S3) and transverse longitude (Figure S4) of sweet cherry fruits, which is consistent with the results of the study on grapes (Melino et al., 2011).Then, we analyzed the sweet cherry fruits in the four selected stage treatment groups by transcriptome and metabolomics, subsequently identifying 678 DEMs and 4217 DEGs. KEGG enrichment analysis showed that various intrinsic qualities of sweet cherry fruits were affected by shading treatment. We focused on the most obvious part of the changes in the fruit phenotype—the fruit color, and found that the sugar metabolism pathway, sugar degradation/gluconeogenesis pathway, phenylpropane synthesis pathway, and flavonoid synthesis pathway related to anthocyanin biosynthesis were significantly enriched. Note that the circadian pathway related to the biological clock of higher plants is also involved, which provides a basis for our follow-up study on the role of light in anthocyanin accumulation in sweet cherry fruits. In the analysis of the anthocyanin biosynthesis pathway, the expressions of *PAL*, *4CL*, *CYP98A*, and *HCT* after shading were significantly down-regulated in the later stage of fruit growth and development (63 DAA). As the middle and upstream genes of the anthocyanin biosynthesis pathway, they have an impact on anthocyanin content. Similar to our findings, the study of A. dahurica revealed that the expression of *4CL* was significantly down-regulated at 90% shade rate (Huang et al., 2022). Likewise, the expression of *4CL* in broccoli was down-regulated after shading treatment (Liu et al., 2020). The transcription level of *PAL* gene in Scutellaria baicalensis was down-regulated under dark conditions (Chen et al., 2010). The expression of *PAL* and *CHS* was also down-regulated with the increase of light intensity in tea plants (Ye et al., 2021). According to Wang et al., light intensity significantly up-regulated the relative expression levels of *CHS*, *FLS*, and *PAL* in bitter malt (Wang et al., 2022a).

WGCNA further narrowed the DEGs range. By setting the MM and GS thresholds, we obtained 66 candidate genes and visualized their correlations. Among the candidate genes, five structural genes (3 *4CL* and 2 *HCT*) were identified as candidate structural genes affecting anthocyanin content in fruit after shading. *4CL* plays an important role in anthocyanin accumulation in fruits. As the upstream gene, different *4CL* isoenzymes can selectively catalyze cinnamic acid, p-coumanyic acid, and other substances to produce the corresponding CoA thioester and then generate anthocyanin through a series of chemical reactions (Wang et al., 2022b). In the study of purple tea, *4CL* was significantly upregulated in the S2_GP (dark grayish purple) stage compared with the S1_RP (reddish purple) and S3_G (medium olive green) stages (Maritim et al., 2021). Similarly, *4CL* expression levels were significantly upregulated in colored Tibetan hulless barley compared to colorless varieties (Xu et al., 2022; Yao et al., 2022). Schulz Dietmar F et al. identified a 4CLSNP marker through GWAS for anthocyanin content and identified it as a candidate gene for influencing anthocyanin concentration (Schulz et al., 2016). *4CL* gene family members are divided into two groups, Types I and II. Type I members are involved in lignin synthesis, and Type II members mostly regulate flavonoid biosynthesis (Wang et al., 2022b). Given the different expression patterns of the three *4Cls* in this study, we speculate that *4CL1* is Type I and *4CL2* and *4CL3* are Type II members. The decrease of anthocyanin content in fruit after shading is more likely related to the down-regulation of *4CL2* and *4CL3* expression levels.

HCT is the midstream gene of anthocyanin biosynthesis. The key step of anthocyanin synthesis in potato is the formation of p-coumaryl-CoA, and the down-regulated expression of *HCT* significantly promotes this process (Tengkun et al., 2019). In the study of Cong et al., *HCT* was present in violet-red skin pinellia but was not detected in the pale yellow skin counterpart (Yin et al., 2023). Similarly, *HCT* expression levels in purple testa peanuts were significantly higher than that in pink counterparts. All these evidences indicate that *HCT* also plays an important role in anthocyanin biosynthesis (Li et al., 2022a).

In addition, we identified four light-responsive proteins (*FKF1*, *LHY*, *RVE1*, and *RVE2*) involved in plant circadian pathways and in the regulation of plant photosynthesis and flowering. Note that few investigations have been conducted on plant anthocyanin synthesis (Joo et al., 2017; Yan et al., 2020; Liu et al., 2023). The only existing studies showed that after overexpression of *RVE8* in Arabidopsis thaliana, the structural genes of the anthocyanin pathway such as *4CL* were up-regulated during the day, but this up-regulation was lost at night, suggesting that *RVE8* may play a specific role in the influence of light on anthocyanin biosynthesis (Pérez-García et al., 2015). In the present study, *RVE1* and *RVE2* were highly positively correlated with *4CL2* and *4CL3*, a finding which is consistent with the results of Pérez-García et al. (Pérez-García et al., 2015). *FKF1*, as a photoperiodic blue light receptor (Imaizumi et al., 2003), has recently been linked to cellulose biosynthesis as a negative regulator (Yuan et al., 2019). However, its role in anthocyanin synthesis has not been reported. In this research, *FKF1* was positively correlated with two anthocyanin components, suggesting that *FKF1* may play an important role in the effect of shading on the sweet cherry fruit color, but its mechanism remains unexplored. This work also suggests that *LHY* seems to play a negative regulatory role, a significant negative correlation with *4CL1*, and a weak correlation with other structural genes and anthocyanin components. Although few reports are available on the anthocyanin synthesis pathway of the *LHY* protein, other related pathways such as glucose metabolism and transport, starch degradation, tricarboxylic acid cycle, and phenylalanine metabolism have been examined (Lu et al., 2005; Nakamichi et al., 2009; Graf et al., 2010; Fenske et al., 2015).

Therefore, we hypothesize that *LHY* and its related regulators may play a role in another regulatory network, which may affect anthocyanin accumulation by influencing the precursor pathways associated with anthocyanin biosynthesis.

Anthocyanin accumulation in fruits is regulated by a variety of transcription factors. In this study, we obtained 57 potential regulatory transcription factors. Among them, *LBD*, *bHLH*, *bZIP*, *ERF*, *G2-like*, *NAC*, *HSF*, *WRKY*, *NF-X1*, *FAR1*, and other transcription factor families are closely related to the anthocyanin biosynthesis structural genes *4CL2*, *4CL3*, *HCT1*, and *HCT2* and two anthocyanin components. *LBD* is a plant-specific transcription factor which mainly affects the development of lateral vegetative organs such as leaves and roots (Liang et al., 2022). In addition, *LBD* plays an important role in plant resistance to stress and disease (Feng et al., 2022; Jiao et al., 2022). Regarding the role of *LBD* in anthocyanin biosynthesis, *LBD* in Arabidopsis is reported to regulate anthocyanin biosynthesis by inhibiting *PAP1* and *PAP2* (Rubin et al., 2009). Rui-Min et al. showed that Arabidopsis plants post *CsLBD39* overexpression were affected by nitrogen stress, and the contents of nitrate and total anthocyanins were significantly reduced compared with those of the wild type (Teng et al., 2022). The ectopic expression of *MdLBD13* gene in apple inhibited anthocyanin biosynthesis in Arabidopsis by down-regulating the expression of *AtPAP1*, *AtCHS*, *AtCHI*, *AtDFR1*, and *AtUFGT* (Li et al., 2017). Moreover, overexpression of *MdLBD13* also inhibited anthocyanin accumulation in the apple callus (Li et al., 2017).

bHLH transcription factors usually affect anthocyanin accumulation by interacting with *MYB* transcription factors, such as in onions, where *AcB2*, which belongs to the *bHLH* IIIf subfamily, interacts with *AcMYB1* to promote transcriptional activation of the anthocyanin biosynthesis structural genes *AcANS* and *AcF3H1* and ultimately increases anthocyanin accumulation (Li et al., 2022c). *MdbZIP44*, a member of the *bZIP* transcription factor family, was identified as a positive regulator of ABA-promoted anthocyanin accumulation (An et al., 2018). The *ERF* family is also involved in transcriptional regulation of anthocyanin biosynthesis. In the study of An et al., *MDERF38* participated in drought stress-induced anthocyanin synthesis by interacting with *MDMYB1* as a positive regulator (An et al., 2020b). *GFR*, a member of the *G2-like* family, further inhibits anthocyanin synthesis genes by inhibiting *bHLH* and *MYB*, thereby negatively regulating anthocyanin accumulation (Petridis et al., 2016). Apple *NAC* family member *MdNAC52* promotes anthocyanin accumulation by binding to promoters of *MdMYB9* and *MdMYB11* (Sun et al., 2019). Similarly, *MdWRKY11*, a member of the *WRKY* family of red flesh apples, is involved in anthocyanin accumulation by influencing the *MYB* transcription factor and the photoresponse factor *MdHY5* (Liu et al., 2019b). FAR1 transcription factor may also be involved in anthocyanin regulation in citrus (Jin et al., 2023). In conclusion, these transcription factors regulate the accumulation of anthocyanins in plants by directly or indirectly affecting the structural genes of anthocyanin biosynthesis.

To analyze the possible mechanism of action between transcription factors and structural genes, we conducted promoter cis-acting elements analysis for the first 2000 bp of the structural genes and photoinducable protein sequences screened in this study. We found that all these genes contained light response elements. Among them, G-box, a motif that has been shown to be necessary and sufficient to induce light-responsive transcription of certain genes (Giuliano et al., 1988), has played an important role in the regulation of photoinduced anthocyanin biosynthesis in pear (Liu et al., 2019a). G-box has been found to exist in candidate structural genes, and the number of G-box motifs is highest in *4CL3* and *HCT2*. In previous studies, promoter sequences of the potato *4CL* gene family members all contain photoresponsive elements. UV stress can promote the up-regulation of *St4CL6* and *St4CL8* but inhibit the expression of *St4CL5* (Zhong et al., 2022). Studies in tea tree also showed that the promoters of two *4CL* genes, *Cs4CL1* and *Cs4CL2*, contained 9 kinds of light response elements, and *Cs4CL2* was induced by UV-B (Li et al., 2022b). In parsley, both *Pc4CL*-1 and *Pc4CL*-2 were transcriptionally activated by UV irradiation (Douglas et al., 1987). This finding presents a novel idea. Thus, we posit that the shading treatment in the present study may affect the binding of regulatory proteins to the G-box motif in the *4CL* and *HCT* promoters and then affect the transcription of these structural genes, ultimately influencing anthocyanin accumulation.

To further narrow down the range of candidate genes that affect the decrease of anthocyanin content in fruits after shading, we removed the fruit bags at 55 DAA to restore light and harvested the fruits at 63 DAA. The transcriptome data showed that the expression levels of several transcription factors and structural genes discussed above were changed to varying degrees after light restoration. Specifically, *LBD*, *NF-X1*, *FAR1*, *ERF4*, *NAC2*, *NAC3*, *HSF4*, and *WRKY4* showed the same change trend as *4CL2*, *4CL3*, and *HCT2*, and their expressions were up-regulated after light restoration (*LBD* was the most strongly up-regulated), which further confirmed the possible positive regulatory role of these genes in anthocyanin biosynthesis. However, the expression levels of *FKF1*, *RVE1*, and *RVE2* proteins decreased at 63 DAA after light restoration. If these light-responsive proteins act as positive regulators in the light-induced anthocyanin biosynthesis of red sweet cherry, we speculate that the regulatory effects of these proteins on anthocyanin biosynthesis may be time-sensitive. In other words, the expression levels of *FKF1*, *RVE1*, and *RVE2* may be up-regulated within a short time after bag-removal and affect the expression levels of related transcription factors (*LBD*, *ERF4*, *NAC2*, *NAC3*, etc.) and structural genes (*4CL* and *HCT*), resulting in the fruit color darkening after light restoration.

In this study, the potential regulatory genes for photoinduced anthocyanin biosynthesis of sweet cherry were screened by combining transcriptome and metabolome analysis. The results showed that light shading had a negative effect on anthocyanin accumulation and changed the expression levels of several related genes in red light sweet cherry. *4CL2*, *4CL3*, and *HCT2* may be the key structural genes affecting anthocyanin content, and their expression levels may be regulated by transcription factors such as *LBD*, *NF-X1*, *FAR1*, *ERF4*, *NAC2*, *NAC3*, *HSF4*, *WRKY4*, *FKF1*, *LHY*, *RVE1*, and *RVE2*. However, the regulatory network of plant growth is intricate, and its specific regulatory modes still need further investigation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ngdc.cncb.ac.cn/gsa/browse/CRA011359, CRA011359.

Author contributions

YZ and RG: designed the whole experiments. YZ and CC: Data determination. YZ, CC, YC, WeT, GK, WaT, HC: analyzed the data. YZ:wrote the manuscript. RG: supervision. All authors read and approved the final manuscript.

Funding

This research was funded by the Sichuan Science and Technology Plan Project (2021YFN0082, 2021ZHIFP0086). 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.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1238624/full#supplementary-material

References

Alcalde-Eon, C., Rivas-Gonzalo, J. C., Muñoz, O., and Escribano-Bailón, M. T. (2013). Schizanthus grahamii and Schizanthus hookeri. Is there any relationship between their anthocyanin compositions and their different pollination syndromes? *Phytochemistry* 85, 62–71. doi: 10.1016/j.phytochem.2012.09.013

An, J. P., Liu, Y. J., Zhang, X. W., Bi, S. Q., Wang, X. F., You, C. X., et al. (2020a). Dynamic regulation of anthocyanin biosynthesis at different light intensities by the BT2-TCP46-MYB1 module in apple. *J. Exp. Bot.* 71 (10), 3094–3109. doi: 10.1093/jxb/eraa056

An, J. P., Yao, J. F., Xu, R. R., You, C. X., Wang, X. F., and Hao, Y. J. (2018). Apple bZIP transcription factor MdbZIP44 regulates abscisic acid-promoted anthocyanin accumulation. *Plant Cell Environ.* 41 (11), 2678–2692. doi: 10.1111/pce.13393

An, J. P., Zhang, X. W., Bi, S. Q., You, C. X., Wang, X. F., and Hao, Y. J. (2020b). The ERF transcription factor MdERF38 promotes drought stress-induced anthocyanin biosynthesis in apple. *Plant J.* 101 (3), 573–589. doi: 10.1111/tpj.14555

Bai, S., Tao, R., Tang, Y., Yin, L., Ma, Y., Ni, J., et al. (2019). BBX16, a B-box protein, positively regulates light-induced anthocyanin accumulation by activating MYB10 in red pear. *Plant Biotechnol. J.* 17 (10), 1985–1997. doi: 10.1111/pbi.13114

Ceccarelli, D., Antonucci, F., Costa, C., Talento, C., and Ciccoritti, R. (2020). An artificial class modelling approach to identify the most largely diffused cultivars of sweet cherry (Prunus avium L.) in Italy. *Food Chem.* 333, 127515. doi: 10.1016/j.foodchem.2020.127515

Chen, C., Chen, H., Yang, W., Li, J., Tang, W., and Gong, R. (2022). Transcriptomic and metabolomic analysis of quality changes during sweet cherry fruit development and mining of related genes. *Int. J. Mol. Sci.* 23 (13), 7402. doi: 10.3390/ijms23137402

Chen, S., Yuan, Y., Luo, Y., Huang, L., Chen, P., and Li, X. (2010). Effects of light on flavonoids accumulation and related gene expression in suspension cultures of Scutellaria baicalensis. *China J. Chin. Mater. Med.* 35 (6), 682–685. doi: 10.4268/cmm20100604

Chen, W., Zhang, M., Zhang, G., Li, P., and Ma, F. (2019). Differential regulation of anthocyanin synthesis in apple peel under different sunlight intensities. *Int. J. Mol. Sci.* 20 (23), 6060. doi: 10.3390/ijms20236060

Davis, P. H. (1975). *Flora of Turkey and the East Aegean Islands* Vol. 4 (Edinburgh: Edinburgh University Press).

Douglas, C., Hoffmann, H., Schulz, W., and Hahlbrock, K. (1987). Structure and elicitor or uv-light-stimulated expression of two 4-coumarate: CoA ligase genes in parsley. *EMBO J.* 6 (5), 1189–1195. doi: 10.1002/j.1460-2075.1987.tb02353.x

Feng, S., Shi, J., Hu, Y., Li, D., Guo, L., Zhao, Z., et al. (2022). Genome-wide analysis of soybean lateral organ boundaries domain gene family reveals the role in phytophthora root and stem rot. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.865165

Fenske, M. P., Hewett Hazelton, K. D., Hempton, A. K., Shim, J. S., Yamamoto, B. M., Riffell, J. A., et al. (2015). Circadian clock gene LATE ELONGATED HYPOCOTYL directly regulates the timing of floral scent emission in Petunia. *Proc. Natl. Acad. Sci.* 112 (31), 9775–9780. doi: 10.1073/pnas.1422875112

Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A., and Cashmore, A. R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci.* 85 (19), 7089–7093. doi: 10.1073/pnas.85.19.7089

Gonçalves, A. C., Nunes, A. R., Falcão, A., Alves, G., and Silva, L. R. (2021). Dietary effects of anthocyanins in human health: A comprehensive review. *Pharmaceuticals* 14 (7), 690. doi: 10.3390/ph14070690

Graf, A., Schlereth, A., Stitt, M., and Smith, A. M. (2010). Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proc. Natl. Acad. Sci.* 107 (20), 9458–9463. doi: 10.1073/pnas.0914299107

Huang, Y., Zhai, Y., Huang, Y., Huang, Y., Liu, K., Zhang, J., et al. (2022). Effects of light intensity on physiological characteristics and expression of genes in coumarin biosynthetic pathway of Angelica dahurica. *Int. J. Mol. Sci.* 23 (24), 15912. doi: 10.3390/ijms232415912

Imaizumi, T., Tran, H. G., Swartz, T. E., Briggs, W. R., and Kay, S. A. (2003). FKF1 is essential for photoperiodic-specific light signaling in Arabidopsis. *Nature* 426 (6964), 302–306. doi: 10.1038/nature02090

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18 (9), 477–483. doi: 10.1016/j.tplants.2013.06.003

Jiao, P., Wei, X., Jiang, Z., Liu, S., Guan, S., and Ma, Y. (2022). ZmLBD2 a maize (Zea mays L.) lateral organ boundaries domain (LBD) transcription factor enhances drought tolerance in transgenic Arabidopsis thaliana. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1000149

Jin, Y., Liao, M., Li, N., Ma, X., Zhang, H., Han, J., et al. (2023). Weighted gene coexpression correlation network analysis reveals the potential molecular regulatory mechanism of citrate and anthocyanin accumulation between postharvest 'Bingtangcheng'and 'Tarocco'blood orange fruit. *BMC Plant Biol.* 23 (1), 296. doi: 10.1186/s12870-023-04309-5

Jiu, S., Guan, L., Leng, X., Zhang, K., Haider, M. S., Yu, X., et al. (2021). The role of VvMYBA2r and VvMYBA2w alleles of the MYBA2 locus in the regulation of anthocyanin biosynthesis for molecular breeding of grape (Vitis spp.) skin coloration. *Plant Biotechnol. J.* 19 (6), 1216–1239. doi: 10.1111/pbi.13543

Joo, Y., Fragoso, V., Yon, F., Baldwin, I. T., and Kim, S. G. (2017). Circadian clock component, LHY, tells a plant when to respond photosynthetically to light in nature. *J. Integr. Plant Biol.* 59 (8), 572–587. doi: 10.1111/jipb.12547

Kim, D., Jeon, S. J., Yanders, S., Park, S. C., Kim, H. S., and Kim, S. (2022). MYB3 plays an important role in lignin and anthocyanin biosynthesis under salt stress condition in Arabidopsis. *Plant Cell Rep.* 41 (7), 1549–1560. doi: 10.1007/s00299-022-02878-7

Leong, S. J., Lu, W. C., and Chiou, T. J. (2018). Phosphite-mediated suppression of anthocyanin accumulation regulated by mitochondrial ATP synthesis and sugars in Arabidopsis. *Plant Cell Physiol.* 59 (6), 1158–1169. doi: 10.1093/pcp/pcy051

Li, X., Cao, L., Jiao, B., Yang, H., Ma, C., and Liang, Y. (2022c). The bHLH transcription factor AcB2 regulates anthocyanin biosynthesis in onion (Allium cepa L.). *Horticult. Res.* 9, uhac128. doi: 10.1093/hr/uhac128

Li, M., Guo, L., Wang, Y., Li, Y., Jiang, X., Liu, Y., et al. (2022b). Molecular and biochemical characterization of two 4-coumarate: CoA ligase genes in tea plant (Camellia sinensis). *Plant Mol. Biol.* 109 (4-5), 579–593. doi: 10.1007/s11103-022-01269-6

Li, H. H., Liu, X., An, J. P., Hao, Y. J., Wang, X. F., and You, C. X. (2017). Cloning and elucidation of the functional role of apple MdLBD13 in anthocyanin biosynthesis and nitrate assimilation. *Plant Cell Tissue Organ Cult. (PCTOC)* 130, 47–59. doi: 10.1007/s11240-017-1203-x

Li, J., Ma, Y., Hu, M., Zhao, Y., Liu, B., Wang, C., et al. (2022a). Multi-omics and miRNA interaction joint analysis highlight new insights into anthocyanin biosynthesis in peanuts (Arachis hypogaea L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.818345

Li, X., Sun, Y., Wang, X., Dong, X., Zhang, T., Yang, Y., et al. (2019). Relationship between key environmental factors and profiling of volatile compounds during cucumber fruit development under protected cultivation. *Food Chem.* 290, 308–315. doi: 10.1016/j.foodchem.2019.03.140

Liang, J., Hou, Z., Liao, J., Qin, Y., Wang, L., Wang, X., et al. (2022). Genome-wide identification and expression analysis of LBD transcription factor genes in Passion Fruit (Passiflora edulis). *Int. J. Mol. Sci.* 23 (9), 4700. doi: 10.3390/ijms23094700

Liu, H., Liu, Z., Wu, Y., Zheng, L., and Zhang, G. (2021). Regulatory mechanisms of anthocyanin biosynthesis in apple and pear. *Int. J. Mol. Sci.* 22 (16), 8441. doi: 10.3390/ijms22168441

Liu, H., Su, J., Zhu, Y., Yao, G., Allan, A. C., Ampomah-Dwamena, C., et al. (2019a). The involvement of PybZIPa in light-induced anthocyanin accumulation via the activation of PyUFGT through binding to tandem G-boxes in its promoter. *Horticult. Res.* 6, 134. doi: 10.1038/s41438-019-0217-4

Liu, W., Wang, Y., Yu, L., Jiang, H., Guo, Z., Xu, H., et al. (2019b). MdWRKY11 participates in anthocyanin accumulation in red-fleshed apples by affecting MYB transcription factors and the photoresponse factor MdHY5. *J. Agric. Food Chem.* 67 (32), 8783–8793. doi: 10.1021/acs.jafc.9b02920

Liu, C., Yao, X., Li, G., Huang, L., and Xie, Z. (2020). Transcriptomic profiling of purple broccoli reveals light-induced anthocyanin biosynthetic signaling and structural genes. *PeerJ* 8, e8870. doi: 10.7717/peerj.8870

Liu, Z., Zhu, X., Liu, W., Qi, K., Xie, Z., Zhang, S., et al. (2023). Characterization of the REVEILLE family in Rosaceae and role of PbLHY in flowering time regulation. *BMC Genomics* 24 (1), 1–16. doi: 10.1186/s12864-023-09144-4

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2– ΔΔCT method. *methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Lu, Y., Gehan, J. P., and Sharkey, T. D. (2005). Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiol.* 138 (4), 2280–2291. doi: 10.1104/pp.105.061903

Ma, D., and Constabel, C. P. (2019). MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* 24 (3), 275–289. doi: 10.1016/j.tplants.2018.12.003

Maritim, T. K., Masand, M., Seth, R., and Sharma, R. K. (2021). Transcriptional analysis reveals key insights into seasonal induced anthocyanin degradation and leaf color transition in purple tea (Camellia sinensis (L.) O. Kuntze). *Sci. Rep.* 11 (1), 1244. doi: 10.1038/s41598-020-80437-4

Melino, V. J., Hayes, M. A., Soole, K. L., and Ford, C. M. (2011). The role of light in the regulation of ascorbate metabolism during berry development in the cultivated grapevine Vitis vinifera L. *J. Sci. Food Agric.* 91 (9), 1712–1721. doi: 10.1002/jsfa.4376

Nakamichi, N., Fukushima, A., Kusano, M., Sakakibara, H., Mizuno, T., and Saito, K. (2009). Linkage between circadian clock and tricarboxylic acid cycle in Arabidopsis. *Plant Signaling Behav.* 4 (7), 660–662. doi: 10.1073/pnas.0900952106

Nguyen, H. M., Putterill, J., Dare, A. P., Plunkett, B. J., Cooney, J., Peng, Y., et al. (2023). Two genes, ANS and UFGT2, from Vaccinium spp. are key steps for modulating anthocyanin production. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1082246

Passeri, V., Koes, R., and Quattrocchio, F. M. (2016). New challenges for the design of high value plant products: stabilization of anthocyanins in plant vacuoles. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00153

Pérez-García, P., Ma, Y., Yanovsky, M. J., and Mas, P. (2015). Time-dependent sequestration of RVE8 by LNK proteins shapes the diurnal oscillation of anthocyanin biosynthesis. *Proc. Natl. Acad. Sci.* 112 (16), 5249–5253. doi: 10.1073/pnas.1420792112

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Petridis, A., Döll, S., Nichelmann, L., Bilger, W., and Mock, H. P. (2016). Arabidopsis thaliana G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation. *New Phytol.* 211 (3), 912–925. doi: 10.1111/nph.13986

Rao, M. J., Duan, M., Yang, M., Fan, H., Shen, S., Hu, L., et al. (2021). Novel insights into anthocyanin metabolism and molecular characterization of associated genes in sugarcane rinds using the metabolome and transcriptome. *Int. J. Mol. Sci.* 23 (1), 338. doi: 10.3390/ijms23010338

Rubin, G., Tohge, T., Matsuda, F., Saito, K., and Scheible, W. R. (2009). Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis. *Plant Cell* 21 (11), 3567–3584. doi: 10.1105/tpc.109.067041

Schulz, D. F., Schott, R. T., Voorrips, R. E., Smulders, M. J., Linde, M., and Debener, T. (2016). Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01798

Song, Z. Y., Chao, F., Zhuo, Z., Ma, Z., Li, W., and Chen, G. (2019b). Identification of hub genes in prostate cancer using robust rank aggregation and weighted gene co-expression network analysis. *Aging (Albany NY)* 11 (13), 4736. doi: 10.18632/aging.102087

Song, T., Li, K., Wu, T., Wang, Y., Zhang, X., Xu, X., et al. (2019a). Identification of new regulators through transcriptome analysis that regulate anthocyanin biosynthesis in apple leaves at low temperatures. *PloS One* 14 (1), e0210672. doi: 10.1371/journal.pone.0210672

Sun, Q., Jiang, S., Zhang, T., Xu, H., Fang, H., Zhang, J., et al. (2019). Apple NAC transcription factor MdNAC52 regulates biosynthesis of anthocyanin and proanthocyanidin through MdMYB9 and MdMYB11. *Plant Sci.* 289, 110286. doi: 10.1016/j.plantsci.2019.110286

Tao, R., Bai, S., Ni, J., Yang, Q., Zhao, Y., and Teng, Y. (2018). The blue light signal transduction pathway is involved in anthocyanin accumulation in 'Red Zaosu'pear. *Planta* 248, 37–48. doi: 10.1007/s00425-018-2877-y

Teng, R. M., Yang, N., Li, J. W., Liu, C. F., Chen, Y., Li, T., et al. (2022). Isolation and characterization of an LBD transcription factor CsLBD39 from tea plant (Camellia sinensis) and its roles in modulating nitrate content by regulating nitrate-metabolism-related genes. *Int. J. Mol. Sci.* 23 (16), 9294. doi: 10.3390/ijms23169294

Tengkun, N., Dongdong, W., Xiaohui, M., Yue, C., and Qin, C. (2019). Analysis of key genes involved in potato anthocyanin biosynthesis based on genomics and transcriptomics data. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00603

Wang, Y., Guo, L., Zhao, Y., Zhao, X., and Yuan, Z. (2022b). Systematic analysis and expression profiles of the 4-coumarate: CoA ligase (4CL) gene family in pomegranate (punica granatum L.). *Int. J. Mol. Sci.* 23 (7), 3509. doi: 10.3390/ijms23073509

Wang, D., Yang, T., Li, Y., Deng, F., Dong, S., Li, W., et al. (2022a). Light intensity—A key factor affecting flavonoid content and expression of key enzyme genes of flavonoid synthesis in tartary buckwheat. *Plants* 11 (16), 2165. doi: 10.3390/plants11162165

Wei, H., Chen, X., Zong, X., Shu, H., Gao, D., and Liu, Q. (2015). Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (Prunus avium L.). *PloS One* 10 (3), e0121164. doi: 10.1371/journal.pone.0121164

Xie, G., Zou, X., Liang, Z., Wu, D., He, J., Xie, K., et al. (2022). Integrated metabolomic and transcriptomic analyses reveal molecular response of anthocyanins biosynthesis in perilla to light intensity. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.976449

Xu, C., Abbas, H. M. K., Zhan, C., Huang, Y., Huang, S., Yang, H., et al. (2022). Integrative metabolomic and transcriptomic analyses reveal the mechanisms of Tibetan hulless barley grain coloration. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1038625

Yan, J., Li, X., Zeng, B., Zhong, M., Yang, J., Yang, P., et al. (2020). FKF1 F-box protein promotes flowering in part by negatively regulating DELLA protein stability under long-day photoperiod in Arabidopsis. *J. Integr. Plant Biol.* 62 (11), 1717–1740. doi: 10.1111/jipb.1297

Yang, T., Ma, H., Li, Y., Zhang, Y., Zhang, J., Wu, T., et al. (2021). Apple MPK4 mediates phosphorylation of MYB1 to enhance light-induced anthocyanin accumulation. *Plant J.* 106 (6), 1728–1745. doi: 10.1111/tpj.15267

Yang, T., Ma, H., Zhang, J., Wu, T., Song, T., Tian, J., et al. (2019). Systematic identification of long noncoding RNA s expressed during light-induced anthocyanin accumulation in apple fruit. *Plant J.* 100 (3), 572–590. doi: 10.1111/tpj.14470

Yao, X., Yao, Y., An, L., Li, X., Bai, Y., Cui, Y., et al. (2022). Accumulation and regulation of anthocyanins in white and purple Tibetan Hulless Barley (Hordeum vulgare L. var. nudum Hook. f.) revealed by combined *de novo* transcriptomics and metabolomics. *BMC Plant Biol.* 22 (1), 1–18. doi: 10.1186/s12870-022-03699-2

Ye, J. H., Lv, Y. Q., Liu, S. R., Jin, J., Wang, Y. F., Wei, C. L., et al. (2021). Effects of light intensity and spectral composition on the transcriptome profiles of leaves in shade grown tea plants (Camellia sinensis L.) and regulatory network of flavonoid biosynthesis. *Molecules* 26 (19), 5836. doi: 10.3390/molecules26195836

Yin, C., Tang, D., Liu, X., Li, Z., Xiang, Y., Gao, K., et al. (2023). Transcriptome analysis reveals important regulatory genes and pathways for tuber color variation in Pinellia ternata (Thunb.) Breit. *Protoplasma* 260 (5), 1–13. doi: 10.1007/s00709-023-01845-7

Yuan, N., Balasubramanian, V. K., Chopra, R., and Mendu, V. (2019). The photoperiodic flowering time regulator FKF1 negatively regulates cellulose biosynthesis. *Plant Physiol.* 180 (4), 2240–2253. doi: 10.1104/pp.19.00013

Yue, M., Jiang, L., Zhang, N., Zhang, L., Liu, Y., Wang, Y., et al. (2022). Importance of faWRKY71 in strawberry (Fragaria× ananassa) fruit ripening. *Int. J. Mol. Sci.* 23 (20), 12483. doi: 10.3390/ijms232012483

Zhang, Y., Jiang, L., Li, Y., Chen, Q., Ye, Y., Zhang, Y., et al. (2018b). Effect of red and blue light on anthocyanin accumulation and differential gene expression in strawberry (Fragaria× ananassa). *Molecules* 23 (4), 820. doi: 10.3390/molecules23040820

Zhang, J., Xu, H., Wang, N., Jiang, S., Fang, H., Zhang, Z., et al. (2018a). The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. *Plant Mol. Biol.* 98, 205–218. doi: 10.1007/s11103-018-0770-5

Zhang, Q., Zhai, J., Shao, L., Lin, W., and Peng, C. (2019). Accumulation of anthocyanins: an adaptation strategy of Mikania micrantha to low temperature in winter. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01049

Zhang, B., Zhu, Z. Z., Qu, D., Wang, B. C., Hao, N. N., Yang, Y. Z., et al. (2021). MdBBX21, a B-box protein, positively regulates light-induced anthocyanin accumulation in apple peel. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.774446

Zhao, L., Sun, J., Cai, Y., Yang, Q., Zhang, Y., Ogutu, C. O., et al. (2022a). PpHYH is responsible for light-induced anthocyanin accumulation in fruit peel of Prunus persica. *Tree Physiol.* 42 (8), 1662–1677. doi: 10.1093/treephys/tpac025

Zhao, X., Zhang, Y., Long, T., Wang, S., and Yang, J. (2022b). Regulation mechanism of plant pigments biosynthesis: anthocyanins, carotenoids, and betalains. *Metabolites* 12 (9), 871. doi: 10.3390/metabo12090871

Zhong, J., Qing, J., Wang, Q., Liu, C., Du, H., Liu, P., et al. (2022). Genome-wide identification and expression analyses of the 4-Coumarate: CoA Ligase (4CL) Gene family in eucommia ulmoides. *Forests* 13 (8), 1253. doi: 10.3390/ijms24021642

Zhou, X., Liu, S., Yang, Y., Liu, J., and Zhuang, Y. (2022). Integrated metabolome and transcriptome analysis reveals a regulatory network of fruit peel pigmentation in eggplant (Solanum melongena L.). *Int. J. Mol. Sci.* 23 (21), 13475. doi: 10.3390/ijms232113475

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 August 2023
DOI 10.3389/fpls.2023.1269286

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China

REVIEWED BY
Kuiju Niu,
Gansu Agricultural University, China
Hao Sun,
Henan Agricultural University, China

*CORRESPONDENCE
Kun Zhang
zk61603@163.com

†These authors have contributed equally to this work

RECEIVED 29 July 2023
ACCEPTED 18 August 2023
PUBLISHED 31 August 2023

CITATION
Zuo Z-F, Li Y, Mi X-F, Li Y-L, Zhai C-Y, Yang G-F, Wang Z-Y and Zhang K (2023) Physiological and lipidomic response of exogenous choline chloride alleviating salt stress injury in Kentucky bluegrass (*Poa pratensis*). *Front. Plant Sci.* 14:1269286. doi: 10.3389/fpls.2023.1269286

COPYRIGHT
© 2023 Zuo, Li, Mi, Li, Zhai, Yang, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Physiological and lipidomic response of exogenous choline chloride alleviating salt stress injury in Kentucky bluegrass (*Poa pratensis*)

Zhi-Fang Zuo†, Yan Li†, Xin-Feng Mi, Yong-Long Li, Chen-Yuan Zhai, Guo-Feng Yang, Zeng-Yu Wang and Kun Zhang*

College of Grassland Science, Qingdao Agricultural University, Qingdao, Shandong, China

Introduction: Choline participates in plant stress tolerance through glycine betaine (GB) and phospholipid metabolism. As a salt-sensitive turfgrass species, Kentucky bluegrass (*Poa pratensis*) is the main turfgrass species in cool-season areas.

Methods: To improve salinity tolerance and investigate the effects of choline on the physiological and lipidomic responses of turfgrass plants under salinity stress conditions, exogenous choline chloride was applied to Kentucky bluegrass exposed to salt stress.

Results: From physiological indicators, exogenous choline chloride could alleviate salt stress injury in Kentucky bluegrass. Lipid analysis showed that exogenous choline chloride under salt-stress conditions remodeled the content of phospholipids, glycolipids, and lysophospholipids. Monogalactosyl diacylglycerol, digalactosyl diacylglycerol, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and lysophosphatidylcholine content were increased and phosphatidic acid content were decreased in plants after exogenous choline chloride under salt treatment. Plant leaf choline content increased, but GB was not detected in exogenous choline chloride treatment plants under nonstress or salt-stress conditions.

Discussion: GB synthesis pathway related genes showed no clear change to choline chloride treatment, whereas cytidyldiphosphate-choline (CDP-choline) pathway genes were upregulated by choline chloride treatment. These results reveal that lipid remodeling through choline metabolism plays an important role in the salt tolerance mechanism of Kentucky bluegrass. Furthermore, the lipids selected in this study could serve as biomarkers for further improvement of salt-sensitive grass species.

KEYWORDS

choline chloride, salt stress, phospholipids, glycolipids, turfgrass

1 Introduction

Salt stress, caused by soil salinization, is an important factor for plant distribution worldwide and greatly affects plant development and productivity (Deinlein et al., 2014). Salt stress primarily causes osmotic and ionic toxicity in plant cells, which induces secondary effects of oxidative stress that greatly damage cellular components, such as membrane lipids (Zhu, 2016). Halophytes and salt-tolerant glycophyte cultivars are tolerant to saline conditions and exhibit various mechanisms, including complex physiological traits, molecular or gene networks, and metabolic pathways (Annunziata et al., 2019). The research about plant salinity stress response, such as through osmotic adjustment (OA), which maintains cell membrane integrity and stability, could provide a combination of molecular tools for developing salt-tolerant varieties in saline environments. Therefore, it is an effective method to elevate plant salinity tolerance through increasing membrane stability (Lv et al., 2021).

Salinity enhances lipid peroxidation, which, in turn, impairs cellular membrane permeability. Phospholipids and glycolipids are two basic lipids in plant cell membranes and function as maintaining membrane structure and regulating plant salt stress tolerance (Mansour, 2013; Chalbi et al., 2015). Phospholipids, like phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), could work as signaling compounds in plants exposed to abiotic stress (Hou et al., 2016; Han and Yang, 2021). In structural phospholipids, PC and PE occupy a large proportion and the remainder consists of phosphatidylglycerol (PG), phosphoinositides (PI), phosphatidylserine (PS), and PA (Furt et al., 2011). The minor lipids PA and phosphatidylinositol bisphosphate (PIP2) act as salt stress-signaling lipids that rapidly accumulate in rice (*Oryza sativa*) leaves, as detected using ^{32}P labeling (Darwish et al., 2009). Platre et al. (2018) showed that PS and PA are required to generate the electrostatic signature of the plant plasma membrane, so their function could be separated based on their varying surface charges. PC is induced by salt stress and might function in the maintenance of membrane structure and function. Liu et al. (2015) reported that PE might be indirectly converted to PC in *Arabidopsis via* a hypothetical methylation pathway involving the synthesis of lysophosphatidylethanolamine (LysoPE), its methylation to lysophosphatidylcholine (LysoPC), and conversion to PC. PC could be synthesized by choline from the cytidyldiphosphate-choline (CDP-choline) pathway (Annunziata et al., 2019). In maize roots, PC levels decreased significantly with the increased expression of some phospholipase genes under saline-alkaline stress, suggesting there was a PC mediated lipid reprogramming under stress condition (Xu et al., 2021).

Glycolipids, like monogalactosyl diacylglycerol (MGDG) and digalactosyl diacylglycerol (DGDG), are major chloroplast thylakoid membranes components and primarily affect the photosystem II (PSII) characteristic (Mizusawa and Wada, 2012). In addition, a higher ratio of DGDGs/MGDGs to PCs/PEs is an indicator of the maintenance of membrane fluidity in response to abiotic stress in different plant species. It has been reported that the DGDG/MGDG ratio is response to salt stress (Guo et al., 2019). In transgenic tobacco (*Nicotiana tabacum*) with *OsMGD* gene, plants had a high salt tolerance, increased MGD activity and higher DGDG/MGDG ratios, which revealed a more stable thylakoid membrane in transgenic plant (Wang et al., 2014). In addition, the variation of membrane lipids compositions and fatty acid desaturation under salt stress condition have been identified as a common strategy for plant *via* affecting membranes mobility and stress signal transduction. (Magdy et al., 1993; Mansour, 2013).

Glycine betaine (GB) could synthesized from choline with two key enzyme, choline monooxygenase (CMO) and betaine aldehyde dehydrogenase (BADH) (Annunziata et al., 2019). In halophyte *Chenopodium quinoa*, choline accumulation could help to increase GB to improve salt tolerance (Pottosin et al., 2014). For some species like kidney bean (*Phaseolus vulgaris*) and wheat (*Triticum aestivum*), GB is a highly abundant compatible solute during the regulation of osmotic stress (Kreslavski et al., 2001; Salama et al., 2011). In *seashore paspalum*, the application of choline induced salt tolerance both in different salt sensitivity cultivars with remodeling the lipid profile and increase the GB content (Gao et al., 2020). However, not all plant species has high levels of GB, such as *Arabidopsis thaliana* and some grass species (Sakamoto and Murata, 2002; Hu et al., 2012). The exogenous treatment of choline or GB has effectively enhanced plant tolerance to abiotic stress, particularly for these low GB level species. In the C3 *Poa* turfgrass, foliar application of GB alleviated the physiological injury caused by drought or salt stress by maintaining membrane stability and active SOD, POD, and APX activity (Yang et al., 2012). In rice, choline priming could mitigate the salt stress during seed germination (Huang et al., 2023). Therefore, other metabolic pathways, such as the phospholipid and glycolipid reprograming through CDP-choline pathway, might play important role in dealing with salt stress in these species.

In this study, we examined the physiological response of Kentucky bluegrass with exogenous choline chloride under salt stress. Characterizing the GB and lipids (mainly phospholipids and glycolipids) metabolism that may be involved in the choline regulation of salt tolerance could provide theoretical basis for the important role of lipid metabolism in CDP-choline pathway during plant under salt stress condition. The selected lipid species in our study could be molecular markers for further salt stress tolerance improvement in salt-sensitive grass species.

2 Materials and methods

2.1 Plant material and treatment

Kentucky bluegrass ('Diva') seeds were sown in plastic pots (length 10 cm, width 7 cm, height 8.5 cm) filled with fritted clay. After 2-month pre-cultivation, uniform-size seeding were transferred to hydroponic conditions (0.5×Hoagland's nutrient solution) in a growth chamber. Seedlings were cultivated for 21 days to adapt to the environment before exposure to the experimental treatments.

Plants were foliar-sprayed with 1 mM choline chloride until it dripped from the plants based on the previous study (Gao et al., 2020). Choline chloride treatment was performed 7 days prior to the start of stress application and continued at 7-day intervals

during the experimental processing. 100 mM sodium chloride (NaCl) was added for the salt stress treatment. The salt stress was starting with 30 mM NaCl, and then gradually increased to 100 mM in two days. Each treatment had four replicates. Physiological measurements were performed 7, 14, and 21 days after treatment with 100 mM NaCl. Leaf was sampled at 14 days after salt stress treatment, then washed with deionized water and stored in -80°C for GB and choline content measurements, gene analysis, and lipid analysis. The plant growth state was photographed after 21 days of salt stress treatment.

2.2 Physiological measurements

The leaf Fv/Fm, electrolyte leakage, relative water content, and chlorophyll content was measured according to the previous study (Zhang et al., 2019; Zhang et al., 2020). The leaf osmotic adjustment was determined following a previous study (Gao et al., 2020). Quantification of GB and choline was performed using HPLC–MS based on a method described previously with modifications (Koc et al., 2002).

2.3 Lipid extraction

Lipid extraction was performed according to previous studies (Zhang et al., 2019). 1 mL lipid extract (methyl alcohol:methylene dichloride=1:1) was added and vortexed for 30 min. Added 300 μL ultra-pure water and vortexed for 1 minute and incubated at 4°C for 10 minutes. After centrifugation at 12000 rpm and 4°C for 3 minutes, 400 μL supernatant evaporated with nitrogen gas and concentrated to dry. Then, the sample was redissolved in 200 μL isopropanol. After filtered with filter membrane, the solution was ready for LC–MS/MS analysis.

2.4 HPLC and ESI–MS/MS conditions

The lipid extracts were analyzed with an LC–ESI- QTRAP/MS system (UPLC, Exion LC AD; MS, QTRAP 6500+ System). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 500°C; ion spray voltage 5500 V (Positive), −4500 V (Negative); Ion source gas 1, gas 2, and curtain gas were set at 45, 55, and 35 psi, respectively. The lipid unsaturation index was calculated as previously described (Su et al., 2009).

2.5 qRT-PCR analysis

Leaf RNA was isolated with the RNA EASY Fast Kit (TIANGEN, China). cDNA synthesis and qRT-PCR analyses was performed with FastKing One Step RT-PCR Kit (TIANGEN, China) on QuantStudio 1 Plus (Applied biosystems, USA). The qRT-PCR primers are listed in Supplementary Table S1. The *PpGADPH* gene was used as a reference control for qRT-PCR analysis according to a previous study (Niu et al., 2017). The gene expression was calculated using the $2^{-\triangle\triangle Ct}$ method (Livak and Schmittgen, 2001). Each qRT-PCR analysis was performed in triplicate.

2.6 Statistical analysis

The data were analyzed with Spss 17.0 (SPSS Inc., Chicago, IL, USA). Fisher's protected least significant difference (LSD) test was performed to calculate the significant differences ($P < 0.05$).

3 Results

3.1 The growth and physiological changes of Kentucky bluegrass

Kentucky bluegrass plants were cultured in nutrient solution were foliar-sprayed with choline chloride and subjected to a salt solution for 21 days (Figure 1). Choline chloride treatment had improved the plant growth with more leaves compared to untreated plant. Under salt stress condition, greener and fewer wilted leaves were observed in the choline chloride treatment plants than untreated plant. Growth characteristics showed that exogenous choline chloride increased the salt tolerance of Kentucky bluegrass.

In addition, physiological indicators were determined to illustrate the mechanism that increased salt tolerance in Kentucky bluegrass after the exogenous application of choline chloride. The Fv/Fm ratio decreased after salt stress treatments (Figure 2A). After 7 days of treatment, the difference of Fv/Fm ratio among the four treatments was not significant. After 14 days of treatment, the Fv/Fm ratio was significantly declined under stress conditions compared to the nonstress condition, whereas choline chloride did not affect the Fv/Fm ratio under stress and nonstress conditions. At 21 days of treatment, plants had a significantly higher Fv/Fm ratio in choline chloride treated plants, whereas salt treatment significantly decreased the Fv/Fm ratio compared to the nonstress plants. Electrolyte leakage were upregulated in both choline chloride untreated and untreated plants after salt stress (Figure 2B). At 7 and 14 days after salt stress, exogenous choline chloride resulted in lower EL (9.62% and 9.37%, respectively) than that of untreated plants. Chlorophyll content in all treatment were decreased during salt stress (Figure 2C) and it was significantly higher in choline chloride treated plants than that in untreated plants 7 (7.03%) and 21 days (10.00%) after salt treatment, whereas exogenous choline chloride significantly increased the chlorophyll content 14 and 21 days after the non-stress treatment. The relative water content in all treatment were decreased during salt stress (Figure 2D). Plants had significantly higher relative water content in choline chloride treated plants than that in untreated plants 7 (3.86%) and 14 days (5.37%) after salt treatment. Salt treatment significantly declined the relative water content compared to the nonstress treatment during days 7–21 of salt stress (Figure 2E). In addition, OA was detected in the leaves of the four treatments, and this value increased during the salt stress treatment. Significantly

10.3389/fpls.2023.1269286

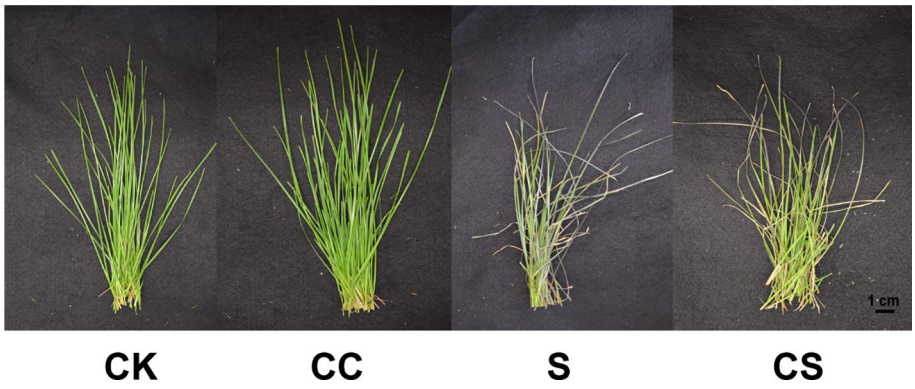


FIGURE 1
Growth of Kentucky bluegrass treated with choline chloride and salt stress. The photographs were taken after 21 days of treatment. CK, control treatment; CC, choline chloride treatment; S, salt treatment; CS, choline chloride and salt treatment, the same as below.

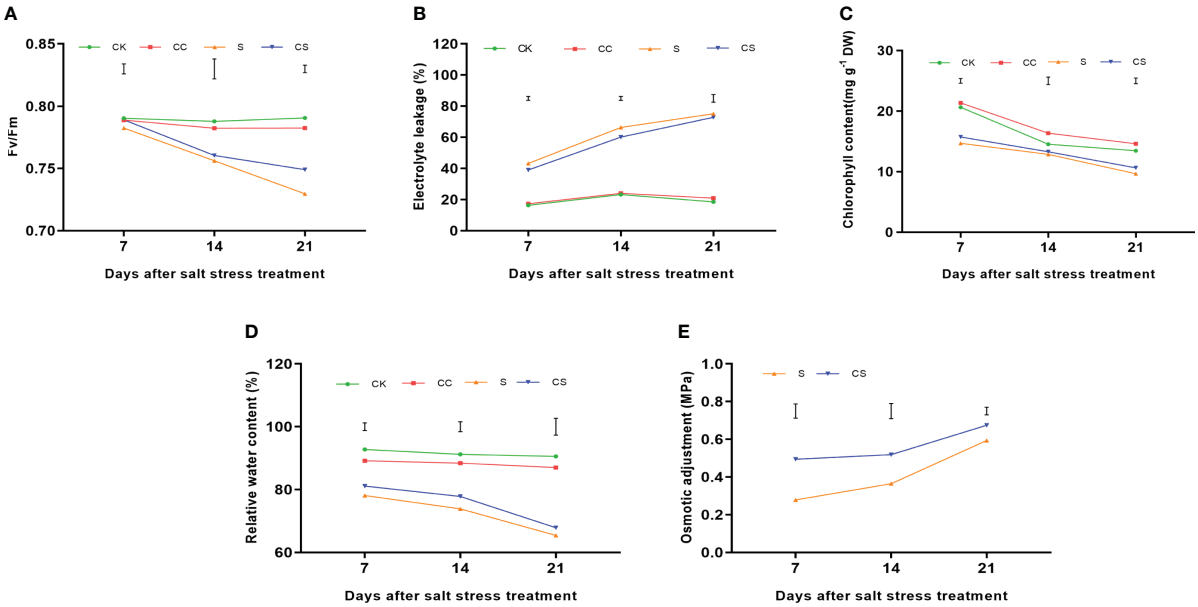


FIGURE 2
The physiological indicators of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** Fv/Fm. **(B)** Electrolyte leakage. **(C)** Chlorophyll content. **(D)** Relative water content. **(E)** Osmotic adjustment. All data are means ± SE for three biological replicates. The vertical bars indicate the values of LSD at *P*-value=0.05.

higher OA was found in choline chloride-treated plants from 7 to 21 days of stress treatment, with increases of 77.61%, 42.54%, and 13.55%, respectively, than in untreated plants.

3.2 The leaf lipidomic changes of Kentucky bluegrass

After 14 days of salt stress, the leaf tissues of Kentucky bluegrass were collected for lipidomic analysis. The leaf lipidomic change profiles are shown in Figure S1. In total, 150 lipid molecular species were tested, including DGDG, MGDG, PA, PC, PE, PG, PI, PS, and LysoPC in Kentucky bluegrass treated with choline chloride and salt stress.

As shown in Figure 3, the total lipid, total glycolipid, and total phospholipid contents significantly decreased after salt stress treatment. The total lipid content increased by 8.43% and 18.41% with exogenous choline chloride under nonstress and salt stress treatment (Figure 3A). The content of total glycolipid was significantly higher in choline chloride -treated plants than that in untreated nonstress control (8.67%) and salt-stressed plants (21.03%) (Figure 3B). For total phospholipids, an increase was found only in choline chloride-treated plants compared to non-treated plants under salt stress conditions (Figure 3C).

In glycolipids, the total DGDG and MGDG contents were both significantly declined with or without choline chloride treatment under salt stress (Figures 4A, B). Exogenous choline chloride significantly increased the total DGDG content under nonstress

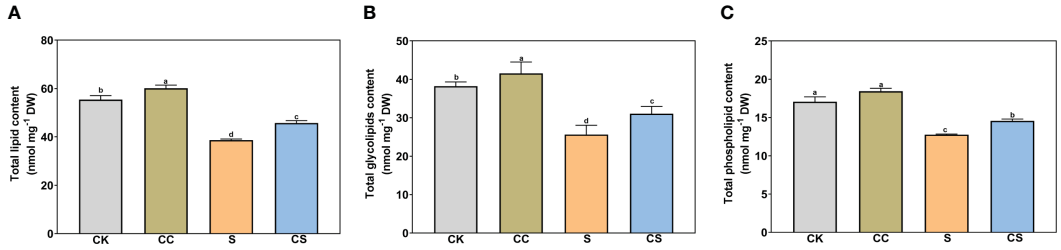


FIGURE 3
The contents of total lipid, glycolipid, and phospholipid of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** Total lipid content. **(B)** Total glycolipid content. **(C)** Total phospholipid content. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$, the same as below.

and salt-stress conditions by 17.37% and 20.15%, respectively, and increased the total MGDG content by 19.76% under salt-stress conditions compared to the control. In phospholipids, the total PA, PC, and PE (Figures 4C–E) and total PG and PS (Figures 5A, C) contents significantly decreased after salt stress treatment with or without choline chloride treatment. Exogenous choline chloride significantly increased the total PC content by 9.73% under salt-stress conditions, decreased the total PA content by 17.33% and 38.70% under nonstress and salt-stress conditions, respectively, and increased the total PS content by 22.76% and 22.73% under nonstress and salt-stress conditions, respectively. The PI content had no significant changes under different treatments (Figure 5B). The lysophospholipid (LysoPC) content also had no significant difference with choline chloride treatment under nonstress conditions, but the difference was significant under salt-stress conditions (S, 0.11 nmol mg^{-1}, CS 0.18 nmol mg^{-1}) (Figure 5D). In addition, exogenous choline chloride significantly increased the DGDG:MGDG ratio compared with that in treated plants under nonstress conditions (Figure 5E). Under salt treatment, the PC:PE ratio was significantly higher (8.68%) in choline chloride -treated plants than in untreated plants (Figure 5F).

3.3 The variation of different lipid molecular species in Kentucky bluegrass

The different lipid specific molecular classes are shown in Figures 6, 7 in Kentucky bluegrass leaves treated with choline chloride and salt stress.

For glycolipid species, the 36:6 lipid species had the highest abundant in all treatments in MGDG and DGDG (Figure 6A, B). The content of 34:3 and 36:6 molecular classes in DGDG were

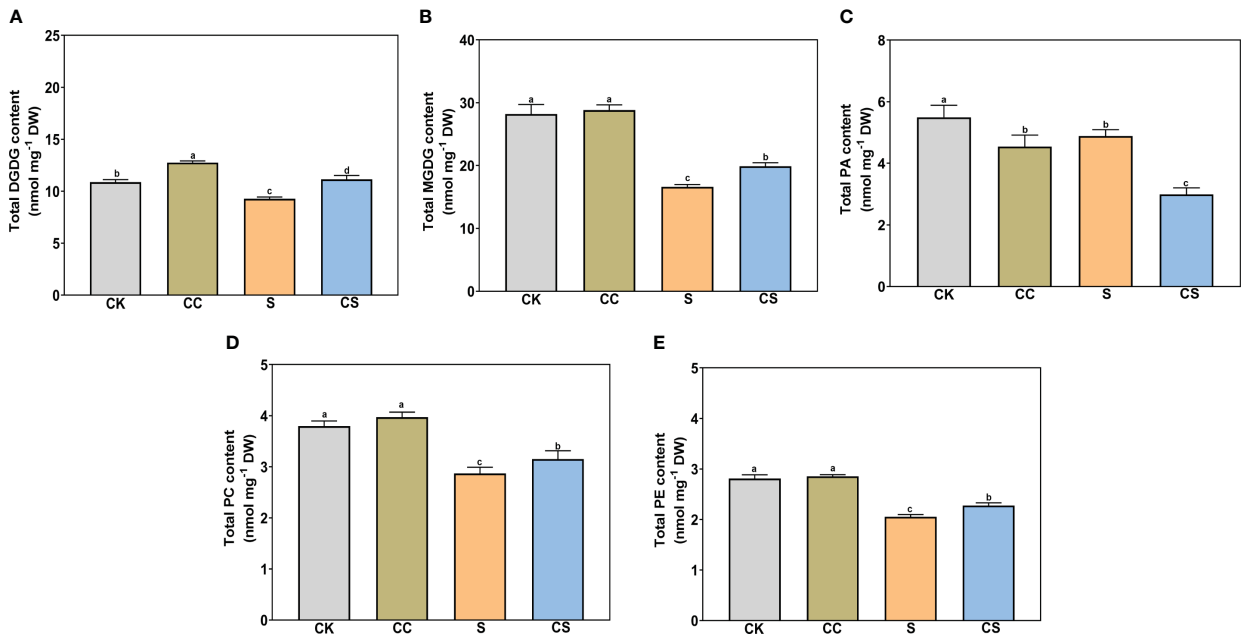


FIGURE 4
The contents of total DGDG, MGDG, PA, PC, and PE of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** Total DGDG content. **(B)** Total MGDG content. **(C)** Total PA content. **(D)** Total PC content. **(E)** Total PE content. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

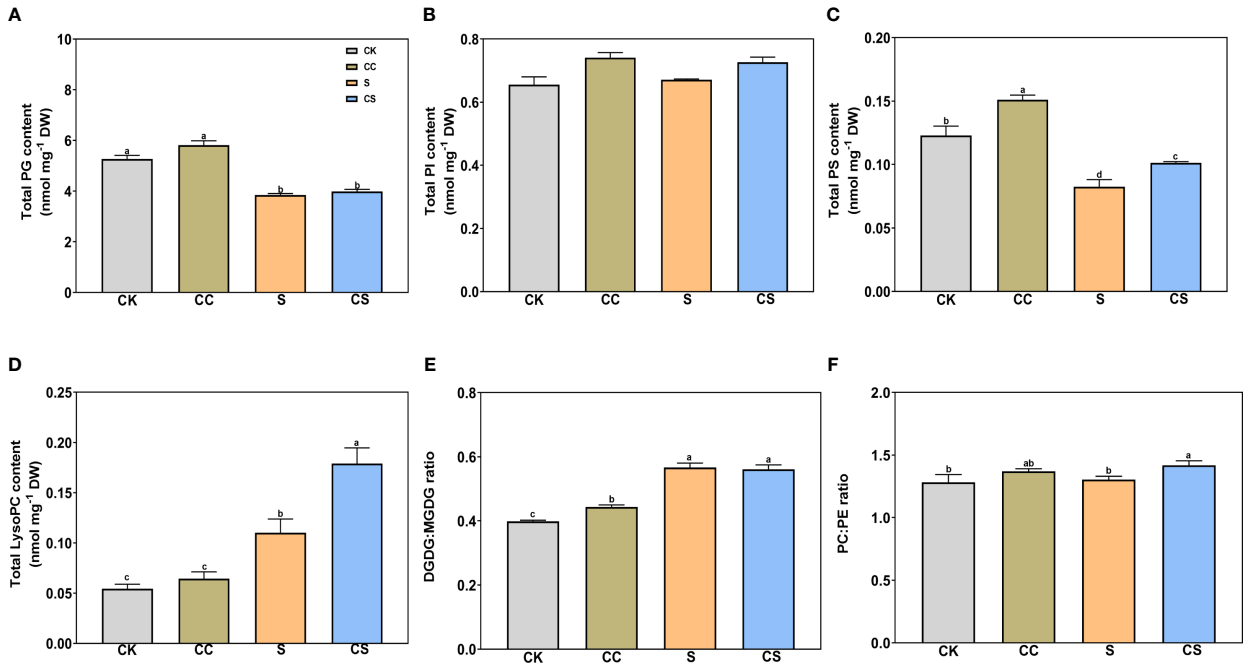


FIGURE 5
The contents of total PG, PI, PS, and LysoPC, and the ratio of DGDG:MGDG and PC:PE of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** Total PG content. **(B)** Total PI content. **(C)** Total PS content. **(D)** Total LysoPC content. **(E)** DGDG:MGDG ratio. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

decreased and the content of the 36:5 molecular class in MGDG was increased, with or without choline chloride treatment under salt stress condition. The DGDG (34:3, 36:6) content was significantly higher in choline chloride-treated plants under non stress conditions, whereas the DGDG (34:3, 36:6, 36:5) content was significantly higher in choline chloride -treated plants under salt and non-stress conditions. Compared to untreated plants, the MGDG (36:6, 36:5) content was significantly increased in choline

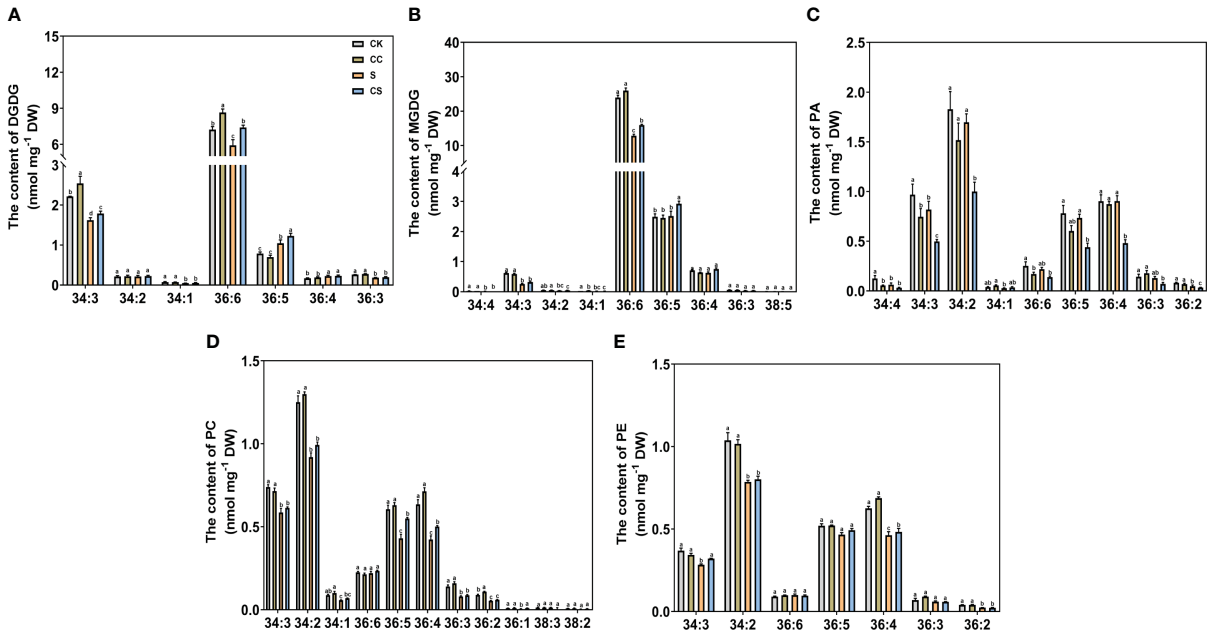


FIGURE 6
The specific molecular species of DGDG, MGDG, PA, PC, and PE of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** The content of DGDG content. **(B)** The content of MGDG content. **(C)** The content of PA content. **(D)** The content of PC content. **(E)** The content of PE content. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

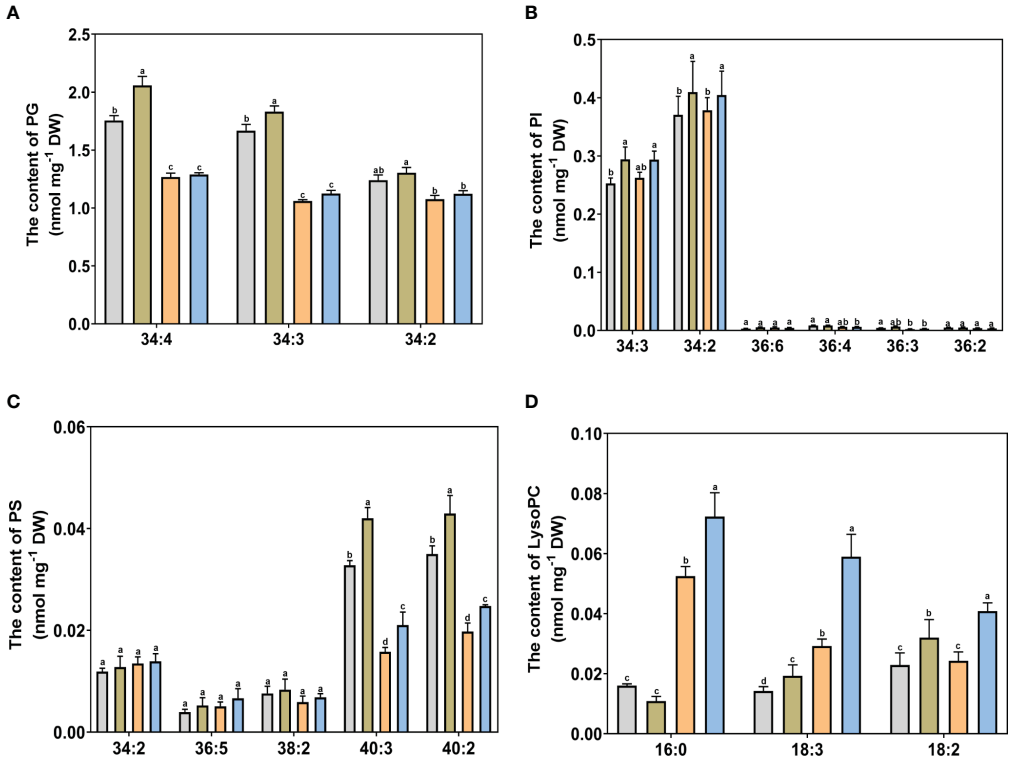


FIGURE 7
The specific molecular species of PG, PI, PS, and LysoPC of Kentucky bluegrass with choline chloride and salt stress treatments. **(A)** The content of PG content. **(B)** The content of PI content. **(C)** The content of PS content. **(D)** The content of LysoPC content. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

chloride -treated plants in under salt-stress. For phospholipid species, under non stress conditions, exogenous choline chloride significantly decreased the PA (34:3, 36:6) content, whereas PA species (34:3, 34:2, 36:5, 36:4, and 36:2) content were lower in choline chloride-treated plants compared nontreated plants under salt stress (Figure 6C). The PC (34:3, 34:2, 36:5, 36:4, 36:3, and 36:2) content decreased significantly after salt stress treatment. Under salt stress, the content of PC (36:5 and 36:4) in choline chloride treated plant were significantly higher that of non-treated plants (Figure 6D). The PE (34:2, 36:4) content was significantly lower under salt stress than that under nonstress conditions, and PE (34:3, 36:4) was higher in choline chloride -treated plants (Figure 6F). For other phospholipid species, a significant decrease in PG (34:4, 34:3, 34:2) and PS (40:3, 40:2) was observed after salt stress and choline chloride treatment significantly increased the PG (34:4, 34:3), PI (34:3, 34:2), and PS (40:3, 40:2) contents under non stress conditions (Figure 7A-C). Under salt stress, PI (34:2) and PS (40:3, 40:2) contents were both greatly increased by choline chloride treatment. For lysophospholipids, the content of LysoPC (18:3, 18:2) significantly increased in choline chloride -treated plants under non stress conditions, whereas that of LysoPC (16:0, 18:3, 18:2) significantly increased in choline chloride -treated plants under salt and non-stress conditions (Figure 7D). The DGDG, PC, and PE unsaturation were significantly increased, whereas the PG unsaturation was significantly decreased with or without choline chloride treatment after salt stress treatment (Table S2). In choline chloride -treated plants, a significantly lower unsaturation (3.02%) was observed in PC under salt stress.

3.4 The leaf GB and choline content determination

To determine the major pathways regulated by choline chloride, GB and choline contents were analyzed under different treatments (Figure 8). The results showed that the GB content were not detected following either salt or choline chloride treatment. Salt stress significantly increased the choline content with or without choline chloride treatment. Exogenous choline chloride significantly increased choline content (4.52% and 7.13%) under nonstress and salt-stress conditions, respectively. These results revealed that exogenous choline chloride could affect endogenous choline content rather than the GB synthesis pathway under stress and nonstress conditions in Kentucky bluegrass.

3.5 The expression of GB synthesis and CPD-choline pathway genes

Furthermore, expression analysis of genes controlling lipid synthesis was performed to determine the molecular mechanisms affected by choline. For genes in the GB synthesis pathway, salt

10.3389/fpls.2023.1269286

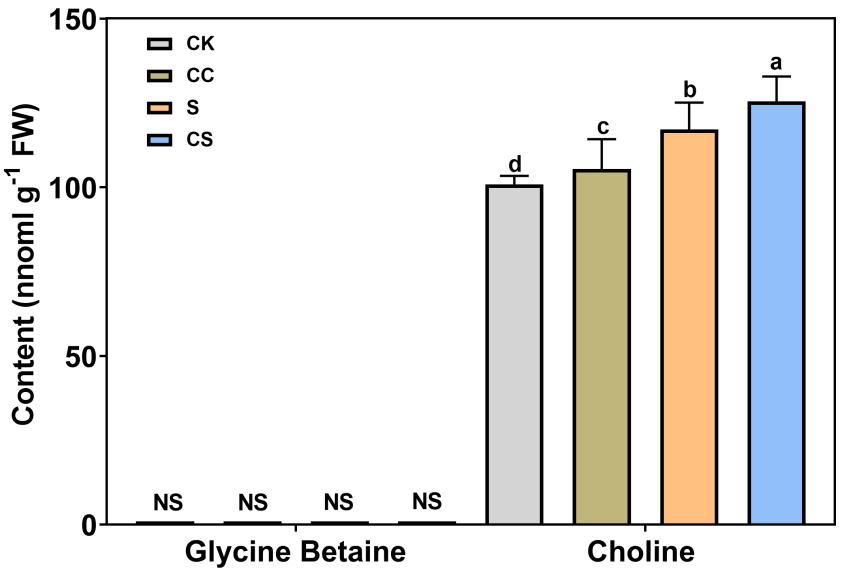


FIGURE 8
Leaf endogenous GB and choline content with choline chloride and salt stress treatments. NS means undetectable. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

stress greatly increased the expression of *CMO* and *BADH1* and decreased the expression of *BADH2* (Figures 9A-C). Exogenous choline chloride upregulated the *CMO* (1.53-fold), *BADH1* (2.97-fold), and *BADH2* (1.74-fold) gene expressions under nonstress conditions. The expression of *CMO* and *BADH1* had no significant changes to exogenous choline chloride after salt stress treatment, whereas *BADH2* gene expression was downregulated by the choline chloride treatment under salt stress. For genes in the CPD-choline pathway, salt stress significantly induced the expression of *Choline kinase* (*CK1*), *CTP-phosphocholine cytidylyltransferase* (*CCT1*), and *Choline/ethanolaminephosphotransferase* (*CEPT*) (Figures 9D-F). Under salt stress, the expression of *CK1* was significantly upregulated (73.19%) by choline chloride treatment. The expression of *CCT1* and *CEPT* gene were significantly higher in choline chloride treated plants under nonstress and salt-stress conditions by 104.24% and 209.87% and 14.64% and 15.21%,

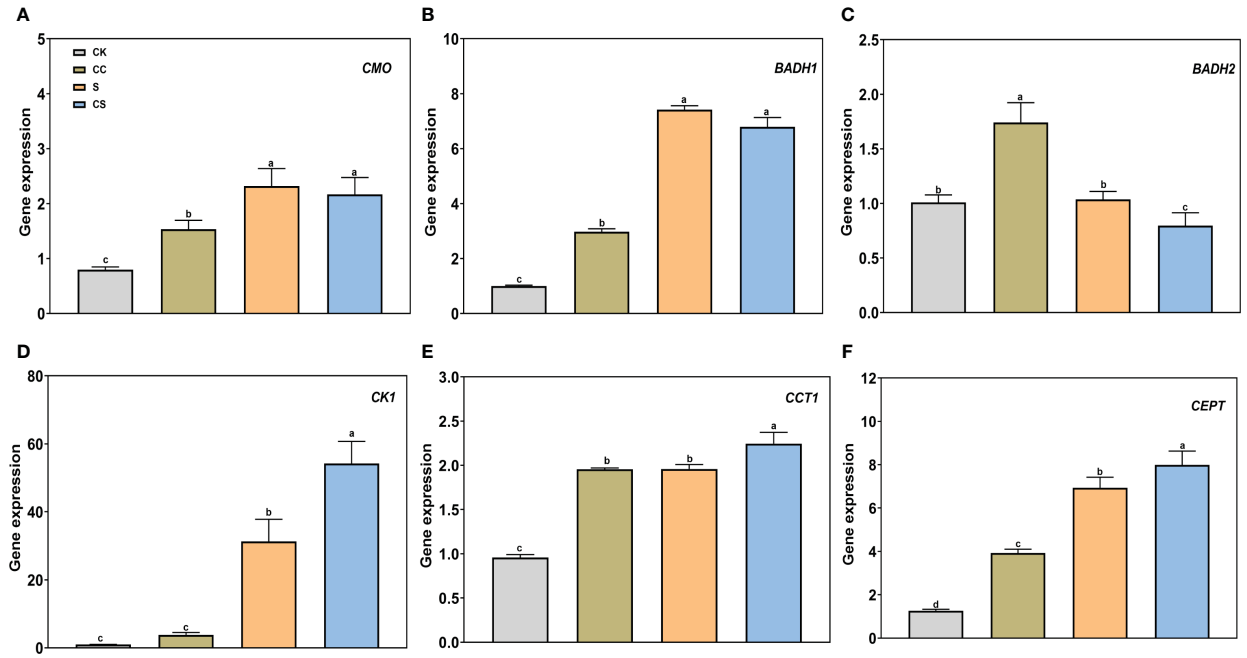


FIGURE 9
The gene expression of GB and CDP-choline metabolism pathway with choline chloride and salt stress treatments. **(A)** CMO. **(B)** BADH1. **(C)** BADH2. **(D)** CK1. **(E)** CCT1. **(F)** CEPT. All data are means ± SE for three biological replicates. Different letter indicated the significance at $P < 0.05$.

respectively. These results revealed that choline chloride regulates salt stress in Kentucky bluegrass *via* choline metabolic pathways and not through the GB synthesis pathway.

4 Discussion

Soil salinity is an important factor that affecting turfgrass growth and utilization. Kentucky bluegrass is sensitive to salt stress, and mitigation of the detrimental effects of salt stress and improvement in its tolerance is urgently required. Choline has been identified play role in plant salt stress tolerance (Incharoensakdi and Karnchanatat, 2003). In the present study, Fv/Fm, relative water content, chlorophyll content, and OA were increased, and EL were decreased in choline chloride -treated plants in salt treated plants. This illustrated that exogenous choline chloride could improve the salt tolerance of Kentucky bluegrass. Membrane remodeling is an important metabolic regulatory mechanism in plants exposed to salt stress (Guo et al., 2022). In soybeans, a short-term salt stress treatment could trigger a dynamic reprogramming in both galactolipids and phospholipids (Liu et al., 2021). Similar results have been reported in potato (*Solanum tuberosum*) (Yu et al., 2019), cotton (*Gossypium herbaceum*) (Liu et al., 2022), and *Carex rigescens* (Hu et al., 2021). As a precursor, choline is participated in two key pathways including GB metabolism pathway and lipid metabolism pathway in plant (Annunziata et al., 2019). In our study, endogenous GB metabolism pathway had not obviously response to choline application, while lipid (galactolipids and phospholipids) remodeling, were occurred with choline treatment, especially under salt stress. These results show that choline-mediated lipid metabolism is the central salt tolerance mechanism in Kentucky bluegrass.

Our data showed that the glycolipid and phospholipid contents were increased with choline chloride treatment. Choline alters the reorganization of lipid composition to protect membrane from the injury caused by salt stress, and similar results have been reported for the seashore paspalum (Gao et al., 2020). Plant glycolipids contribute to the formation of stacked thylakoid membranes and DGDG and MGDG content have been reported to decrease after salt stress in rice (Wang et al., 2014) and *Suaeda salsa* (Sui et al., 2010). In our study, an increase in the content of DGDG and MGDG was observed under salt stress with choline chloride treatment. DGDG and MDGD are likely to act as stabilizers in the organization of thylakoid membranes and are indispensable components of the light-harvesting complex II in plant photosynthesis (Guo et al., 2019). Furthermore, plants can alter the thylakoid membrane lipids component to resist abiotic stress. The content of MGDG 36:6, 36:5, and DGDG 34:3, 36:6, and 36:5 was significantly higher in choline chloride treated plants under salt stress, suggesting that these glycolipid species make a difference in choline mediated salt tolerance improvement. Similarly, the levels of 18:2-containing glycolipids (like MGDG-36:5 and DGDG-36:5) were upregulated by different abiotic stresses in tall fescue (*Festuca arundinacea*) (Zhang et al., 2019), *C. rigescens* (Hu et al., 2021), and rice (Wang et al., 2019). These results suggest that unsaturated 18:2-containing glycolipids might be significant in abiotic stress responses in plants.

Under salt stress condition, the PC content was significantly higher in choline chloride treated plants. PC participate in abiotic stress tolerance in plants (Tasseva et al., 2004). Studies have confirmed membrane PC has positive role in plant salt tolerance in *Arabidopsis thaliana* (Qiao et al., 2018), cordgrass (*Spartina patens*) (Wu et al., 2005) and tomato (*Lycopersicon esculentum*) (Kerkeb et al., 2001). In addition, in *Catharanthus roseus* and *Mesembryanthemum crystallinum* which had high salt tolerence (Elkahoui et al., 2004; Barkla et al., 2018), PC species (36:5 and 36:4) increased after salt stress compared to those in salt-sensitive plants, suggesting that they are involved in salt stress adaptation (Magdy et al., 1994). In the present study, the PC unsaturation index decreased with exogenous choline chloride. In PC-specific molecular species, the abundance of 16:0 acyl chains were higher than that of 18:3/18:2 acyl chains after choline chloride treatment. High lipid saturation contributes to plant cell membrane stability and integrity under abiotic stress (Li et al., 2020), indicating that these saturated lipid species may play an important role in choline-mediated salt stress tolerance in Kentucky bluegrass.

PE and PC are the major phospholipid components of extraplastidial membranes, and PE can regulate stress signal transduction under hyperosmotic stress *via* phospholipase C (Pokotylo et al., 2014; Singh et al., 2015). Under salt stress, the PE content was significantly increased in choline chloride treated plants. Studies have reported that PE increases in association with PC, resulting in abiotic stress due to increased membrane fluidity (Higashi and Saito, 2019). The choline group of PC is a bilayer-forming lipid, whereas the ethanolamine group of PE is a non-lamellar-forming lipid, which affects the formation of the bilayer phase and preserves membrane fluidity (Toumi et al., 2008; Narayanan et al., 2018). Therefore, the PC/PE ratio may reflect a cellular response to maintain membrane stability and balance between the membrane and storage lipids (Narayanan et al., 2018). In the present study, the PC/PE ratio increased in response to salt stress with exogenous choline chloride, illustrating that the elevated abundance of PC rather than PE with choline chloride treatment could help maintain the membrane integrity and fluidity in Kentucky bluegrass.

PA has been reported to crucial for salt stress responses by helping maintain ion homeostasis (Yao and Xue, 2018; Liu et al., 2019). In addition, PA is a major intermediate that guiding the PC and PE biosynthesis from the CDP-choline and CDP-ethanolamine pathways, respectively (Carman and Han, 2009). The PA content and specific molecular species decreased under salt stress in Kentucky bluegrass, and the result was the opposite of the PC and PE variation. Under abiotic stress conditions, phospholipase D could hydrolyze of PE and PC and produce PA in plants (Jiang et al., 2019). Therefore, these results suggest that choline can improve salt tolerance in Kentucky bluegrass by promoting PA downstream of PC and PE synthesis through the CDP-choline pathway.

The phosphatidylserine (PS) content can be regulated by abiotic stress in plants (Higashi and Saito, 2019). PS includes a large amount of lipid species, such as PS (40:3), which contains very-long-chain fatty acids and has been reported to increase in heat stress-susceptible wheat genotype plants (Narayanan et al., 2016).

In the present study, choline chloride treatment increased the PS (40:3 and 40:2) content under salt stress conditions. However, the function of the increased PS content under salt stress is unknown. A previous study reported that PS decarboxylase (PSD) converts PS to PE through endoplasmic reticulum (Yamaoka et al., 2011). Therefore, the changes in PS in plants exposed to salinity stress with choline chloride treatment may rely on the PSD-mediated metabolic pathway and, together with PE, may play a role in membrane stability.

LysoPC is a lipid class with only one fatty acyl chain on the glycerol moiety of glycerolipids, which is surface-active, alters membrane fluidity, and affects membrane receptor functions (Zhang et al., 2018). In the present study, choline chloride treatment increased LysoPC content under salt stress conditions. In *Schizochytrium* sp., LysoPC lipid species were significantly enriched in the choline metabolism pathway, contributing to improved salt stress tolerance (Jiang et al., 2019). However, the role of LysoPC in stress tolerance in plants remains unclear. Plant Ca^{2+} plays a significant role in cell membrane integrity, and endogenous Ca^{2+} levels may decrease during salt stress (Hashemipetroudi et al., 2022). A previous study found that the content of LysoPC increased in response to Ca^{2+} deprivation (IsHizuKA et al., 2000), demonstrating that LysoPC might play an essential role in Ca^{2+} maintaining plasma membrane stability.

In summary, choline chloride increased the salt tolerance of Kentucky bluegrass by increasing Fv/Fm, relative water content, chlorophyll content, and OA and reducing leaf EL. Lipidomic results showed that the content of total MGDG, DGDG, PC, PE, PS, and LysoPC were increased, and PA content was decreased with exogenous choline chloride when exposed to salt stress. The leaf GB content and synthesis related genes had no response to exogenous choline chloride, indicating that choline triggered lipid metabolism was a primary mechanism in Kentucky bluegrass response to salt stress condition The specific lipid molecular species of glycolipids (MGDG 36:6, 36:5, and DGDG 34:3, 36:6, 36:5), phospholipids (PC 36:5, 36:4, PE 34:3, 34:2, 36:5, 36:4, PI 34:2, and PS 40:3, 40:2), and lysophospholipids (LysoPC 16:0, 18:3, 18:2) were increased exposed to exogenous choline chloride. Our results provide lipid biomarkers for further improvement of salt-sensitive grass species.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZZ: Writing – original draft, Writing – review and editing, Software, Data curation, Formal Analysis, Validation. YL: Validation, Writing – review and editing, Data curation, Formal Analysis, Software. XM: Data curation, Software, Writing – review and editing, Formal Analysis. YL: Writing – review and editing, Investigation, Validation. CZ: Validation, Writing – review and editing, Data curation. GY: Writing – review and editing, Resources, Supervision. ZW: Supervision, Writing – review and editing, Funding acquisition. KZ: Funding acquisition, Supervision, Writing – review and editing, Software, Writing – original draft.

Funding

The work was supported by the First-Class Grassland Science Discipline Program of Shandong Province, China and Start-Up Foundation for High Talents of Qingdao Agricultural Universtiy (no.665/1121011).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1269286/full#supplementary-material

References

Annunziata, M. G., Ciarmiello, L. F., Woodrow, P., Dell'Aversana, E., and Carillo, P. (2019). Spatial and temporal profile of glycine betaine accumulation in plants under abiotic stresses. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00230

Barkla, B. J., Garibay-Hernández, A., Melzer, M., Rupasinghe, T. W. T., and Roessner, U. (2018). Single cell-type analysis of cellular lipid remodelling in response to salinity in the epidermal bladder cells of the model halophyte *Mesembryanthemum crystallinum*. *Plant Cell Environ.* 41 (10), 2390–2403. doi: 10.1111/pce.13352

Carman, G. M., and Han, G. S. (2009). Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. *J. Biol. Chem.* 284 (5), 2593–2597. doi: 10.1074/jbc.R800059200

Chalbi, N., Martínez-Ballesta, M. C., Youssef, N. B., and Carvajal, M. (2015). Intrinsic stability of Brassicaceae plasma membrane in relation to changes in proteins and lipids as a response to salinity. *J. Plant Physiol.* 175, 148–156. doi: 10.1016/j.jplph.2014.12.003

Darwish, E., Testerink, C., Khalil, M., El-Shihy, O., and Munnik, T. (2009). Phospholipid signaling responses in salt-stressed rice leaves. *Plant Cell Physiol.* 50 (5), 986–997. doi: 10.1093/pcp/pcp051

Deinlein, U., Stephan, A. B., Horie, T., Luo, W., Xu, G., and Schroeder, J. I. (2014). Plant salt-tolerance mechanisms. *Trends Plant Sci.* 19 (6), 371–379. doi: 10.1016/j.tplants.2014.02.001

Elkahoui, S., Smaoui, A., Zarrouk, M., Ghrir, R., and Limam, F. (2004). Salt-induced lipid changes in *Catharanthus roseus* cultured cell suspensions. *Phytochemistry* 65 (13), 1911–1917. doi: 10.1016/j.phytochem.2004.06.021

Furt, F., Simon-Plas, F., and Mongrand, S. (2011). Lipids of the plant plasma membrane. *Plant plasma membrane* 19, 3–30. doi: 10.1007/978-3-642-13431-9_1

Gao, Y., Li, M., Zhang, X., Yang, Q., and Huang, B. (2020). Up-regulation of lipid metabolism and glycine betaine synthesis are associated with choline-induced salt tolerance in halophytic seashore paspalum. *Plant Cell Environ.* 43 (1), 159–173. doi: 10.1111/pce.13657

Guo, Q., Liu, L., and Barkla, B. J. (2019). Membrane lipid remodeling in response to salinity. *Int. J. Mol. Sci.* 20 (17), 4264. doi: 10.3390/ijms20174264

Guo, Q., Liu, L., Rupasinghe, T. W. T., Roessner, U., and Barkla, B. J. (2022). Salt stress alters membrane lipid content and lipid biosynthesis pathways in the plasma membrane and tonoplast. *Plant Physiol.* 189 (2), 805–826. doi: 10.1093/plphys/kiac123

Han, X., and Yang, Y. (2021). Phospholipids in salt stress response. *Plants (Basel)* 10 (10), 2024. doi: 10.3390/plants10102204

Hashemipetroudi, S. H., Ahmadian, G., Fatemi, F., Nematzadeh, G., Yamchi, A., and Kuhlmann, M. (2022). Ion content, antioxidant enzyme activity and transcriptional response under salt stress and recovery condition in the halophyte grass *Aeluropus littoralis*. *BMC Res. Notes* 15 (1), 201. doi: 10.1186/s13104-022-06090-4

Higashi, Y., and Saito, K. (2019). Lipidomic studies of membrane glycerolipids in plant leaves under heat stress. *Prog. Lipid Res.* 75, 100990. doi: 10.1016/j.plipres.2019.100990

Hou, Q., Ufer, G., and Bartels, D. (2016). Lipid signalling in plant responses to abiotic stress. *Plant Cell Environ.* 39 (5), 1029–1048. doi: 10.1111/pce.12666

Hu, Q., Cui, H., Ma, C., Li, Y., Yang, C., Wang, K., et al. (2021). Lipidomic metabolism associated with acetic acid priming-induced salt tolerance in *Carex rigescens*. *Plant Physiol. Biochem.* 167, 665–677. doi: 10.1016/j.plaphy.2021.08.045

Hu, L., Hu, T., Zhang, X., Pang, H., and Fu, J. (2012). Exogenous glycine betaine ameliorates the adverse effect of salt stress on perennial ryegrass. *J. Am. Soc. Hortic. Sci.* 137 (1), 38–46. doi: 10.21273/JASHS.137.1.38

Huang, L., Yu, M., Feng, N., Di., Z., Jie, Y., and Li, Y. (2023). Effects of seed initiators choline chloride and indole-3-butyric acid potassium on growth and physiological characteristics of rice seedlings under salt Stress. *J. Nuc Agr Sci.* 37 (1), 196–206. doi: 10.11869/j.issn.1000-8551.2023.01.0196

Incharoensakdi, A., and Karnchanatat, A. (2003). Salt stress enhances choline uptake in the halotolerant cyanobacterium *Aphanothece halophytica*. *Biochim. Biophys. Acta* 1621 (1), 102–109. doi: 10.1016/s0304-4165(03)00052-7

IsHizuKA, K., Kanayama, A., Satsu, H., Miyamoto, Y., Furihata, K., and Shimizu, M. (2000). Identification of a taurine transport inhibitory substance in sesame seeds. *Bioscience biotechnology Biochem.* 64 (6), 1166–1172. doi: 10.1271/bbb.64.1166

Jiang, J. Y., Zhu, S., Zhang, Y., Sun, X., Hu, X., Huang, H., et al. (2019). Integration of lipidomic and transcriptomic profiles reveals novel genes and regulatory mechanisms of Schizochytrium sp. in response to salt stress. *Bioresour Technol.* 294, 122231. doi: 10.1016/j.biortech.2019.122231

Kerkeb, L., Donaire, J. P., Venema, K., and Rodríguez-Rosales, M. P. (2001). Tolerance to NaCl induces changes in plasma membrane lipid composition, fluidity and H^+-ATPase activity of tomato calli. *Physiol. Plant* 113 (2), 217–224. doi: 10.1034/j.1399-3054.2001.1130209.x

Koc, H., Mar, M. H., Ranasinghe, A., Swenberg, J. A., and Zeisel, S. H. (2002). Quantitation of choline and its metabolites in tissues and foods by liquid chromatography/electrospray ionization-isotope dilution mass spectrometry. *Anal. Chem.* 74 (18), 4734–4740. doi: 10.1021/ac025624x

Kreslavski, V., Balakhnina, T., Khristin, M., and Bukhov, N. (2001). Pre-treatment of bean seedlings with choline compounds increases the resistance of photosynthetic apparatus to UV-B radiation and elevated temperatures. *Photosynthetica* 39, 363–368. doi: 10.1023/A:1015174108937

Li, J., Liu, L. N., Meng, Q., Fan, H., and Sui, N. (2020). The roles of chloroplast membrane lipids in abiotic stress responses. *Plant Signal Behav.* 15 (11), 1807152. doi: 10.1080/15592324.2020.1807152

Liu, X., Ma, D., Zhang, Z., Wang, S., Du, S., Deng, X., et al. (2019). Plant lipid remodeling in response to abiotic stresses. *Environ. Exp. Bot.* 165, 174–184. doi: 10.1016/j.envexpbot.2019.06.005

Liu, Y., Wang, L., Li, X., and Luo, M. (2022). Detailed sphingolipid profile responded to salt stress in cotton root and the GhIPCS1 is involved in the regulation of plant salt tolerance. *Plant Sci.* 316, 111174. doi: 10.1016/j.plantsci.2021.111174

Liu, Y., Wang, G., and Wang, X. (2015). Role of aminoalcoholphosphotransferases 1 and 2 in phospholipid homeostasis in Arabidopsis. *Plant Cell* 27 (5), 1512–1528. doi: 10.1105/tpc.15.00180

Liu, A., Xiao, Z., Wang, Z., Lam, H. M., and Chye, M. L. (2021). Galactolipid and phospholipid profile and proteome alterations in soybean leaves at the onset of salt stress. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.644408

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Lv, S., Tai, F., Guo, J., Jiang, P., Lin, K., Wang, D., et al. (2021). Phosphatidylserine synthase from salicornia europaea is involved in plant salt tolerance by regulating plasma membrane stability. *Plant Cell Physiol.* 62 (1), 66–79. doi: 10.1093/pcp/pcaa141

Magdy, M., Mansour, F., Lee-Stadelmann, O. Y., and Stadelmann, E. J. (1993). Salinity stress and cytoplasmic factors. A comparison of cell permeability and lipid partiality in salt sensitive and salt resistant cultivars and lines of *Triticum aestivum* and *Hordeum vulgare*. *Physiologia Plantarum* 88 (1), 141–148. doi: 10.1111/j.1399-3054.1993.tb01771.x

Magdy, M., Mansour, F., van Hasselt, P. R., and Kuiper, P. J. (1994). Plasma membrane lipid alterations induced by NaCl in winter wheat roots. *Physiologia Plantarum* 92 (3), 473–478. doi: 10.1111/j.1399-3054.1994.tb08838.x

Mansour, M. M. F. (2013). Plasma membrane permeability as an indicator of salt tolerance in plants. *Biol. plantarum* 57, 1–10. doi: 10.1007/s10535-012-0144-9

Mizusawa, N., and Wada, H. (2012). The role of lipids in photosystem II. *Biochim. Biophys. Acta* 1817 (1), 194–208. doi: 10.1016/j.bbabio.2011.04.008

Narayanan, S., Prasad, P. V. V., and Welti, R. (2018). Alterations in wheat pollen lipidome during high day and night temperature stress. *Plant Cell Environ.* 41 (8), 1749–1761. doi: 10.1111/pce.13156

Narayanan, S., Tamura, P. J., Roth, M. R., Prasad, P. V., and Welti, R. (2016). Wheat leaf lipids during heat stress: I. High day and night temperatures result in major lipid alterations. *Plant Cell Environ.* 39 (4), 787–803. doi: 10.1111/pce.12649

Niu, K., Shi, Y., and Ma, H. (2017). Selection of Candidate Reference Genes for Gene Expression Analysis in Kentucky Bluegrass (*Poa pratensis* L.) under Abiotic Stress. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00193

Platre, M. P., Noack, L. C., Doumane, M., Bayle, V., Simon, M. L. A., Maneta-Peyret, L., et al. (2018). A combinatorial lipid code shapes the electrostatic landscape of plant endomembranes. *Dev. Cell* 45 (4), 465–480.e411. doi: 10.1016/j.devcel.2018.04.011

Pokotylo, I., Kolesnikov, Y., Kravets, V., Zachowski, A., and Ruelland, E. (2014). Plant phosphoinositide-dependent phospholipases C: variations around a canonical theme. *Biochimie* 96, 144–157. doi: 10.1016/j.biochi.2013.07.004

Pottosin, I., Bonales-Alatorre, E., and Shabala, S. (2014). Choline but not its derivative betaine blocks slow vacuolar channels in the halophyte Chenopodium quinoa: implications for salinity stress responses. *FEBS Lett.* 588 (21), 3918–3923. doi: 10.1016/j.febslet.2014.09.003

Qiao, K., Wang, M., Takano, T., and Liu, S. (20181772). Overexpression of acyl-coA-binding protein 1 (ChACBP1) from saline-alkali-tolerant chlorella sp. Enhances stress tolerance in *arabidopsis*. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01772

Sakamoto, A., and Murata, N. (2002). The role of glycine betaine in the protection of plants from stress: clues from transgenic plants. *Plant Cell Environ.* 25 (2), 163–171. doi: 10.1046/j.0016-8025.2001.00790.x

Salama, K. H., Mansour, M. M., and Hassan, N. S. (2011). Choline priming improves salt tolerance in wheat (*Triticum aestivum* L.). *Aust. J. Basic Appl. Sci.* 5 (11), 126–132. doi: 10.13140/2.1.4228.9606

Singh, A., Bhatnagar, N., Pandey, A., and Pandey, G. K. (2015). Plant phospholipase C family: Regulation and functional role in lipid signaling. *Cell Calcium* 58 (2), 139–146. doi: 10.1016/j.ceca.2015.04.003

Su, K., Bremer, D. J., Jeannotte, R., Welti, R., and Yang, C. (2009). Membrane lipid composition and heat tolerance in cool-season turfgrasses, including a hybrid bluegrass. *J. Am. Soc. Hortic. Sci.* 134 (5), 511–520. doi: 10.21273/JASHS.134.5.511

Sui, N., Li, M., Li, K., Song, J., and Wang, B.-S. (2010). Increase in unsaturated fatty acids in membrane lipids of *Suaeda salsa* L. enhances protection of photosystem II under high salinity. *Photosynthetica* 48, 623–629. doi: 10.1007/s11099-010-0080-x

Tasseva, G., Richard, L., and Zachowski, A. (2004). Regulation of phosphatidylcholine biosynthesis under salt stress involves choline kinases in Arabidopsis thaliana. *FEBS Lett.* 566 (1-3), 115–120. doi: 10.1016/j.febslet.2004.04.015

Toumi, I., Gargouri, M., Nouairi, I., Moschou, P., Salem-Fnayou, A. B., Mliki, A., et al. (2008). Water stress induced changes in the leaf lipid composition of four grapevine genotypes with different drought tolerance. *Biol. Plantarum* 52, 161–164. doi: 10.1007/s10535-008-0035-2

Wang, Y., Jiang, D., Hou, J., and Chen, G. (2019). Physiological characterization and thylakoid ultrastructure analysis in super high-yield hybrid rice leaves under drought stress. *Photosynthetica* 57 (3), 890–896. doi: 10.32615/ps.2019.106

Wang, S., Uddin, M. I., Tanaka, K., Yin, L., Shi, Z., Qi, Y., et al. (2014). Maintenance of chloroplast structure and function by overexpression of the rice MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE gene leads to enhanced salt tolerance in tobacco. *Plant Physiol.* 165 (3), 1144–1155. doi: 10.1104/pp.114.238899

Wu, J., Seliskar, D. M., and Gallagher, J. L. (2005). The response of plasma membrane lipid composition in callus of the halophyte Spartina patens (*Poaceae*) to salinity stress. *Am. J. Bot.* 92 (5), 852–858. doi: 10.3732/ajb.92.5.852

Xu, X., Zhang, J., Yan, B., Wei, Y., Ge, S., Li, J., et al. (2021). The adjustment of membrane lipid metabolism pathways in maize roots under saline-alkaline stress. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.635327

Yamaoka, Y., Yu, Y., Mizoi, J., Fujiki, Y., Saito, K., Nishijima, M., et al. (2011). PHOSPHATIDYLSERINE SYNTHASE1 is required for microspore development in *Arabidopsis thaliana*. *Plant J.* 67 (4), 648–661. doi: 10.1111/j.1365-313X.2011.04624.x

Yang, Z., Yu, J., Merewitz, E., and Huang, B. (2012). Differential effects of abscisic acid and glycine betaine on physiological responses to drought and salinity stress for two perennial grass species. *J. Am. Soc. Hortic. Sci.* 137 (2), 96–106. doi: 10.21273/JASHS.137.2.96

Yao, H. Y., and Xue, H. W. (2018). Phosphatidic acid plays key roles regulating plant development and stress responses. *J. Integr. Plant Biol.* 60 (9), 851–863. doi: 10.1111/jipb.12655

Yu, Y., Kou, M., Gao, Z., Liu, Y., Xuan, Y., Liu, Y., et al. (2019). Involvement of phosphatidylserine and triacylglycerol in the response of sweet potato leaves to salt stress. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01086

Zhang, K., Cui, H., Li, M., Xu, Y., Cao, S., Long, R., et al. (2020). Comparative time-course transcriptome analysis in contrasting *Carex rigescens* genotypes in response to high environmental salinity. *Ecotoxicol Environ. Saf.* 194, 110435. doi: 10.1016/j.ecoenv.2020.110435

Zhang, X., Xu, Y., and Huang, B. (2019). Lipidomic reprogramming associated with drought stress priming-enhanced heat tolerance in tall fescue (*Festuca arundinacea*). *Plant Cell Environ.* 42 (3), 947–958. doi: 10.1111/pce.13405

Zhang, T., Yang, J., Sun, Y., Kang, Y., Yang, J., and Qi, Z. (2018). Calcium deprivation enhances non-selective fluid-phase endocytosis and modifies membrane lipid profiles in Arabidopsis roots. *J. Plant Physiol.* 226, 22–30. doi: 10.1016/j.jplph.2018.04.002

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167 (2), 313–324. doi: 10.1016/j.cell.2016.08.029

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 September 2023
DOI 10.3389/fpls.2023.1242836

Check for updates

OPEN ACCESS

EDITED BY
Xiaoxu Li,
Anhui Agricultural University, China

REVIEWED BY
Babar Shahzad,
University of Tasmania, Australia
Allah Ditta,
Nuclear Institute for Agriculture and Biology, Pakistan
Jinhao Sun,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Muhammad Sajjad
muhammadsajjad.pbg@hotmail.com
Meng-Zhu Lu
lumz@caf.ac.cn

RECEIVED 19 June 2023
ACCEPTED 14 August 2023
PUBLISHED 15 September 2023

CITATION
Sajjad M, Ahmad A, Riaz MW, Hussain Q, Yasir M and Lu M-Z (2023) Recent genome resequencing paraded COBRA-*Like* gene family roles in abiotic stress and wood formation in Poplar.
Front. Plant Sci. 14:1242836.
doi: 10.3389/fpls.2023.1242836

COPYRIGHT
© 2023 Sajjad, Ahmad, Riaz, Hussain, Yasir and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Recent genome resequencing paraded COBRA-*Like* gene family roles in abiotic stress and wood formation in Poplar

Muhammad Sajjad[1]*, Adeel Ahmad[2], Muhammad Waheed Riaz[1,3], Quaid Hussain[4], Muhammad Yasir[5] and Meng-Zhu Lu[1]*

[1]State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A & F University, Hangzhou, China, [2]State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan, China, [3]Zhejiang Provincial Key Laboratory of Resource Protection and Innovation of Traditional Chinese Medicine, Zhejiang A&F University, Hangzhou, China, [4]Shenzhen Key Laboratory of Marine Bioresource and Eco-Environmental Science, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, [5]The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, College of Advanced Agricultural Sciences, Zhejiang A&F University, Hangzhou, China

A cell wall determines the mechanical properties of a cell, serves as a barrier against plant stresses, and allows cell division and growth processes. The COBRA-*Like* (COBL) gene family encodes a putative glycosylphosphatidylinositol (GPI)-anchored protein that controls cellulose deposition and cell progression in plants by contributing to the microfibril orientation of a cell wall. Despite being studied in different plant species, there is a dearth of the comprehensive global analysis of COBL genes in poplar. Poplar is employed as a model woody plant to study abiotic stresses and biomass production in tree research. Improved genome resequencing has enabled the comprehensive exploration of the evolution and functional capacities of PtrCOBLs (Poplar COBRA-Like genes) in poplar. Phylogeny analysis has discerned and classified PtrCOBLs into two groups resembling the *Arabidopsis* COBL family, and group I genes possess longer proteins but have fewer exons than group II. Analysis of gene structure and motifs revealed PtrCOBLs maintained a rather stable motif and exon–intron pattern across members of the same group. Synteny and collinearity analyses exhibited that the evolution of the COBL gene family was heavily influenced by gene duplication events. PtrCOBL genes have undergone both segmental duplication and tandem duplication, followed by purifying selection. Promotor analysis flaunted various phytohormone-, growth- and stress-related cis-elements (e.g., MYB, ABA, MeJA, SA, AuxR, and ATBP1). Likewise, 29 Ptr-miRNAs of 20 families were found targeting 11 PtrCOBL genes. PtrCOBLs were found localized at the plasma membrane and extracellular matrix, while gene ontology analysis showed their involvement in plant development, plant growth, stress response, cellulose biosynthesis, and cell wall biogenesis. RNA-seq datasets depicted the bulk of PtrCOBL genes expression being found in plant stem tissues and leaves, rendering mechanical strength and rejoinders to environmental cues. PtrCOBL2, 3, 10, and 11 manifested the highest expression in vasculature and abiotic stress, and resemblant expression trends were upheld by qRT-PCR. Co-expression network analysis identified PtrCOBL2 and PtrCOBL3 as hub genes

across all abiotic stresses and wood developing tissues. The current study reports regulating roles of PtrCOBLs in xylem differentiating tissues, tension wood formation, and abiotic stress latency that lay the groundwork for future functional studies of the PtrCOBL genes in poplar breeding.

KEYWORDS

cell wall, COBRA-*Like*, genome-wide, abiotic stress, tension wood, RNA-Seq

1 Introduction

Forests are greatly important for this world, particularly in an era of huge climate change, which is bringing about drastic changes around the globe, e.g., temperature hikes, droughts, and floods (Rudel et al., 2020). Secondly, the growing global population is causing a surge in demand for wood and tree-based products. To address these issues, we need to understand the regulatory mechanisms in tree growth, development, and stress tolerance so that we can produce genetically better trees with improved wood properties and stress resilience (De Frenne et al., 2021).

The cell wall is a flagship component of plant cells that provides mechanical strength and plasticity for expansion to the cells. As a result of cell division and expansion, plant cells differentiate into numerous ultimate shapes and forms throughout their life span (Coleman et al., 2021). The cell wall determines the shape of a plant cell, which is achieved through a combination of directed cell division and cell expansion. The form of a plant cell is essential to its function, and its shape can be maneuvered in response to biotic and abiotic signals (Brady et al., 2007; Yoshida et al., 2021). The cell wall of a plant is a dynamic, fibrillar nexus. After a plant cell has reached its ultimate shape and the primary cell wall has been established, the secondary cell wall forms during the development of woody tissues (Xie et al., 2018; Zhang et al., 2018). Cellulose is a microfibrillar polymer that is found in high concentrations in plant cell walls. Cellulose synthase (CESA) proteins, which dwell in the plasma membrane inside cellulose synthase complexes (CSCs), are responsible for the formation of these microfibrils (McFarlane et al., 2014; Gritsch et al., 2015). COBLs are found to be associated with cellulose crystallizations and decisively co-expressed with cellulose synthase complex (CSC) (Sangi et al., 2021; Li Z. et al., 2022). Cellulose synthases (CESAs), KORRIGAN (KOR), NAC transcription factor, chitinase-like genes (CTLs), fasciclin-like arabinogalactan genes (FLAs), and MYB transcription factors are only a few of the genes that have been shown to be crucial for cellulose biosynthesis during the past several decades (Liu et al., 2013; Yoshida et al., 2021). Despite these significant advances, the molecular process of cellulose production and deposition is still poorly understood. Veritably, it is still unclear how plants naturally organize the molecules in their cellulose framework. Little is known regarding the proteins and the mechanisms that regulate cellulose crystallinity in plants (Somerville, 2006; Niu et al., 2015); this is where COBL proteins come into play. But there has not been enough research done on COBRA-*Like* genes in tree species like poplar to draw any firm conclusions. COBRA belong to a plant-specific multigene (COB-like) family and encode glycosylphosphatidylinositol (GPI)-anchored proteins embodying the hydrophilic region at the C-terminal, typically having an ω-attachment site for GPI modification accompanying an N-terminal secretion signal (Liu et al., 2013; Yang et al., 2021). COB protein is localized in vesicles, in the Golgi apparatus, and lastly at the cell surface, conforming to a classic GPI secretion pathway. The microtubule arrangement determines where the COB protein is distributed (Sato et al., 2010; Zhang et al., 2016).

The COBL family members have been identified to intercede in several developmental and physiological processes, including stem strength, pollen tube growth, pathogen resistance, and root-hair growth (Ko et al., 2006; Li et al., 2013; Yang et al., 2021; Li et al., 2022). Many mutants with aberrant cell expansion and cellulose deposition have been investigated to learn about the molecular processes that govern the direction and process of cell expansion during plant development (Ko et al., 2006; Hochholdinger et al., 2008). In *Arabidopsis*, cobra-mutants like *cob-6* are unable to produce cellulose, which leads to defective growth (Sato-Izawa et al., 2020). A decrease in the frequency of mis-regulated genes and a suppression of symptoms are seen in *cob-6* mutants with mutations in MEDIATOR16 (MED16) (Sorek et al., 2015). This finding suggests a role for MED16 in the transcriptional response to a breach in the cell wall (Sorek et al., 2015). *Arabidopsis* COBL2 controls seed mucilage ray shaping, adhesion, and crystalline cellulose deposition into the cell walls. In contrast, *cobl2* triggers impaired cell wall assembly and morphology (Ben-Tov et al., 2018). COBL10 likely regulates pollen tube growth by the deposition of cellulose microfibrils and apical pectin (Li et al., 2013). *cobl4* mutants exhibit phenotypes with impaired secondary cell wall (SCW) formation (Liu et al., 2013). *AtCOBL9* is mandatory for root hair growth whereas *AtCOBL4* is found to be vital for SCW cellulose generation in *Arabidopsis* (Dai et al., 2009). In rice, *BC1* (*Brittle Culm1*) participation is required for cellulose biogenesis and crystallization. While *BC1* mutation causes enfeebled stem strength. In maize, *Brittle Stalk2* (BK2) mutation begets defective mechanical strength (Li et al., 2003; Julius et al., 2021). The epidermal cells of a tomato mutant with the *SlCOBRA-like* gene silenced lacked cuticles, with aberrant shapes and an erratic size distribution. Plants with loss-of-function COBL genes, such as *BRITTLE CULM1* in sorghum (*SbBC1*) and *TmBr1* in diploid wheat, induce the brittle

phenotype (Li et al., 2018). One study also reported that plant cells interpret the *COBRA* null mutation's effects on cell elongation as a hint of an impending pathogen assault and adapt their growth accordingly (Ko et al., 2006). The *cob-5* mutant has aberrant cell development across the whole plant body and agglomerates compounds that are released in reaction to stress, e.g., anthocyanins and callose. Further investigation revealed that *cob-5* plants increased JA concentration thus dramatically increasing the regulation of their cellular defense mechanisms (Ko et al., 2006; Gao et al., 2013). These findings provide credence to the hypothesis that modifications in cell wall integrity are closely linked to the control of plant defense mechanisms (Yoshida et al., 2021).

Poplar is a model plant in tree research and a valuable forest resource to study wood formation mechanisms and stress response (Liu et al., 2015). The COBL gene family modulates wood developing tissues (secondary xylem, phloem, xylem vessel, and fiber differentiation) involved in stem girth and stress response (Zaheer et al., 2021). In *Populus*, one study (Ye et al., 2009) found 18 COBRA genes, but an inclusive and systematically analyzed study is lacking. Recent and advanced *Populus* genome sequencing can greatly improve our understanding about the functions, evolution, gene distribution, and expansion of the COBL gene family in poplar (Tuskan et al., 2006; Zhang et al., 2022). Here, we present a comprehensive study on the COBL gene family in poplar. In the current study, chromosomal collinearity with different plant species and expression patterns under abiotic stress and in different tissues was evaluated. Furthermore, miRNA regulation and gene ontology (cellular, molecular, and biological processes) of COBL genes were investigated. PtrCOBL genes were identified regulating tension-wood formation and co-expression links among these genes were explored. The current study will better equip the researchers to use the key PtrCOBL genes in cutting-edge genome editing procedures that will assist their functional assessment and may result in superior genotypes with improved wood quality and stress tolerance capacity.

2 Materials and methods

2.1 Evolutionary relationships and search for COBLs

Poplar and *Arabidopsis* COBRA-*Like* family protein sequences were replevied from the Phytozome13 and TAIR online databases, respectively (https://phytozome.jgi.doe.gov/pz/portal.html and https://www.arabidopsis.org/ (Accessed on 01 Sep 2022)) (Xu et al., 2017). After retrieving the COBL domain (PF04833) from Pfam (http://pfam.xfam.org/ (Retrieved on 01 Sep 2022)), we used it as a query to search other plant species protein databases using HMMER 3.0 to identify the genes that contained this domain (Potter et al., 2018). The predicted genes were verified using the SMART database (http://smart.embl-heidelberg.de/) and the NCBI CD search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to look for conserved domains and motifs. To determine the COBL proteins' chemical and physical properties, the ExPASy database (http://web.expasy.org/protparam/) was employed. For protein sequence alignment, ClustalW (Larkin et al., 2007) and DNAMAN (http://www.lynnon.com/) software were used. For the phylogenetic tree construction, muscle alignment was utilized coupled with MEGA-X (http://www.megasoftware.net/mega-x/) and its neighbor-joining (NJ) approach. The numbers at the tree branches indicate the percentage (%) of 1,000 bootstrap values (Kumar et al., 2018).

2.2 Gene structure prediction, domain, and motif analyses

The genomic sequences of poplar were obtained from the Phytozome13 database. To examine the COBL family gene-structure (exon, intron, and UTR), we utilized the web-based GSDS program (http://gsds.cbi.pku.edu.cn/). The conserved motifs were analyzed with default settings using the online freeware MEME (http://meme-suite.org/tools/meme), and the motifs were visualized using the TBtools program (Chen et al., 2020). The Conserved Domains Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) was used to identify COBLs conserved domains. The TBtools program was used to draw the domain structure with default parameters (Chen et al., 2020).

2.3 PtrCOBLs mapping on chromosomes and gene duplication

The Phytozome and PopGenIE (http://popgenie.org/chromosome-diagram) databases were mined for chromosomal location data of PtrCOBLs (Leng et al., 2021). Every single gene was mapped on the representing chromosome based on its position report. Whole genome duplication (WGD) occurrences were analyzed using MCScanX. The synteny and collinearity analyses were visualized between poplar and five other plant species (*Arabidopsis*, *eucalyptus*, cotton, *T. cacao*, and *V. vinifera*) using TBtools with default settings (Chen et al., 2020; Yang et al., 2021).

2.4 Analysis of cis-regulatory elements and sub-cellular localization

To carry out promotor analysis, 2KB upstream sequences (from the start codon) of COBL genes were submitted to online freeware, Plant-CARE (Lescot et al., 2002) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The subcellular localization of COBL genes was predicted using the CELLO web server (http://cello.life.nctu.edu.tw/).

2.5 Putative miRNA targeting COBLs and analysis of their functions

The psRNATarget website was used to estimate miRNA target sites. The CDS sequences of all the COBL genes were submitted to

the website (https://www.zhaolab.org/psRNATarget/) under default parameters. Moreover, these miRNAs and target sites were verified with another tool (Raza et al., 2022). The interactive network of miRNAs with their sites was developed utilizing Cytoscape software (v3.9). All of the PtrCOBL protein sequences were presented to the eggNOG v4.0 for review of gene ontology (GO) and KEGG (Kyoto encyclopedia of genes and genomics) annotations (Bano et al., 2021; Raza et al., 2022).

2.6 Gene co-expression network analysis

The PtrCOBL gene expression was measured in terms of fragments per kilobase of exon per million mapped (FPKM). Differentially expressed COBL genes were studied in vegetative organs, vasculature (wood-making tissues), and in response to abiotic stresses using transcriptome data (NCBI (ID: GSE153793) ENA (ID: ERP016242)) from NCBI (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/) (Raza et al., 2022). The heat map was produced using TBtool software and the FPKM data were used to conduct co-expression network analysis. Cytoscape was used to generate a co-expression network out of the collected data (Shannon et al., 2003; Chen et al., 2020).

2.7 Bioinformatic analysis of potential distinct features of COBLs

The PlantRegMap online database (http://plantregmap.gao-lab.org/) (Raza et al., 2022) was employed to procure GO information. Additionally, GPI modification motif (ω site for dissociation), the signal peptide, and hydrophobic plot were envisaged via the SignalP 4.1 Server (http://genome.cbs.dtu.dk/services/SignalP/), KYTE DOOLITTLE HYDROPATHY PLOT (http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm), and the big-PI predictor (http://mendel.imp.ac.at/sat/gpi/gpi_server.html), correspondingly (Roudier et al., 2002; Niu et al., 2015).

2.8 RNA-seq datasets to explore the expression arrays of PtrCOBLs

Tissue preferential expression data of PtrCOBL genes in five different tissues (root, node, internode, mature expanding leaf, and young leaf) were adopted from PopGenIE (http://popgenie.org) and used to develop optical images (Sundell et al., 2015; Xu et al., 2017).

For further evaluation of the expression profiles of PtrCOBLs across abiotic stresses (cold, salt, drought, and heat), we selected previously reported transcriptome expression data of four stem samples, from PSDX (Poplar stem differentiating xylem) database (http://forestry.fafu.edu.cn/db/SDX) (Wang H. et al., 2021). For heat stress, plants were treated at 39°C for 7 days. For cold stress, plants were subjected to 4°C (night) and 12°C (day) for 7 days. For drought treatment, watering was withheld until the moisture of soil reached 0.1 m3/m3 and was maintained at the level of 0.06 – 0.1 m3/m3. For drought stress, plants were grown for 7 days after the initial leaf wilting point. For salinity stress, plants were treated with 100 mM sodium chloride solution for 7 days.

For expression analysis of PtrCOBLs in tension-wood formation, a publicly available transcriptome dataset was downloaded from NCBI (ID: GSE153793). To induce tension wood formation, 6-month old poplar plants were subjected to 90^0 angle mechanical bending (vertical treatment), then tissues from stem differentiating xylem (SDX) were taken at 3 and 7 days post-treatment (Liu et al., 2021).

To investigate the PtrCOBL genes expression assay all through cambium growth and wood development (Sundell et al., 2017), transcriptome data with high spatial-resolution for the whole poplar plant was employed, including the cambium, secondary phloem, and xylem, available at ENA (ID: ERP016242) (http://www.ebi.ac.uk/ena/). Each replication tree has 25-28 samples collected from 15-mm-thick sections, underlying sections were taken across the cambial meristem and SCW developing xylem (http://aspwood.popgenie.org).

2.9 Plant materials, RNA extraction, and qRT-PCR analysis

Poplar (*Populus trichocarpa*) plants were endowed by the Zhejiang A & F University. Poplar seedlings (6 months old) were grown in the growth room at 16.0 h light, 20-25°C, and 70% air humidity. To examine expression patterns across different tissues, the plant samples were taken directly without treatment. The samples were snap-frozen in liquid nitrogen and kept at -80°C until additional analysis could be performed. In order to get consistent findings, all the experiments were conducted with three biological replicates, and three separate sets of samples were used in each experiment.

Plant total RNA was isolated out of roots, stems, and leaves using the RNAprep Pure Kit from TIANGEN Biotech (Beijing, China) following the given methodology. The PrimeScriptTM RT Reagent Kit (Takara Bio, Dalian, China) was used to isolate the first strand of cDNA after removal of genomic DNA. The qRT-PCR machine, ABI 7500 Real Time PCR System (Applied Biosystems, Foster City, CA, USA), was run using UltraSYBR Mixture (Low ROX) (CWBIO, Beijing, China). Each amplification reaction included 10.0 μL of 2x UltraSYBR Mixture (with ROX), 4.0 ng of cDNA template, and 0.2 μM of forward and reverse primers in a total volume of 20 μL. The qRT-PCR primers listed in (Supplementary Table S1) were all created using Primer Premier 5.0 (Premier Biosoft, Palo Alto, CA, USA) following these parameters for the reaction: 95°C for 2 min followed by 40 cycles of 95°C for 15 s and 60°C for 20 s. The $2^{-\Delta\Delta CT}$ method was used to analyze the expression levels of PtrCOBL genes in different tissues. The relative expression variations of PtrCOBL genes were analyzed by using the PtrACTIN reference gene (Supplementary Table S1). The t-test was used to determine the statistically significant variations in the expression levels. Three separate experiments were conducted, and the results are shown as mean standard errors (p value: * p 0.05, ** p 0.01, *** p 0.001). Each sample was evaluated with three biological replicates.

3 Results

3.1 Identification of COBL genes and sequence analysis in poplar

In order to study the roles and physiogenomics of COBLs in poplar, an *in silico* study was carried out to identify the COBL-gene family members. The Basic Local Alignment Search Tool algorithms (BLASTp) and Hidden Markov Model (HMM) profile search were employed to retrieve the COBL members. Subsequently, these candidate COBL genes were further confirmed by SMART and NCBI-CDD tests, to confirm the genes possessing the (PF04833) domain. A total of 14 putative COBL genes were identified in the poplar genome v4.1 (*P. trichocarpa*) and were named as PtrCOBL1-PtrCOBL14 on the basis of their chromosomal locations. To clearly understand the characteristics of the COBL family in poplar, we analyzed the gene length, transcriptional sequence length, coding sequence (CDS) length, the position of the conserved domain, amino acid (AA) length, protein molecular weight (MW), grand average of hydropathicity (GRAVY), and isoelectric point (PI) of the proteins encoded by these genes (Table 1).

Out of these 14 PtrCOBL genes, PtrCOBL13 was the shortest with 412 amino acids (AA) followed by PtrCOBL14 and PtrCOBL5 with 414 and 431 AA, respectively. PtrCOBL1 was the longest with 664 amino acids. The MW of PtrCOBL proteins was 22.200 to 68.678 kDa, the GRAVY of the proteins was – 46.052 (PtrCOBL13) to – 74.164 (PtrCOBL7), and PI was 5.24 (PtrCOBL6) to 9.49 (PtrCOBL13). Generally, identifying the molecular characteristics of genes is helpful in studying their specific biological roles.

Based on the presence of the Pfam domain and similarity with query sequences, protein sequence alignment was carried out (Figure 1). Group I (similar to AtCOB) of *Populus* displayed the most conserved manner in their domains, while group II (similar to AtCOBL7) of *Populus* was the least conserved. Considering the number of *Populus* COBL genes in each group, group I was more conserved than group II. Poplar COBL genes overwhelmingly incorporated the CCV motif, N-terminal signal peptide, and ω site (GPI attachment cleavage site). PtrCOBL1 and PtrCOBL8 exceptionally showed the lack of an N-terminal signal peptide, while PtrCOBL1, PtrCOBL10, PtrCOBL13, and PtrCOBL14 were devoid of the ω site (Figure 1; Table 1).

TABLE 1 Summary of COBL genes in Poplar.

Gene Name	Chr.	Strand	CDS (bp)	Protein Length	Mol. weight (KDa) Protein	pI	GRAVY	Cellular Localization	Potential ω-Site Position	CCVS Motif	Signal Peptide
PtrCOBL1	1	forward	1995	664	73.8821	8.72	-0.272	Extra-cellular	None	Yes	No
PtrCOBL2	4	forward	1362	453	50.74724	9.11	-0.145	Plasma Membrane	N(426)	Yes	Yes
PtrCOBL3	4	Reverse	1302	433	48.80238	8.92	-0.085	Plasma Membrane	S(411)	Yes	Yes
PtrCOBL4	4	forward	1335	444	49.66873	6.07	-0.113	Plasma Membrane	G(417)	Yes	Yes
PtrCOBL5	4	Reverse	1296	431	48.80097	8.22	-0.262	Lysosome	A(413)	Yes	Yes
PtrCOBL6	10	Reverse	1953	650	70.93764	5.24	-0.127	Extra-cellular	N(426)	Yes	Yes
PtrCOBL7	11	forward	1992	663	74.16447	8.87	-0.322	Extra-Cellular	S(634)	Yes	Yes
PtrCOBL8	12	Reverse	1317	438	48.73394	8.42	-0.049	Plasma Membrane	N(410)	Yes	No
PtrCOBL9	14	forward	1923	640	70.96131	5.28	-0.253	Plasma Membrane	S(618)	Yes	Yes
PtrCOBL10	15	forward	1374	457	50.98067	8.83	-0.107	Plasma Membrane	none	Yes	Yes
PtrCOBL11	15	Reverse	1299	432	48.40889	9	-0.185	Lysosome	N(408)	Yes	Yes
PtrCOBL12	15	Reverse	1320	439	49.43289	8.77	-0.117	Plasma Membrane	S(415)	Yes	Yes
PtrCOBL13	17	forward	1239	412	46.05201	9.49	-0.263	Extra-cellular	none	Yes	Yes
PtrCOBL14	17	forward	1245	414	46.14007	9.11	-0.206	Extra-cellular	none	Yes	Yes

10.3389/fpls.2023.1242836

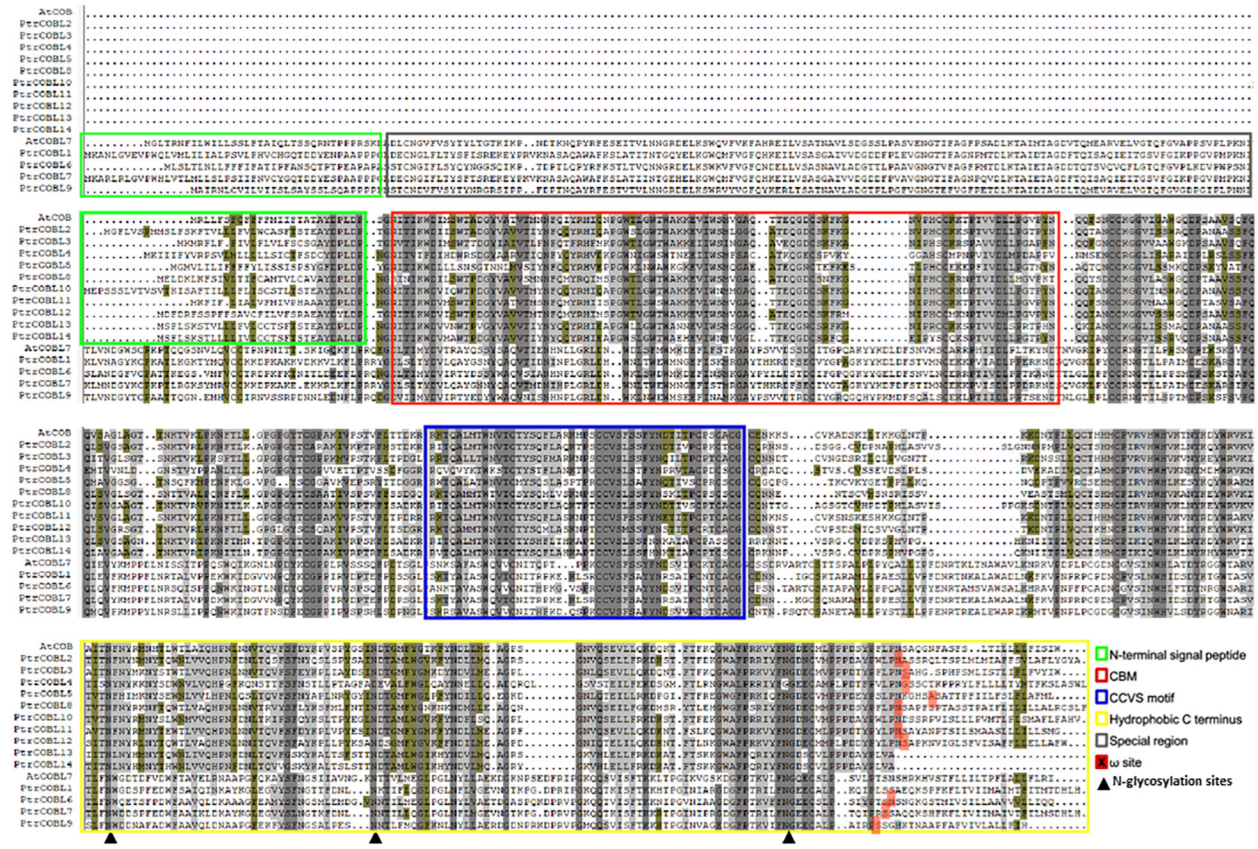


FIGURE 1
Multiple sequence alignment of COBL proteins. AtCOB-like group I proteins and AtCOBL7 group II proteins were aligned. The green boxes represent the N-terminal signal peptide, the red box denotes the cellulose binding motif, the blue box represents the CCVS-motif, the yellow box signifies the hydrophobic C-terminus, the gray box represents the special region found in group I, the solid red filled boxes epitomize ω sites (GPI attachment cleavage sites), and the triangle sign represents N-glycosylation sites in the sequence.

3.2 Evolutionary and phylogenetic analyses of COBLs

To assess the evolutionary relationships among different plant species, a phylogenetic tree was constructed. For this purpose, full length protein sequences of *Arabidopsis*, poplar, rice, *Vitis vinifera*, *Theobroma cacao*, and cotton were retrieved. Using the multiple sequences alignment with the MEGA-X muscle program, a neighbor-joining tree was constructed (Figure 2). In this phylogenetic tree, protein sequences of 12 (*Arabidopsis thaliana*), 14 (*Oryza sativa*), 14 (*P. trichocarpa*), 15 (*Theobroma cacao*), 33 (*Gossypium hirsutum*), and 10 (*Vitis vinifera*) were used. Based on comparison with model plant *Arabidopsis* and analysis of COBL proteins' diversity, these proteins were categorized into two subgroups, i.e., an AtCOB group and an AtCOBL7 group, designated as group 1 and group 2, respectively. Consistent with orthologs in *A. thaliana* and other species, the COBL family members were clustered into two subgroups GroupI and GroupII, phylogenetically related to AtCOBRA and AtCOBL7 in *A. thaliana*, respectively. Each subgroup had COBLs from six species, and a total of 6 genes from *Arabidopsis*, 4 genes from poplar, 12 genes from cotton, 3 genes from rice, 5 genes from *T. cacao*, and 4 genes from *V. vinifera* were found in group 1, whereas a total of 6 genes from *Arabidopsis*, 10 genes from poplar, 21 genes from cotton, 8 genes from rice, 11 genes from *T. cacao*, and 6 genes from *V. vinifera* were found in group 2. The phylogenetic analysis indicated that the COBLs were descendants of an ancient duplication that occurred even before the separation of monocots and dicots. By the high bootstrap values of the internal branches, it was possible to deduce that there are real homologs with likely similar functions.

3.3 Gene structure, conserved domain, and motifs of PtrCOBLs

Gene structure and motif diversity is a mechanism that promotes the evolution of gene families. To structurally characterize PtrCOBLs, gene structure, conserved motif, and domain analyses were performed (Figures 3A–D). The total number of exons ranged from two to six in these genes, and this range varied between groups. As described earlier, COBL genes can be distinguished into two groups. The gene structure analysis showed that, despite being different gene lengths and exon positions, genes in the same group possess the same number of exons. This indicates the number of exons among the majority of PtrCOBL genes are highly conserved (Figure 3D). About 72% of genes occupying group II possess six

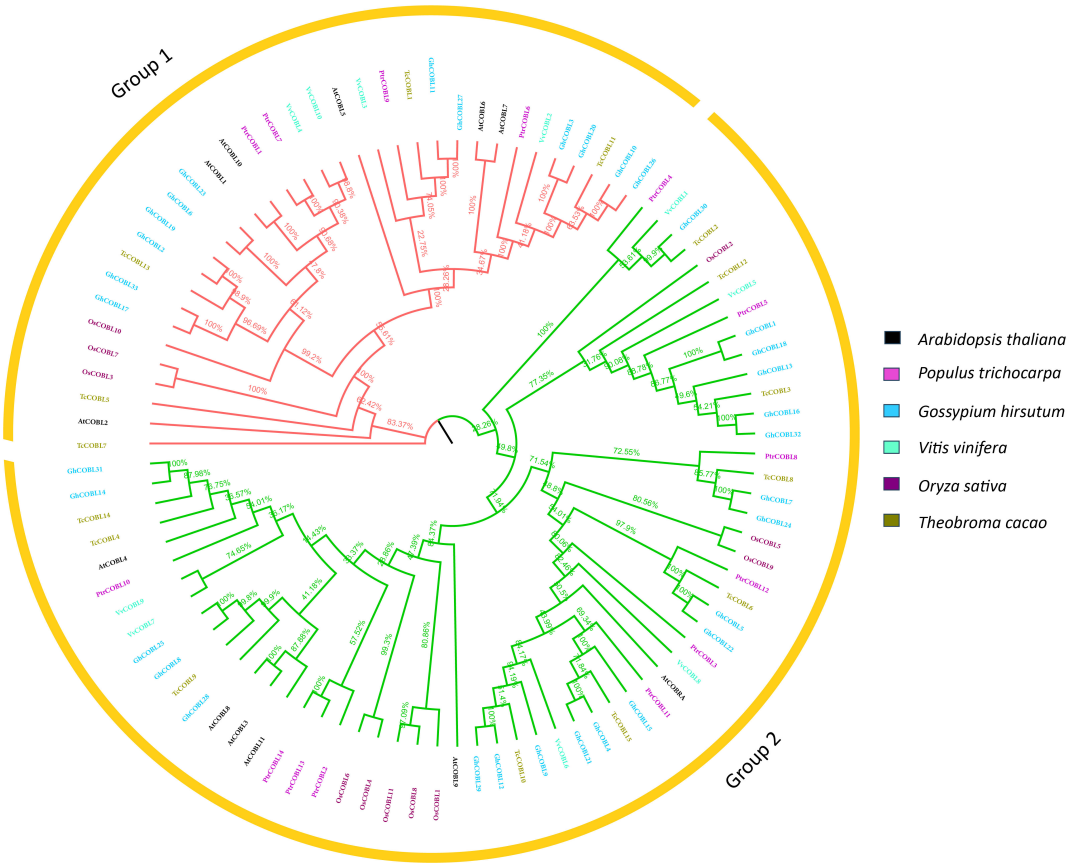


FIGURE 2
Phylogenetic analysis. The evolutionary relationship among six plant species, *Arabidopsis*, poplar, cotton, grape, rice, and cocoa. Muscle alignment was used to construct the unrooted, neighbor-joining tree.

exons, the largest number of exons, whereas PtrCOBL1, PtrCOBL6, PtrCOBL7, and PtrCOBL9 contain two exons, representing group I. This organization and structural conservation of genes supports the results of phylogenetic analysis (Figure 3A).

To identify motif composition and motif diversity among PtrCOBLs, the entire protein sequences of all the genes were submitted to the online MEME software. The average length of proteins appeared to be larger in group II in comparison with group I. Overall, 10 motifs were assigned and designated as motif 1-10 (Figure 3B). Interestingly, all the proteins contain the same number of eight motifs except PtrCOBL4. Motifs 9 and 10 are only found in PtrCOBL1, PtrCOBL2, PtrCOBL3, and PtrCOBL4, but motif 3 is missing in all these genes. Motif 6 is missing in PtrCOBL1, PtrCOBL2, and PtrCOBL4 but motifs 3 and 6 are only found in PtrCOBL 5 to 14. Notably, motifs 1, 4, 5, 7, and 8 were found in all the proteins, indicating these motifs are the essential components of PtrCOBL proteins. Domain verification tools indicated all the COBRA-*Like* genes of poplar contain the COBRA domain (Figure 3C).

3.4 Chromosomal localization

Through genome annotation, the COBL family genes were mapped to poplar chromosomes. Uneven distribution of genes on 8 chromosomes out of 19 chromosomes was observed in the COBL genes of poplar (Supplementary Figure 1). No gene was residing on the 2nd, 3rd, 5th, 6th, 7th, 8th, 9th, 13th, 16th, 18th, and 19th number chromosomes. All the 14 COBL genes were found asymmetrically distributed on just eight (1st, 4th, 10th, 11th, 12th, 14th, 15th, 17th) chromosomes. Only one gene was allocated to chromosomes 1, 10, 11, 12, and 14, while two genes were allocated to chromosome 2, followed by chromosomes 15 and 4 containing three and four genes, respectively. The PtrCOBL genes were sprinkled around different positions on the chromosomes: PtrCOBL1 was located from (446 88589.44693374) on Chrom-1, PtrCOBL2 was located from (110 02597.11006405) on chrom-4, PtrCOBL3 was located from (11 006313.11008482) on chrom-4, PtrCOBL4 was located from (1854 7098.18551009) on chrom-4, PtrCOBL5 was located from (22 495396.22499058) on chrom-4, PtrCOBL6 was located from (120146.123684) on chrom-10, and PtrCOBL7 was located from (16 766946.16771575) on chrom-11. Similarly, PtrCOBL8 was located from (6499312.6501945) on chrom-12, PtrCOBL9 was located from (8418045.8421086) on chrom-14, PtrCOBL10 was located from (8 383270.8386458) on chrom-15, PtrCOBL11 was located from (8386506.8389407) on chrom-15, PtrCOBL12 was located from (8390106.8393117) on chrom-15, PtrCOBL13 was located from (11038917.11042151) on chrom-17, and PtrCOBL14 was located from (11045798.11049523) on chrom-17.

10.3389/fpls.2023.1242836

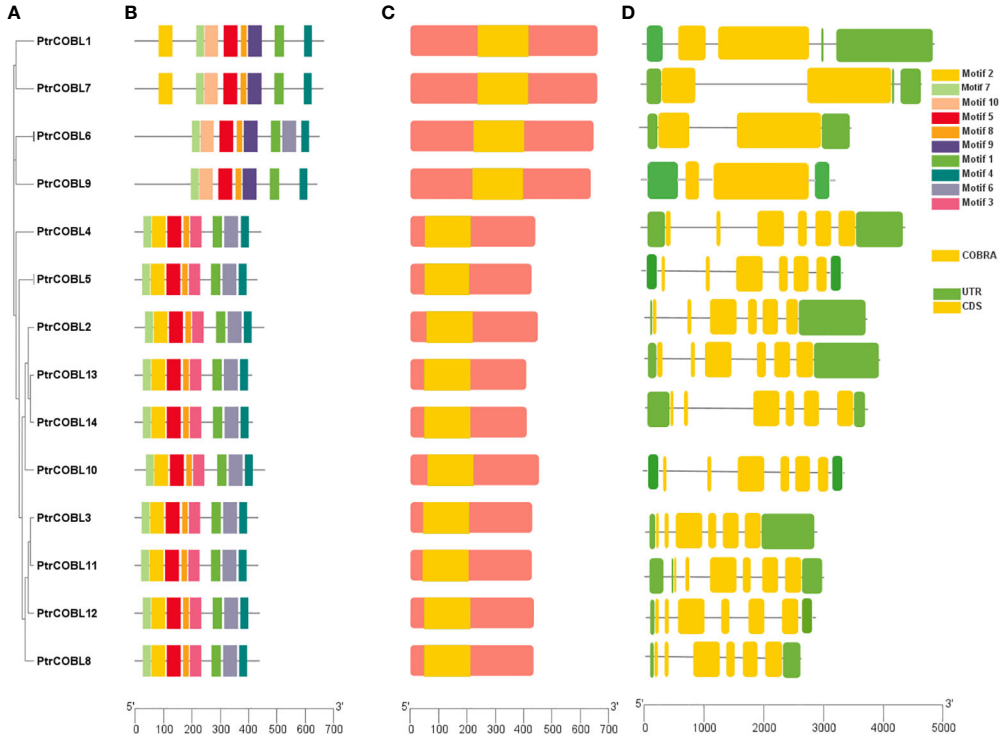


FIGURE 3
Motif and gene structure characterization. **(A)** Phylogenetic tree, **(B)** protein conserved motifs, **(C)** conserved domain analysis, with yellow indicating the *COBRA* domain, and **(D)** gene structure.

The majority of PtrCOBL genes were located on chromosomes 4 and 15, which together harbor 50% of all the COBRA-*Like* genes from the *Populus trichocarpa* genome (four and three genes, respectively). Interestingly, we found that there were four PtrCOBL genes on chromosome 4 of the *Populus* linkage group (PtrCOBL2 to PtrCOBL5). In the same manner, three COBLs were on chromosome 15 (PtrCOBL10 to PtrCOBL12) of the linkage group, forming a relatively large tandem duplication region; it thus seemed that chromosomes 4 and 15 were a hot spot for studying the distribution of PtrCOBL genes. In addition, there were five fragment repeats located at the ends or near the ends of chromosomes 1, 10, 11, 12, and 14, respectively (Supplementary Figure 1).

3.5 Collinearity and synteny analysis among plant species

Circos analysis was employed to identify duplication events within the poplar genome. Three sorts of events were observed in plant species gene duplication, including tandem (a duplicated chromosomal region within 200 kb containing two or more genes and occurring adjacent to the original), segmental (genes located on duplicated chromosomal blocks 1 kb to 400 kb that occur at multiple locations throughout the genome), and whole genome duplication (the process which generates additional copies of the genome). The gene duplication events may influence the gene function diversity among plant species. In total, six paralog pairs were observed, where PtrCOBL1 was a paralog pair of PtrCOBL7 and PtrCOBL12 was a paralog pair of PtrCOBL3. A third paralog pair comprised PtrCOBL4 and PtrCOBL5. Likewise, PtrCOBL6 and PtrCOBL9 formed a fourth paralog pair, and the fifth represented PtrCOBL8 and PtrCOBL10. The sixth paralog pair was PtrCOBL13 and PtrCOBL14 (Figure 4; Supplementary Figure 1).

Tandem duplication events were observed in two paralogous gene pairs, i.e., PtrCOBL4/PtrCOBL5 and PtrCOBL13/PtrCOBL14, whereas a series of segmental duplication events were observed in this study that likely increased the number of COBL genes in poplar as compared to *Arabidopsis*. PtrCOBL1 and PtrCOBL7 experienced segmental duplication. Similarly, PtrCOBL6 and PtrCOBL9 were also tandemly duplicated, as were PtrCOBL8 and PtrCOBL10. Furthermore, PtrCOBL3/PtrCOBL12 also indicated segmental duplication event occurrence (Figure 4). In order to further study gene duplication, we constructed five PtrCOBL comparative synteny maps (Figure 5). Among these maps, the strength of correlation with PtrCOBL genes in ascending order was given as: *Arabidopsis* (17), T. *cacao* (15), grape (14), *Eucalyptus* (15), and cotton (18). Comparative collinear mapping between monocotyledons and dicotyledons depicted that the tandem duplication events and the segmental duplication events may not only be the primary mechanism of gene family expansion in plant evolution but may also make a significant contribution to the diversity of gene families.

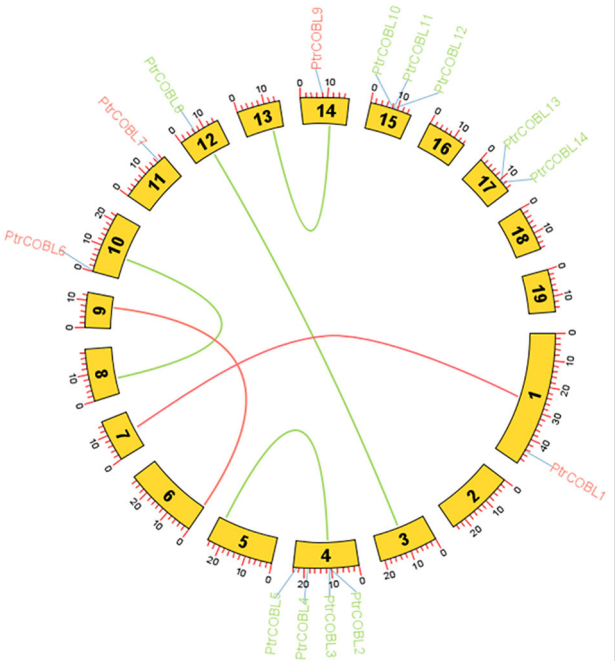


FIGURE 4
Fragment duplication of the poplar COBL genes. The outer circle blocks represent the chromosome numbers and gene density on the chromosome; the red and green lines represent the fragment duplicate gene pair. The length of each arc corresponds to the length of the chromosome (Mb).

3.6 Cis-acting elements of promotor and GO annotation

The studies on the non-coding regions of genes are becoming of prime importance. Cis-acting regulatory elements (CRE) in the promoter region of the genes influence transcription and thus regulate the functions of the genes (Figure 6A). The upstream region (of 2000 bp) from the site of initiation (ATG) was uploaded to the PlantCare database. *In silico* promotor analysis resulted in the identification of 21 CRE sites in the COBL genes; along with common CREs there were plenty of cis-elements identified related to low temperature response (LTR), MeJA responsiveness, ABA responsiveness, Salicylic acid responsiveness, Gibberellin responsive, Auxin responsive, MYB binding site, MYB drought inducibility, MYBHv1 binding site, AT rich DNA binding protein (ATBP-1), anoxic inducibility, anaerobic induction, stress responsive, and elicitor-mediated activation and defense. Importantly, regulatory motifs for meristem-expression, wound responsiveness, seed specific regulation, root specific regulation, endosperm expression, circadian control, and zein metabolism regulation were also identified (Supplementary Table S2). Out of all these CREs, ABA and auxin-related regulatory motifs were abundantly found in the COBL genes of poplar. Conjointly, the results indicated that the functional expression of PtrCOBL genes is controlled by cis-regulatory elements linked to plant growth and development, plant stress response, and hormonal response.

To brainstorm functions of COBL family genes, gene ontology (GO) annotation and enrichment analysis was enacted on the basis of their biological processes (BP), molecular functions (MF), and cellular component (CC) sets. The results revealed that several GO terms were enriched in the BP, MF, and CC (Figure 6B). The 14 COBL proteins were assigned with 48 GO terms, where BP terms were extensively enriched followed by MF and CC. Among BP enrichment, analysis suggested that 25 terms were highly enriched: cellulose microfibril organization (GO:0010215), cell growth (GO:0016049), plant-type cell wall organization or biogenesis (GO:0071669), cellular component organization or biogenesis (GO:0071840), extracellular matrix organization (GO:0030198), extracellular structure organization (GO:0043062), cellular component biogenesis (GO:0044085), cellular component assembly (GO:0022607), external encapsulating structure organization (GO:0045229), and single-organism cellular process (GO:0044763) were abundant. Moreover, vegetative to reproductive phase transition of meristem (GO:0010228), response to temperature stimulus (GO:0009266), response to osmotic stress (GO:0006970), and response to abiotic stimulus (GO:0009628) were also among the annotated terms. In the MP enrichment, polysaccharide binding (GO:0030247), chaperone binding (GO:0051087), carbohydrate binding (GO:0030246), and hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553) were enriched. In the CC set, anchored component of membrane (GO:0031225), intrinsic component of membrane (GO:0031224), membrane (GO:0016020), membrane part (GO:0044425), and cell junction (GO:0030054) were enriched. This GO analysis indicated that COBL genes are extensively involved in plant growth and development related functions at BP, MF, and cellular levels.

3.7 Tissue-specific and qRT-PCR expression profiles of COBL genes

To investigate the potential functions of the PtrCOBL genes in the developmental processes of *P. trichocarpa*, the tissue-specific expression pattern of PtrCOBLs was analyzed (Figures 7A–N). The visual representations of plant tissues were created using the Plant Genome Integrative Explorer. Herein, PtrCOBL1, 4, 5, and 7 were highly expressed in young leaves (YL) followed by PtrCOBL13 and 14, where gene expression was abundant in YL. PtrCOBL3, 6, 10, and 11 were found highly expressed in stems (S) followed by PtrCOBL2, where gene expression was abundant in stem (node and internode) zones. Similarly, PtrCOBL5, 8, 9, 12, 13, and 14 followed by PtrCOBL7 were found to be highly expressed in mature leaves (ML). PtrCOBL2, 9, and 12 were discovered to be predominantly active in the root area.

Then, RT-qPCR was used to validate the accuracy of microarray data results. Overall, microarray data results were corroborated by RT-qPCR, but there were some inconsistent expression patterns found in the roots, stems, and leaves. It is noteworthy that PtrCOBL1, 7, 9, and 12 in stem tissues and PtrCOBL8 in YL showed slightly different expression as compared to RT-qPCR results. PtrCOBL1, 7, 9, and 12 demonstrated more expression in stem area, whereas PtrCOBL8 expressed more in YL as compared to the microarray data, which may be due to variations in their testing

10.3389/fpls.2023.1242836

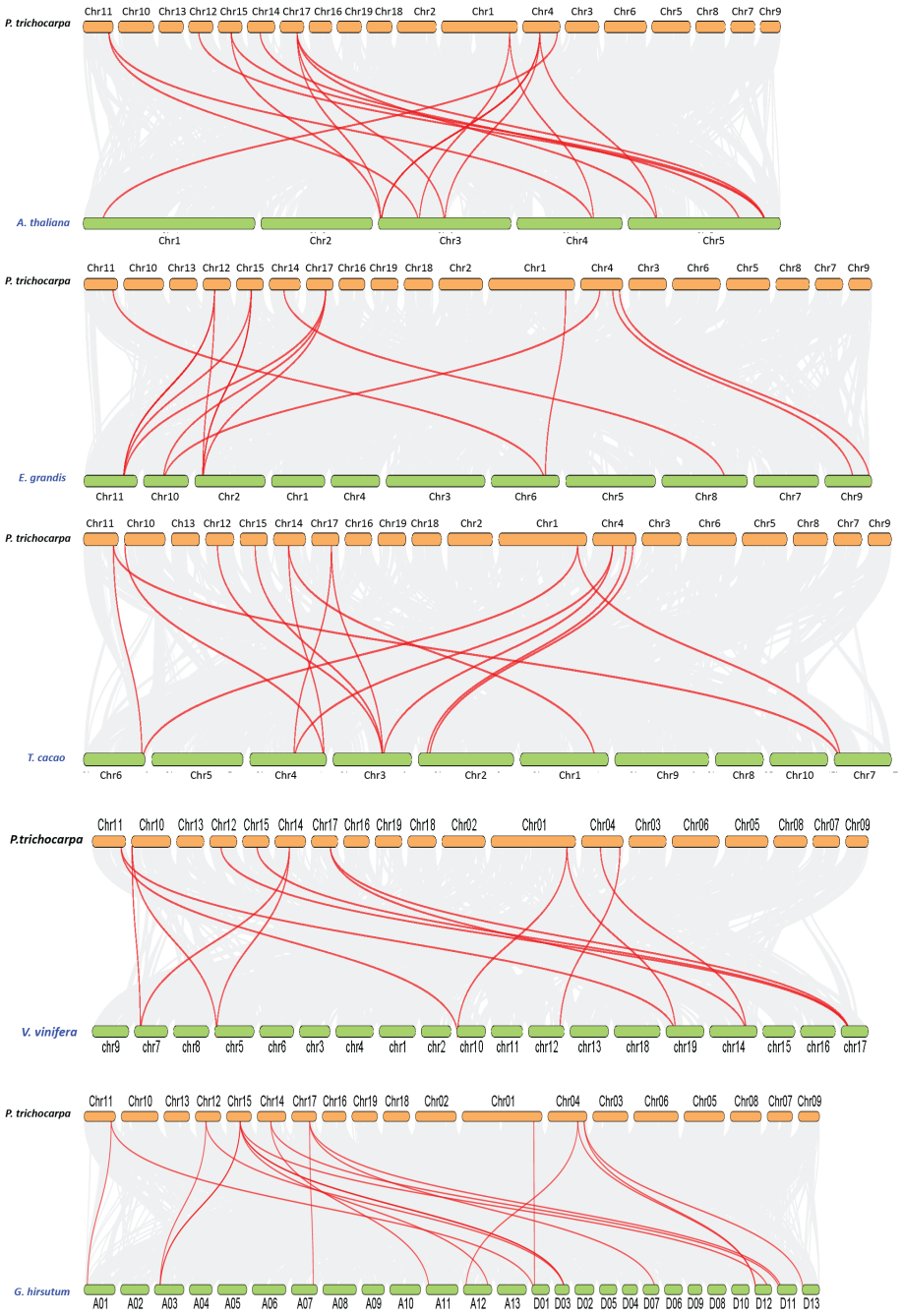


FIGURE 5
Collinearity of the COBL genes from poplar and other five plant species. Gray lines in the background indicate the collinear blocks within poplar and other plant genomes, while the orthologs between two species are connected with red lines.

materials and conditions (e.g., environmental factors and sample gathering intervals).

3.8 Whole-genome mining of miRNAs directing COBL genes

Numerous studies have reported a significant impact of microRNAs on fine-tuning plant growth and development. To explore the miRNA-targeted post-transcriptional modification of COBL genes, 29 putative miRNAs (belonging to 20 different families) were identified regulating 11 genes (Figure 8; Supplementary Figure 2). To provide insight into miRNA regulation, (Figure 8) depicts the target sites on PtrCOBL7, PtrCOBL8, PtrCOBL9, and PtrCOBL14. Detailed report of all the miRNA associated genes and their targeting sites are presented in Supplementary Figure 2. Global analysis showed that PtrCOBL8 was targeted by seven miRNAs (Ptr-miR159a, b, c; Ptr-miR319a, b, c; Ptr-miR6479). In the same

10.3389/fpls.2023.1242836

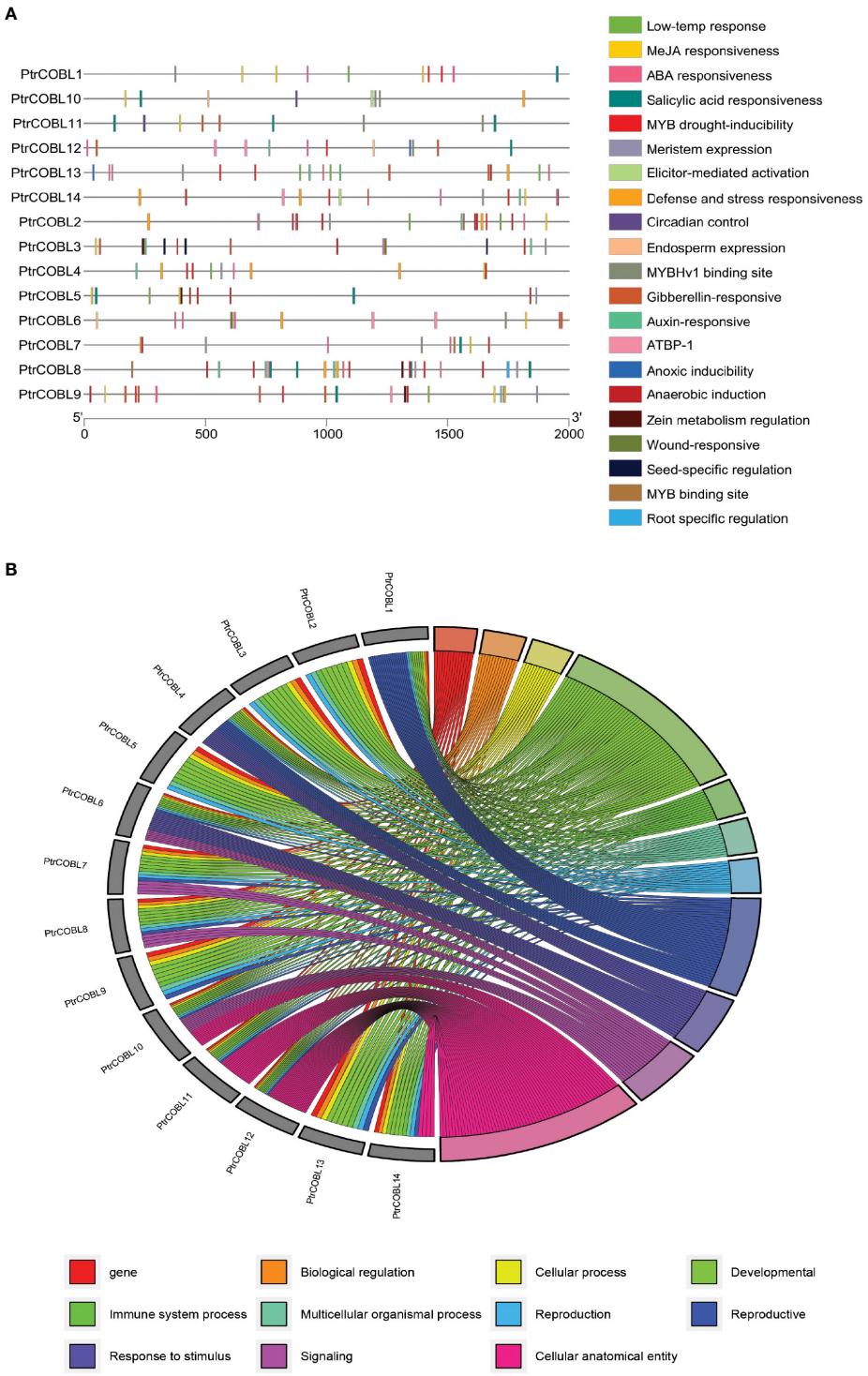


FIGURE 6
Upstream cis-regulator elements and gene ontology analysis: **(A)** Predicted cis-elements in COBLs promoters. Diverse colors were used for representing different cis-elements, as given on the right side. **(B)** Gene ontology annotation of COBLs in poplar; different colors represent different biological, molecular, and cellular processes.

manner, PtrCOBL9 was targeted by Ptr-miR6431, Ptr-miR2825, and Ptr-miR394. PtrCOBL4 was targeted by Ptr-miR7816, Ptr-miR2840, and Ptr-miR1450. PtrCOBL1 was found to be directed by Ptr-miR159, Ptr-miR399, and Ptr-miR7826. PtrCOBL12 was targeted by Ptr-miR156 and Ptr-miR6462; PtrCOBL7 was targeted by Ptr-miR2820 and Ptr-miR7816. Interestingly, Ptr-miR159 targeted the most, nine sites of three different genes, i.e., PtrCOBL1, 8, and 10, whereas PtrCOBL5 was targeted by Ptr-

10.3389/fpls.2023.1242836

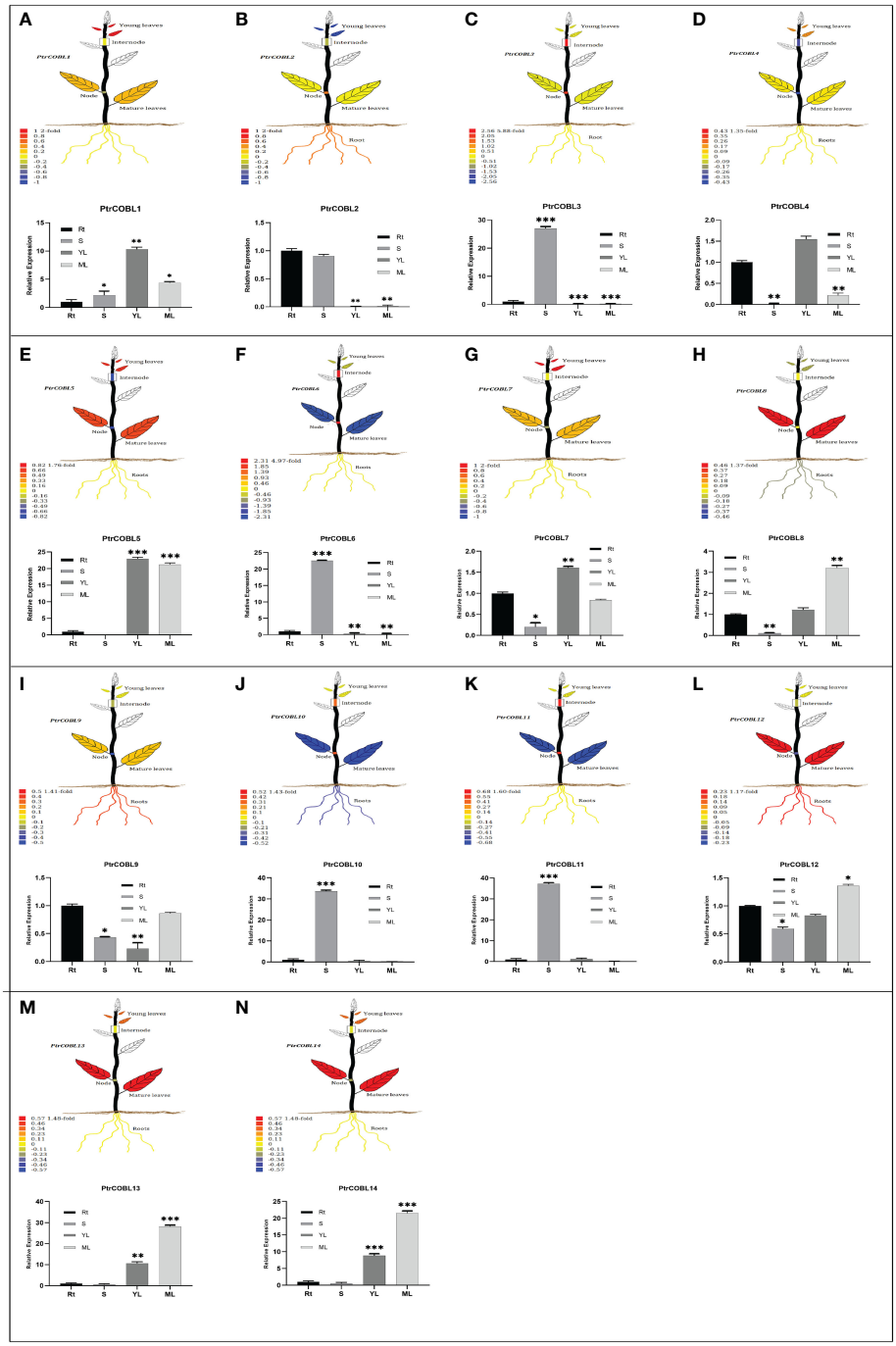


FIGURE 7
Tissue-specific expression profiles of PtrCOBLs. The PopGenIE v3 database (accession number: GSE6422) was utilized to generate visual plant images. Different colors of the plant organs indicate the expression abundance of the respective genes **(A-N)**. The bar graphs were generated employing RT-qPCR to verify relative expressions. The data represent the means ± SE from three independent experiments, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

miR530 and Ptr-miR2829, while PtrCOBL6 was targeted by Ptr-miR2838 and Ptr-miR3627. Ptr-miR7816 targeted two different genes, PtrCOBL4 and PtrCOBL7. Furthermore, PtrCOBL3 and PtrCOBL14 were found to be targeted by only one miRNA. Therefore, validation of the expression profiling of these predicted miRNAs and their target genes is required to evaluate their biological functions in the poplar genome.

3.9 Expression profiling and co-expression network analysis for abiotic stress

The PtrCOBL genes were surveyed against multiple abiotic stimuli to investigate their ability to withstand stress. Employing FPKM values (publicly available RNAseq data), expression patterns of COBL genes toward the endurance of abiotic stresses in poplar

10.3389/fpls.2023.1242836

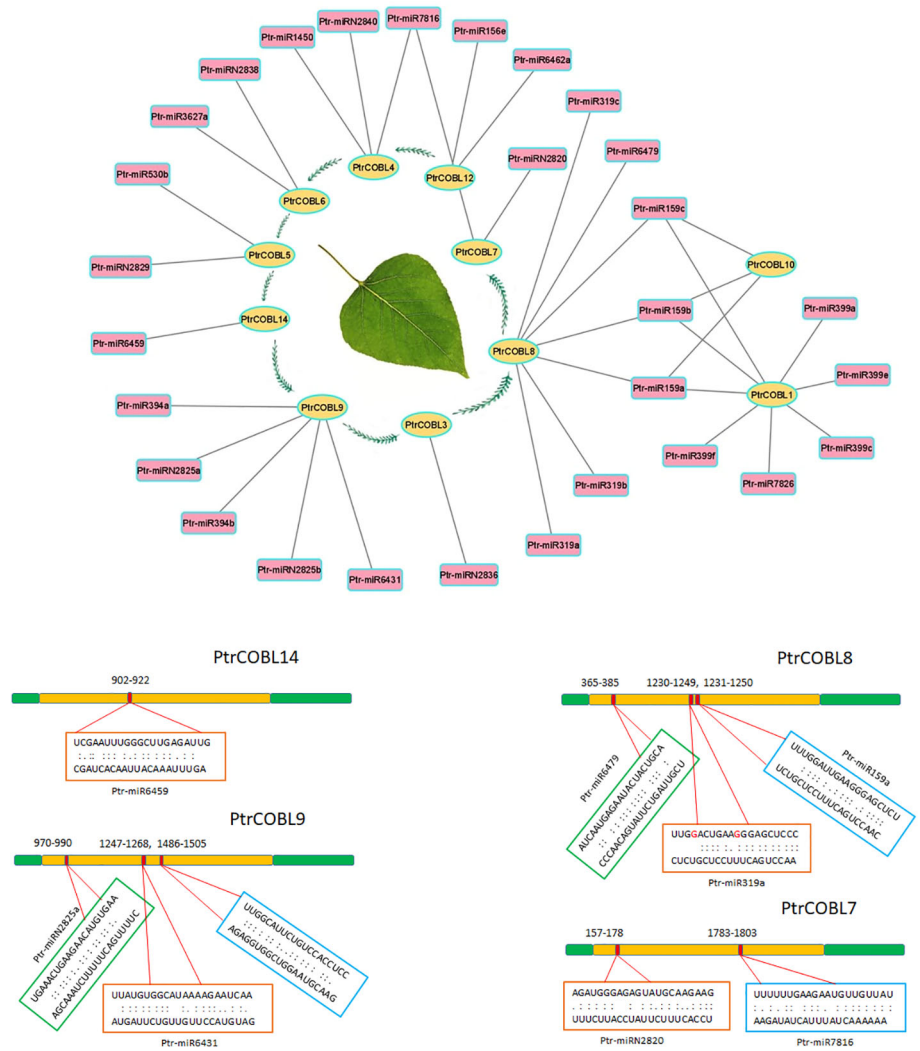


FIGURE 8
mi-RNA targeting network and their target sites in poplar COBLs. Genes are located in the oval shapes, while targeting miRNAs are in the rectangular shapes. Lines connect different genes to their putative miRNAs. Multiple miRNAs sharing the same target site are represented once.

was observed. Almost half of the genes were found to be expressed in all abiotic stresses, i.e., drought, heat, cold, and salt stress (Figure 9A). PtrCOBL10 along with PtrCOBL11 showed the highest expression among all the genes, consistent with all the stresses. Notably, PtrCOBL10 demonstrated higher expression in heat and salt stress compared with cold and drought stress. Conversely, PtrCOBL11 showed perpetually higher expression among all the abiotic stresses. The same went for PtrCOBL2 and PtrCOBL3, where PtrCOBL3 showed high expression in cold and drought stress in comparison to heat and salt stress. PtrCOBL6 demonstrated constant expression except in the drought stress, where higher expression was observed, suggesting PtrCOBL6 expression was related to drought response. PtrCOBL12 also showed an unswerving expression pattern. Contrastingly, PtrCOBL8 exhibited the least expression in abiotic stresses among all the genes. These results suggested that COBL genes take part in abiotic stress response mechanisms.

Having said that, to identify hub genes, the co-expression network analyses were performed based on the correlation coefficient of their respective expression data. Co-expression network analysis showed positive as well as negative correlations among all the abiotic stresses. In the case of cold stress, 46 gene pairs were positively correlated while 45 gene pairs were negatively correlated (Figure 9B). In regard to drought stress, 51 gene pairs were found to be positively correlated while 40 gene pairs were found to be negatively correlated (Figure 9E). As far as heat stress was concerned, there were 44 positively correlated gene pairs and 47 negatively correlated gene pairs. (Figure 9C). Moreover, in salt stress, 53 gene pairs were found to be negatively correlated while 38 gene pairs were found to be positively correlated (Figure 9D). Considering the high expression and large number of correlation links, hub genes were decided. Interestingly, PtrCOBL2 and PtrCOBL3 turned out to be hub genes in each stress group.

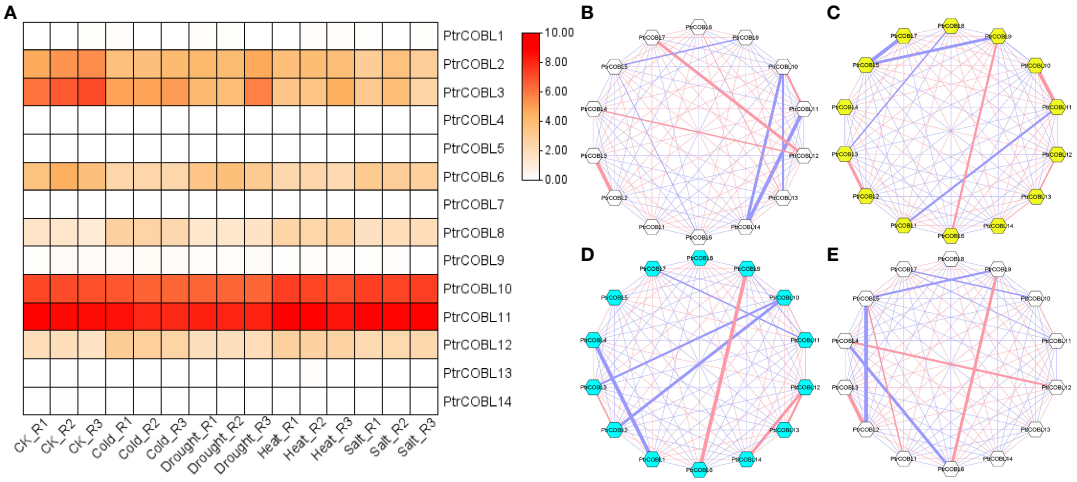


FIGURE 9
Expression profiling and co-expression network analysis for abiotic stress. Heat map of expression patterns. **(A)** The co-relation networks of 14 COBL genes in **(B)** cold, **(C)** heat, **(D)** salt, and **(E)** drought were analyzed based on Pearson correlation of the selected COBL genes obtained from transcriptomic data. The red and blue edges show positive and negative correlations, respectively. However, the thickness of each edge shows the strength of the correlation for each pair.

The cold stress network showed that PtrCOBL2 and PtrCOBL3 shared a strong positive correlation. Likewise, PtrCOBL10 and PtrCOBL11 were linked with a great positive correlation. In contrast, PtrCOBL10 had a strong but negative correlation with PtrCOBL13 and PtrCOBL14, and similarly PtrCOBL11 also exhibited a strong but negative correlation with PtrCOBL14. All these correlation links were consistent with expression patterns shown in the heatmap. In the case of heat stress, PtrCOBL2 and PtrCOBL3 shared a strong and positive correlation and, similarly, PtrCOBL10 and PtrCOBL11 were also found to share strong and positive correlation. On the flip side, some interactions were not found to be in cooperation in the expression data, e.g., PtrCOBL3 had a negative correlation with PtrCOBL8 while in the salt stress network, PtrCOBL12 had a positive correlation with PtrCOBL13 and PtrCOBL14. In the same line, PtrCOBL10 had a negative correlation with PtrCOBL2 and PtrCOBL3. This may serve as a foundational guide for further research into how the PtrCOBL genes interact.

3.10 Expression profiling and co-expression network analysis for wood formation

Transcriptome data (publicly available) were used to investigate the putative functions of PtrCOBL genes in wood formation, where the expression of PtrCOBL genes was as follows: PtrCOBL1 along with PtrCOBL12 displayed moderate expression. On the flip side, PtrCOBL1, 4, 5, 7, 13, and 14 presented minute levels of expression in early xylem, mature xylem, phloem, pith, and other wood fiber tissues (Figure 10A). COBL 2, 3, 6, 9, 10, 11, and 12 showed a good level of expression in all the tissue sections. Overall, the entire list of COBL genes showed expression in xylem, cambium, and phloem development tissues. This analysis gave an approximate measure of the xylem area containing living vessels, fibers, and other tissues taking part in stem growth and plant girth (Figure 10A). To further explore the role of PtrCOBLs in developing tissues, expression profiles were investigated under tension-wood formation. Interestingly, the majority of COBLs presented good expressions; COBL 8 and 12 showed a minute level of expression whereas COBL 2, 3, 6, 9, and 10 and PtrCOBL11 divulged higher levels of expression. This expression indicated a pivotal role of COBLs in tension wood developing tissue (Figure 10B).

Moreover, co-expression network analysis was done to investigate the interactions among all COBL genes during tension wood development (Figure 10C). PtrCOBL5 and PtrCOBL13 showed strong positive interaction; likewise PtrCOBL6 and PtrCOBL10 showed strong and positive interaction. PtrCOBL4 and 10 showed strong but negative interaction, as PtrCOBL7 and 8 showed negative interaction, whereas PtrCOBL 8 with 12 and COBL5 with 13 also showed negative interaction. Similarly, PtrCOBL9 with 13 showed negative interaction. Interestingly, PtrCOBL13 and 14 were found to be negatively interacted. PtrCOBL12 and 13 showed a total of 100 interactions with the majority being positive ones. We identified hub genes based on the large number of correlation hits and high expression. Intriguingly, two genes (PtrCOBL2 and PtrCOBL3) stood out as hub genes, suggesting that these two genes play an important role in the development of wood in the poplar plant.

4 Discussion

The world's population is expected to reach roughly 9.2 billion by 2050, and the current rate of increase in plant yield and biomass is not enough to fulfill their demands (Zaheer et al., 2021). It is generally known that abiotic stress hinders plant growth, development, and production, and climate change is projected to

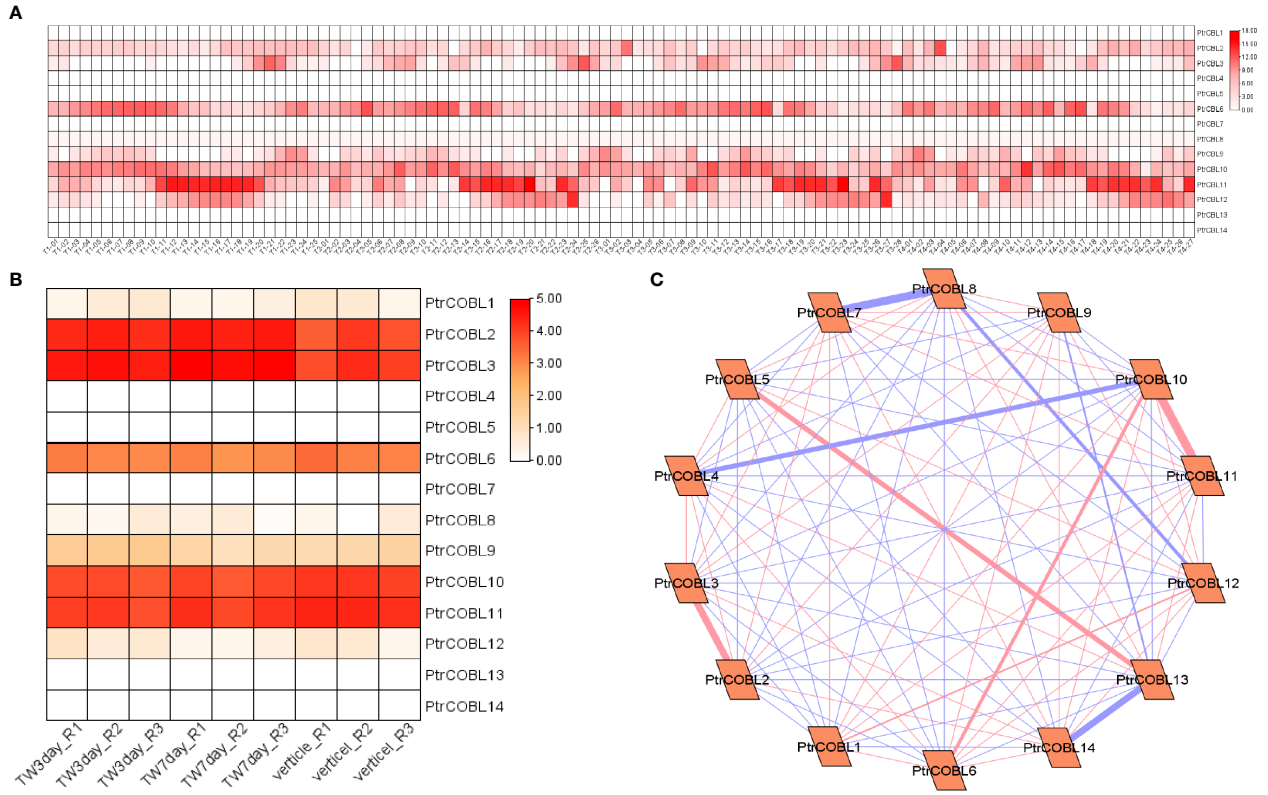


FIGURE 10
Expression profiling and co-expression network analysis for wood formation. **(A)** Heat map of the expression patterns of COBLs in all the wood forming tissues. **(B)** Heatmap of the expression patterns of COBLs in tension-wood forming tissues. **(C)** The co-relation networks of 14 COBL genes in tension-wood formation were analyzed based on Pearson correlation of the selected COBL genes obtained from transcriptomic data. The red and blue edges show the positive and negative correlations, respectively. However, the thickness of each edge shows the strength of the correlation for each pair.

worsen severe weather events, making the situation even more precarious (Liu et al., 2015). In this scenario, the importance of forest trees has already increased: tree plants are simply a lifeline for earth. To develop more rigorous and resilient plants we need to improve our comprehension about the rudimentary processes, functions, and evolutionary mechanisms in plants (Leng et al., 2021). The advent of next-generation sequencing techniques has revolutionized the plant genomic landscape by making data from an endless variety of dimensions available. There is now the possibility of doing more accurate genome-wide analysis because to the availability of recently re-sequenced high-quality poplar genome assemblies (Liu et al., 2015; Sangi et al., 2021) (Figure 11).

4.1 Identification and evolution

The COBRA gene family has been reported in many other plant species. Even the moss *Physcomitrella patens* has been shown to contain members of the COBRA family. The ancestor of *Arabidopsis* already had members of this family (Roudier et al., 2002). There are 14 members in group I of the COBL genes in poplar, and they are most closely related to the COB genes in the *Arabidopsis* genome. Whilst group II proteins, which consist of four members, are alike to the AtCOBL7 subfamily, they are longer in size than subfamily I proteins. When comparing groups I and II, the main difference is the 170 more N-terminal amino acids found in group I (Brady et al., 2007). The differentiation between the two clades is supported by the differences in gene structure and the number of exons and introns. After the break between eurosid I (*Populus*) and eurosid II (*Arabidopsis*), there may have been divergence that led to this (Roudier et al., 2002). The COBL genes encode for proteins that are anticipated to have a cellulose binding site, a Cys-Cys-Val-Ser motif (CCVS), a possible N-glycosylation site, and an N-terminal peptide signal. Cleavage of the proteins at the GPI anchor modification sites (ω sites) and subsequent addition of the GPI anchor structure through amide bonding occurs at the c-terminus (Roudier et al., 2002; Niu et al., 2015) (Figure 1; Table 1). Furthermore, phylogenetic research revealed that both monocot and dicot COBLs are present within the two groups (Figure 2). This study provides more evidence that members of the COBL family are descended from a duplication that happened prior to the split between monocots and dicots (Tuskan et al., 2006; Ye et al., 2009). The number of COBL genes is larger in dicots because of repeated gene duplications. The very similar gene structures across PtrCOBL genes provide strong evidence for the extensive conservation of this gene family (Figure 3). Weak selection pressure and a higher intron abundance were also thought to have allowed introns to evolve into unique functional roles (Zaheer et al., 2021).

10.3389/fpls.2023.1242836

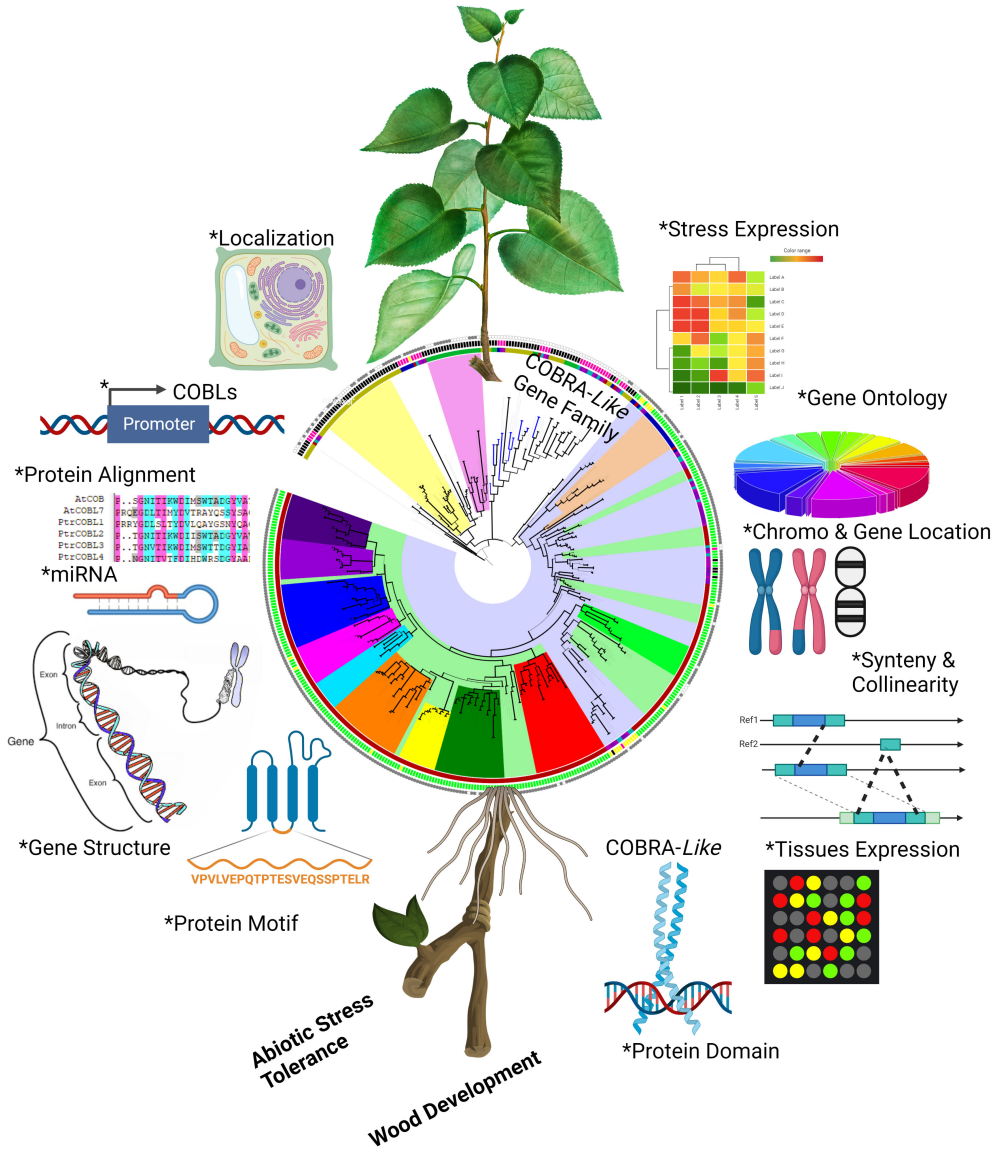


FIGURE 11
Summary of the comprehensive study of COBL genes in poplar. Asterisk symbol indicate different analyses were performed to get insight into the characterization of PtrCOBL genes. The figure was constructed using the BioRender tool.

4.2 Gene structure and intron exon genome expansion

COBL genes play a vital role in cell growth, wood fiber, and cellulose synthesis. The function and biological significance of the COBL proteins in a wide variety of plants, *Arabidopsis*, brassica, soyabean, rice, cotton, wheat, *L. Chinense*, and *Populus*, have been the subject of study (Dai et al., 2009; Ye et al., 2009; Niu et al., 2015; Sangi et al., 2021; Yang et al., 2021; Zaheer et al., 2021; Qiu et al., 2023). There has not been a methodical and comprehensive study of COBL genes in poplar. Studies on eukaryotic diversification have shown extensive intron gain or loss. An identical exon–intron configuration was found while analyzing duplicated genes (Figure 3). We discovered that all COBL genes have the same number of introns except COBL 1, 7, 6, and 9 (paralog pairs; 1-7, 6-9), which have only one intron, indicating that COBL protein in poplar was recently evolved with fewer introns, corroborating that the number of introns has been decreasing with time, meaning that more recently developed species have fewer introns than more ancient ones (Zaheer et al., 2021). In the same manner, all the COBL have same number of exons (CDS), except COBL 1, 7, 6, and 9. Notably the COBL1, 7, 6, and 9 group is alike and the COB group is alike (Figure 3A). Significant insertion/deletion events may affect the exon–intron structure. In plant evolution, introns have a major impact on the development of new species (Roudier et al., 2002; Brady et al., 2007).

4.3 WGD collaborated to COBL gene family expansion

Evidence suggests that, compared to monocots, the number of family members in eudicots has expanded. Several potential explanations have been proposed for this expansion, including whole-genome duplication and segmental duplication (Tang et al., 2008). Whole-genome duplication (WGD) occurred twice in the poplar genome, first at 8-13 Mya and again around 60-65 Mya (Tuskan et al., 2006; Sundell et al., 2015). Only two paralogous gene pairs (PtrCOBL4/PtrCOBL5 and PtrCOBL13/PtrCOBL14) were found tandemly arranged, while all the other paralogous genes pairs were scattered on the different chromosomes (Figure 4; Supplementary Figure 1), revealing the expansion of the COBL gene family in poplar, primarily due to segmental duplication rather than singletons or tandem duplication (Bano et al., 2021; Yang et al., 2021). This confirms prior findings that segmental duplication has been predominated in the poplar genome. Based on our data, we were able to determine that only two paralog pairs originated from the first WGD and four paralog pairs originated from the second WGD (Cheng et al., 2021; He et al., 2021). We hypothesize that the recent WGD duplication may be the primary mechanism for the growth and functional diversity of COBL genes in poplar (Figures 4, 5). To analyze the non-synonymous/synonymous mutation ratio of PtrCOBL genes, the Ka/Ks ratios were determined. Pseudogenes emerge from neutral selection when the Ka/Ks ratio is 1, purifying selection favors duplicated genes with a Ka/Ks ratio below 1, and positive selection of rapid evolution is shown with a Ka/Ks ratio above 1. All the COBLs have Ka/Ks ratios below 1, suggesting that the COBL gene family was subject to purifying selection over the course of evolution (Niu et al., 2015; Bano et al., 2021; Kabir et al., 2022). The Ka/Ks ratios for PtrCOBL12-PtrCOBL3 and PtrCOBL8-PtrCOBL10 were the lowest among the paralogous genes, indicating these genes were able to keep their original roles despite their shorter divergence dates. The highest Ka/Ks value and very divergent expression patterns were found for the paralog pair PtrCOBL13-PtrCOBL14, providing strong evidence for their functional divergence after duplication (Sangi et al., 2021; Yang et al., 2021).

4.4 Cellular localization and promotor analysis of PtrCOBLs

The subcellular location of genes play a significant role in determining their function (Wang et al., 2022). The COBRA-*Like* proteins have highly conserved structures. Evidence indicates that, similar to COBL proteins, cellulose synthase complexes (CSCs) are assembled in the endoplasmic reticulum (ER) or the Golgi apparatus. The N-terminal signal peptide and C-terminal GPI anchoring motif of COBRA-*Like* proteins are hypothesized to direct the protein to the outer leaflet of the plasma membrane (McFarlane et al., 2014; Sangi et al., 2021). It is unclear whether proteins encoded by COBRA genes that lack any of these patterns are functional. After a protein has been localized to the outer leaflet of the plasma membrane, the GPI anchor motif is usually disrupted (Zaheer et al., 2021). It is thus likely that COBRA's ability to bind cellulose is crucial for its proper localization at the plasma membrane-cum-cell wall interface (Roudier et al., 2002; Zaheer et al., 2021). The majority of genes showed a strong localization signal in Golgi apparatus coupled with plasma membrane and/or vacuole, in agreement with the results (Qiu et al., 2023). Particularly PtrCOBL2 is dual-targeted to plasma membrane and/or vacuoles whereas PtrCOBL4 and PtrCOBL13 are extracellular- and vacuole-specific (Table 1).

Gene expression control occurs primarily via transcriptional regulation, which requires interaction between transcription factors and promoter binding sites (Wang et al., 2022). Specific and consistent cis-acting elements in promoter regions are required for different external signals to activate inducible promoters; for example, auxin-induced promoters typically contain cis-acting elements of the AuxRE, and light-induced promoters typically contain cis-acting elements rich in the AT-rich motif, the I-box, the G-box, and the GT1-motif (Kabir et al., 2022; Wang et al., 2022). Similarly, drought-induced promoters included the cis-acting elements CACG and CATGTG. We isolated sections of potential genes' upstream promoters for this analysis (Figure 6A). Rigorous appearance of MeJA, Auxin, SA, and ABA response elements in COBL promotor regions implied their involvement in stress and plant development (Yu et al., 2021; Qiu et al., 2023). Phytohormones have been proven in previous research to ensure that plant cell walls remain unaltered. MeJA has a mediatory role in the regulation of genes in response to plant damage. Auxin controls cell wall thickness by stimulating membrane permeability (Ye et al., 2009). Strawberry fruit softens and ripens with the help of ethylene, which controls the production of pectin (Iqbal et al., 2022; Kabir et al., 2022). Interestingly, meristem and wound response related cis-elements prove COBLs' direct role in wood development and tension wood forming tissue in poplar (Supplementary Table S2). In plants, the TATA-box has been shown to have a role in the regulation of transcription via the mediation of miRNA (Kabir et al., 2022).

4.5 Global analysis of micro-RNAs in poplar

MicroRNAs (miRNAs) are a class of single-stranded, non-coding RNAs that play a role in post-transcriptional gene regulation (Wang et al., 2022). Genome-wide investigation has espied 29 miRNAs of 20 families (Figure 8; Supplementary Figure 2) belonging to 108 total miRNA families currently reported in *P. Trichocarpa* (Guo et al., 2022). Using our methodology, miRNA specific target sites were determined in the genes of PtrCOBL family (Figure 8; Supplementary Figure 2). Most of these miRNAs have been reported with high expression in xylem and leaf area, indicating their key role in plant growth and development and stress response. For instance, many plant families, including angiosperms, mosses, and lycopods, share a highly preserved version of MIR159 because of its essential role in plant growth, with high expression in xylem tissues and leaf (Guo et al., 2022; Kabir et al., 2022). The miR159 targets the MYB and

NAC genes to fine-tune the juvenile development. miR159 interacts with ABI5 shape the vegetative development (Guo et al., 2021). Ptr-mRNA319, Ptr-miR1450, and Ptr-miR6462 are involved in the regulation of secondary growth in plants, with high expression in xylem (Wang R. et al., 2021). Ptr-miR6479 is involved in oxidation reduction and cell development. Ptr-miR6431 is reported to be ABA responsive and involved in stress and growth-related functions. Ptr-miR394, Ptr-miR156, Ptr-miR530, Ptr-miR6459, and Ptr-miR3627 have shown high expression both in leaf and xylem while Ptr-miR7816, Ptr-miR399, and Ptr-miR7826 are selectively expressed in various vegetative tissues like leaf area (Kuang et al., 2018; Wang R. et al., 2021). One unanticipated finding was that we found some unique miRNAs (Ptr-miR2820, miR2825, Ptr-miR2829, Ptr-miR2836, Ptr-miR2838, and Ptr-miR2840) that have not been reported in previous poplar studies. Our findings will pave the way for more studies into the biological roles of Ptr-miRNAs and the target sites in poplar.

4.6 Expression profiles and likely role of PtrCOBLs in abiotic stress and wood formation

The COBL gene family is speculated to have a role in both the plant's stress response and its ability to produce wood tissues (Yang et al., 2021). High expression levels of these genes have been seen in secondary growth, tension wood, and abiotic stressors, as bolstered by transcriptome datasets. Five different plant organs (Figure 7) and 27 different tissue types (Figure 10) have different PtrCOBLs expression patterns, as shown by the expression profiling (Sundell et al., 2015; Sundell et al., 2017). In continuation of the trailing results, Cobra-*Like* genes (PtrCOBL2, 3, 10, and 11) registered a positive expression response under cold, heat, drought, and salinity stresses (Figure 9). Similar results were reported in *Arabidopsis* (Ko et al., 2006), wheat (Zaheer et al., 2021), *S. spontaneum* (Qiu et al., 2023), and rice (Sun et al., 2022). PtrCOBL2, 3, 6, and 10 and PtrCOBL11 flaunted high representation in stem (node, internode) regions. Studies have shown that stems with increased mechanical strength were more resistant to lodging and disease, which also indicates more biomass production. COBL2 is involved in cell differentiation and is mandatory for cellulose deposition in *Arabidopsis* (Ben-Tov et al., 2018; Niu et al., 2018); accordingly, the PtrCOBL6 and 10 in the AtCOBL2 subgroup showed higher expression abiotic stress response and in wood formation process. Contrary to the expectations, PtrCOBL1, 7, and 9, belonging to the same group, showed the lowest expression, implying an unequal evolution event among the orthologous groups (Yang et al., 2021). PtrCOBL2 and the *BC1* gene, the orthologs of AtCOBL4 found in poplar, rice, and sorghum, control the cellulose content of the secondary cell wall. AtCOBL10 and 11 are involved in pollen tube cell elongation, so showed the highest expression in abiotic stress and cell expansion (Niu et al., 2018; Qiu et al., 2023). Members of the COB group, *BC1* in rice and *BC2* in maize, and *Bna9*, 35, and 41 in rapeseed, were found to be involved in cell wall biosynthesis rendering stem-breaking resistance (Li et al., 2003; Li et al., 2018; Yang et al., 2021). Similar results were found in the current study: PtrCOBL2, 3, 8, 10, 11, and 12 exhibited a higher protein abundance in wood developing tissues. Overall, the AtCOB group was found to be highly involved in cell wall deposition and rendering mechanical strength, in accordance with the previously reported studies (Yang et al., 2021).

For further exploration in this study, correlation coefficients of the expression data were used to conduct co-expression network analysis (Iqbal et al., 2022). Notably, PtrCOBL2 and PtrCOBL3 were the two genes characterized as hub genes in abiotic stress as well as in wood formation-related PtrCOBLs (Figures 9B–E, 10C). Therefore, PtrCOBL2 and PtrCOBL3 could be considered as critical genes in further studies regarding plant stress and secondary growth in poplar. However, some discrepancies in the expression patterns of paralogous gene pairs were discerned (Sangi et al., 2021), e.g., PtrCOBL6/PtrCOBL9 possess a similar structure and conserved motif but showed different expression profiles. PtrCOBL6 has a higher response in wood forming tissues and abiotic stress, while PtrCOBL9 is totally absent in abiotic stress and has lower response in wood forming tissues. Interestingly, PtrCOBL6 has high expression in stems whereas PtrCOBL9 has high expression in roots. As a result, we hypothesize that PtrCOBL genes went through duplication events that may have caused gene mutations, which in turn alter gene function and expression patterns. This phenomenon has also been observed in the ERF family, NAC family, and XTH gene family (Liu et al., 2015; Cheng et al., 2021).

The quantitative polymerase chain reaction (RT-qPCR) was employed to examine the expression of PtrCOBL genes in poplar across a range of tissue types to determine their potential role in controlling cell expansion and abiotic stress. The expression of most genes was consistent, except for PtrCOBL6, PtrCOBL9, and PtrCOBL12 wherein better expressions were found in roots, stems, and YLs, respectively, in comparison to microarray data. This discrepancy could be the result of differences in their experimental settings. Gene ontology (GO) enrichment has bolstered this study (Figure 6B). The presumed functions of PtrCOBLs have been fortified with following terms: cellulose microfibril organization, anchored component of membrane, plant-type cell wall organization or biogenesis, cell growth, plant-type secondary cell wall biogenesis, anchored component of plasma membrane, polysaccharide binding, carbohydrate binding, external encapsulating structure organization, extracellular matrix organization, response to abiotic stimulus, and growth. It has been demonstrated that the PtrCOBL genes are predominantly implicated in mechanisms relating to cell distension, biological growth, stress response, and wood development (Li et al., 2018; Zaheer et al., 2021). The findings of the current study suggest that the PtrCOBL genes will be useful in the future development of better abiotic stress-resistant woody plants. Nevertheless, additional research is needed to be done on the functional dissection and potential of COBL genes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

MS conceived the idea, conducted the analyses, investigation, and wrote the original draft. AA provided the methodology and software. MR and MY contributed the plant materials and to data analysis. QH conducted the literature survey and visualization. MS and M-ZL undertook manuscript review and editing. M-ZL performed supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the National Key Research and Development Program of China "International Cooperation in Science and Technology Innovation between Governments" (2019YFE0119100).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1242836/full#supplementary-material

References

Bano, N., Fakhrah, S., Mohanty, C. S., and Bag, S. K. (2021). Genome-wide identification and evolutionary analysis of *Gossypium* tubby-like protein (TLP) gene family and expression analyses during salt and drought stress. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.667929

Ben-Tov, D., Idan-Molakandov, A., Hugger, A., Ben-Shlush, I., Günl, M., Yang, B., et al. (2018). The role of COBRA-LIKE 2 function, as part of the complex network of interacting pathways regulating *Arabidopsis* seed mucilage polysaccharide matrix organization. *Plant J.* 94 (3), 497–512. doi: 10.1111/tpj.13871

Brady, S. M., Song, S., Dhugga, K. S., Rafalski, J. A., and Benfey, P. N. (2007). Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant Physiol.* 143 (1), 172–187. doi: 10.1104/pp.106.087262

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cheng, Z., Zhang, X., Yao, W., Gao, Y., Zhao, K., Guo, Q., et al. (2021). Genome-wide identification and expression analysis of the xyloglucan endotransglucosylase/hydrolase gene family in poplar. *BMC Genomics* 22 (1), 804. doi: 10.1186/s12864-021-08134-8

Coleman, H. D., Brunner, A. M., and Tsai, C.-J. (2021). Synergies and entanglement in secondary cell wall development and abiotic stress response in trees. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.639769

Dai, X., You, C., Wang, L., Chen, G., Zhang, Q., and Wu, C. (2009). Molecular characterization, expression pattern, and function analysis of the OsBC1L family in rice. *Plant Mol. Biol.* 71 (4), 469–481. doi: 10.1007/s11103-009-9537-3

De Frenne, P., Lenoir, J., Luoto, M., Scheffers, B. R., Zellweger, F., Aalto, J., et al. (2021). Forest microclimates and climate change: Importance, drivers and future research agenda. *Global Change Biol.* 27 (11), 2279–2297. doi: 10.1111/gcb.15569

Gao, Y., Bian, L., Shi, J., Xu, J., Xi, M., and Wang, G. (2013). Expression of a conifer COBRA-*like* gene ClCOBL1 from Chinese fir (*Cunninghamia lanceolata*) alters the leaf architecture in tobacco. *Plant Physiol. Biochem.* 70, 483–491. doi: 10.1016/j.plaphy.2013.06.013

Gritsch, C., Wan, Y., Mitchell, R. A. C., Shewry, P. R., Hanley, S. J., and Karp, A. (2015). G-fibre cell wall development in willow stems during tension wood induction. *J. Exp. Bot.* 66 (20), 6447–6459. doi: 10.1093/jxb/erv358

Guo, C., Jiang, Y., Shi, M., Wu, X., and Wu, G. (2021). *ABI5* acts downstream of miR159 to delay vegetative phase change in *Arabidopsis*. *New Phytol.* 231 (1), 339–350. doi: 10.1111/nph.17371

Guo, Z., Kuang, Z., Zhao, Y., Deng, Y., He, H., Wan, M., et al. (2022). PmiREN2.0: from data annotation to functional exploration of plant microRNAs. *Nucleic Acids Res.* 50 (D1), D1475–D1482. doi: 10.1093/nar/gkab811

He, F., Shi, Y.-J., Zhao, Q., Zhao, K.-J., Cui, X.-L., Chen, L.-H., et al. (2021). Genome-wide investigation and expression profiling of polyphenol oxidase (PPO) family genes uncover likely functions in organ development and stress responses in *Populus trichocarpa*. *BMC Genomics* 22 (1), 1–15. doi: 10.1186/s12864-021-08028-9

Hochholdinger, F., Wen, T.-J., Zimmermann, R., Chimot-Marolle, P., da Costa e Silva, O., Bruce, W., et al. (2008). The maize (*Zea mays* L.) roothairless3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.: Cell Mol. Biol.* 54 (5), 888–898. doi: 10.1111/j.1365-313X.2008.03459.x

Iqbal, A., Huiping, G., Xiangru, W., Hengheng, Z., Xiling, Z., and Meizhen, S. (2022). Genome-wide expression analysis reveals involvement of asparagine synthetase family in cotton development and nitrogen metabolism. *BMC Plant Biol.* 22 (1), 122–122. doi: 10.1186/s12870-022-03454-7

Julius, B. T., McCubbin, T. J., Mertz, R. A., Baert, N., Knoblauch, J., Grant, D. G., et al. (2021). Maize Brittle Stalk2-Like3, encoding a COBRA protein, functions in cell wall formation and carbohydrate partitioning. *Plant Cell* 33 (10), 3348–3366. doi: 10.1093/plcell/koab193

Kabir, N., Zhang, X., Liu, L., Qanmber, G., Zhang, L., Wang, Y. X., et al. (2022). RAD gene family analysis in cotton provides some key genes for flowering and stress tolerance in upland cotton *G. hirsutum*. *BMC Genomics* 23 (1), 1–14. doi: 10.1186/s12864-021-08248-z

Ko, J.-H., Kim, J. H., Jayanty, S. S., Howe, G. A., and Han, K.-H. (2006). Loss of function of COBRA, a determinant of oriented cell expansion, invokes cellular defence responses in *Arabidopsis thaliana*. *J. Exp. Bot.* 57 (12), 2923–2936. doi: 10.1093/jxb/erl052

Kuang, Z., Wang, Y., Li, L., and Yang, X. (2018). miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. *Bioinformatics* 35 (14), 2521–2522. doi: 10.1093/bioinformatics/bty972

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi: 10.1093/molbev/msy096

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23 (21), 2947–2948. doi: 10.1093/bioinformatics/btm404

Leng, X., Wang, H., Zhang, S., Qu, C., Yang, C., Xu, Z., et al. (2021). Identification and Characterization of the APX Gene Family and Its Expression Pattern under Phytohormone Treatment and Abiotic Stress in *Populus trichocarpa*. *Genes* 12 (3), 1–21. doi: 10.3390/genes12030334

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Li, S., Ge, F.-R., Xu, M., Zhao, X.-Y., Huang, G.-Q., Zhou, L.-Z., et al. (2013). Arabidopsis COBRA-LIKE 10, a GPI-anchored protein, mediates directional growth of pollen tubes. *Plant J.* 74 (3), 486–497. doi: 10.1111/tpj.12139

Li, P., Liu, Y., Tan, W., Chen, J., Zhu, M., Lv, Y., et al. (2018). *Brittle culm* 1 encodes a COBRA-like protein involved in secondary cell wall cellulose biosynthesis in Sorghum. *Plant Cell Physiol.* 60 (4), 788–801. doi: 10.1093/pcp/pcy246

Li, Y., Qian, Q., Zhou, Y., Yan, M., Sun, L., Zhang, M., et al. (2003). *BRITTLE CULM*1, which encodes a COBRA-like protein, affects the mechanical properties of rice plants. *Plant Cell* 15 (9), 2020–2031. doi: 10.1105/tpc.011775

Li, Z., Zhou, T., Sun, P., Chen, X., Gong, L., Sun, P., et al. (2022). *COBL9* and *COBL7* synergistically regulate root hair tip growth via controlling apical cellulose deposition. *Biochem. Biophys. Res. Commun.* 596, 6–13. doi: 10.1016/j.bbrc.2022.01.096

Liu, B., Liu, J., Yu, J., Wang, Z., Sun, Y., Li, S., et al. (2021). Transcriptional reprogramming of xylem cell wall biosynthesis in tension wood. *Plant Physiol.* 186 (1), 250–269. doi: 10.1093/plphys/kiab038

Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., et al. (2013). *Brittle Culm*1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. *PloS Genet.* 9 (8), e1003704. doi: 10.1371/journal.pgen.1003704

Liu, Q., Wang, Z., Xu, X., Zhang, H., and Li, C. (2015). Genome-Wide analysis of C2H2 zinc-Finger family transcription factors and their responses to abiotic stresses in poplar (*Populus trichocarpa*). *PloS One* 10 (8), e0134753. doi: 10.1371/journal.pone.0134753

McFarlane, H. E., Döring, A., and Persson, S. (2014). The cell biology of cellulose synthesis. *Annu. Rev. Plant Biol.* 65, 69–94. doi: 10.1146/annurev-arplant-050213-040240

Niu, E., Fang, S., Shang, X., and Guo, W. (2018). Ectopic expression of *GhCOBL9A*, a cotton glycosyl-phosphatidyl inositol-anchored protein encoding gene, promotes cell elongation, thickening and increased plant biomass in transgenic *Arabidopsis*. *Mol. Genet. Genomics* 293 (5), 1191–1204. doi: 10.1007/s00438-018-1452-3

Niu, E., Shang, X., Cheng, C., Bao, J., Zeng, Y., Cai, C., et al. (2015). Comprehensive analysis of the COBRA-like (COBL) gene family in *Gossypium* identifies two COBLs potentially associated with fiber quality. *PloS One* 10 (12), e0145725. doi: 10.1371/journal.pone.0145725

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46 (W1), W200–W204. doi: 10.1093/nar/gky448

Qiu, C., Chen, J., Wu, W., Liao, B., Zheng, X., Li, Y., et al. (2023). Genome-wide analysis and abiotic stress-responsive patterns of COBRA-*like* gene family in *Liriodendron chinense*. *Plants (Basel Switzerland)* 12 (8), 1616. doi: 10.3390/plants12081616

Raza, A., Sharif, Y., Chen, K., Wang, L., Fu, H., Zhuang, Y., et al. (2022). Genome-wide characterization of ascorbate peroxidase gene family in peanut (*Arachis hypogea* L.) revealed their crucial role in growth and multiple stress tolerance. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.962182

Roudier, F., Schindelman, G., DeSalle, R., and Benfey, P. N. (2002). The COBRA family of putative GPI-anchored proteins in Arabidopsis. A new fellowship in expansion. *Plant Physiol.* 130 (2), 538–548. doi: 10.1104/pp.007468

Rudel, T. K., Meyfroidt, P., Chazdon, R., Bongers, F., Sloan, S., Grau, H. R., et al. (2020). Whither the forest transition? Climate change, policy responses, and redistributed forests in the twenty-first century. *Ambio* 49 (1), 74–84. doi: 10.1007/s13280-018-01143-0

Sangi, S., Araújo, P. M., Coelho, F. S., Gazara, R. K., Almeida-Silva, F., Venancio, T. M., et al. (2021). Genome-Wide Analysis of the COBRA-*Like* Gene Family Supports Gene Expansion through Whole-Genome Duplication in Soybean (*Glycine max*). *Plants* 10 (1), 167. doi: 10.3390/plants10010167

Sato, K., Ito, S., Fujii, T., Suzuki, R., Takenouchi, S., Nakaba, S., et al. (2010). The carbohydrate-binding module (CBM)-like sequence is crucial for rice CWA1/BC1 function in proper assembly of secondary cell wall materials. *Plant Signaling Behav.* 5 (11), 1433–1436. doi: 10.4161/psb.5.11.13342

Sato-Izawa, K., Nakamura, S.-I., and Matsumoto, T. (2020). Mutation of rice *bc1* gene affects internode elongation and induces delayed cell wall deposition in developing internodes. *Plant Signaling Behav.* 15 (5), e1749786. doi: 10.1080/15592324.2020.1749786

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22 (1), 53–78. doi: 10.1146/annurev.cellbio.22.022206.160206

Sorek, N., Szemenyei, H., Sorek, H., Landers, A., Knight, H., Bauer, S., et al. (2015). Identification of MEDIATOR16 as the *Arabidopsis* COBRA suppressor MONGOOSE1. *Proc. Natl. Acad. Sci. U. States America* 112 (52), 16048–16053. doi: 10.1073/pnas.1521675112

Sun, X., Xiong, H., Jiang, C., Zhang, D., Yang, Z., Huang, Y., et al. (2022). Natural variation of *DROT1* confers drought adaptation in upland rice. *Nat. Commun.* 13 (1), 4265. doi: 10.1038/s41467-022-31844-w

Sundell, D., MannapPeruma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjödin, A., et al. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytol.* 208 (4), 1149–1156. doi: 10.1111/nph.13557

Sundell, D., Street, N. R., Kumar, M., Mellerowicz, E. J., Kucukoglu, M., Johnsson, C., et al. (2017). AspWood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *Plant Cell* 29 (7), 1585–1604. doi: 10.1105/tpc.17.00153

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320 (5875), 486–488. doi: 10.1126/science.1153917

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691

Wang, H., Liu, S., Dai, X., Yang, Y., Luo, Y., Gao, Y., et al. (2021). PSDX: A comprehensive multi-omics association database of *Populus trichocarpa* with a focus on the secondary growth in response to stresses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.655565

Wang, R., Reng, M., Tian, S., Liu, C., Cheng, H., Liu, Y., et al. (2021). Transcriptome-wide identification and characterization of microRNAs in diverse phases of wood formation in *Populus trichocarpa*. *G3* 11 (8), 1–19. doi: 10.1093/g3journal/jkab195

Wang, H., Umer, M. J., Liu, F., Cai, X., Zheng, J., Xu, Y., et al. (2022). Genome-wide identification and characterization of *CPR5* genes in *Gossypium* reveals their potential role in trichome development. *Front. Genet.* 13. doi: 10.3389/fgene.2022.921096

Xie, M., Zhang, J., Tschaplinski, T. J., Tuskan, G. A., Chen, J.-G., and Muchero, W. (2018). Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01427

Xu, Z., Gao, L., Tang, M., Qu, C., Huang, J., Wang, Q., et al. (2017). Genome-wide identification and expression profile analysis of *CCH* gene family in *Populus*. *PeerJ* 5, e3962. doi: 10.7717/peerj.3962

Yang, Q., Wang, S., Chen, H., You, L., Liu, F., and Liu, Z. (2021). Genome-wide identification and expression profiling of the COBRA-*like* genes reveal likely roles in stem strength in rapeseed (*Brassica napus* L.). *PloS One* 16 (11), e0260268. doi: 10.1371/journal.pone.0260268

Ye, X., Kang, B.-G., Osburn, L. D., and Cheng, Z.-M. (2009). The COBRA gene family in *Populus* and gene expression in vegetative organs and in response to hormones and environmental stresses. *Plant Growth Regul.* 58 (2), 211–223. doi: 10.1007/s10725-009-9369-9

Yoshida, K., Sakamoto, S., and Mitsuda, N. (2021). In planta cell wall engineering: from mutants to artificial cell walls. *Plant Cell Physiol.* 62 (12), 1813–1827. doi: 10.1093/pcp/pcab157

Yu, D., Janz, D., Zienkiewicz, K., Herrfurth, C., Feussner, I., Chen, S., et al. (2021). Wood formation under severe drought invokes adjustment of the hormonal and transcriptional landscape in poplar. *Int. J. Mol. Sci.* 22 (18), 9899. doi: 10.3390/ijms22189899

Zaheer, M., Rehman, S. U., Khan, S. H., Shahid, S., Rasheed, A., Naz, R., et al. (2021). Characterization of new COBRA like (COBL) genes in wheat (*Triticum aestivum*) and their expression analysis under drought stress. *Mol. Biol. Rep.* 49 (2), 1379–1387. doi: 10.1007/s11033-021-06971-0

Zhang, Y., Nikolovski, N., Sorieul, M., Vellosillo, T., McFarlane, H. E., Dupree, R., et al. (2016). Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in *Arabidopsis*. *Nat. Commun.* 7, 11656. doi: 10.1038/ncomms11656

Zhang, Y.-J., Ren, L.-L., Lin, X.-Y., Han, X.-M., Wang, W., and Yang, Z.-L. (2022). Molecular evolution and functional characterization of chitinase gene family in *Populus trichocarpa*. *Gene* 822, 146329. doi: 10.1016/j.gene.2022.146329

Zhang, J., Xie, M., Tuskan, G. A., Muchero, W., and Chen, J.-G. (2018). Recent advances in the transcriptional regulation of secondary cell wall biosynthesis in the woody plants. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01535

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 November 2023
DOI 10.3389/fpls.2023.1260414

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Deepak Singla,
Punjab Agricultural University, India
M. Kamran Azim,
Mohammad Ali Jinnah University, Pakistan

*CORRESPONDENCE
Vineet K. Sharma
vineetks@iiserb.ac.in

RECEIVED 17 July 2023
ACCEPTED 22 September 2023
PUBLISHED 17 November 2023

CITATION
Chakraborty A, Mahajan S, Bisht MS and Sharma VK (2023) Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties.
Front. Plant Sci. 14:1260414.
doi: 10.3389/fpls.2023.1260414

COPYRIGHT
© 2023 Chakraborty, Mahajan, Bisht and Sharma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties

Abhisek Chakraborty, Shruti Mahajan, Manohar S. Bisht and Vineet K. Sharma*

MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, India

Syzygium cumini, also known as jambolan or jamun, is an evergreen tree widely known for its medicinal properties, fruits, and ornamental value. To understand the genomic and evolutionary basis of its medicinal properties, we sequenced *S. cumini* genome for the first time from the world's largest tree genus *Syzygium* using Oxford Nanopore and 10x Genomics sequencing technologies. We also sequenced and assembled the transcriptome of *S. cumini* in this study. The tetraploid and highly heterozygous draft genome of *S. cumini* had a total size of 709.9 Mbp with 61,195 coding genes. The phylogenetic position of *S. cumini* was established using a comprehensive genome-wide analysis including species from 18 Eudicot plant orders. The existence of neopolyploidy in *S. cumini* was evident from the higher number of coding genes and expanded gene families resulting from gene duplication events compared to the other two sequenced species from this genus. Comparative evolutionary analyses showed the adaptive evolution of genes involved in the phenylpropanoid-flavonoid (PF) biosynthesis pathway and other secondary metabolites biosynthesis such as terpenoid and alkaloid in *S. cumini*, along with genes involved in stress tolerance mechanisms, which was also supported by leaf transcriptome data generated in this study. The adaptive evolution of secondary metabolism pathways is associated with the wide range of pharmacological properties, specifically the anti-diabetic property, of this species conferred by the bioactive compounds that act as nutraceutical agents in modern medicine.

KEYWORDS

genome sequencing, genome assembly and annotation, adaptive evolution, secondary metabolism, medicinal properties

Introduction

Syzygium cumini, also known as jamun, jambolan, or black plum, is a tropical tree belonging to the Myrtaceae plant family. It is native to the Indian subcontinent and South-East Asia, and is known for its wide range of medicinal properties and typical purple-black berries (Srivastava and Chandra, 2013). *Syzygium*, the clove genus, is the world's largest tree genus with 1,193 recognized species. They occupy various habitats, medium to large-sized, typically sub-canopy trees, and thus affect the ecosystems of a wide range of organisms. Some of the other species from *Syzygium* genus are - *S. caryophyllatum*, *S. aromaticum*, *S. aqueum, S. grande*, S. *myrtifolium*, etc., which are used as spices or fruits in pharmacology and horticulture industry (Low et al., 2022).

S. cumini is an evergreen tree with 30 meters of height, 3.6 meters of girth, and up to 15 meters of bole, and can live more than 100 years (Dagadkhair, 2017). This *Syzygium* species is widely cultivated in tropical countries for its edible fruit ("Jamun"), which has significant economic importance (Madani et al., 2021). *S. cumini* was introduced in several tropical and sub-tropical regions of the world for its commercial applications, such as Southern Africa, West Indies, California, and Israel (Chaudhary and Mukhopadhyay, 2012). The purple-black colored fruits of *S. cumini* are rich in anthocyanin, polyphenol, and tannin content and possess high nutrient values and medicinal properties (Ghosh et al., 2017). Besides this, other parts of the tree, such as wood, leaf, flower, seed, and bark also have various economic and medicinal properties (Chaudhary and Mukhopadhyay, 2012).

All the plant parts of *S. cumini* have therapeutic properties, which are used in various treatments since the Ayurvedic era (Dagadkhair, 2017). The extracts from seed, bark, fruit, leaf, and flower of this species contain various phytochemicals such as glucoside jambolin, flavonoids including anthocyanin, terpenoids, and alkaloids (e.g., jambosine), which confer medicinal properties including anti-allergic, anti-oxidant, anti-diarrhoeal, anti-microbial, anti-inflammatory, anti-cancer, and others (Chaudhary and Mukhopadhyay, 2012; Srivastava and Chandra, 2013; Dagadkhair, 2017; Kumar et al., 2023). Specifically, the fruit seed extracts of *S. cumini* have well-known anti-diabetic properties conferred by the glucoside jambolin present in the seeds (Chaudhary and Mukhopadhyay, 2012; Dagadkhair, 2017). Besides, the leaves and bark extracts of *S. cumini* also have anti-diabetic potential due to the presence of phyenylpropanoids (Srivastava and Chandra, 2013; Correia et al., 2023). Further, *S. cumini* fruits also possess anti-hyperlipidemic, hepatoprotective, anti-ulcer, anti-arthritic, anti-fertility, and anti-pyretic activities (Srivastava and Chandra, 2013). Different parts of this tree are also rich in compounds containing isoquercetin, myrecetin, kaemferol, and others (Ayyanar and Subash-Babu, 2012).

Syzygium, the tree genus with the highest number of species, is characterized by rapid speciation events, which resulted in a wide range of ecological and morphological diversity within the genus. A previous study has indicated that an ancient pan-Myrtales Whole Genome Duplication (WGD) event might have contributed to the early stages of diversification in the Myrtales plant order (Low et al., 2022). However, whole genome sequencing of only two *Syzygium* species has been performed until now (Low et al., 2022; Ouadi et al., 2022), and the whole genome and transcriptome assembly of *S. cumini* was not available.

Therefore, in this study, the genome sequencing of *S. cumini* was performed using 10x Genomics linked reads and Oxford Nanopore long reads to assemble its nuclear genome (710 Mbp) and chloroplast genome (158 Kbp). We also report the transcriptome assembly of *S. cumini* for the first time in this study. We inferred *S. cumini* genome to be tetraploid in this study, whereas previous studies have also shown the existence of intraspecific polyploidy (ranging from 2x to 6x) in *S. cumini* species (Ohri, 2015; Sharma et al., 2020). The phylogenetic position of *S. cumini* was resolved with respect to 17 other Eudicot orders, and comparative evolutionary analyses showed the key plant secondary metabolism pathways, such as the phenylpropanoid-flavonoid (PF) biosynthesis pathway, were adaptively evolved in this *Syzygium* species, which are responsible for the immense medicinal properties of this tree.

Materials and methods

Genome and transcriptome sequencing

The clean leaves of *S. cumini* were used for DNA-RNA extraction and species identification (Supplementary Figure 1). DNA extraction was performed using Carlson lysis buffer (Supplementary Text). Quality check and quantification of the extracted DNA were carried out using Nanodrop 8000 spectrophotometer and Qubit 2.0 fluorometer, respectively. Species identification was performed using amplification and Sanger sequencing of the two marker genes - *ITS2* (Internal Transcribed Spacer) and *MatK* (Maturase K), followed by BLASTN of the gene sequences against NCBI non-redundant (nt) database (Supplementary Text). The extracted DNA was used to prepare the 10x Genomics library on the Chromium controller instrument using Chromium Genome Library and Gel Bead Kit (10x Genomics), and sequenced on Illumina NovaSeq 6000 instrument. Further, the DNA was purified using Ampure XP magnetic beads (Beckman Coulter, USA), which was used to prepare the Nanopore library with SQK-LSK109 and SQK-LSK110 library preparation kit (Oxford Nanopore Technologies, UK) for sequencing on a MinION Mk1C sequencer.

The RNA was extracted following a similar method that was used for *Syzygium longifolium* species with a few modifications (Supplementary Text) (Soewarto et al., 2019). Extracted RNA was washed and purified using a RNeasy mini kit (Qiagen, CA, USA). The RNA quality was diluted ten times and Qubit 2.0 fluorometer was used for quantification using a qubit ss RNA HS kit (Life Technologies, United States). Quality of the RNA was evaluated using High Sensitivity D1000 ScreenTape on Agilent 2200 TapeStation (Agilent, Santa Clara, CA). The RNA library was prepared using TruSeq Stranded Total RNA Library Preparation kit with the Ribo-Zero Plant workflow (Illumina Inc., CA, USA). The transcriptome library was sequenced on Illumina NovaSeq 6000 instrument to generate 150 bp paired-end reads.

Genome assembly

Genomic characteristics such as genome size, genomic ploidy, and heterozygosity content were predicted using the 10x Genomics short reads before genome assembly. For this, barcode sequences were filtered using proc10xG (https://github.com/ucdavis-bioinformatics/proc10xG). The barcode-filtered reads were further processed using Trimmomatic v0.39 with the filtering parameters - "TRAILING:20 LEADING:20 MINLEN:60 SLIDINGWINDOW:4:20" (Bolger et al., 2014). The filtered short reads were used to estimate the ploidy level of *S. cumini* genome using Smudgeplot v0.2.2 (Ranallo-Benavidez et al., 2020). Further, these pre-processed reads were used to construct the k-mer count-based histogram with Jellyfish v2.2.10 (Marçais and Kingsford, 2011), which was used to predict the genome size and heterozygosity content with GenomeScope v2 (Ranallo-Benavidez et al., 2020).

Oxford Nanopore long reads were processed to remove adapter sequences using Porechop v0.2.4, which were further error-corrected and *de novo* assembled using Canu v2.2 (Koren et al., 2017). The resultant genome assembly was polished thrice with Pilon v1.23 using the pre-processed 10x Genomics short reads to fix any assembly errors (Walker et al., 2014). Scaffolding was performed with the quality-filtered RNA-Seq reads (filtered using Trimmomatic v0.39), barcode-processed 10x Genomics reads (processed using Longranger basic v2.2.0), and error-corrected Nanopore reads using AGOUTI v0.3, ARCS v1.1.2, and LINKS v1.8.6, respectively (Warren et al., 2015; Zhang et al., 2016; Yeo et al., 2018). Gap-closing of this scaffolded assembly was carried out using the error-corrected Nanopore long reads using LR_Gapcloser (three iterations) (Xu et al., 2018). Finally, the pre-processed 10x Genomics reads were used to again polish the assembly using Pilon v1.23 (Walker et al., 2014), and scaffolds with lengths of ≥5 Kb were retained to construct the final genome assembly of *S. cumini*.

After constructing the final genome assembly, the genomic ploidy level was further verified using nQuire (Weib et al., 2018). The pre-processed linked reads were mapped onto the assembled genome using BWA-MEM v0.7.17 (Li, 2013), and using these alignments, base frequencies were modeled using a Gaussian Mixture Model in nQuire. Log-likelihood values were estimated for each fixed model using the denoised base frequency distribution. The fixed model with the lowest Δlog-likelihood value compared to the free model was considered as the predicted ploidy level.

To assess the genome assembly quality, the pre-processed 10x Genomics linked reads, error-corrected Nanopore reads and quality-filtered RNA-Seq reads were mapped onto the assembled genome to calculate the read mapping percentage using BWA-MEM v0.7.17 (Li, 2013), Minimap v2.17 (Li, 2018), and HISAT v2.2.1 (Kim et al., 2015), respectively. BUSCO v5.4.4 was used to check the presence of single-copy orthologous genes in the final genome assembly with embryophyta_odb10 dataset (Simão et al., 2015). Further, LTR Assembly Index (LAI) score was also calculated to evaluate the genome assembly quality using GenomeTools v1.6.1 and LTR_retriever v2.9.0 (Gremme et al., 2013; Ou and Jiang, 2018).

To identify sequence variation in the *S. cumini* genome assembly, the pre-processed linked reads were mapped using BWA-MEM followed by variant calling using BCFtools "mpileup" v1.9 with the parameters - depth ≥30, variant sites quality ≥30, and mapping quality ≥50 (Li, 2013; Danecek et al., 2021).

Chloroplast genome assembly

The chloroplast genome of *S. cumini* species was assembled with the pre-processed 10x Genomics data using GetOrganelle v1.7.7.0 with embplant_pt as seed database (Jin et al., 2020). The chloroplast genome was annotated using GeSeq in CHLOROBOX with chloroplast genomes of other *Syzygium* species (*S. aromaticum*, *S. forrestii*, *S. jambos*, and *S. malaccense*) available in NCBI RefSeq database as reference sequences (Tillich et al., 2017; Greiner et al., 2019).

Single nucleotide variant (SNV) analysis was performed using the chloroplast genome constructed in this study with three previously reported *S. cumini* chloroplast genomes with the NCBI accessions - GQ870669.3 (Asif et al., 2013), NC_053327.1, and MN095412.1. The chloroplast genomes were aligned using MAFFT v7.310 (Katoh and Standley, 2013), and the alignments were used to identify the single nucleotide variants using DnaSP v6 with a sliding window of 600 bp and a step size of 200 bp (Rozas et al., 2017).

Genome annotation

The whole genome assembly of *S. cumini* was used for constructing a *de novo* repeat library with RepeatModeler v2.0.3 (Flynn et al., 2020), which was used to soft-mask the *S. cumini* genome with RepeatMasker v4.1.2 (http://www.repeatmasker.org). The repeat-masked genome was used to identify the coding genes with MAKER v3.01.04 genome annotation pipeline using AUGUSTUS as the *ab initio* gene predictor (Stanke et al., 2006; Campbell et al., 2014). For evidence-based alignments, *de novo* transcriptome assembly of *S. cumini* constructed in this study using Trinity v2.14.0 (Haas et al., 2013), and protein sequences of the two sequenced species from *Syzygium* genus - *S. aromaticum* and *S. grande* (Low et al., 2022; Ouadi et al., 2022) and other species from Myrtales order (*Eucalyptus grandis* and *Corymbia citriodora*) available in Ensembl plants release 56 were used (Bolser et al., 2016). A high-confidence coding gene set was constructed with the filtering criteria of AED value <0.5 and coding gene length ≥150 nucleotides.

The completeness of this coding gene set was evaluated using BUSCO v5.4.4 with embryophyta_odb10 database (Simão et al., 2015). Gene expression values (TPM) were also estimated by mapping the quality-filtered RNA-Seq data of *S. cumini* onto the coding gene set (nucleotides) using Kallisto v0.48.0 (Bray et al., 2016).

The genome assembly of *S. cumini* was used for prediction of non-coding RNAs. *de novo* prediction of rRNA and tRNA was performed using Barrnap v0.9 (https://github.com/tseemann/barrnap) and tRNAscan-SE v2.0.9 (Chan et al., 2021), respectively. miRNA sequences were identified using BLASTN (sequence identity 80% and e-value 10^{-9}) against the miRBase database (Griffiths-Jones et al., 2008).

Collinearity analysis

MCScanX was used to analyze the intra-species collinearity for *S. cumini* species using the BLASTP homology alignments of coding genes and GFF annotations (Wang et al., 2012). Further, inter-species collinear blocks were identified between *S. cumini* and *S. grande*, *S. grande* and *S. aromaticum*, and *S. aromaticum* and *S. cumini* using previously available data (Low et al., 2022; Ouadi et al., 2022). Gene duplication analysis was also performed for the three *Syzygium* species using MCScanX.

Phylogenetic analysis

For constructing the species phylogenetic tree, one Eudicot species from each available plant order (except Myrtales) in Ensembl plants release 56 was considered (Bolser et al., 2016). For Myrtales order, both the species available in Ensembl plants release 56 – *E. grandis* and *C. citriodora* were considered, along with *S. grande* and *S. aromaticum* from previous studies (Low et al., 2022; Ouadi et al., 2022). The selected species from other 17 Eudicot plant orders were – *Arabidopsis thaliana* (Brassicales), *Actinidia chinensis* (Ericales), *Beta vulgaris* (Caryophyllales), *Citrus clementina* (Sapindales), *Cucumis sativus* (Cucurbitales), *Coffea canephora* (Gentianales), *Cynara cardunculus* (Asterales), *Daucus carota* (Apiales), *Gossypium raimondii* (Malvales), *Juglans regia* (Fagales), *Kalanchoe fedtschenkoi* (Saxifragales), *Rosa chinensis* (Rosales), *Populus trichocarpa* (Malpighiales), *Sesamum indicum* (Lamiales), *Solanum tuberosum* (Solanales), *Vigna radiata* (Fabales), and *Vitis vinifera* (Vitales). Alongside, *Zea mays* was considered as an outgroup species.

Proteome files of these 23 species with the longest isoforms for each protein were used for orthogroups construction with OrthoFinder v2.5.4 (Emms and Kelly, 2019). The orthogroups were filtered to extract the fuzzy one-to-one orthogroups using KinFin v1.1 (Laetsch and Blaxter, 2017). Only those orthogroups comprising sequences from all 23 species were considered, and each orthogroup was processed to include only one longest sequence per species. The resultant orthogroups were individually aligned using MAFFT v7.310 (Katoh and Standley, 2013), and filtered and concatenated with BeforePhylo v0.9.0 (https://github.com/qiyunzhu/BeforePhylo). The species phylogenetic tree was constructed with this concatenated alignment using maximum likelihood-based RAxML v8.2.12 with 100 bootstrap values and "PROTGAMMAAUTO" model (Stamatakis, 2014).

Analysis of gene family evolution

Proteome files of 23 species with the longest isoforms for each protein were used to analyze the expansion/contraction of gene families with CAFÉ v5 (Mendes et al., 2020). Homology-based search results obtained from All-versus-All BLASTP with the protein sequences of all 23 species were clustered using MCL v14.137 (Van Dongen and Abreu-Goodger, 2012). Gene families containing genes from <2 species of the specified clades and ≥100 gene copies for ≥1 species were filtered out as per the suggestions for performing CAFÉ analysis. An ultrametric species phylogenetic tree across the 23 species was constructed using the calibration point for *S. cumini* and *B. vulgaris* (118 years), as reported in the TimeTree database v5 (Kumar et al., 2022). The ultrametric species tree and the filtered gene families were used in the two-lambda (λ) model implemented in CAFÉ v5, where species from Myrtales order were indicated separate λ-value compared to the other species.

Identification of secondary metabolite biosynthesis-related gene families and biosynthetic gene clusters

For identification of secondary metabolism-related gene families, the protein sequences of the genes involved in secondary metabolite biosynthesis pathways were downloaded from UniProt or KEGG databases for *Arabidopsis thaliana* or other closely related Eudicot species (Kanehisa, 2002; Bateman et al., 2023). The protein sequences were mapped against the *S. cumini* protein sequences using BLASTP with e-value 10^{-5}, and the annotations were further verified from the assigned KO (Kegg Orthology) IDs. These secondary metabolite biosynthesis genes of *S. cumini* were identified in the filtered gene families used for CAFÉ analysis, and these families were further analyzed for gene family expansion/contraction. Biosynthetic gene clusters (BGCs) were identified in *S. cumini* genome using plantiSMASH v1.0 with CD-HIT filtering cut-off 0.5 and "–inclusive", "–borderpredict", "–all-orfs", and "–smcogs" parameters (Kautsar et al., 2017).

Identification of evolutionary signatures in *S. cumini* genes

Comparative analysis was performed to identify evolutionary signatures in *S. cumini* genes across 13 Eudicot species including *S. cumini*. Four other species from Myrtales order itself (*S. grande*, *S. aromaticum*, *E. grandis*, and *C. citriodora*), and species from its closer plant orders were considered for the analysis. Species from other plant orders used in these analyses were – *V. radiata* (order Fabales), *C. sativus* (order Cucurbitales), *J. regia* (order Fagales), *R. chinensis* (order Rosales), *P. trichocarpa* (order Malpighiales), *A. thaliana* (order Brassicales), *G. raimondii* (order Malvales), and *C. clementina* (order Sapindales).

Unique amino acid substitution with functional impact

Protein sequences of the 13 species were used for orthogroups construction with OrthoFinder v2.5.4 (Emms and Kelly, 2019). Orthogroups comprising sequences from the 13 species were extracted, and each orthogroup was filtered to retain the longest sequence for each species. The resultant orthogroups were individually aligned with MAFFT v7.310, and from these multiple sequence alignments *S. cumini* genes were identified that showed

different amino acids in positions where the other species had the same amino acid. In this analysis, gaps in the alignments and ten positions around the gaps were not considered. Further, impact of the unique substitutions on the protein function was predicted with SIFT using UniProt as a reference database (Ng and Henikoff, 2003).

Higher nucleotide divergence

The protein sequence alignments for the orthogroups obtained in the previous step were used for orthogroup-specific phylogenetic tree construction with RAxML v8.2.12 using 100 bootstrap values and "PROTGAMMAAUTO" model (Stamatakis, 2014). Using the gene phylogenetic trees, *S. cumini* genes showing greater branch length values compared to the genes from other species were identified using "adephylo" package in R, and were termed as the genes with higher nucleotide divergence (Jombart and Dray, 2010).

Positive selection

The orthogroups constructed across 13 species (nucleotide sequences) were individually aligned with MAFFT v7.310 (Katoh and Standley, 2013). The resultant multiple sequence alignments and the species tree of 13 species (constructed with RAxML) were used to detect the positively selected genes in *S. cumini* with a branch-site model in "codeml" of PAML v4.10.6 (Yang, 2007). Likelihood-ratio test (LRT) and chi-square analysis were performed, and genes qualifying against the null model with FDR-corrected p-values of < 0.05 were identified as the genes showing positive selection in *S. cumini*. Further, Bayes Empirical Bayes analysis was performed to detect the genes with codon sites under positive selection (with >95% probability) for the foreground lineage.

S. cumini genes with more than one of the evolutionary signatures – unique substitution with functional impact, positive selection, and higher nucleotide divergence were termed as the genes showing multiple signatures of adaptive evolution (MSA) (Agaba et al., 2016; Jaiswal et al., 2021).

Functional annotation

S. cumini coding gene set was annotated against publicly available databases - Swiss-Prot database using BLASTP (e-value 10^{-5}), NCBI-nr database using BLASTP (e-value 10^{-5}), and Pfam-A database using HMMER v3.1 (e-value 10^{-5}) (Bairoch and Apweiler, 2000; Bateman, 2004; Finn et al., 2011). *S. cumini* coding gene set, including the expanded gene families and the genes with evolutionary signatures, were annotated using eggNOG-mapper v2.1.9 and KAAS v2.1 genome annotation servers (Moriya et al., 2007; Huerta-Cepas et al., 2017). Over-representation analysis using WebGestalt web server was performed to assign Gene Ontology (GO) categories to the MSA genes of *S. cumini* (Liao et al., 2019).

Gene structure analysis

The key genes associated with phenylpropanoid-flavonoid (PF) biosynthesis pathway and terpenoid biosynthesis pathway were identified in *S. cumini* genome from the functional annotation of the coding genes. Gene families were identified from the CAFÉ analysis, and the longest gene for each gene family was extracted. The genes were mapped separately onto *S. cumini* genome constructed in this study and the previously available *S. aromaticum* (Ouadi et al., 2022) and *S. grande* (Low et al., 2022) genomes using Exonerate v2.4.0 (https://github.com/nathanweeks/exonerate) to construct the exon-intron structures (Chakraborty et al., 2023), and for a comparative analysis across the three *Syzygium* species.

Identification of plant disease susceptible genes in *S. cumini* genome

To identify the disease susceptible genes (S-genes) in *S. cumini* genome, the coding genes were mapped against the DSP (Disease Susceptibility Genes in Plants) database consisting of 448 S-genes using BLASTN with query coverage 80%, sequence identity 80%, and e-value of 10^{-9} (Kaur et al., 2023).

Results

Genome assembly

Species identification was performed using *matK* and *ITS2* marker gene sequencing, which showed 99.65% and 99.7% sequence similarity (the best hits), respectively, with *S. cumini* gene sequences available in NCBI non-redundant nucleotide (nt) database. 120.7 Gb of 10x Genomics data and 14.4 Gb Oxford Nanopore data (read N50 = 10.9 Kb) were generated for genome assembly. Based on the predicted genome size of 730.3 Mbp (using GenomeScope), the genomic data corresponded to 165.3x and 19.7x sequencing coverage for 10x Genomics and Nanopore reads, respectively (Supplementary Table 1). Additionally, 15.1 Gb RNA-Seq data was also sequenced from the *S. cumini* leaf tissue.

S. cumini genome contained 3.25% heterozygosity (Supplementary Figure 2) and was inferred as a tetraploid genome since the distribution of base frequencies at the variable sites showed the smallest Δlog-likelihood value for the tetraploid fixed model (Supplementary Figure 3A). The heterozygous k-mer pair distribution showed that 87% of the k-mers represented the total coverage of k-mer pair 4n (Supplementary Figure 3B) (Ranallo-Benavidez et al., 2020).

S. cumini genome assembly constructed using Canu (Koren et al., 2017) had a size of 706.9 Mbp with 9,704 contigs, N50 value of 102.1 Kb, and 95.4% BUSCO completeness. However, after three rounds of assembly polishing, the BUSCO completeness was improved to 98.1%. After genome assembly scaffolding and other post-processing steps, the final genome assembly had a size of 709.9 Mbp containing 7,702 scaffolds with an N50 value of 179.2 Kb, and the largest scaffold size of 1.6 Mb. The improvement in the genome assembly statistics after each assembly process step is mentioned in Supplementary Table 2.

S. cumini genome showed the presence of 98.3% complete BUSCOs (64.6% complete and single-copy, and 33.7% complete and duplicated) (Supplementary Table 3). The genome assembly also had an LAI (LTR Assembly Index) score of 11.69. Further, 97.83% of barcode-filtered 10x Genomics reads, 93.45% error-corrected Nanopore reads, and 95.25% quality-filtered RNA-Seq reads were mapped onto the genome assembly. A total of 6,184,849 base positions (0.87%) in the genome assembly had sequence variations.

Chloroplast genome assembly

The chloroplast genome assembly of *S. cumini* showed a circular genome of 158,509 bases with 83 protein-coding genes (Supplementary Figure 4). The assembled genome size was comparable to the *S. malaccense* chloroplast genome (Tao et al., 2020). However, the previously reported *S. cumini* chloroplast genome size (Asif et al., 2013) was larger compared to this study, and also larger than the other two *S. cumini* chloroplast genomes available at NCBI database (NCBI accessions: NC_053327.1 and MN095412.1) (Supplementary Table 4). The comparative statistics of the sizes of Large Single Copy (LSC), Small Single Copy (SSC), and Inverted Repeat (IR) regions of *S. cumini* chloroplast genomes in this study and the previous study (Asif et al., 2013) and the chloroplast genome of *S. malaccense* (Tao et al., 2020) are mentioned in Supplementary Table 4. The LSC region in the *S. cumini* chloroplast genome provided by Asif et al., 2013 was larger compared to the other genomes available at NCBI database, the genome constructed in this study, and the *S. malaccense* chloroplast genome (Tao et al., 2020).

SNV analysis between the *S. cumini* chloroplast genome sequences constructed in this study and the previous report (Asif et al., 2013) showed the presence of 978 positions with nucleotide variations and a nucleotide diversity (π) value of 0.00620. LSC (π = 0.00837) and SSC (π = 0.00843) regions showed higher intraspecies variability compared to the IR-a (π = 0.00169) and IR-b (π = 0.00169) regions, similar to other studies (Silva et al., 2019). However, the SNV analysis of *S. cumini* chloroplast genome sequence constructed in this study with two other *S. cumini* chloroplast genomes available at NCBI database revealed the presence of a lower number of single nucleotide variants - 282 SNVs with π = 0.00179 (NCBI accession - NC_053327.1) and 197 SNVs with π = 0.00125 (NCBI accession - MN095412.1) (Supplementary Figure 5).

S. cumini chloroplast genome constructed in this study had 83 protein-coding genes, same as the previous study (Asif et al., 2013). Additionally, three pseudogenes were present in *S. cumini* chloroplast genome (Asif et al., 2013), among which *ycf15* could not be annotated in our study using GeSeq in CHLOROBOX tool (Tillich et al., 2017). However, we could map the *ycf15* gene sequence of *S. cumini* (Asif et al., 2013) in the chloroplast genome of this study at the same genomic location with 100% identity and 100% query coverage using BLASTN.

Genome annotation

A *de novo* repeat library consisting of 2,521 sequences for *S. cumini* genome was used to repeat mask 51.51% of the genome assembly. Among the repeat classes, 49.31% were interspersed repeats, including 8.09% Gypsy and 5.37% Copia elements (Supplementary Table 5). Using the repeat-masked genome assembly, coding genes were predicted in *S. cumini* genome using the MAKER pipeline (Campbell et al., 2014).

A total of 204,525 transcripts were assembled with an N50 value of 2,313 bp (Supplementary Table 6), which were used as empirical evidence along with the protein sequences of species from Myrtales order during coding genes prediction. 74,657 coding genes were predicted, among which 62,971 were retained (84.35%) after AED-based filtering. Further, length-based filtering was performed to retain 61,195 coding genes in the final high-confidence gene set with an average CDS length of 1,106.2 bp. 32,888 of the coding genes (53.74%) showed relatively higher gene expression (TPM values > 1). Distribution of the coding genes in various KEGG pathways and COG categories are mentioned in Supplementary Tables 7, 8.

This coding gene set showed the presence of 92.8% BUSCOs (78.7% complete and 14.1% fragmented) (Supplementary Table 3). 92.38% of the genes were annotated using any of the three publicly available databases – NCBI-nr, Pfam-A, and Swiss-Prot (Supplementary Table 9). 1,174 tRNAs decoding standard amino acids, 702 rRNAs, and 176 miRNAs were also identified in the assembled genome of *S. cumini*. Further, seven disease susceptible genes (S-genes) were identified in the *S. cumini* coding gene set (Supplementary Table 10).

Collinearity and orthologous gene clustering

Synteny analysis showed the presence of 16.76% intra-species collinearity in *S. cumini* genome. Further, 90.55% of the coding genes were originated by duplication events, whereas, gene duplication analysis using previously available data of *Syzygium* species showed a lesser percentage of duplicated genes in *S. grande* (75.94%) and *S. aromaticum* (85.06%) (Low et al., 2022; Ouadi et al., 2022). Inter-species collinearity analysis showed a higher percentage of collinear genes between *S. grande* and *S. aromaticum* than *S. cumini* and *S. grande*, and *S. cumini* and *S. aromaticum* (Supplementary Table 11). Further, a higher percentage of *S. cumini* genes and a higher number of collinear blocks were present in the inter-species collinear blocks constructed between *S. cumini* and *S. grande*, compared to *S. cumini* and *S. aromaticum* (Supplementary Table 11). 17,882 *S. cumini* genes (29.22%) were present in the inter-species collinear blocks constructed with both *S. aromaticum* and *S. grande*, indicating their conserveness. The distribution of the 17,882 genes in KEGG pathways is mentioned in Supplementary Table 12.

Gene clustering among *S. cumini* and four other species from Myrtales order showed a large number of species-specific gene

clusters in *S. cumini* (2,891 clusters) compared to other species (Figure 1). 3,980 gene clusters were common between *S. cumini* and *S. grande*, and 839 common gene clusters were identified between *S. cumini* and *S. aromaticum*. Genes included in the species-specific gene clusters of *S. cumini* (15,721 genes) were involved in various KEGG pathways mentioned in Supplementary Table 13.

Phylogenetic position of *S. cumini*

1,465 one-to-one fuzzy orthogroups were identified across 23 species spanning 18 Eudicot plant orders. Filtered and concatenated sequence alignments of the orthogroups containing 1,248,870 alignment positions were used to construct the species phylogenetic tree with *Zea mays* as the outgroup species.

In the phylogenetic tree, *S. cumini* was found in a position closer to *S. grande* (in the same clade) compared to *S. aromaticum* (Figure 2), which can further be explained by a higher number of collinear blocks and a higher number of shared gene clusters present between *S. cumini* and *S. grande*, compared to *S. cumini* and *S. aromaticum* (Figure 1; Supplementary Table 11). Among all the core Eudicot species in our phylogenetic tree, the species from the Saxifragales plant order (*K. fedtschenkoi*) diverged the earliest. The relative phylogenetic positions of the Eudicot orders were similar to the previous studies (Soltis et al., 2000; Chakraborty et al., 2022).

Gene family evolution

A total of 17,366 filtered gene families were identified across 23 species. Among these, 4,587 gene families were expanded, and 473 gene families were contracted in *S. cumini* species. The number of expanded gene families was much higher than that of *S. grande* and *S. aromaticum* (Figure 2). Among the expanded gene families, 41 families were highly expanded (>25 expanded genes) in *S. cumini* (Supplementary Table 14). The highly expanded gene families were involved in secondary metabolism-related pathways, such as phenylpropanoid and flavonoid biosynthesis (Supplementary Table 15).

Genes with evolutionary signatures

8,583 orthogroups across 13 species were constructed to identify the *S. cumini* genes with evolutionary signatures. 1,630 genes were

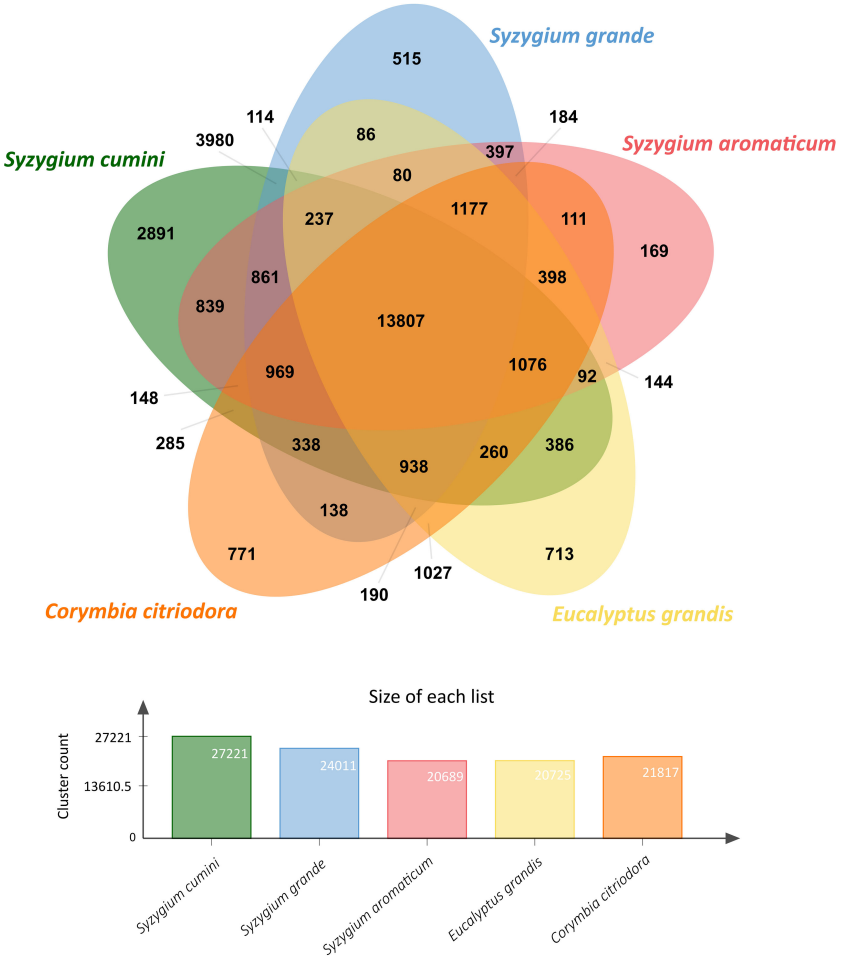


FIGURE 1
Orthologous gene clusters among *S. cumini* and other species from Myrtales plant order.

10.3389/fpls.2023.1260414

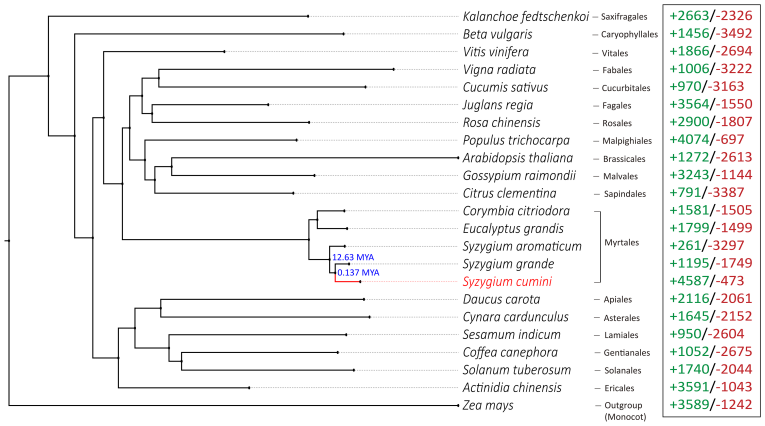


FIGURE 2
Phylogenetic position of *S. cumini* with respect to Eudicot species from Myrtales and 17 other plant orders. *Zea mays* was used as an outgroup species. Numbers mentioned in the nodes denote the divergence times of *Syzygium* species obtained from TimeTree v5 database (Kumar et al., 2022). Numbers in green and red represent the number of expanded and contracted gene families in each species, respectively.

positively selected (p-value < 0.05), 1,113 genes had unique amino acid substitution with functional impact, and 135 genes showed higher nucleotide divergence (Supplementary Tables 16–18). Among these genes, 430 genes had more than one signature of adaptive evolution (MSA genes). 333 of the MSA genes were also supported by gene expression (TPM > 1) data obtained in this study (Supplementary Data 1). GO categories of the *S. cumini* MSA genes are mentioned in Supplementary Table 19.

Adaptive evolution of genes involved in secondary metabolism pathways

Plant secondary metabolites are derived from the primary metabolites and mainly function in the interaction of plants with their environment, abiotic and biotic stress tolerance, and are responsible for the medicinal properties of plants. The main classes of plant secondary metabolites are terpenoids, phenolic compounds, and alkaloids. The phenylpropanoid-flavonoid (PF) biosynthesis pathway is the key pathway for producing a wide range of phenolic compounds, such as flavonol, lignin, and anthocyanin (Taheri et al., 2022). Flavonoids and phenylpropanoids were the most abundant bioactive compounds in the fruit extracts of *S. cumini*, showing a wide range of pharmacological activities and can be used as preventive measures in many diseases, including type-2 diabetes (Correia et al., 2023).

PF biosynthesis pathway

The shikimate pathway is responsible for the production of the precursors of phenylpropanoids along with tannins and other major groups of phenolic compounds (Salminen and Karonen, 2011). Genes involved in the shikimate pathway showed evolutionary signatures – *DAHPS* was identified as an MSA gene (TPM > 1), and *EPSPS* showed unique substitution with functional impact. Six (*DAHPS*, *DHQD*, *SDH*, *SK*, *EPSPS*, and *CS*) of the seven genes involved in the shikimate pathway also showed gene family expansion in *S. cumini* (Figure 3; Supplementary Table 20).

In the common phenylpropanoid pathway (conversion of phenylalanine to *p*-coumaroyl CoA), all three genes *PAL*, *C4H* and *4CL* showed gene family expansion. *PAL* and *4CL* genes were highly expressed (TPM > 1), and *4CL* also showed unique substitution with functional impact. In the downstream phenylpropanoid pathway, *HCT*, *COMT*, *CAD*, *CCR*, and peroxidase genes had highly expanded gene families (with >25 expanded genes). *CCR* and *F5H* were identified as MSA genes (TPM > 1). Peroxidase and *CAD* genes showed unique substitution with functional impact. *C3'H* showed all three evolutionary signatures and gene family expansion, and *CCoAOMT* gene family was contracted (Figure 4; Supplementary Table 20).

p-coumaroyl CoA, formed in the phenylpropanoid pathway, is also a precursor to flavonoid (e.g., flavonol and anthocyanin) and lignin biosynthesis. Enzymes involved in the biosynthesis steps were also adaptively evolved in *S. cumini* species. *FLS* and *CHS* were found among the MSA genes (TPM > 1), and gene families of these two genes showed high expansion and contraction, respectively. Gene families of *CHI*, *F3'H*, and *F3H* were expanded, and *F3H* also had unique substitution with functional impact. All these genes had high gene expression (TPM > 1). Further, gene families of *DFR* and *ANS*, the two major enzymes for anthocyanin biosynthesis, were highly expanded in *S. cumini* (Figure 4).

Terpenoid biosynthesis pathway

Terpenoids are another important class of plant secondary metabolites. Fruits and flowers of *S. cumini* are rich in terpenoids (Chaudhary and Mukhopadhyay, 2012; Correia et al., 2023), and the terpenoids present in the *S. cumini* leaves can be used to treat inflammatory diseases (Siani et al., 2013). In support of this, adaptive evolution in the genes involved in terpenoid backbone biosynthesis pathways was observed. *AACT*, *HMGR*, *MK*, *MDD*, *GPPS*, and *GGPPS* had expanded gene families (Supplementary Table 20). Further, *HMGS* was positively selected, and *GPPS*

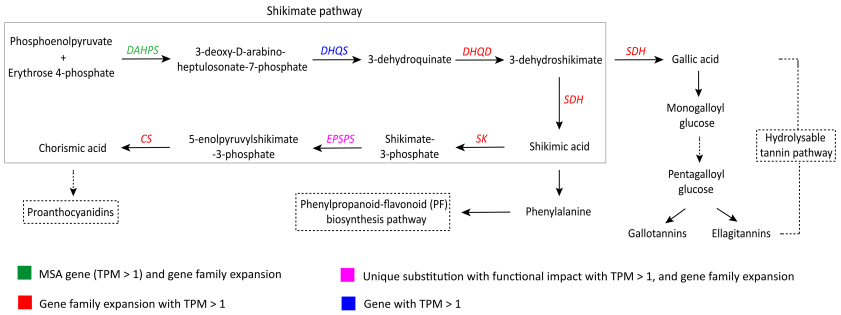


FIGURE 3
Adaptive evolution of the shikimate pathway in *S. cumini* (Salminen and Karonen, 2011; Mora et al., 2022). *DAHPS*, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; *DHQS*, Dehydroquinate synthase; *DHQD*, 3-dehydroquinate dehydratase; *SDH*, Shikimate dehydrogenase; *SK*, Shikimate kinase; *EPSPS*, 5-enolpyruvylshikimate 3-phosphate synthase; *CS*, Chorismate synthase.

showed higher nucleotide divergence in *S. cumini*. *FOLK* gene, responsible for forming farnesyl diphosphate from farnesol, was found among the MSA genes (TPM > 1) along with gene family expansion. Further, the terpenoid backbone biosynthesis pathway is the precursor to the formation of monoterpenes, triterpenes, sesquiterpenes, other terpenoid-quinones, carotenoids, etc. Enzymes involved in these pathways such as *FDFT1*, *SQLE*, neomenthol dehydrogenase, *menA*, *HST*, and *crtB* also showed adaptive evolution in *S. cumini* species (Figure 5).

Using STRING database (v11.5) (Szklarczyk et al., 2021), protein-protein interaction was examined in the genes belonging to phenylpropanoid-flavonoid (PF) pathway and terpenoid biosynthesis pathway. Only the MSA genes (with TPM > 1) and the genes in highly expanded genes families of *S. cumini* were considered for the above analysis. Based on the protein-protein interaction evidence available on the STRING database, two clusters were formed by the terpenoid and phenolic compounds biosynthesis-related genes, and an association between these clusters was

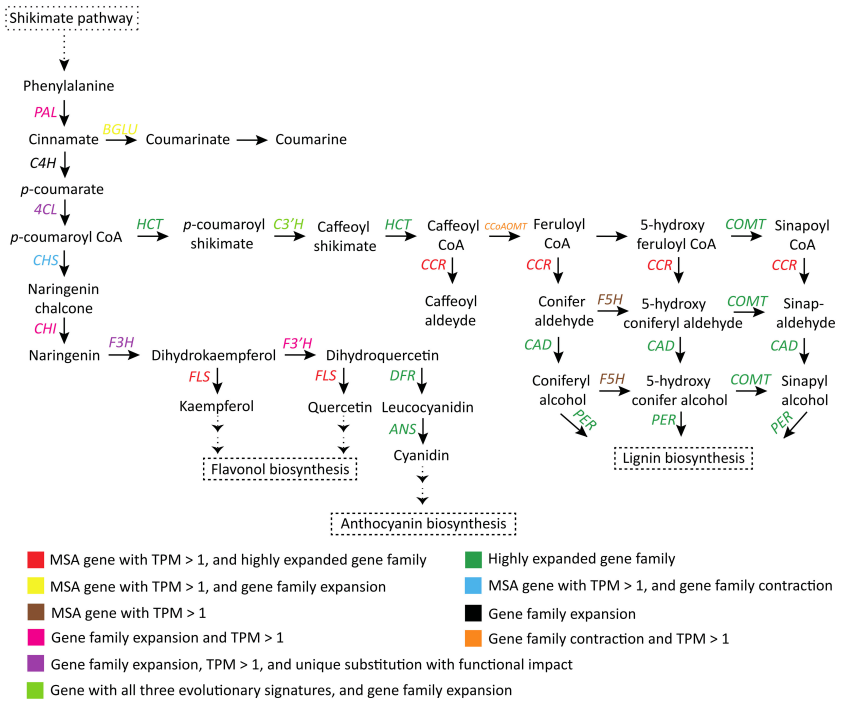


FIGURE 4
Adaptive evolution of the phenylpropanoid-flavonoid (PF) biosynthesis pathway (Yadav et al., 2020; Taheri et al., 2022). *PAL*, Phenylalanine ammonia lyase; *C4H*, Cinnamic acid 4-hydroxylase; *4CL*, 4-hydroxycinnamoyl-CoA ligase; *HCT*, Shikimate O-hydroxycinnamoyltransferase; *C3'H*, p-coumaroyl shikimate 3'-hydroxylase; *CCoAOMT*, Caffeoyl-CoA O-methyltransferase; *COMT*, Caffeic acid 3-O-methyltransferase; *CCR*, Cinnamoyl-CoA reductase; *F5H*, Ferulate-5-hydroxylase; *CAD*, Cinnamyl-alcohol dehydrogenase; *PER*, Peroxidase; *CHS*, Chalcone synthase; *CHI*, Chalcone isomerase; *F3H*, Flavanone 3-hydroxylase; *F3'H*, Flavonoid 3'-monooxygenase; *FLS*, Flavonol synthase; *DFR*, Dihydroflavonol 4-reductase; *ANS*, Anthocyanidin synthase.

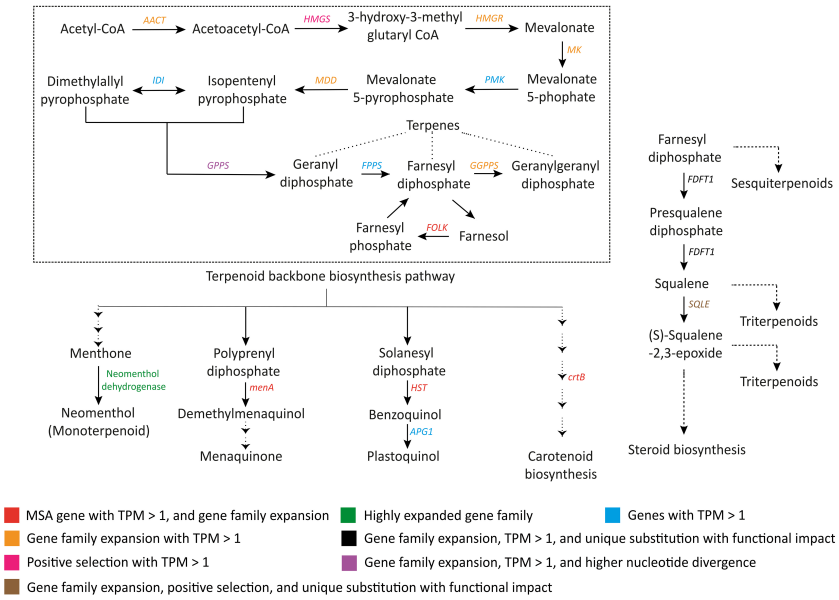


FIGURE 5
Evolutionary signatures in terpenoid and other terpenoid-quinone biosynthesis genes. *AACT*, Acetoacetyl-CoA thiolase; *HMGS*, HMG-CoA synthase; *HMGR*, HMG-CoA reductase; *MK*, Mevalonate kinase; *PMK*, Phosphomevalonate kinase; *MDD*, Mevalonate-5-diphosphate decarboxylase; *IDI*, Isopentenyl diphosphate isomerase; *GPPS*, Geranyl diphosphate synthase; *FPPS*, Farnesyl diphosphate synthase; *GGPPS*, Geranylgeranyl diphosphate synthase; *FOLK*, Farnesol kinase; *menA*, 1;4-dihydroxy-2-naphthoate polyprenyltransferase; *HST*, Homogentisate solanesyltransferase; *APG1*, MPBQ/MSBQ methyltransferase; *crtB*, 15-cis-phytoene synthase; *FDFT1*, Farnesyl-diphosphate farnesyltransferase 1; *SQLE*, Squalene monooxygenase.

observed mediated by *CHS* (involved in PF biosynthesis pathway) and *crtB* (involved in terpenoid biosynthesis) indicating a functional relationship between the two pathways (Figure 6).

Alkaloid and other secondary metabolites biosynthesis

The pharmacological activities of the alkaloids provide essential health benefits through the fruits and other plant parts of *S. cumini* (Srivastava and Chandra, 2013). Genes (*TAT*, *TYDC*, *NCS*, *6OMT*, *CNMT*, *4OMT*, *BBE*, *SOMT*, *CAS*, *STOX*, and *CoOMT*) involved in the benzylisoquinoline alkaloid (BIA) biosynthesis pathway, a diverse group of alkaloids with numerous medicinal properties, showed gene family expansion (Figure 7, Supplementary Table 20). *GOT2* gene involved in isoquinoline and tropane alkaloid biosynthesis pathways was found among the MSA genes (with TPM > 1) (Zhou and Chen, 2022). Further, *AAE* gene family and another 'GDSL' lipolytic family involved in indole alkaloid biosynthesis were highly expanded (Ruppert et al., 2005).

Among other notable adaptively evolved genes involved in secondary metabolites biosynthesis, *F6H* (highly expanded gene family) functions in scopoletin biosynthesis (Kai et al., 2008), *PRR1* (MSA gene with TPM > 1) helps in lignan biosynthesis (Hemmati et al., 2007), and *BX1* (MSA gene with TPM > 1) is involved in glucoside (via benzoxazinoid biosynthesis pathway) production (Frey et al., 2009).

Plant BGCs in *S. cumini* genome

Plants produce immensely diverse specialized metabolites, such as secondary metabolites, that function in ecological interactions and possess nutritional and medicinal importance. Genes encoding these biosynthetic pathways are often clustered in a genomic locus known as the biosynthetic gene cluster (BGC) (Kautsar et al., 2017). 39 BGCs were identified in *S. cumini* genome containing a total of 562 plant biosynthetic genes, which were involved in KEGG pathways such as phenylpropanoid biosynthesis, cell cycle, plant hormone signal transduction, terpenoid biosynthesis, etc. (Figure 8; Supplementary Table 21). Among the *S. cumini* genes in the BGCs, 29 key genes involved in secondary metabolites biosynthesis were present in one or more copies, and 35 genes showed at least one of the three signatures of adaptive evolution, namely, positive selection, higher nucleotide divergence, and unique substitution with functional impact (Supplementary Tables 22, 23). However, not all secondary metabolite biosynthesis pathway genes are present in clusters; therefore, the biosynthetic genes that are not present in clusters will not be identified in the BGCs (Shi and Xie, 2014).

Adaptive evolution of genes associated with stress tolerance mechanisms

MSA genes of *S. cumini* were also involved in various biotic (such as pathogen resistance and defense against herbivores) and abiotic (such as ROS scavenging, heat, drought, and salinity, etc.) stress tolerance mechanisms (Supplementary Data 2). Among the key genes with MSA involved in biotic stress tolerance responses, *GI* downregulates salicylic acid accumulation and alters the phenylpropanoid pathway, thus reducing PR (Pathogenesis-Related) gene expression and negatively affecting biotic defense

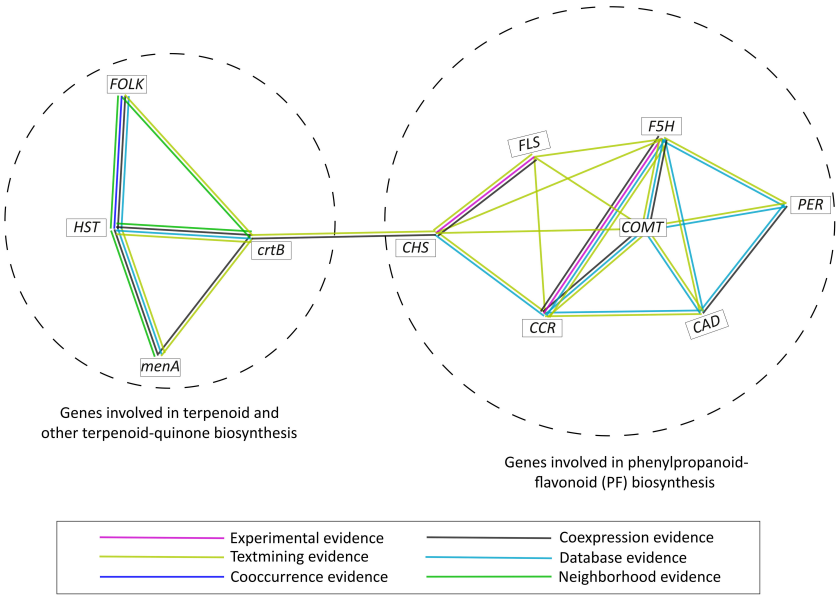


FIGURE 6
Protein-protein interaction among the MSA genes (TPM > 1) and highly expanded gene families involved in phenylpropanoid-flavonoid (PF) biosynthesis and terpenoid biosynthesis pathways in *S. cumini*.

responses (Kundu and Sahu, 2021). *BSK* provides resistance against bacterial and fungal pathogens by playing a role in pattern-triggered immunity (PTI) (Majhi et al., 2019), *NPR1* is a crucial regulator of salicylic acid signaling and triggers immune responses by inducing PR genes (Chen et al., 2019), *MPK3* responds to biotic stress by upregulating jasmonic acid signaling and negatively regulating salicylic acid accumulation (Jagodzik et al., 2018), *PIK1* also acts in pathogen recognition and activation of defense responses (Romeis, 2001).

Among the major genes with MSA involved in abiotic stress tolerance responses, *ABF* regulates the expression of abscisic acid-responsive genes to provide salinity, drought, and osmotic stress tolerance to plants (Feng et al., 2019), *MPAO* facilitates oxidative burst-mediated programmed cell death to aid plant defense responses (Yoda et al., 2006), *KUP* K^+ transporter family is involved in potassium deficiency and salt and drought stress response (Yang et al., 2020), Heat shock transcription factor (Hsf) regulates oxidative stress response by directly sensing the reactive oxygen species (ROS) (Miller and Mittler, 2006). Besides these, *LOX* confers abiotic (drought, salinity, etc.) and biotic stress tolerance (Viswanath et al., 2020), and *CNGC* has multifaceted functions in plants, such as pathogen resistance and abiotic (salt, drought, cold, etc.) stress tolerance (Guo et al., 2018).

Discussion

In this study, we performed whole genome sequencing of *S. cumini* species and constructed a draft genome assembly for the first time. It is only the third and till date the largest genome to be sequenced from the largest tree genus containing approximately 1,200 species. *S. cumini* was previously reported to show intraspecific polyploidy compared to *S. aromaticum* and *S. grande* (Ohri, 2015). Our analyses using two independent approaches to estimate the genomic ploidy also confirmed the tetraploidy in *S. cumini* genome. Further, the genome was found to be highly heterozygous (3.25%), and a combination of polyploidy and high heterozygosity increases the genomic complexity in this species. Polyploidy causes difficulty in haplotype resolving (Kyriakidou et al., 2018) and a higher percentage of allelic differences (1% or above) also poses a challenge in genome assembly (Asalone et al., 2020). Despite of this genomic complexity, we could successfully construct the whole genome assembly of *S. cumini* with the assembled genome size close to the predicted genome size.

We used multiple approaches such as BUSCO assessment, LAI score estimation, and read mapping percentage calculation to evaluate the genome assembly quality. 98.3% complete BUSCOs in the genome assembly suggest a near-complete genome assembly. LAI score of 11.69 indicates that the genome assembly constructed in this study can be considered as a "Reference" quality assembly (Ou et al., 2018). LAI score of *S. cumini* genome constructed in this study was also similar to the other chromosome-scale plant genome assemblies such as *Angelica sinensis* (Han et al., 2022). A high percentage of mapped reads onto the genome assembly further attests to the assembly quality. Further, the usage of strict parameters, AED cut-off <0.5, and coding gene length ≥150 bp in the MAKER pipeline underscores the quality of the high-confidence coding genes. The presence of 92.8% BUSCOs in the coding gene set also suggests the near-completeness of the genome annotation performed in this study. Further, the complete structures (exon-intron number and gene length) of the *S. cumini* key genes involved in PF biosynthesis and terpenoid biosynthesis pathways could be identified, which was similar to the other two high-quality genome assemblies of *Syzygium* species, that also attests to the quality of *S.*



FIGURE 7
Adaptive evolution of genes involved in the benzylisoquinoline alkaloid (BIA) biosynthesis pathway (Deng et al., 2018). *GOT2*, Aspartate aminotransferase; *TAT*, Tyrosine aminotransferase; *TYDC*, Tyrosine decarboxylase; *NCS*, (S)-Norcoclaurine synthase; *6OMT*, Norcoclaurine 6-O-methyltransferase; *CNMT*, (S)-Coclaurine N-methyltransferase; *NMCH*, (S)-N-Methylcoclaurine 3'-hydroxylase; *4OMT*, 3'-Hydroxy-N-methylcoclaurine 4'-O-methyltransferase; *BBE*, Berberine bridge enzyme; *SOMT*, (S)-Scoulerine 9-O-methyltransferase; *CAS*, (S)-Canadine synthase; *STOX*, (S)-Tetrahydroprotoberberine oxidase; *CoOMT*, Columbamine O-methyltransferase.

cumini genome assembly (Supplementary Tables 24, 25, Supplementary Text).

We also noted a high percentage of complete and duplicated BUSCOs (D score) in the genome assembly and coding gene set of *S. cumini* (Supplementary Table 3), which is perhaps due to the additional neopolyploidy event following the Pan-Myrtales WGD event in *S. cumini* species (tetraploid) compared to the other *Syzygium* species that remained at the same ploidy level (Low et al., 2022). This event could also be the reason for an increased genome size of *S. cumini* compared to *S. grande* and *S. aromaticum* (Low et al., 2022; Ouadi et al., 2022). The increase in genome size also appears to be due to an expansion in copy number (37% higher) of LTR-RT repeat elements in *S. cumini* genome compared to *S. grande*, and an overall 6.4% and 8.1% higher repeat content compared to *S. grande* and *S. aromaticum* genomes, respectively (Supplementary Table 5) (Low et al., 2022; Ouadi et al., 2022; Zhu et al., 2023). The neopolyploidy event in *S. cumini* might also be the cause of a greater number of coding genes (61,195 genes), a greater fraction of genes (90.55%) originated from duplicated events, a higher number of gene clusters (27,221), and a higher number of species-specific gene clusters (2,891) observed in *S. cumini* species compared to *S. grande* and *S. aromaticum* (Figure 1). Duplicated genes may either undergo deletion or pseudogenization due to relaxed selection pressure (Wang et al., 2018) or acquire novel functions (Panchy et al., 2016), which could also be the case in *S. cumini* species as observed in the adaptive evolution of secondary metabolism pathways.

Further, the consideration of species from Myrtales order (including two other *Syzygium* species) and species from its closer phylogenetic orders for comparative analysis to identify the genes with evolutionary signatures in *S. cumini* helped to reduce the false positives that could have resulted from the greater genetic distance of the selected species. The genes with evolutionary signatures identified from genomic analyses were also supported by transcriptomic analysis. It is important to note that MSA genes and the genes from highly expanded gene families were found to be majorly involved in secondary metabolite biosynthesis pathways such as phenylpropanoids, flavonoids, alkaloids, and terpenoids, which are responsible for the medicinal properties of this species.

Phenylpropanoids play essential roles in plant development, response to abiotic and biotic stress signals, and biosynthesis of a

10.3389/fpls.2023.1260414

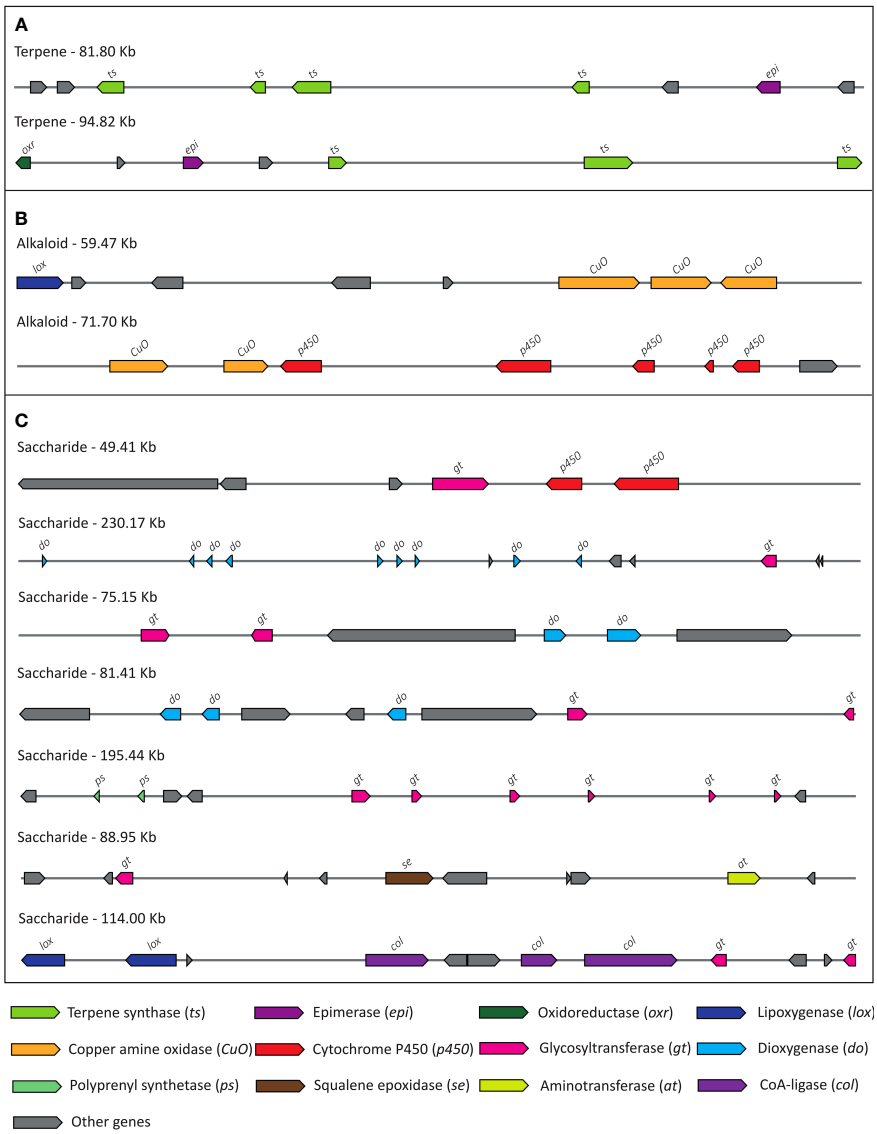


FIGURE 8
The biosynthetic gene clusters (BGCs) annotated in *S. cumini* genome. **(A)** Terpene gene clusters, **(B)** Alkaloid gene clusters, **(C)** Saccharide gene clusters.

broad spectrum of secondary metabolites (Vogt, 2010). Phenylpropanoid-derived metabolites contribute to the biosynthesis of several other secondary metabolites, such as lignin and lignan, isoflavonoid, coumarin, stilbene, anthocyanin, isoquercetin, myrecetin, and kaemferol, which confer numerous pharmacological properties in *S. cumini* species (Ayyanar and Subash-Babu, 2012; Chaudhary and Mukhopadhyay, 2012; Srivastava and Chandra, 2013). One particular class of flavonoids - anthocyanin, is responsible for the purple-black color of the fruits of *S. cumini* and their health benefits (Chaudhary and Mukhopadhyay, 2012). Phenolic compounds (e.g., catechin, gallic acid, etc.) extracted from *S. cumini* seeds have immense potential as anti-diabetic and anti-oxidant agents, that have found commercial significance as nutraceutical ingredients in modern medicine and can be used as a substitution of allopathic remedies for chronic diseases such as type-2 diabetes (Mahindrakar and Rathod, 2021; Kumar et al., 2023). One of the main findings of this study was the identification of evolutionary signatures and gene family evolution of all the key *S. cumini* genes involved in the PF biosynthesis pathway (Figure 4). It is an important finding because the evolutionary signatures and evolution in gene families have been recognized as critical mechanisms shaping natural variation for species adaptation, which might also be the case in this species (Guo, 2013; Chakraborty et al., 2021). Further, 741 genes were present in the expanded gene families of the PF biosynthesis pathway, among which 98.9% of the genes originated from different modes of duplication, which function in increasing the dosage of gene products and in accelerating the metabolic flux for rate-limiting steps in such biosynthetic pathways (Conant and Wolfe, 2007; Bekaert et al., 2011). Taken together, the adaptive evolution of PF biosynthesis pathway and its precursor shikimate pathway in *S. cumini* could be responsible for their numerous

therapeutic properties, specifically the anti-diabetic property conferred by the seeds and leaves (Srivastava and Chandra, 2013).

Notably, the comparative evolutionary analyses revealed 14 key genes involved in the biosynthesis of terpenoids and other terpenoid-quinone compounds to show adaptive evolution in *S. cumini* (Figure 5; Supplementary Data 1). Terpenoids are a structurally diverse class of secondary metabolites responsible for plant defense responses against herbivores and pathogens, and are abundant in *S. cumini* fruits contributing to the anti-oxidant and anti-inflammatory properties (Cheng et al., 2007). Other terpenoid-quinone compounds also function in plant stress tolerance responses (Liu and Lu, 2016) and show pharmacological activities (Gordaliza, 2012). Thus, the adaptive evolution of terpenoid biosynthesis pathway could explain the anti-inflammatory properties of *S. cumini* leaves and seeds conferred by the terpenoids (Siani et al., 2013; Srivastava and Chandra, 2013).

Among the other classes of secondary metabolites, alkaloids present in different plant parts of *S. cumini* are pharmaceutically diverse secondary metabolites with curative properties against many diseases (Ziegler and Facchini, 2008). Alkaloids, along with flavonoids and tannins also confer anti-arthritic property to the *S. cumini* seeds (Srivastava and Chandra, 2013). Glucosides are also critical secondary metabolites for plant defense responses and possess therapeutic properties (Bennett and Wallsgrove, 1994). *S. cumini* seed extracts contain alkaloid jambosine, and glucoside jambolin that prevents the conversion step of starch into sugar (anti-diabetic), which is the most significant therapeutic property of this species (Ayyanar and Subash-Babu, 2012). In this study, genes related to alkaloid and glucoside biosynthesis (e.g., *GOT2*, *AAE*, *BX1*, etc.) showed adaptive evolution in *S. cumini* that emphasizes the genomic basis for its pharmacological properties.

It is important to mention that the secondary metabolites are produced and regulated in response to various abiotic and biotic stresses, and aid in better survival of the plants and confer their medicinal properties (Isah, 2019). Here, we also noted that various biotic and abiotic stress tolerance response genes displayed multiple signatures of adaptive evolution in *S. cumini* (Supplementary Data 2). Further, KEGG pathways related to phenylpropanoid, flavonoid, terpenoid, alkaloid biosynthesis, plant hormone signal transduction, and plant-pathogen interaction were found in all the gene sets that showed the evolutionary signatures and gene family expansion (Supplementary Tables 15–18; Supplementary Data 1). Taken together, it is tempting to speculate that the adaptive evolution of major plant secondary metabolism pathways in *S. cumini* species confers unprecedented anti-diabetic, anti-oxidant, anti-inflammatory, and other pharmacological properties of this tree. Further, the whole genome sequence of *S. cumini* will facilitate future genomic, evolutionary, and ecological studies on the world's largest tree genus.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author. Raw reads obtained from the genome and transcriptome sequencing have been deposited in NCBI SRA database under the BioProject accession - PRJNA982613 and BioSample accession - SAMN35711647.

Author contributions

AC: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. SM: Methodology, Validation, Writing – original draft, Writing – review & editing. MSB: Data curation, Investigation, Writing – review & editing. VKS: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

AC and SM thank Council of Scientific and Industrial Research (CSIR) for the research fellowship. MSB thanks Ministry of Education, Govt. of India for Prime Minister Research Fellowship (PMRF). The authors also thank the Sanger sequencing and NGS facilities at IISER Bhopal and the intramural research funds provided by IISER Bhopal.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1260414/full#supplementary-material

References

Agaba, M., Ishengoma, E., Miller, W. C., McGrath, B. C., Hudson, C. N., Bedoya Reina, O. C., et al. (2016). Giraffe genome sequence reveals clues to its unique morphology and physiology. *Nat. Commun.* 7, 1–8. doi: 10.1038/ncomms11519

Asalone, K. C., Ryan, K. M., Yamadi, M., Cohen, A. L., Farmer, W. G., George, D. J., et al. (2020). Regional sequence expansion or collapse in heterozygous genome assemblies. *PloS Comput. Biol.* 16 (7), e1008104. doi: 10.1371/journal.pcbi.1008104

Asif, H., Khan, A., Iqbal, A., Khan, I. A., Heinze, B., and Azim, M. K. (2013). The chloroplast genome sequence of Syzygium cumini (L.) and its relationship with other angiosperms. *Tree Genet. Genomes* 9, 867–877. doi: 10.1007/S11295-013-0604-1/FIGURES/6

Ayyanar, M., and Subash-Babu, P. (2012). Syzygium cumini (L.) Skeels: A review of its phytochemical constituents and traditional uses. *Asian Pac. J. Trop. Biomed.* 2, 240. doi: 10.1016/S2221-1691(12)60050-1

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28 (1), 45–48. doi: 10.1093/nar/28.1.45

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al (2004). The Pfam protein families database. *Nucleic Acids Res.* 32 (suppl_1), D138–D141. doi: 10.1093/nar/gkh121

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., et al. (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi: 10.1093/NAR/GKAC1052

Bekaert, M., Edger, P. P., Chris Pires, J., and Conant, G. C. (2011). Two-phase resolution of polyploidy in the arabidopsis metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell* 23, 1719–1728. doi: 10.1105/TPC.110.081281

Bennett, R. N., and Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytol.* 127, 617–633. doi: 10.1111/J.1469-8137.1994.TB02968.X

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolser, D., Staines, D. M., Pritchard, E., and Kersey, P. (2016). "Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomics data," in *Methods in molecular biology*. (Springer, United States: Humana Press). 115–140. doi: 10.1007/978-1-4939-3167-5_6

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 345 (34), 525–527. doi: 10.1038/nbt.3519

Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinf.* 48 (1), 4–11. doi: 10.1002/0471250953.bi0411s48

Chakraborty, A., Bisht, M. S., Saxena, R., Mahajan, S., Pulikkan, J., and Sharma, V. K. (2023). Genome sequencing and *de novo* and reference-based genome assemblies of Bos indicus breeds. *Genes Genomics* 1, 1–10. doi: 10.1007/S13258-023-01401-W

Chakraborty, A., Mahajan, S., Bisht, M. S., and Sharma, V. K. (2022). Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa species reveal evolutionary mechanisms of longevity. *iScience* 25, 105100. doi: 10.1016/J.ISCI.2022.105100

Chakraborty, A., Mahajan, S., Jaiswal, S. K., and Sharma, V. K. (2021). Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* 41 (4), 1–12. doi: 10.1038/s42003-021-02720-y

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). TRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49 (16), 9077–9096. doi: 10.1093/nar/gkab688

Chaudhary, B., and Mukhopadhyay, K. (2012). Syzygium cumini (L.) skeels: a potential source of nutraceuticals. *Int. J. Pharm. Biol. Sci.* 2, 46–53.

Chen, J., Mohan, R., Zhang, Y., Li, M., Chen, H., Palmer, I. A., et al. (2019). NPR1 promotes its own and target gene expression in plant defense by recruiting CDK8. *Plant Physiol.* 181, 289–304. doi: 10.1104/PP.19.00124

Cheng, A. X., Lou, Y. G., Mao, Y. B., Lu, S., Wang, L. J., and Chen, X. Y. (2007). Plant terpenoids: Biosynthesis and ecological functions. *J. Integr. Plant Biol.* 49, 179–186. doi: 10.1111/J.1744-7909.2007.00395.X

Conant, G. C., and Wolfe, K. H. (2007). Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* 3, 129. doi: 10.1038/MSB4100170

Correia, V. T., da, V., Silva, V. D. M., Mendonça, H., de, O. P., Ramos, A. L. C. C., et al. (2023). Efficiency of Different Solvents in the Extraction of Bioactive Compounds from Plinia cauliflora and Syzygium cumini Fruits as Evaluated by Paper Spray Mass Spectrometry. *Mol.* 28, 2359. doi: 10.3390/MOLECULES28052359

Dagadkhair, A. C. (2017). Jamun (Syzygium cumini) Skeels: A Traditional Therapeutic Tree and its Processed Food Products. *Int. J. Pure Appl. Biosci.* 5, 1202–1209. doi: 10.18782/2320-7051.4011

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi: 10.1093/gigascience/giab008

Deng, X., Zhao, L., Fang, T., Xiong, Y., Ogutu, C., Yang, D., et al. (2018). Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. *Hortic. Res.* 51 (5), 1–16. doi: 10.1038/s41438-018-0035-0

Emms, D. M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20 (1), 1–14. doi: 10.1186/s13059-019-1832-y

Feng, R. J., Ren, M. Y., Lu, L. F., Peng, M., Guan, X., Zhou, D. B., et al. (2019). Involvement of abscisic acid-responsive element-binding factors in cassava (Manihot esculenta) dehydration stress response. *Sci. Rep.* 91 (9), 1–12. doi: 10.1038/s41598-019-49083-3

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39 (suppl_2), W29–W37. doi: 10.1093/nar/gkr367

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117 (17), 9451–9457. doi: 10.1073/pnas.1921046117

Frey, M., Schullehner, K., Dick, R., Fiesselmann, A., and Gierl, A. (2009). Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70, 1645–1651. doi: 10.1016/J.PHYTOCHEM.2009.05.012

Ghosh, P., Pradhan, R. C., Mishra, S., Patel, A. S., and Kar, A. (2017). Physicochemical and nutritional characterization of jamun (Syzygium Cuminii). *Curr. Res. Nutr. Food Sci.* 5, 25–35. doi: 10.12944/CRNFSJ.5.1.04

Gordaliza, M. (2012). Synthetic strategies to terpene quinones/hydroquinones. *Mar. Drugs* 10, 358. doi: 10.3390/MD10020358

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Gremme, G., Steinbiss, S., and Kurtz, S. (2013). Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 10 (3), 645–656, doi: 10.1109/TCBB.2013.68

Griffiths-Jones, S., Saini, H. K., Van Dongen, S., and Enright, A. J. (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* 36 (suppl_1), D154–D158. doi: 10.1093/nar/gkm952

Guo, Y. L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. *Plant J.* 73, 941–951. doi: 10.1111/TPJ.12089

Guo, J., Islam, M. A., Lin, H., Ji, C., Duan, Y., Liu, P., et al. (2018). Genome-wide identification of cyclic nucleotide-gated ion channel gene family in wheat and functional analyses of TaCNGC14 and TaCNGC16. *Front. Plant Sci.* 9. doi: 10.3389/FPLS.2018.00018/BIBTEX

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512. doi: 10.1038/nprot.2013.084

Han, X., Li, C., Sun, S., Ji, J., Nie, B., Maker, G., et al. (2022). The chromosome-level genome of female ginseng (Angelica sinensis) provides insights into molecular mechanisms and evolution of coumarin biosynthesis. *Plant J.* 112, 1224–1237. doi: 10.1111/TPJ.16007

Hemmati, S., Schmidt, T. J., and Fuss, E. (2007). (+)-Pinoresinol/(–)-lariciresinol reductase from Linum perenne Himmelszelt involved in the biosynthesis of justicidin B. *FEBS Lett.* 581, 603–610. doi: 10.1016/J.FEBSLET.2007.01.018

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34 (8), 2115–2122. doi: 10.1093/molbev/msx148

Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52, 39. doi: 10.1186/s40659-019-0246-3

Jagodzik, P., Tajdel-Zielinska, M., Ciesla, A., Marczak, M., and Ludwikow, A. (2018). Mitogen-activated protein kinase cascades in plant hormone signaling. *Front. Plant Sci.* 9, 1387. doi: 10.3389/fpls.2018.01387

Jaiswal, S. K., Mahajan, S., Chakraborty, A., Kumar, S., and Sharma, V. K. (2021). The genome sequence of Aloe vera reveals adaptive evolution of drought tolerance mechanisms. *iScience* 24, 102079. doi: 10.1016/j.isci.2021.102079

Jin, J. J., Yu, W., Yang, J. B., Song, Y., Depamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 1–31. doi: 10.1186/s13059-020-02154-5

Jombart, T., and Dray, S. (2010). Adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics* 26 (15), 1907–1909. doi: 10.1093/bioinformatics/btq292

Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., et al. (2008). Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. *Plant J.* 55, 989–999. doi: 10.1111/J.1365-313X.2008.03568.X

Kanehisa, M. (2002). The KEGG database. *Novartis Found. Symp.* 247, 91–103. doi: 10.1002/0470857897.CH8. Subramaniam.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010

Kaur, S., Bishnoi, R., Priyadarshini, P., Singla, D., and Chhuneja, P. (2023). DSP: database of disease susceptibility genes in plants. *Funct. Integr. Genomics* 23, 1–4. doi: 10.1007/S10142-023-01132-X/FIGURES/1

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., and Medema, M. H. (2017). plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* 45, W55–W63. doi: 10.1093/NAR/GKX305

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.* 27 (5), 722–736. doi: 10.1101/gr.215087.116

Kumar, S., Sharma, S., Kumar, V., Sharma, A., Kaur, R., and Saini, R. (2023). Jamun (Syzygium cumini (L.) Skeels): The conventional underutilized multifunctional plant-an exotic gleam into its food and functional significance. *Ind. Crops Prod.* 191, 115873. doi: 10.1016/J.INDCROP.2022.115873

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., et al. (2022). TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39, 1–6. doi: 10.1093/molbev/msac174

Kundu, P., and Sahu, R. (2021). GIGANTEA confers susceptibility to plants during spot blotch attack by regulating salicylic acid signalling pathway. *Plant Physiol. Biochem.* 167, 349–357. doi: 10.1016/J.PLAPHY.2021.02.006

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. V. (2018). Current strategies of polyploid plant genome sequence assembly. *Front. Plant Sci.* 871. doi: 10.3389/FPLS.2018.01660/BIBTEX

Laetsch, D. R., and Blaxter, M. L. (2017). KinFin: Software for taxon-aware analysis of clustered protein sequences. *G3 Genes Genomes Genet.* 7 (10), 3349–3357. doi: 10.1534/g3.117.300233

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997. arXiv preprint*, United States. doi: 10.48550/arXiv.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47 (W1), W199–W205. doi: 10.1093/nar/gkz401

Liu, M., and Lu, S. (2016). Plastoquinone and ubiquinone in plants: Biosynthesis, physiological function and metabolic engineering. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01898

Low, Y. W., Rajaraman, S., Tomlin, C. M., Ahmad, J. A., Ardi, W. H., Armstrong, K., et al. (2022). Genomic insights into rapid speciation within the world's largest tree genus Syzygium. *Nat. Commun.* 131 (13), 1–15. doi: 10.1038/s41467-022-32637-x

Madani, B., Mirshekari, A., Yahia, E. M., Golding, J. B., Hajivand, S., and Dastjerdy, A. M. (2021). Jamun (Syzygium cumini L. Skeels): A promising fruit for the future. *Hortic. Rev.* 48, 275–306. doi: 10.1002/9781119750802.CH5

Mahindrakar, K. V., and Rathod, V. K. (2021). Antidiabetic potential evaluation of aqueous extract of waste Syzygium cumini seed kernel's by in *vitro* α-amylase and α-glucosidase inhibition. *Prep. Biochem. Biotechnol.* 51, 589–598. doi: 10.1080/10826068.2020.1839908

Majhi, B. B., Sreeramulu, S., and Sessa, G. (2019). BRASSINOSTEROID-SIGNALING KINASE5 associates with immune receptors and is required for immune responses. *Plant Physiol.* 180, 1166–1184. doi: 10.1104/PP.18.01492

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27 (6), 764–770. doi: 10.1093/bioinformatics/btr011

Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022

Miller, G., and Mittler, R. (2006). Could heat shock transcription factors function as hydrogen peroxide sensors in plants? *Ann. Bot.* 98, 279–288. doi: 10.1093/AOB/MCL107

Mora, J., Pott, D. M., Osorio, S., and Vallarino, J. G. (2022). Regulation of plant tannin synthesis in crop species. *Front. Genet.* 13. doi: 10.3389/FGENE.2022.870976

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35 (suppl_2), W182–W185. doi: 10.1093/nar/gkm321

Ng, P. C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31 (13), 3812–3814. doi: 10.1093/nar/gkg509

Ohri, D. (2015). How small and constrained is the genome size of angiosperm woody species. *Silvae Genet.* 64, 20–32. doi: 10.1515/sg-2015-0002

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 46 (21), e126–e126. doi: 10.1093/nar/gky730

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* 176 (2), 1410–1422. doi: 10.1104/pp.17.01310

Ouadi, S., Sierro, N., Goepfert, S., Bovet, L., Glauser, G., Vallat, A., et al. (2022). The clove (Syzygium aromaticum) genome provides insights into the eugenol biosynthesis pathway. *Commun. Biol.* 51 (5), 1–13. doi: 10.1038/s42003-022-03618-z

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/PP.16.00523

Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11 (1), 1–10. doi: 10.1038/s41467-020-14998-3

Romeis, T. (2001). Protein kinases in the plant defence response. *Curr. Opin. Plant Biol.* 4, 407–414. doi: 10.1016/S1369-5266(00)00193-X

Rozas, J., Ferrer-mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-onsins, E., et al. (2017). DnaSP 6 : DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34 (12), 3299–3302. doi: 10.1093/molbev/msx248

Ruppert, M., Woll, J., Giritch, A., Genady, E., Ma, X., and Stöckigt, J. (2005). Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plant-virus expression system. *Planta* 222, 888–898. doi: 10.1007/s00425-005-0031-0

Salminen, J. P., and Karonen, M. (2011). Chemical ecology of tannins and other phenolics: we need a change in approach. *Funct. Ecol.* 25, 325–338. doi: 10.1111/J.1365-2435.2010.01826.X

Sharma, N., Singh, B., Wani, M. S., Gupta, R. C., and Habeeb, T. H. (2020). Morphological, cytological, and chemotypic variation of essential oils in syzygium cumini (L.) skeels. *Anal. Chem. Lett.* 10, 609–619. doi: 10.1080/22297928.2020.1862703

Shi, M.-Z., and Xie, D.-Y. (2014). Biosynthesis and metabolic engineering of anthocyanins in arabidopsis thaliana. *Recent Pat. Biotechnol.* 8, 47–60. doi: 10.2174/1872208307666131218123538

Siani, A. C., Souza, M. C., Henriques, M. G. M. O., and Ramos, M. F. S. (2013). Anti-inflammatory activity of essential oils from *Syzygium cumini* and *Psidium guajava*. *Pharmaceutical Biology* 51 (7), 881–887. doi: 10.3109/13880209.2013.768675

Silva, S. R., Pinheiro, D. G., Penha, H. A., Płachno, B. J., Michael, T. P., Meer, E. J., et al. (2019). Intraspecific Variation within the Utricularia amethystina Species Morphotypes Based on Chloroplast Genomes. *Int. J. Mol. Sci.* 20 (24), 6130. doi: 10.3390/IJMS20246130

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Soewarto, J., Hamelin, C., Bocs, S., Mournet, P., Vignes, H., Berger, A., et al. (2019). Transcriptome data from three endemic Myrtaceae species from New Caledonia displaying contrasting responses to myrtle rust (Austropuccinia psidii). *Data Br.* 22, 794. doi: 10.1016/J.DIB.2018.12.080

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., et al. (2000). Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn. Soc.* 133 (4), 381–461. doi: 10.1006/bojl.2000.0380

Srivastava, S., and Chandra, D. (2013). Pharmacological potentials of Syzygium cumini: a review. *J. Sci. Food Agric.* 93, 2084–2093. doi: 10.1002/JSFA.6111

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.* 34 (suppl_2), W435–W439. doi: 10.1093/nar/gkl200

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/NAR/GKAA1074

Taheri, S., Teo, C. H., Heslop-Harrison, J. S., Schwarzacher, T., Tan, Y. S., Wee, W. Y., et al. (2022). Genome assembly and analysis of the flavonoid and phenylpropanoid biosynthetic pathways in fingerroot ginger (Boesenbergia rotunda). *Int. J. Mol. Sci.* 23, 7269. doi: 10.3390/IJMS23137269/S1

Tao, L., Shi, Z. G., and Long, Q. Y. (2020). Complete chloroplast genome sequence and phylogenetic analysis of *Syzygium malaccense*. *Mitochondrial DNA Part B* 5 (3), 3567–3568. doi: 10.1080/23802359.2020.1829132

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391

Van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Bacterial molecular networks: Methods and protocols*, 281–295. doi: 10.1007/978-1-61779-361-5_15

Viswanath, K. K., Varakumar, P., Pamuru, R. R., Basha, S. J., Mehta, S., and Rao, A. D. (2020). Plant lipoxygenases and their role in plant physiology. *J. Plant Biol.* 63, 83–95. doi: 10.1007/S12374-020-09241-X/FIGURES/3

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/MP/SSP106

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9 (11), e112963. doi: 10.1371/journal.pone.0112963

Wang, P., Moore, B. M., Panchy, N. L., Meng, F., Lehti-Shiu, M. D., and Shiu, S. H. (2018). Factors influencing gene family size variation among related species in a plant family, solanaceae. *Genome Biol. Evol.* 10, 2596. doi: 10.1093/GBE/EVY193

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7), e49. doi: 10.1093/nar/gkr1293

Warren, R. L., Yang, C., Vandervalk, B. P., Behsaz, B., Lagman, A., Jones, S. J., et al. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* 4 (1), s13742–13015-10076-13743. doi: 10.1186/s13742-015-0076-3

Weib, C. L., Pais, M., Cano, L. M., Kamoun, S., and Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinf.* 19 (1), 1–8. doi: 10.1186/s12859-018-2128-z

Xu, G. C., Xu, T. J., Zhu, R., Zhang, Y., Li, S. Q., Wang, H. W., et al. (2019). LR-Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* 8 (1), giy157. doi: 10.1093/gigascience/giy157

Yadav, V., Wang, Z., Wei, C., Amo, A., Ahmed, B., Yang, X., et al. (2020). Phenylpropanoid pathway engineering: an emerging approach towards plant defense. *Pathog.* 9, 312. doi: 10.3390/PATHOGENS9040312

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24 (8), 1586–1591. doi: 10.1093/molbev/msm088

Yang, T., Lu, X., Wang, Y., Xie, Y., Ma, J., Cheng, X., et al. (2020). HAK/KUP/KT family potassium transporter genes are involved in potassium deficiency and stress responses in tea plants (Camellia sinensis L.): Expression and functional analysis. *BMC Genomics* 21, 1–18. doi: 10.1186/S12864-020-06948-6/FIGURES/9

Yeo, S., Coombe, L., Warren, R. L., Chu, J., and Birol, I. (2018). ARCS: Scaffolding genome drafts with linked reads. *Bioinformatics* 34 (5), 725–731. doi: 10.1093/bioinformatics/btx675

Yoda, H., Hiroi, Y., and Sano, H. (2006). Polyamine oxidase is one of the key elements for oxidative burst to induce programmed cell death in tobacco cultured cells. *Plant Physiol.* 142, 193. doi: 10.1104/PP.106.080515

Zhang, S. V., Zhuo, L., and Hahn, M. W. (2016). AGOUTI: Improving genome assembly and annotation using transcriptome data. *Gigascience* 5 (1), s13742–13016-10136-13743. doi: 10.1186/s13742-016-0136-3

Zhou, P., and Chen, M. (2022). Exploration of the mechanisms of differential indole alkaloid biosynthesis in dedifferentiated and cambial meristematic cells of catharanthus roseus using transcriptome sequencing. *Front. Genet.* 13. doi: 10.3389/FGENE.2022.867064/BIBTEX

Zhu, S., Zhang, X., Ren, C., Xu, X., Comes, H. P., Jiang, W., et al. (2023). Chromosome-level reference genome of Tetrastigma hemsleyanum (Vitaceae) provides insights into genomic evolution and the biosynthesis of phenylpropanoids and flavonoids. *Plant J.* 114 (4), 805–823. doi: 10.1111/TPJ.16169

Ziegler, J., and Facchini, P. J. (2008). Alkaloid biosynthesis: metabolism and trafficking. *Annual Review of Plant Biology* 59, 735–769. doi: 10.1146/ANNUREV.ARPLANT.59.032607.092730

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 December 2023
DOI 10.3389/fpls.2023.1282661

Check for updates

OPEN ACCESS

EDITED BY
Yunpeng Cao,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Junxiang Zhang,
Shenyang Agricultural University, China
Xiangpeng Leng,
Qingdao Agricultural University, China

*CORRESPONDENCE
Arnaud G. Bovy
arnaud.bovy@wur.nl
Jianjun Zhao
jjz1971@aliyun.com
Xueping Chen
chenxueping@hebau.edu.cn

†These authors have contributed equally to this work

RECEIVED 24 August 2023
ACCEPTED 24 November 2023
PUBLISHED 11 December 2023

CITATION
Luo L, Molthoff J, Li Q, Liu Y, Luo S, Li N, Xuan S, Wang Y, Shen S, Bovy AG, Zhao J and Chen X (2023) Identification of candidate genes associated with less-photosensitive anthocyanin phenotype using an EMS mutant (*pind*) in eggplant (*Solanum melongena* L.). *Front. Plant Sci.* 14:1282661. doi: 10.3389/fpls.2023.1282661

COPYRIGHT
© 2023 Luo, Molthoff, Li, Liu, Luo, Li, Xuan, Wang, Shen, Bovy, Zhao and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of candidate genes associated with less-photosensitive anthocyanin phenotype using an EMS mutant (*pind*) in eggplant (*Solanum melongena* L.)

Lei Luo[1†], Jos Molthoff[2†], Qiang Li[1], Ying Liu[3], Shuangxia Luo[1], Na Li[1], Shuxin Xuan[1], Yanhua Wang[1], Shuxing Shen[1], Arnaud G. Bovy[2*], Jianjun Zhao[1*] and Xueping Chen[1*]

[1]State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory of Vegetable Germplasm Innovation and Utilization of Hebei, Collaborative Innovation Center of Vegetable Industry in Hebei, International Joint R & D Center of Hebei Province in Modern Agricultural Biotechnology, College of Horticulture, Hebei Agricultural University, Baoding, China, [2]Plant Breeding, Wageningen University and Research, Wageningen, Netherlands, [3]Horticulture and Product Physiology, Wageningen University and Research, Wageningen, Netherlands

Eggplant (*Solanum melongena* L.) is a highly nutritious and economically important vegetable crop. However, the fruit peel of eggplant often shows poor coloration owing to low-light intensity during cultivation, especially in the winter. The less-photosensitive varieties produce anthocyanin in low light or even dark conditions, making them valuable breeding materials. Nevertheless, genes responsible for anthocyanin biosynthesis in less-photosensitive eggplant varieties are not characterized. In this study, an EMS mutant, named *purple in the dark* (*pind*), was used to identify the key genes responsible for less-photosensitive coloration. Under natural conditions, the peel color and anthocyanin content in *pind* fruits were similar to that of wildtype '14-345'. The bagged *pind* fruits were light purple, whereas those of '14-345' were white; and the anthocyanin content in the *pind* fruit peel was significantly higher than that in '14-345'. Genetic analysis revealed that the less-photosensitive trait was controlled by a single dominant gene. The candidate gene was mapped on chromosome 10 in the region 7.72 Mb to 11.71 Mb. Thirty-five differentially expressed genes, including 12 structural genes, such as *CHS*, *CHI*, *F3H*, *DFR*, *ANS*, and *UFGT*, and three transcription factors *MYB113*, *GL3*, and *TTG2*, were identified in *pind* using RNA-seq. Four candidate genes *EGP21875* (myb domain protein 113), *EGP21950* (unknown protein), *EGP21953* (CAAX amino-terminal protease family protein), and *EGP21961* (CAAX amino-terminal protease family protein) were identified as putative genes associated with less-photosensitive anthocyanin biosynthesis in *pind*. These findings may clarify the molecular mechanisms underlying less-photosensitive anthocyanin biosynthesis in eggplant.

KEYWORDS

eggplant, EMS mutants, anthocyanin, less-photosensitive, MutMap

1 Introduction

Anthocyanins are natural flavonoid pigments, which are responsible for blue, purple and red colors in plant tissues, such as flowers, fruits, seed coats, leaves, and stems. In addition to their role in coloring, anthocyanins attract pollinators and seed disperser and protect plants from biotic and abiotic stresses (Fan et al., 2016; Sivankalyani et al., 2016; Sun et al., 2018; Naing and Kim, 2021). Anthocyanins also show beneficial effects on human health, including antioxidation, antimutagenicity, cardiovascular disease prevention, liver protection, and the inhibition of tumor cell metastasis (Pojer et al., 2013; Liu et al., 2016).

Anthocyanins are produced through the phenylpropanoid pathway, which has been studied in many plants, such as petunia and *Arabidopsis* (Krol et al., 1990; Dooner et al., 1991; Timothy and Holton, 1995; Baudry et al., 2004). The anthocyanin biosynthesis is catalyzed stepwise by a series of enzymes, including phenylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumarate coenzyme A ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonoid 3′-hydroxylase (F3′H), flavonoid 3′,5′-hydroxylase (F3′5′H), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), and UDP-glucose: flavonoid-3-O-glycosyl transferase (UFGT). The expression of these structural genes was primarily regulated by three classes of transcription factors (MYB, bHLH, and WD40), which always formed a MYB-bHLH-WD40 (MBW) complex. (Gonzalez et al., 2008; Albert et al., 2014; Xu et al., 2015; Shan et al., 2019; Li et al., 2020). Among them, MYB TFs have been identified to be the major determinant regulator in the MBW complex (Gonzalez et al., 2008; Tang et al., 2020). In *Arabidopsis*, MYB TFs TT2 (TRANSPARENT TESTA 2), PAP1/PAP2 (the production of anthocyanin pigment 1/2), MYB113 and MYB114, bHLH TFs TT8 (transparent testa8), and GL3 (glabra3), WD40 repeat protein TTG1 (TRANSPARENT TESTA GLABRA1) are the main components of the MBW complexes that regulate anthocyanin production (Zhang et al., 2003; Baudry et al., 2004).

The biosynthesis of anthocyanins is often affected by environmental factors such as light (Sun et al., 2020; Zhao et al., 2022) and temperature (Gao-Takai et al., 2019; Ryu et al., 2020). Most plants accumulate anthocyanin in a light-induced manner (Meng and Liu, 2015; Kim et al., 2021; Ma et al., 2021a), and anthocyanin concentration tends to rise as the light intensity increase (Li et al., 2018; Bai et al., 2019; Hong et al., 2019). In light signal transduction, phytochromes and their downstream factors, such as COP1 (CONSTITUTIVE PHOTOMORPHOGENIC1), HY5 (LONG HYPOCOTYL 5), and other TFs participate in light-induced anthocyanin biosynthesis (Maier et al., 2013; Meng and Liu, 2015; Hoai Nguyen, 2020; Bhatia et al., 2021; Ma et al., 2021b). In this biological process, expression of structural genes and regulatory genes, except WD40, is strongly dependent upon the existence of light (Takos et al., 2006; Bai et al., 2017; Hong et al., 2019; Liu et al., 2019). Interestingly, less- and non-photosensitive biosynthesis of anthocyanins has been reported in cherry, chrysanthemum, mango, turnip, grape and eggplant (Huang et al., 2016; Yang et al., 2017; Guo et al., 2018; He et al., 2019; Zha et al., 2019; Shi et al., 2021), although the underlying mechanism is still unclear.

Eggplant (*Solanum melongena* L.) is a globally cultivated vegetable crop with high economic benefits (Saini and Kaushik, 2019; Oladosu et al., 2021). The color of eggplant fruit is an important quality trait, with purple varieties exhibiting high levels of anthocyanins in the fruit peel (Nayanathara et al., 2016; Nino-Medina et al., 2017). However, low light during cultivation often leads to poor coloration, which reduces their visual quality and commercial value (Li et al., 2018). Therefore, eggplant genotypes with less- and non-photosensitive anthocyanin biosynthesis are valuable for the breeding of low-light tolerant varieties. The purple peel under the calyx is a good indicator of less- and non-photosensitive coloration (Xiang et al., 2015; He et al., 2019). Although some quantitative trait loci (QTL) that are responsible for purple peel under calyx have been identified in eggplant (Chen, 2015; Toppino et al., 2016; Mangino et al., 2022) and markers have been developed (Xiang et al., 2015; Zhang et al., 2021b), the underlying causal genes conferring less-photosensitive biosynthesis of anthocyanin in eggplant remain unknown.

In this study, a less-photosensitive anthocyanin biosynthesis mutant *purple in the dark* (*pind*) in eggplant was discovered. Under bagging conditions, the *pind* fruit color was light purple, whereas the wild type was white, indicating that anthocyanin biosynthesis in the *pind* mutant was less dependent on light. Phenotypic and genetic analysis, gene mapping, and transcriptomic analysis were performed to identify the causative genes conferring less-photosensitive anthocyanin biosynthesis in the fruit peel of *pind*. Taken together, our results provided novel insight into less-photosensitive anthocyanin accumulation in eggplant fruit peel.

2 Materials and methods

2.1 Materials and population construction

The mutant was identified from M_1 line in an ethyl methanesulfonate (EMS) mutagenized population of eggplant variety '14-345', which is white under the calyx. After two generations of self-fertilization, we obtained a homozygous mutant and named it "*pind*" (Figure 1A). The *pind* mutant and wildtype '14-345' were crossed to produce F_1 plants. Subsequently, F_2 and BC_1 populations were generated from F_1 plants via self- and backcross with '14-345' or *pind*, respectively. They were used as segregating populations for genetic analysis and mapping the candidate gene. In addition, *pind* and a cultivated eggplant '18-305' (Figure S1), which is white under the calyx, were crossed to obtain another F_2 population ('$305F_2$'). The '$305F_2$' population was also used for gene mapping. All plants were grown in a plastic greenhouse in the experimental fields of Hebei Agricultural University, Baoding, China. During the cultivation, the temperature inside the greenhouse was ranged from 18 to 32°C, the air humidity ranged from 50 to 70%, and the daytime light intensity ranged from 40k to 55k lx.

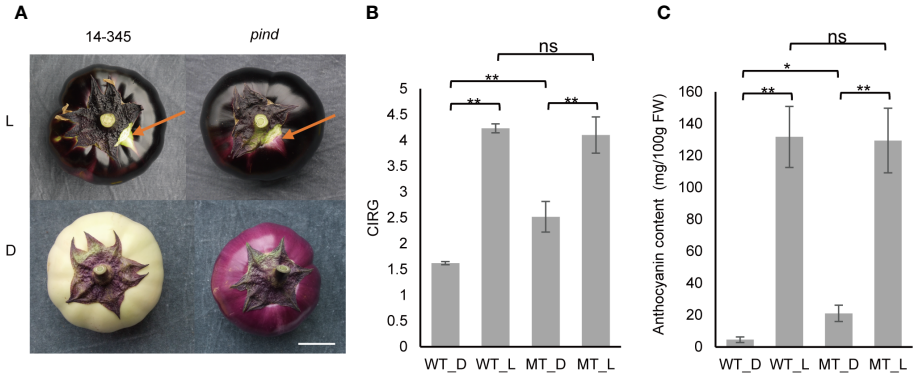


FIGURE 1
Effect of bagging on the peel color of '14-345' (wild type) and *pind* (mutant) fruits. **(A)** Fruit color of '14-345' (WT) and *pind* (MT) under natural light (L) and dark (D) conditions. The peel color under the calyx is indicated with arrows. White bar = 4cm. **(B)** CIRG (color index of red grape) value and **(C)** total anthocyanin content of the peel of '14-345' and *pind* under natural and dark conditions. Data marked with *, ** and ns indicate P <0.05, P<0.01 and no significance, respectively.

2.2 Phenotypic investigation and measurement of anthocyanin content and color index

The fruit of the plants were bagged on the 5th day after flowering and the fruit color and anthocyanin content were investigated on the 14th day under bagging condition. Color differences were assessed by the color index of red grapes (CIRG). A CR-400 colorimeter (Konica Minolta, Chroma Meter, Osaka, Japan) was used to measure the values of L*, a*, and b*. CIRG was calculated with the equation CIRG = [(180-H)/(L* + C*)], H=arctan (b*/a*), C=$(a^{*2} + b^{*2})^{0.5}$ (Carreno et al., 1995; Zhang et al., 2008).

The total anthocyanin content in the peel was quantified using the pH differential spectroscopic method (Cheng and Patrick, 1991).

One-way analysis of variance (ANOVA) was conducted on the CIRG value and anthocyanin content, and significant differences between groups were assessed by Duncan's multiple range tests ($p < 0.05$) using SPSS 16.0 Statistics (SPSS Inc., Chicago, IL, USA).

2.3 Whole genome sequence of bulked DNA

In the F_2 population under bagging conditions, 30 plants with purple fruit and 30 plants with white fruits were selected to construct the mixed pools 'P' and 'G', respectively. Genome DNA was extracted using the CTAB method.

Bulk segregant analysis (BSA)-based sequencing and MutMap analysis were used to map the candidate genes. The qualified DNA was randomly broken into fragments with a length of 350bp, and libraries were built using a TruSeq Library Construction Kit, followed by Illumina PE150 sequencing. BWA software was used to map the clean data to the eggplant reference genome (Li et al., 2021). The UnifiedGenotyper module of GATK3.8 software (Mckenna et al., 2010) was used to detect SNPs, and VariantFiltration was used to filter SNP detection. The SNP-index value with '14-345' as the reference was calculated to identify the key genes. The candidate intervals were determined with the threshold value: SNP-index (P) > 0.67, SNP-index (G) < 0.1, and Δ(SNP-index) > 0.6.

2.4 Obtaining recombinants

To narrow down the candidate region, the genotype of individuals in the F_2 and '305F_2' populations was detected using Kompetitive allele specific PCR (KASP) technology to screen recombinants. The primers used for genotype verification of the individual plants are listed in Table S1.

2.5 RNA extraction, library construction, and RNA-sequencing

Total RNA was extracted using an Eastep Super Kit (Shanghai Promega, Shanghai, China). The extracted RNA was treated with DNase to remove the genome DNA and the integrity and quantity were examined using 1% agarose gels and a Nanodrop 2000c Spectrophotometer (Thermo Nanodrop Technologies, Wilmington, DE, USA). A total of 1 μg RNA per sample was used to construct RNA-seq libraries. After qualification, mRNA was isolated from total RNA using oligo-magnetic beads. The mRNA was interrupted, reverse transcribed into cDNA, and then the cDNA was purified with an AMpure XP system (Beckman Coulter, Beverly, MA, USA). Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Boston, MA, USA). Twenty-four libraries were sequenced on the Illumina Hiseq platform (Illumina, San Diego, CA, USA). The library construction and sequencing were completed by Novogene (Beijing, China).

2.6 RNA sequencing data analysis

Raw data (raw reads) in fastq format were filtered to obtain clean reads, and the Q20, Q30, and GC content of the clean data

were calculated. The clean data were aligned to the eggplant reference genome (Li et al., 2021) using Hisat2 v2.0.5. software (Mortazavi et al., 2008). The featureCounts v.5.0-p3 tool in the subread software was used to count the reads mapped to each gene (Liao et al., 2014) and the expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM) value was calculated based on the read count and the length of the gene. The differential expression analysis was performed using the DESeq2 R package (1.20.0), and the genes with the criteria $|\log_2(\text{fold change})| > 1$ and padj≤ 0.05 were considered DEGs (Love et al., 2014). The relative expression level $\log_2$ (ratios) of all DEGs was clustered by the K-means method. Plant TFs were predicted by hmmscan based on iTAK software (Perez-Rodriguez et al., 2010). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment was analyzed with clusterProfile software and a padj of less than 0.05 was considered significantly enriched (Yu et al., 2012).

2.7 Quantitative real-time PCR analysis

Fourteen genes were chosen for the validation of RNA-seq using qRT-PCR. Primer Premier 5.0 software was used to design the primers, which are listed in Table S2. A total of 1 μg RNA per sample was reverse transcribed using a PrimeScript™ reagent Kit with gDNA Eraser (TaKaRa, Beijing, China) in 20 μL of reaction mixture. The qRT-PCR was performed on a LightCycler® 96 instrument (Roche, Basel, Switzerland), using THUNDERBIRD SYBR qPCR Mix (TOYOBO, Shanghai, China) with the following reactions: 95°C 2 min; 95°C 30s, 60°C 10s, and 68°C 10s for 40 cycles. The PCR products were quantified by $2^{-\triangle\triangle CT}$ method (Livak and Schmittgen, 2001) with normalization to the expression level of *SmGAPDH* (*EGP1067575*). Significant differences between groups were assessed by Student's ***t***-test ($p < 0.05$) using SPSS 16.0 Statistics (SPSS Inc., Chicago, IL, USA).

3 Results

3.1 Phenotypic characterization of *pind*

The *pind* mutant was obtained from eggplant '14-345' mutagenized population with EMS. The coloration of the '14-345' fruit peel is photosensitive with white-colored fruit under the calyx. Interestingly, a light purple pigmentation was observed under the calyx of *pind*, indicating that the coloration of *pind* was less photosensitive. To further characterize the less-photosensitive coloration in *pind*, the fruit was bagged to mimic dark conditions. Compared to the white fruit of '14-345', the *pind* fruit was light purple (Figure 1A). Under natural conditions, the color index, CIRG value, of '14-345' was similar to that of *pind*, whereas the CIRG value of the *pind* mutant was significantly higher than that of '14-345' under bagging conditions (Figure 1B).

The anthocyanin contents in the peel of '14-345' and *pind* were also quantified. The anthocyanin content in the peel of '14-345' and *pind* was comparable under natural conditions. However, the level of anthocyanin in the bagged eggplants decreased significantly in both '14-345' and *pind* compared with natural conditions. Anthocyanin content in the peel of bagged *pind* (21.1 mg/100 g FW) was significantly higher than that in the bagged '14-345' (4.5 mg/100 g FW) (Figure 1C). These results indicated that light was required for the coloration of fruit peel in '14-345', but less so for *pind*.

3.2 Genetic analysis of *pind*

To investigate the inheritance of *pind*, segregating populations were constructed. All the F_1 plants from the crosses of *pind* and '14-345' had purple fruits under bagging conditions, which was phenocopied *pind*. Out of the F_2 plants, 116 had purple bagged fruits, while 37 had white bagged fruits. A chi-squared test revealed that this segregation pattern agreed with the 3:1 Mendelian segregation ratio ($P=0.82>0.05$, $\chi^2 = 0.05<\chi^2_{0.05,1} = 3.84$). In the W-$BC_1$ population (the backcross progeny of F_1 with '14-345'), 30 plants displayed white wild type fruit, whereas the remaining 23 plants produced purple fruit under bagging conditions. The segregation pattern fitted the ratio 1:1 ($P=0.34>0.05$, $\chi^2 = 0.93<\chi^2_{0.05,1} = 3.84$). All M-$BC_1$ plants, the backcross progeny of F_1 with *pind*, exhibited purple fruits under bagging conditions (Table 1). These results suggested that the mutant gene was a single dominant allele.

3.3 Identification of the candidate genomic region

To identify the genomic region responsible for less-photosensitive anthocyanin biosynthesis in the fruit peel of *pind*, BSA-based genome resequencing and MutMap analysis were performed. Quality control resulted in 96,777,013 and 90,196,644 high-quality reads and 98.67% and 98.64% of them were mapped on the eggplant genome (Tables S3, S4), from which 1421190 SNPs were detected in total. Based on the Δ(SNP-index), a single significant interval in a 12.2 Mb region (from 2.6 to 14.8 Mb) on chromosome 10 was identified as a candidate region of the mutation (Figure 2A). There were 2,058 SNPs, including 22 non-synonymous mutation sites.

To identify the causal mutation in this region, several molecular markers were designed to uniformly cover the preliminary mapping interval. Fifteen recombinant plants were identified and the candidate gene was mapped to an 8.2 Mb region (from 6.6 Mb to 14.8 Mb) between markers M5 and M10 (Figure 2B).

To further map the SNP related to less-photosensitive anthocyanin biosynthesis, another F_2 population, '305F_2' (*pind* × '18-305'), was used to obtain recombinants. The '305F_2' population consisted of 408 plants, of which 305 had purple fruit and 103 had white fruit under bagging conditions (Table S5). Thirty-four recombinants were identified and the candidate gene was finally mapped to a 4.0 Mb region between the markers M12 and M9 (Figure 2C), from 7.7 Mb~11.7 Mb on chromosome 10. A total of 224 SNPs that were associated with 75 genes (Table S6), were detected in the candidate interval.

10.3389/fpls.2023.1282661

TABLE 1 The segregation of the fruit color in dark condition.

Generation	Numbers of plants	*pind* phenotype	WT phenotype	Expected ratio	P	χ^{2} [c]
14-345	15	0	15			
pind	45	45	0			
F_1	14	14	0			
F_2	153	116	37	3:1	0.82	0.05
M-BC_1 [a]	59	59	0			
W-BC_1 [b]	53	23	30	1:1	0.34	0.93

[a]M-BC_1 represents the backcross progeny of F_1 with *pind*.
[b]W-BC_1 represents the backcross progeny of F_1 with 14-345.
[c]$\chi^2_{0.05}$ = 3.84, df=1.

3.4 Transcriptome comparison between *pind* and '14-345' in response to dark treatment

To identify the candidate gene responsible for less-photosensitive anthocyanin biosynthesis in *pind*, RNA-seq analysis was performed on the fruits of *pind* (MT) and '14-345' (WT) under bagging (D) and natural (L) conditions for two stages, 2days (2d) and 5 days (5d) after bagging. For both stages, the fruits of the '14-345' bagged were white, whereas the fruits of *pind* were light purple under bagged conditions. The fruit peel color of '14-345' and *pind* at the two stages was almost the same under natural conditions (Figure 3A).

To identify the key differentially expressed genes (DEGs) between the fruit peel of '14-345' and *pind*, pairwise comparisons between bagged '14-345' and *pind* at each stage (MT_D_2d vs. WT_D_2d, MT_D_5d vs. WT_D_5d), bagged and natural '14-345' at each stage (WT_D_2d vs. WT_L_2d, WT_D_5d vs. WT_L_5d), and bagged and natural *pind* at each stage (MT_D_2d vs. MT_L_2d, MT_D_5d vs. MT_L_5d) were analyzed. A total of 4,338 DEGs were detected in the above-mentioned six combinations (Figure 3B).

A total of 415 and 191 genes were significantly induced and inhibited in the peel of '14-345' under bagging conditions on 2d, respectively, compared with natural conditions (WT_D_2d vs. WT_L_2d). KEGG enrichment analysis showed that the pathways 'flavonoid biosynthesis' and 'phenylpropanoid biosynthesis' were significantly enriched (Figure 3C). There were 557 DEGs in the 'WT_D_5d vs. WT_L_5d' combination, in which the pathways 'flavonoid biosynthesis' and 'phenylpropanoid biosynthesis' were also

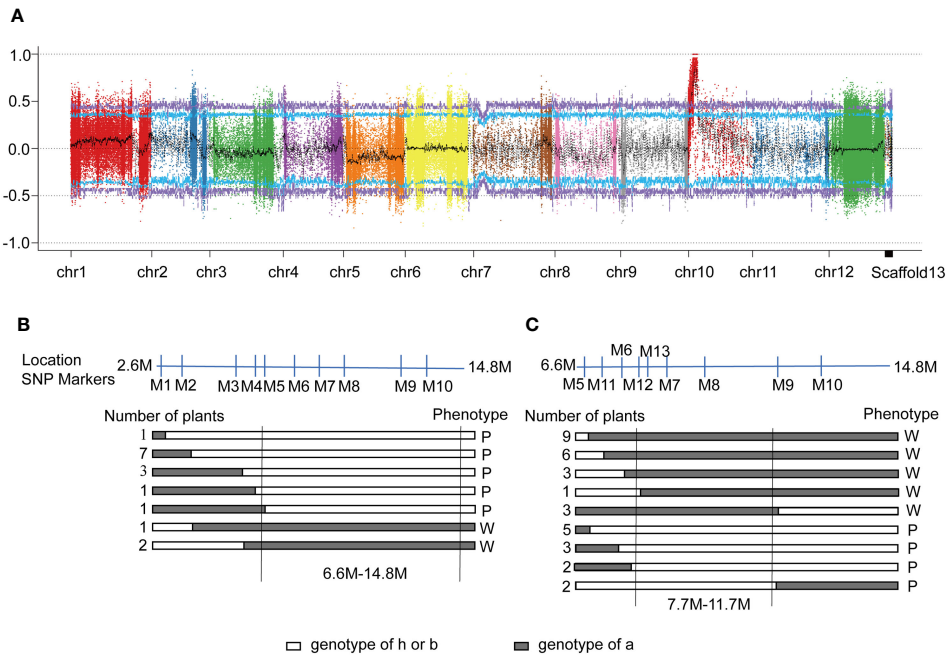


FIGURE 2
Mapping and identification of the candidate gene. **(A)** The distribution of Δ(SNP-index) on all chromosomes. X-axis: Chromosome ID; Y-axis: Δ(SNP-index). Each colored dot represents an Δ(SNP-index) value of an SNP site. Schematic representation of allelic segregation of recombinants in F_2. **(B)** and '$305F_2$' **(C)**. a: Genotype of wildtype '14-345' **(B)** or '18-305' **(C)**; b: Genotype of the *pind* mutant; h: Heterozygote of '14-345' and *pind*. P: Purple fruit peel under bagging conditions; W: White fruit peel under bagging conditions.

10.3389/fpls.2023.1282661

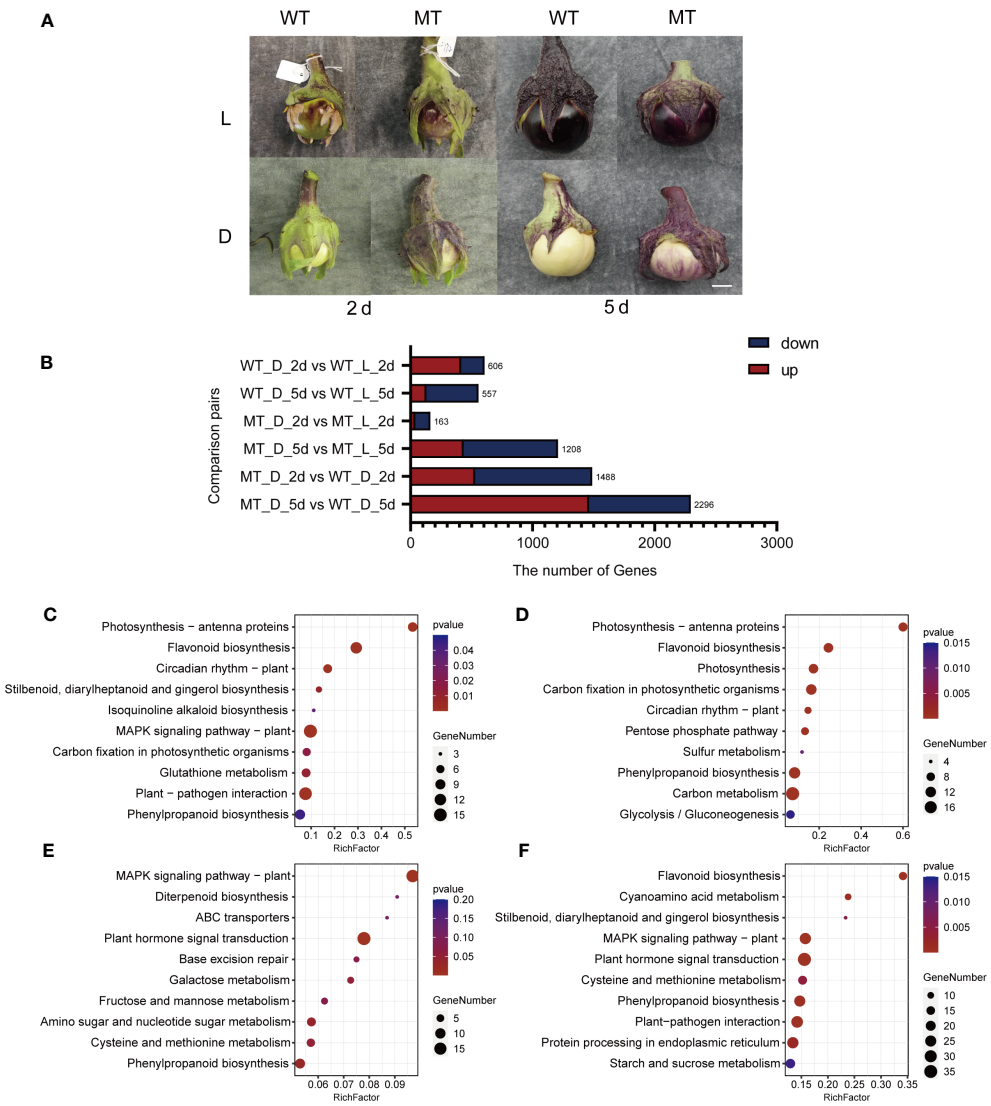


FIGURE 3
Overview of DEGs identified by RNA-seq analysis in the peel of '14-345' and *pind* fruits under bagging conditions at two developmental stages. **(A)** The fruit color of '14-345'(WT) and *pind* (MT) under natural (L) and dark (D) conditions at 2 days (2d) and 5 days (5d) under bagging conditions. Scale bar represents 1 cm. **(B)** The number of DEGs between samples in each combination. **(C–F)** KEGG pathway enrichment of DEGs in 'WT_D_2d vs. WT_L_2d', 'WT_D_5d vs.WT_L_5d', 'WT_D_2d vs. MT_D_2d', and 'WT_D_5d vs. MT_D_5d'.

significantly enriched (Figure 3D). The decreased expression of the eight flavonoid biosynthetic genes, including *CHS* (*EGP24357*, *EGP16216*), *CHI* (*EGP24232*, *EGP22200*), *F3H* (*EGP30923*), *F3'5'H* (*EGP32037*), *DFR* (*EGP31016*), and *ANS* (*EGP18904*) (Figures S2A, B), indicated that the anthocyanin biosynthetic pathway was repressed in '14-345' under dark conditions.

A total of 1,488 DEGs were detected in the 'MT_D_2d vs. WT_D_2d' combination, which were significantly enriched in the 'MAPK signal pathway' and 'plant signal transduction pathway' (Figure 3E). The 'MT_D_5d vs. WT_D_5d' combination produced 1,463 DEGs, and the enrichment analysis revealed that among all of the pathways identified, the 'flavonoid biosynthesis pathway' were the most significantly enriched (Figure 3F). The expression of structural genes *4CL* (*EGP10904*), *CHS* (*EGP24357*), *CHI* (*EGP24232*, *EGP22200*), *F3H* (*EGP30923*), *F3'5'H* (*EGP32037*), *DFR* (*EGP31016*), and *ANS* (*EGP18904*) (Figure S2C) involved in anthocyanin biosynthesis was significantly higher in the mutant than in the wild type at 5d under bagging conditions.

3.5 Analysis of DEGs for less-photosensitive anthocyanin biosynthesis in *pind*

Due to the difference in fruit color between '14-345' and *pind* under bagged conditions (MT_D vs. WT_D) and the differences in the color of '14-345' fruit between bagged and natural conditions (WT_D vs. WT_L), we screened for common DEGs in these combinations. The Venn diagram clearly illustrated the relationships of DEGs between different comparisons, and 238

DEGs (Figure 4A; Table S7) and 145 DEGs (Figure 4B; Table S8) were commonly shared by 'MT_D vs. WT_D' and 'WT_D vs. WT_L' on 2d and 5d, respectively. Moreover, 35 DEGs were commonly shared by these four combinations (Figure 4C; Table S9), which may participate in anthocyanin biosynthesis in *pind* under dark conditions. Therefore, these 35 DEGs, including 12 structural genes and three TFs, were used for further analysis.

Under natural conditions, there was no significant difference in the expression of structural genes between '14-345' and *pind* fruit peel. Bagging inhibited the transcripts of most of the structural genes, including *4CL*, *CHS*, *CHI*, *F3H*, *F3′5′H*, *DFR*, and *ANS* in '14-345' while not in *pind*, as the transcriptional level of structural genes were extremely higher in *pind* than in '14-345' (Figure 4D). These data suggest that '14-345' and *pind* have different responses to bagging in terms of anthocyanin biosynthesis. The highly expressed structural genes in *pind* under dark conditions showed *pind* was less photosensitive.

TFs can regulate the expression of structural genes by directly binding to *cis*-regulatory elements in the promoter of the genes, which play important roles in plant growth and development (Chen et al., 2019; Jiang et al., 2022; Liu et al., 2023). In this study, MYB TF *SmMYB113* (*EGP21875*), WRKY TF *SmTTG2* (*EGP21679*), and bHLH TF *SmGL3* (*EGP16978*) were identified from the 35 DEGs as the putative main regulators of the anthocyanin biosynthesis in a less-photosensitive manner in the fruit peel of *pind*. Under natural conditions, the expression of *SmMYB113*, *SmTTG2*, and *SmGL3* was comparable between '14-345' and *pind*. Under bagging conditions, *SmMYB113*, *SmTTG2*, and *SmGL3* was down-regulated both in '14-345' and *pind* compared to natural conditions (Figure 4E). Notably, the expression levels of *MYB113*, *TTG2*, and *GL3* were significantly increased in the peel of *pind* compared to '14-345' under bagged conditions (Figure 4E).

To confirm the reliability of the RNA-seq data, six structural genes (*C4H*, *CHI*, *F3H*, *DFR*, *ANS*, and *5GT*), five TFs (*MYB113*, *GL2*, *GL3*, *TT8*, and *TTG2*), and three genes involved in light signaling (*COP1*, *UVR8*, and *CRY3*) were selected to analyze expression profiles by qRT-PCR. The transcript abundances of these genes determined using qRT-PCR agreed with those determined from transcriptome sequencing (Figure 5), indicating the reliability of the RNA-seq results.

3.6 Candidate gene analysis

Based on the gene mapping analysis, there were 75 genes in the candidate region (Table S6), of which only nine (Figure 6) were expressed in the eggplant peel according to the transcriptome analysis. Twenty-seven SNPs were detected to be associated with these nine genes, and only one non-synonymous SNP (within *EGP21953*) was identified. The gene *EGP21953* (Table 2) had an amino acid residue substitution at aa-262 (Ala-to-Tyr mutation). According to the annotated information, *EGP21953* encodes a

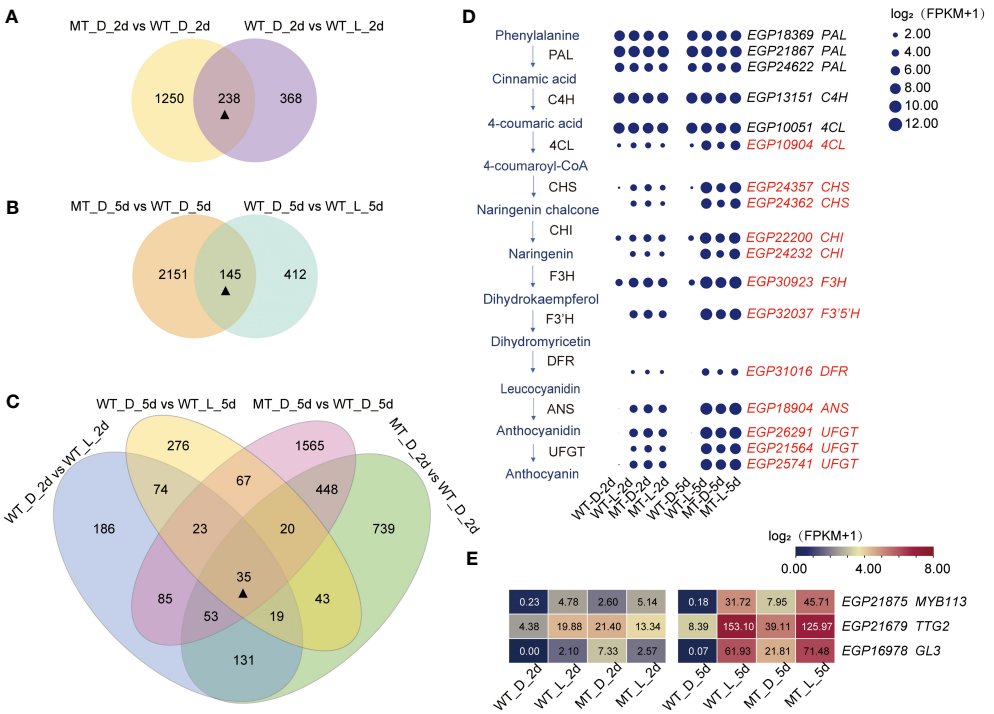


FIGURE 4
Identification of the DEGs related to less-photosensitive anthocyanin biosynthesis in *pind*. **(A–C)** Venn diagram of the different comparisons of DEGs. **(D)** Expression pattern analysis of structural genes involved in less-photosensitive anthocyanin biosynthesis in fruit peel of *pind*. Note: Red gene ID represents candidate DEGs related to less-photosensitive anthocyanin biosynthesis in the fruit peel of *pind*. The size of the circle represents $\log_2$ (FPKM+1); **(E)** Expression pattern analysis of TFs involved in less-photosensitive anthocyanin biosynthesis in the fruit peel of *pind*. The number in the box indicates the FPKM value.

10.3389/fpls.2023.1282661

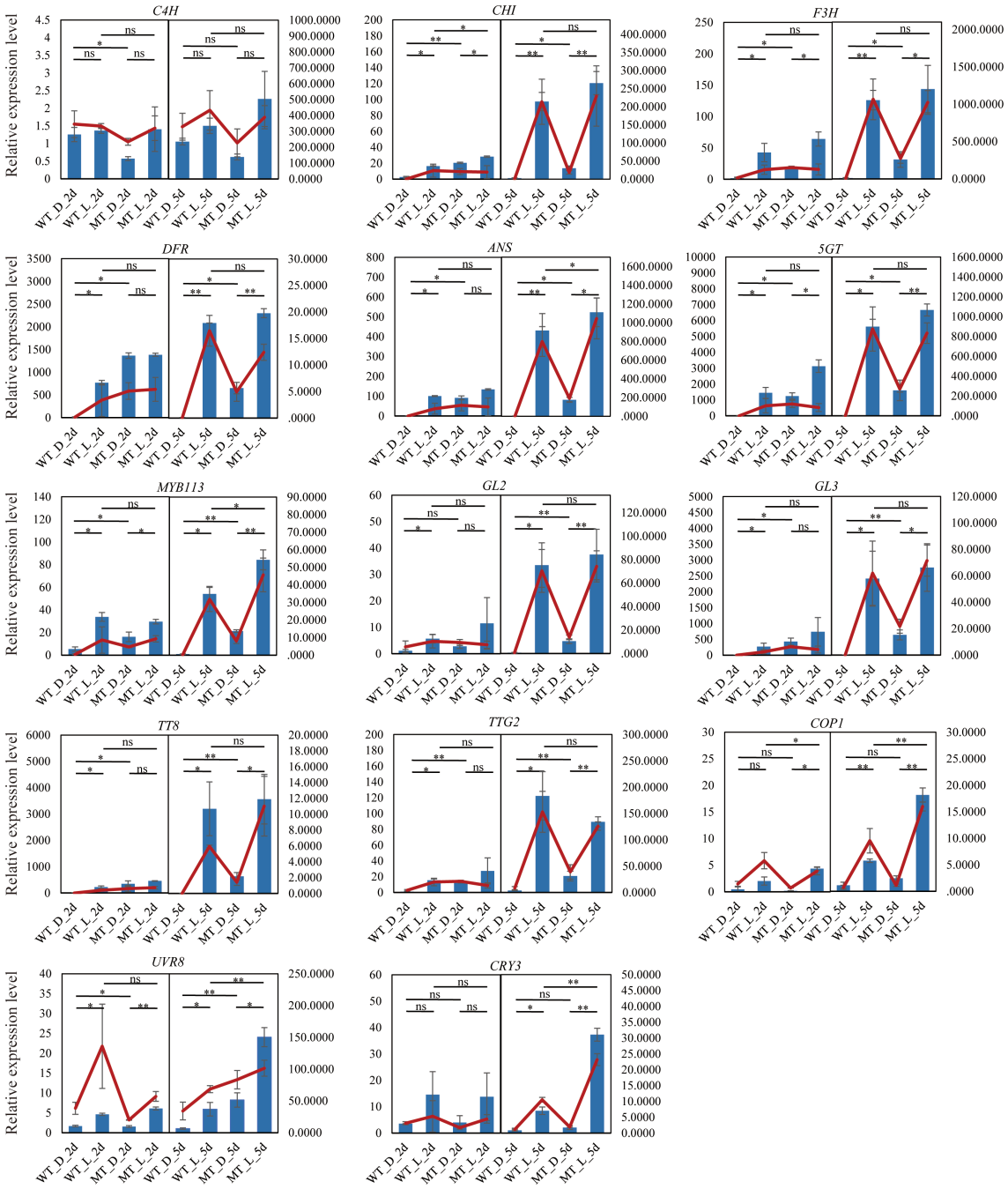


FIGURE 5
Validation of RNA-seq results using qRT-PCR. Y-axis: relative expression. The blue bar and red line represent the data from qRT-PCR and RNA-seq data, respectively. Asterisks indicate significant differences determined by the Student's *t*-test (*$P < 0.05$, **$P < 0.01$). ns, no significance.

CAAX amino-terminal protease family protein. However, there was no significant difference in *EGP21953* expression between '14-345' and *pind*, neither in natural nor in bagged conditions.

Among the other eight genes, whose related SNPs were detected in intergenic regions or introns, only *EGP21875*, *EGP21950*, and *EGP21961*, were differentially expressed between the peel of *pind* and '14-345' (Figure 6; Table 2). *EGP21875* encodes an MYB TF, "*SmMYB113*", which has been shown to participate in the regulation of anthocyanin biosynthesis (Zhang et al., 2014; Jiang et al., 2016; He et al., 2019; Yang et al., 2022). The expression of *SmMYB113* was inhibited in the fruit peel of '14-345' under bagged conditions. Interestingly, *SmMYB113* showed significantly higher expression in the peel of *pind* than in '14-345' at both 2d and 5d stages, pinpointing the activation of SmMYB113 in *pind* in darkness. There was one SNP 22.9 kb upstream of the start codon and one SNP 13 kb downstream of the stop codon of *SmMYB113*, respectively. Because the intergenic SNPs play a potentially important role in phenotype variation in plants by altering gene expression (Schwartz et al., 2009; Ding et al., 2012; Li et al., 2012; Wang et al., 2015), these two SNPs in the intergenic region associated with *SmMYB113* may affect its expression.

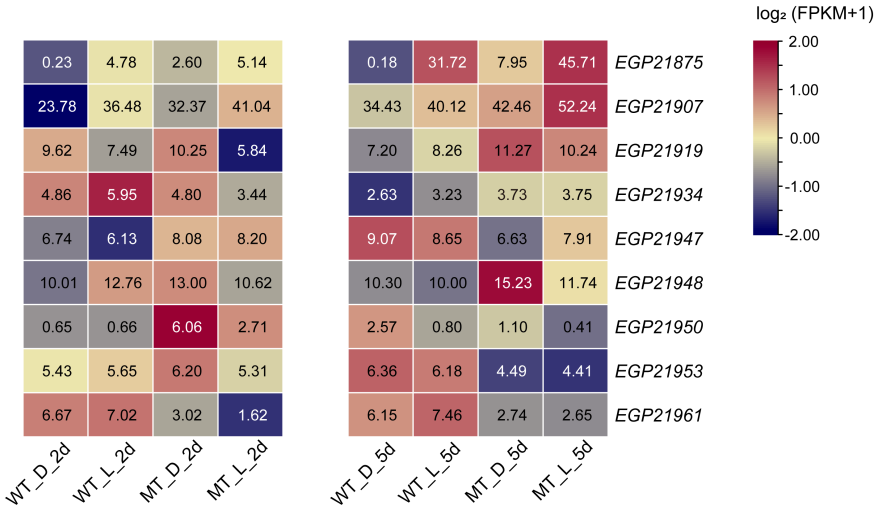


FIGURE 6
Expression pattern analysis of the genes in the candidate region. The number in the box indicates the FPKM value.

EGP21950 encoded an unknown functional protein, and there was one SNP 859 bp upstream of the start codon of *EGP21950*. The expression of *EGP21950* was significantly up-regulated in the fruit peel of *pind* in comparison to '14-345' at 2d under bagged conditions, but its expression level showed no significant difference between '14-345' and *pind* under natural conditions and at 5d under bagged conditions. In addition to *EGP21953*, *EGP21961* also encoded a CAAX amino-terminal protease family protein, and its associated SNP was about 7 kb upstream of the initiation codon. *EGP21961* was down-regulated significantly in the peel of *pind* compared to '14-345', in both natural and bagged conditions.

4 Discussion

4.1 Gene identification related to less-photosensitive anthocyanin biosynthesis in eggplant

Anthocyanin is an important flavonoid type pigment in flowers and fruits, with light being an essential regulator of anthocyanin accumulation (Takos et al., 2006; Meng and Liu, 2015; Bai et al., 2017; Kim et al., 2021; Ma et al., 2021a). Proper anthocyanin accumulation in fruits or flowers can be hindered by low-light conditions (caused by weather and climate, the cultivation facility, or plant morphology) (Hong et al., 2015; Cao et al., 2016; Zhang et al., 2021a; Guo et al., 2022). Therefore, understanding the mechanism of less-and non-photosensitive anthocyanin biosynthesis and creating the less- and non-photosensitive cultivars will provide valuable resources for the breeding of low-light tolerant varieties

In this study, we discovered a less-photosensitive anthocyanin biosynthesis mutant (*pind*) in eggplant, and a candidate gene that was physically mapped to 7.7 ~11.7 Mb on chromosome 10 (Figure 2C). This result was consistent with previous reports, showing that the candidate genes that may correspond to loci controlling fruit peel color under calyx in eggplants were mapped to chromosome 10 (Zhang et al., 2021b; Mangino et al., 2022; Qiao et al., 2022). In addition to chromosome 10, the candidate genes controlling the fruit peel color under calyx in eggplants were also found on other chromosomes, including chromosome 1, 3, 4, and 11 (Mangino et al., 2022; Qiao et al., 2022). However, the above studies did not distinguish materials as less-photosensitive or non-photosensitive materials. When comparing the phenotypic differences between the less-photosensitive or non-photosensitive

TABLE 2 Candidate genes related to less-photosensitive anthocyanin synthesis in the fruit peel of *pind*.

Gene ID	SNP loci	substitution type	SNP category	Distance	Predicted function
EGP21875	7808145 7770881	C->A T->C	Intergenic region in upstream, Intergenic region in downstream	22893bp 13006bp	myb domain protein 113
EGP21950	10889680	T->C	Upstream	859bp	–
EGP21953	11268541	C->T	Non-synonymous		CAAX amino terminal protease family protein
EGP21961	11685483	T->A	Intergenic region in upstream	6998bp	CAAX amino terminal protease family protein

materials, it was observed that under bagging conditions, fruit coloration became significantly lighter for less-photosensitive materials (Figure 1A), whereas there was no significant difference for non-photosensitive materials (He et al., 2019; He et al., 2022). Recent studies have found that the candidate gene related to non-photosensitive anthocyanin biosynthesis in eggplant is *SmFTSH10* (He et al., 2022), which was physically close to, but outside the candidate interval 7.7Mb-11.7Mb (Figure 3C) for less-photosensitive anthocyanin biosynthesis in this study, indicating less- and non-photosensitive anthocyanin biosynthesis in eggplant may be controlled by different locus.

4.2 Candidate genes related to less-photosensitive anthocyanin biosynthesis in *pind*

Through MutMap and transcriptome analysis, we obtained four candidate genes: *EGP21875*, *EGP21950*, *EGP21953*, *and EGP21961* (Table 2). *EGP21953* and *EGP21961* are homologs of the gene *AT1G14270.1* in *Arabidopsis*, which is a CAAX amino-terminal protease family member. However, at present, there is no literature or report indicating that the CAAX amino-terminal protease family proteins under investigation exert regulatory control over anthocyanin biosynthesis. *EGP21950* encodes a protein with an unknown function, that contained a DUF616 (Protein of unknown function) domain (Figure S3). Additionally, proteins encoded by structural genes and regulatory genes didn't possess CAAX prenyl endopeptidase-like domains or UDF616 domains (Figure S3), indicating their low likelihood to bind with CAAX amino-terminal protease family proteins or DUF616 proteins. Therefore, we speculate that *EGP21950*, *EGP21953*, or *EGP21961* had minimal potential in contributing to the genetic basis of less-photosensitivity.

The gene *EGP21875* encoded an MYB TF SmMYB113, homologous to *AtMYB113* in *Arabidopsis*. In this study, *SmMYB113* was hardly expressed in the fruit peel of wild type '14-345' under bagged conditions, whereas its expression was considerably up-regulated in the *pind* mutant compared to '14-345'. It was reported that overexpression of *SmMYB113* (also known as *SmMYB1*) resulted in the up-regulation of *SmCHS*, *SmCHI*, *SmF3H*, *SmANS*, and other genes, and a high level of anthocyanin accumulation (Zhang et al., 2014; Yang et al., 2022). Thus, anthocyanin-pigmentation in *pind* in the dark may be related to the mutations of *SmMYB113*. Collectively, we speculated that *EGP21875* (*SmMYB113*) was the best candidate gene for the regulation of less-photosensitive anthocyanin biosynthesis in the fruit peel of *pind*.

In addition, the SNPs related to *EGP21875* (*SmMYB113*) were more than 13kb from the coding region. Some studies have shown that a non-coding region far from the gene promoter may change the activity of neighboring genes. For example, the *Teosinte Branched 1* (*TB1*) gene is mainly responsible for a major-effect QTL that controls morphological differences in plant architecture between maize (*Zea mays* subsp. *mays*) and its wild relative, teosinte (*Z. mays* subsp. *mexicana* and subsp. *parviglumis*) (Clark et al., 2004; Studer et al., 2011); the expression of *tb1* was altered by a sequence >41kb upstream of *tb1* (Clark et al., 2006). In addition, *booster1* (*b1*) was a regulator gene responsible for the biosynthesis of flavonoid pigments in maize; a 6-kb region at about 100kb upstream of the transcription start site of gene *b1* was found to be a hepta-repeat enhancer, which may increase the transcription initiation rate (Stam et al., 2002). The non-coding region located distally from the gene promoter could be considered as a putative *cis*-regulatory element that modulates gene expression through recruitment of transcription factors (Stadhouders et al., 2014; Visser et al., 2014; Roberts et al., 2016). Studies have shown that natural variation in *cis*-regulatory regions of genes played important roles in phenotypic variations by altering gene expression (Frary et al., 2000; Schwartz et al., 2009; Muños et al., 2011; Van Der Knaap et al., 2014; Chu et al., 2019). Taken together, we speculated that the position of the SNPs in the intergenic region of *EGP21875* (*SmMYB113*) might coincide with a regulatory element that controls the *SmMYB113* expression level.

4.3 The activation of MYB113 may be necessary for anthocyanin accumulation in *pind* peel under dark conditions

Under bagging conditions, '14-345' fruits exhibited white coloration while the *pind* mutants displayed purple coloration. Consequently, under bagging conditions, the structural genes were expressed at a minimal level in the peel of '14-345' but showed higher expression in *pind* (Figure 4D). It is well known that the expression of structural genes is directly regulated by transcription factors, among which MYB TFs, bHLH TFs, and WD40 are key modulators (Shan et al., 2019; Li et al., 2020). In this study, expression of *SmMYB113*, *SmGL3* (bHLH) and *SmTTG2* (WRKY) was significantly higher in *pind* peel than '14-345' in darkness (Figure 4E). Moreover, bHLH TF *SmTT8* was inactive in '14-345' under bagged conditions, but strongly activated in *pind* on 5d (Table S8; Figure 5). Therefore, under bagging conditions, the inactive transcription factors, such as SmMYB113, SmGL3, SmWRY44 and SmTT8, cannot stimulate the expression of structural genes, which is the underlying reason for the inability of '14-345' to produce anthocyanins under bagging conditions (Figure 7A).

In *Arabidopsis*, AtMYB113 interacts with AtTT8 or AtGL3 and AtTTG1, forming the MBW complex, to regulate the expression of late biosynthesis genes (LBGs) (Ramsay and Glover, 2005). The MYB TF is the main determinant of the MBW complex, and MYBs control the expression of early biosynthesis genes (EBGs) and LBGs (Stracke et al., 2007; Gonzalez et al., 2008; Tang et al., 2020). In tomato, after the overexpression of *SlANT1* (a homolog of *AtMYB113*), the expression of the structural genes, and the bHLH gene *SlAN1* (a homolog of *AtTT8*) was significantly up-regulated (Mathews et al., 2003; Kiferle et al., 2015). SlAN1 probably activates the expression of structural genes through the MYB-AN1-WD40 complex, whereas SlJAF13 (a homolog of AtGL3) regulates the transcription of *SlAN1* through the MYB-SlJAF13-WD40 complex (Montefiori et al., 2015; Liu et al., 2018). Therefore, we speculated

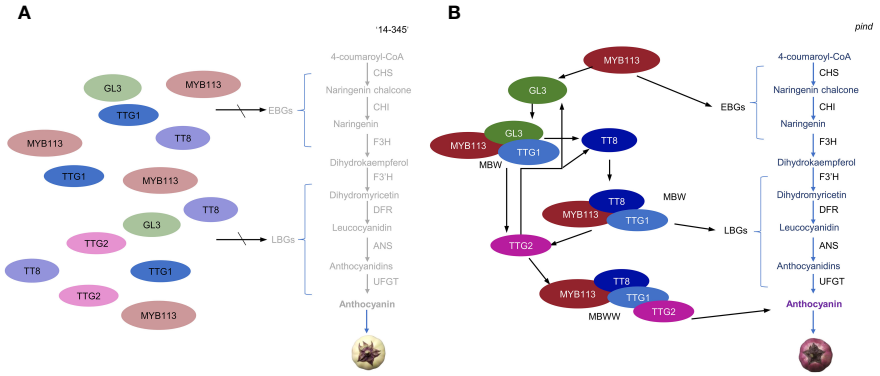


FIGURE 7
A hypothetical model of the mechanism for anthocyanin biosynthesis in the eggplant peel of '14-345' **(A)** and *pind* **(B)** under dark conditions. **(A)** Under bagging conditions, in '14-345', the inactivated state of transcription factors, SmMYB113, SmGL3, SmTTG2 and SmTT8, cannot promote the expression of structural genes, consequently resulting in the inhibition of structural gene expression and subsequently impacting the biosynthesis of anthocyanins. **(B)** TF MYB113 promotes the expression of EBGs and activates the expression of *GL3*. MYB113, and GL3 form a MBW complex with the WD40 protein TTG1 to promote the expression of *TT8*. Then MYB113, TT8, and TTG1 will form a new MBW complex to promote the expression of LBGs. *TTG2* is activated by MBW and forms the MBWW complex, and the MBWW complex promotes the expression of genes related to anthocyanin transport. The highly expressed structural genes in *pind* under dark conditions led to accumulation of anthocyanin. The positive regulation of genes is marked in black arrows. The gray font represents genes with very low expression, while the lighter-colored circles indicate transcription factors with low activity.

that under bagged conditions, the up-regulation of EBGs was mainly due to the activation of SmMYB113 in the peel of *pind*; the LBGs were directly regulated by the MBW complex composed of SmMYB113, SmTT8, and WD40 (SmTTG1), whereas the expression of *SmTT8* was activated by the MBW complex SmMYB113-SmGL3-SmTTG1; and *SmGL3* was activated by SmMYB113 (Figure 7B).

In addition, WRKY44 was reported to regulate anthocyanin accumulation. In tobacco, the coloration of flowers was observed to increase upon overexpression of *TTG2*, whereas a decrease in flower coloration was achieved through silencing *TTG2* (Li et al., 2017). Anthocyanin accumulation was induced in tobacco leaves after transient overexpression of kiwifruit *WRKY44* (Peng et al., 2020). WRKY44 was reported to interact with BMW to regulate the corresponding biological processes (Pesch et al., 2014; Gonzalez et al., 2016; Verweij et al., 2016; Lloyd et al., 2017). In summary, it was suggested that under bagging conditions, the expression of *SmTTG2* was up-regulated by the MBW complex and SmTTG2 could bind to the MBW complex to promote the expression of structural genes in *pind* (Figure 7B).

Based on the expression patterns and putative function of the genes in other plants, we proposed a working model that describes the regulatory mechanism of anthocyanin biosynthesis in the peel of '14-345' and *pind* under dark conditions (Figure 7). It is indicated that the activation of *SmMYB113*, *SmTT8*, *SmTTG2* and *SmTT8* is the necessary factor for anthocyanin accumulation in the dark in eggplant. Considering that SmMYB113 was the most upstream in the pathway, it is postulated that SmMYB113 was the first transcription factor to be activated and serves as the most crucial regulator. Combined with gene mapping analysis, it is speculated that the activation of *SmMYB113* in *pind* under bagged conditions may be related to the SNP of the intergenic region of *EGP21875*.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1014581, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1013920.

Author contributions

LL: Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing. JM: Conceptualization, Methodology, Software, Writing – original draft. QL: Conceptualization, Visualization, Writing – review & editing. YL: Conceptualization, Methodology, Writing – review & editing. SL: Data curation, Investigation, Writing – review & editing. NL: Visualization, Writing – review & editing. SX: Visualization, Writing – review & editing. YW: Visualization, Writing – review & editing. SS: Funding acquisition, Supervision, Writing – review & editing. AB: Conceptualization, Supervision, Writing – review & editing. JZ: Funding acquisition, Supervision, Writing – review & editing. XC: Data curation, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by S&T Program of Hebei (Grant No. 21326309D), S&T Program of Hebei (Grant No. 20592901D), and the Post-graduate's Innovation Fund Project of Hebei (Grant No. CXZZBS2019098).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1282661/full#supplementary-material

References

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell.* 26, 962–980. doi: 10.1105/tpc.113.122069

Bai, S., Sun, Y., Qian, M., Yang, F., Ni, J., Tao, R., et al. (2017). Transcriptome analysis of bagging-treated red Chinese sand pear peels reveals light-responsive pathway functions in anthocyanin accumulation. *Sci. Rep-UK.* 7, 1–15. doi: 10.1038/s41598-017-00069-z

Bai, S., Tao, R., Tang, Y., Yin, L., Ma, Y., Ni, J., et al. (2019). BBX16, a B-box protein, positively regulates light-induced anthocyanin accumulation by activating MYB10 in red pear. *Plant Biotechnol. J.* 17, 1985–1997. doi: 10.1111/pbi.13114

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Bhatia, C., Gaddam, S. R., Pandey, A., and Trivedi, P. K. (2021). COP1 mediates light-dependent regulation of flavonol biosynthesis through HY5 in *Arabidopsis*. *Plant Sci.* 303, 110760. doi: 10.1016/j.plantsci.2020.110760

Cao, L., Xu, X., Chen, S., and Ma, H. (2016). Cloning and expression analysis of *Ficus carica anthocyanidin synthase 1* gene. *Sci. Hortic-Amsterdam.* 211, 369–375. doi: 10.1016/j.scienta.2016.09.015

Carreno, J., Martinez, A., Almela, L., and Fernandezlopez, J. A. (1995). Proposal of an index for the objective evaluation of the color of red table grapes. *Food Res. Int.* 28, 373–377. doi: 10.1016/0963-9969(95)00008-A

Chen, L., Hu, B., Qin, Y., Hu, G., and Zhao, J. (2019). Advance of the negative regulation of anthocyanin biosynthesis by MYB transcription factors. *Plant Physiol. Bioch* 136, 178–187. doi: 10.1016/j.plaphy.2019.01.024

Chen, X. (2015). "Study on genetic diversity and genetic linkage map construction in eggplant," in *[dissertation/doctor's thesis]* (Baoding (Hebei)]: Hebei Agricultural University).

Cheng, G. W., and Patrick, J. B. (1991). Activity of phenylalanine ammonia-lyase (PAL) and concentrations of anthocyanins and phenolics in developing strawberry fruit. *J. Am. Soc. Hortic. Sci.* 116, 865–869. doi: 10.21273/JASHS.116.5.865

Chu, Y. H., Jang, J. C., Huang, Z., and van der Knaap, E. (2019). Tomato locule number and fruit size controlled by natural alleles of *lc* and *fas*. *Plant Direct.* 3, e00142. doi: 10.1002/pld3.142

Clark, R. M., Linton, E., Messing, J., and Doebley, J. F. (2004). Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *P Natl. Acad. Sci. U.S.A.* 101, 700–707. doi: 10.1073/pnas.2237049100

Clark, R. M., Wagler, T. N., Quijada, P., and Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38, 594–597. doi: 10.1038/ng1784

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U S A.* 109, 2654–2659. doi: 10.1073/pnas.1121374109

Dooner, H. K., Robbins, T. P., and Jorgensen, R. A. (1991). Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* 25, 99. doi: 10.1146/annurev.ge.25.120191.001133

Fan, X., Fan, B., Wang, Y., and Yang, W. (2016). Anthocyanin accumulation enhanced in *Lc*-transgenic cotton under light and increased resistance to bollworm. *Plant Biotechnol. Rep.* 10, 1–11. doi: 10.1007/s11816-015-0382-3

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. V. D., Cong, B., et al. (2000). *fw2. 2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science.* 289, 85–88. doi: 10.1126/science.289.5476.85

Gao-Takai, M., Katayama-Ikegami, A., Matsuda, K., Shindo, H., Uemae, S., and Oyaizu, M. (2019). A low temperature promotes anthocyanin biosynthesis but does not accelerate endogenous abscisic acid accumulation in red-skinned grapes. *Plant Sci.* 283, 165–176. doi: 10.1016/j.plantsci.2019.01.015

Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., et al. (2016). TTG2 controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev. Biol.* 419, 54–63. doi: 10.1016/j.ydbio.2016.03.031

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

Guo, X., Shakeel, M., Wang, D., Qu, C., Yang, S., Ahmad, S., et al. (2022). Metabolome and transcriptome profiling unveil the mechanisms of light-induced anthocyanin synthesis in rabbiteye blueberry (*vaccinium ashei*: Reade). *BMC Plant Biol.* 22, 1–14. doi: 10.1186/s12870-022-03585-x

Guo, X., Wang, Y., Zhai, Z., Huang, T., Zhao, D., Peng, X., et al. (2018). Transcriptomic analysis of light-dependent anthocyanin accumulation in bicolored cherry fruits. *Plant Physiol. Biochem.* 130, 663–677. doi: 10.1016/j.plaphy.2018.08.016

He, Y., Chen, H., Zhou, L., Liu, Y., and Chen, H. (2019). Comparative transcription analysis of photosensitive and non-photosensitive eggplants to identify genes involved in dark regulated anthocyanin synthesis. *BMC Genomics* 20, 678. doi: 10.1186/s12864-019-6023-4

He, Y., Li, S., Dong, Y., Zhang, X., Li, D., Liu, Y., et al. (2022). Fine mapping and characterization of the dominant gene *SmFTSH10* conferring non-photosensitivity in eggplant (*Solanum melongena* L.). *Theor. Appl. Genet.* 135 (7), 2187–2196. doi: 10.1007/s00122-022-04078-z

Hoai Nguyen, N. (2020). HY5, an integrator of light and temperature signals in the regulation of anthocyanins biosynthesis in *Arabidopsis*. *AIMS Mol. Science.* 7, 70–81. doi: 10.3934/molsci.2020005

Hong, Y., Li, M., and Dai, S. (2019). iTRAQ-based protein profiling provides insights into the mechanism of light-induced anthocyanin biosynthesis in chrysanthemum (*Chrysanthemum× morifolium*). *Genes-Basel.* 10, 1024–1042. doi: 10.3390/genes10121024

Hong, Y., Tang, X., Huang, H., Zhang, Y., and Dai, S. (2015). Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in chrysanthemum. *BMC Genomics* 16, 202. doi: 10.1186/s12864-015-1428-1

Huang, H., Li, Y., and Dai, S. (2016). Investigation of germplasm in chrysanthemum cultivars with light-independent coloration. *II Int. Symposium Germplasm Ornamentals* 1185, 55–64. doi: 10.17660/ActaHortic.2017.1185.9

Jiang, H., Zhou, L. J., Gao, H. N., Wang, X. F., Li, Z. W., and Li, Y. Y. (2022). The transcription factor MdMYB2 influences cold tolerance and anthocyanin accumulation by activating SUMO E3 ligase MdSIZ1 in apple. *Plant Physiol.* 189 (4), 2044–2060. doi: 10.1093/plphys/kiac211

Jiang, M., Ren, L., Lian, H., Liu, Y., and Chen, H. (2016). Novel insight into the mechanism underlying light-controlled anthocyanin accumulation in eggplant (*Solanum melongena* L.). *Plant Sci.* 249, 46–58. doi: 10.1016/j.plantsci.2016.04.001

Kiferle, C., Fantini, E., Bassolino, L., Povero, G., Spelt, C., Buti, S., et al. (2015). Tomato R2R3-MYB proteins SlANT1 and SlAN2: same protein activity, different roles. *PLoS One* 10, e0136365. doi: 10.1371/journal.pone.0136365

Kim, M. J., Kim, P., Chen, Y., Chen, B., Yang, J., Liu, X., et al. (2021). Blue and UV-B light synergistically induce anthocyanin accumulation by co-activating nitrate reductase gene expression in Anthocyanin fruit (*Aft*) tomato. *Plant Biol.* 23, 210–220. doi: 10.1111/plb.13141

Krol, A. R. V. D., Mur, L. A., Beld, M., Moi, J. N. M., and Stuitje, A. R. (1990). Flavonoid genes in Petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell.* 2, 291–299. doi: 10.1105/tpc.2.4.291

Li, S. F., Allen, P. J., Napoli, R. S., Browne, R. G., Pham, H., and Parish, R. W. (2020). MYB–bHLH–TTG1 regulates *Arabidopsis* seed coat biosynthesis pathways directly and indirectly via multiple tiers of transcription factors. *Plant Cell Physiol.* 61, 1005–1018. doi: 10.1093/pcp/pcaa027

Li, P., Chen, X., Sun, F., and Dong, H. (2017). Tobacco TTG2 and ARF8 function concomitantly to control flower coloring by regulating anthocyanin synthesis genes. *Plant Biol.* 19, 525–532. doi: 10.1111/plb.12560

Li, J., He, Y.-J., Zhou, L., Liu, Y., Jiang, M., Ren, L., et al. (2018). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics* 19, 1–12. doi: 10.1186/s12864-018-4587-z

Li, D., Qian, J., Li, W., Yu, N., Gan, G., Jiang, Y., et al. (2021). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol. Ecol. Resour* 21 (4), 1274–1286. doi: 10.1111/1755-0998.13321

Li, X., Zhu, C., Yeh, C. T., Wu, W., Takacs, E. M., Petsch, K. A., et al. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Res.* 22, 2436–2444. doi: 10.1101/gr.140277.112

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, W., Mei, Z., Yu, L., Gu, T., Li, Z., Zou, Q., et al. (2023). The ABA-induced NAC transcription factor MdNAC1 interacts with a bZIP-type transcription factor to promote anthocyanin synthesis in red-fleshed apples. *Hortic. Res.* 10 (5), uhad049. doi: 10.1093/hr/uhad049

Liu, C., Sun, J., Lu, Y., and Bo, Y. (2016). Effects of anthocyanin on serum lipids in dyslipidemia patients: a systematic review and meta-analysis. *PloS One* 11, e0162089. doi: 10.1371/journal.pone.0162089

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in *Solanaceous* vegetables: *A review*. *Front. Chem.* 6, 52–68. doi: 10.3389/fchem.2018.00052

Liu, B., Wang, L., Wang, S., Li, W., Liu, D., Guo, X., et al. (2019). Transcriptomic analysis of bagging-treated 'Pingguo' pear shows that *MYB4-like1*, *MYB4-like2*, *MYB1R1* and *WDR* involved in anthocyanin biosynthesis are up-regulated in fruit peels in response to light. *Sci. Hortic-Amsterdam.* 244, 428–434. doi: 10.1016/j.scienta.2018.09.040

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Lloyd, A., Brockman, A., Aguirre, L., Campbell, A., Bean, A., Cantero, A., et al. (2017). 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.* 58, 1431–1441. doi: 10.1093/pcp/pcx075

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Ma, Y., Ma, X., Gao, X., Wu, W., and Zhou, B. (2021b). Light induced regulation pathway of anthocyanin biosynthesis in plants. *Int. J. Mol. Sci.* 22, 11116–11131. doi: 10.3390/ijms222011116

Ma, A., Wang, D., Lu, H., Wang, H., Qin, Y., Hu, G., et al. (2021a). LcCOP1 and LcHY5 control the suppression and induction of anthocyanin accumulation in bagging and debagging litchi fruit pericarp. *Sci. Hortic-Amsterdam.* 287, 110281–110290. doi: 10.1016/j.scienta.2021.110281

Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., et al. (2013). 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.* 74, 638–651. doi: 10.1111/tpj.12153

Mangino, G., Arrones, A., Plazas, M., Pook, T., Prohens, J., Gramazio, P., et al. (2022). Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. *Front. Plant Sci.* 13, 847789. doi: 10.3389/fpls.2022.847789

Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., et al. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *Plant Cell.* 15, 1689–1703. doi: 10.1105/tpc.012963

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meng, L. S., and Liu, A. (2015). Light signaling induces anthocyanin biosynthesis via AN3 mediated COP1 expression. *Plant Signal Behav.* 10, e1001223. doi: 10.1080/15592324.2014.1001223

Montefiori, M., Brendolise, C., Dare, A. P., Lin-Wang, K., Davies, K. M., Hellens, R. P., et al. (2015). In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. *J. Exp. Bot.* 66, 1427–1436. doi: 10.1093/jxb/eru494

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., et al. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near *WUSCHEL*. *Plant Physiol.* 156, 2244–2254. doi: 10.1104/pp.111.173997

Naing, A. H., and Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiol. Plant* 172, 1711–1723. doi: 10.1111/ppl.13373

Nayanathara, A. R., Aalolam, K. P., and Reshma, J. K. (2016). Evaluation of total phenol, flavonoid and anthocyanin content in different varieties of eggplant. *Emer Life Sci. Res.* 2, 63–65.

Nino-Medina, G., Urias-Orona, V., Muy-Rangel, M. D., and Heredia, J. B. (2017). Structure and content of phenolics in eggplant (*Solanum melongena*). *S Afr J. Bot.* 111, 161–169. doi: 10.1016/j.sajb.2017.03.016

Oladosu, Y., Rafii, M. Y., Arolu, F., Chukwu, S. C., Salisu, M. A., Olaniyan, B. A., et al. (2021). Genetic diversity and utilization of cultivated eggplant germplasm in varietal improvement. *Plants (Basel).* 10, 1714–1734. doi: 10.3390/plants10081714

Peng, Y., Thrimawithana, A. H., Cooney, J. M., Jensen, D. J., Espley, R. V., and Allan, A. C. (2020). The proanthocyanin-related transcription factors MYBC1 and WRKY44 regulate branch points in the kiwifruit anthocyanin pathway. *Sci. Rep.* 10, 14161. doi: 10.1038/s41598-020-70977-0

Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G., Rensing, S. A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38, D822–D827. doi: 10.1093/nar/gkp805

Pesch, M., Dartan, B., Birkenbihl, R., Somssich, I. E., and Hulskamp, M. (2014). *Arabidopsis* TTG2 regulates *TRY* expression through enhancement of activator complex-triggered activation. *Plant Cell.* 26, 4067–4083. doi: 10.1105/tpc.114.129379

Pojer, E., Mattivi, F., Johnson, D., and Stockley, C. S. (2013). The case for anthocyanin consumption to promote human health: a review. *Compr. Rev. Food Sci. F.* 12, 483–508. doi: 10.1111/1541-4337.12024

Qiao, J., Liu, J., Li, S., and Wang, L. (2022). Prediction of fruit color genes under the calyx of eggplant based on genome-wide resequencing in an extreme mixing pool (in Chinese). *Acta Hortic. Sinica.* 49, 613. doi: 10.16420/j.issn.0513-353x.2021-0018

Ramsay, N. A., and Glover, B. J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011

Roberts, A. R., Vecellio, M., Chen, L., Ridley, A., Cortes, A., Knight, J. C., et al. (2016). An ankylosing spondylitis-associated genetic variant in the *IL23R-IL12RB2* intergenic region modulates enhancer activity and is associated with increased Th1-cell differentiation. *Ann. Rheumatol. Dis.* 75, 2150–2156. doi: 10.1136/annrheumdis-2015-208640

Ryu, S., Han, J. H., Cho, J. G., Jeong, J. H., Lee, S. K., and Lee, H. J. (2020). High temperature at véraison inhibits anthocyanin biosynthesis in berry skins during ripening in 'Kyoho'grapevines. *Plant Physiol. Biochem.* 157, 219–228. doi: 10.1016/j.plaphy.2020.10.024

Saini, D. K., and Kaushik, P. (2019). Visiting eggplant from a biotechnological perspective: A review. *Sci. Hortic-Amsterdam.* 253, 327–340. doi: 10.1016/j.scienta.2019.04.042

Schwartz, C., Balasubramanian, S., Warthmann, N., Michael, T. P., Lempe, J., Sureshkumar, S., et al. (2009). *Cis*-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics.* 183 723-732, 721SI–727SI. doi: 10.1534/genetics.109.104984

Shan, X., Li, Y., Yang, S., Gao, R., Zhou, L., Bao, T., et al. (2019). A functional homologue of *Arabidopsis TTG1* from *Freesia* interacts with bHLH proteins to regulate anthocyanin and proanthocyanidin biosynthesis in both Freesia hybrida and *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 141, 60–72. doi: 10.1016/j.plaphy.2019.05.015

Shi, B., Wu, H. X., Zheng, B., Qian, M. J., Gao, A. P., and Zhou, K. B. (2021). Analysis of light-independent anthocyanin accumulation in mango (*Mangifera indica* L.). *Horticulturae.* 7, 423–434. doi: 10.3390/horticulturae7110423

Sivankalyani, V., Feygenberg, O., Diskin, S., Wright, B., and Alkan, N. (2016). Increased anthocyanin and flavonoids in mango fruit peel are associated with cold and pathogen resistance. *Postharvest Biol. Tec.* 111, 132–139. doi: 10.1016/j.postharvbio.2015.08.001

Stadhouders, R., Aktuna, S., Thongjuea, S., Aghajanirefah, A., Pourfarzad, F., Van Ijcken, W., et al. (2014). *HBS1L-MYB* intergenic variants modulate fetal hemoglobin via long-range *MYB* enhancers. *J. Clin. Invest.* 124, 1699–1710. doi: 10.1172/JCI71520

Stam, M., Belele, C., Dorweiler, J. E., and Chandler, V. L. (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* 16, 1906–1918. doi: 10.1101/gad.1006702

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* 43, 1160–1163. doi: 10.1038/ng.942

Sun, L., Li, S., Tang, X., Fan, X., Zhang, Y., Jiang, J., et al. (2020). Transcriptome analysis reveal the putative genes involved in light-induced anthocyanin accumulation in grape 'Red Globe'(*V. vinifera* L.). *Gene* 728, 144284. doi: 10.1016/j.gene.2019.144284

Sun, X. H., Zhou, T. T., Wei, C. H., Lan, W. Q., Zhao, Y., Pan, Y. J., et al. (2018). Antibacterial effect and mechanism of anthocyanin rich Chinese wild blueberry extract

on various foodborne pathogens. *Food Control.* 94, 155–161. doi: 10.1016/j.foodcont.2018.07.012

Takos, A. M., Jaffe, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142, 1216–1232. doi: 10.1104/pp.106.088104

Tang, B., Li, L., Hu, Z., Chen, Y., Tan, T., Jia, Y., et al. (2020). Anthocyanin accumulation and transcriptional regulation of anthocyanin biosynthesis in purple pepper. *J. Agr Food Chem.* 68, 12152–12163. doi: 10.1021/acs.jafc.0c02460

Timothy, A., and Holton, E. C. C. (1995). genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell.* 7, 1071–1083. doi: 10.2307/3870058

Toppino, L., Barchi, L., Lo Scalzo, R., Palazzolo, E., Francese, G., Fibiani, M., et al. (2016). Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* L.). *Front. Plant Sci.* 7, 256. doi: 10.3389/fpls.2016.00256

Van Der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., et al. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Front. Plant Sci.* 5, 227. doi: 10.3389/fpls.2014.00227

Verweij, W., Spelt, C. E., Bliek, M., De Vries, M., Wit, N., Faraco, M., et al. (2016). Functionally similar WRKY proteins regulate vacuolar acidification in Petunia and hair development in *Arabidopsis*. *Plant Cell.* 28, 786–803. doi: 10.1105/tpc.15.00608

Visser, M., Palstra, R. J., and Kayser, M. (2014). Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby *BNC2* pigmentation gene. *Hum. Mol. Genet.* 23, 5750–5762. doi: 10.1093/hmg/ddu289

Wang, H., Niu, Q. W., Wu, H. W., Liu, J., Ye, J., Yu, N., et al. (2015). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J.* 84, 404–416. doi: 10.1111/tpj.13018

Xiang, C. C., Zhang, W. W., Luo, S. X., Zhao, J. J., Zhang, T. T., Shen, S. X., et al. (2015). Genetic analysis on fruit color under calyx and correlation analysis of SSR maker in eggplant (in Chinese). *J. Agric. Univ. Hebei.* 38, 50–55. doi: 10.13320/j.cnki.jauh.2015.0083

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Yang, J. F., Chen, Y. Z., Kawabata, S., Li, Y. H., and Wang, Y. (2017). Identification of light-independent anthocyanin biosynthesis mutants induced by Ethyl Methane Sulfonate in turnip "Tsuda" (*Brassica rapa*). *Int. J. Mol. Sci.* 18, 1288–1303. doi: 10.3390/ijms18071288

Yang, G., Li, L., Wei, M., Li, J., and Yang, F. (2022). *SmMYB113* is a key transcription factor responsible for compositional variation of anthocyanin and color diversity among eggplant peels. *Front. Plant Sci.* 13, 843996. doi: 10.3389/fpls.2022.843996

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 16, 284–287. doi: 10.1089/omi.2011.0118

Zha, Q., Xi, X. J., He, Y., and Jiang, A. L. (2019). Bagging affecting sugar and anthocyanin metabolism in the ripening period of grape berries. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca.* 47, 1194–1205. doi: 10.15835/nbha47411577

Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development.* 130, 4859–4869. doi: 10.1242/dev.00681

Zhang, X. T., He, Y. J., Liu, Y., and Chen, H. Y. (2021b). Development and validation of molecular marker for photosensitivity of anthocyanin production in eggplant fruit (in Chinese). *J. Nanjing Agric. University.* 44, 637–645.

Zhang, Y., Hu, Z., Chu, G., Huang, C., Tian, S., Zhao, Z., et al. (2014). Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). *J. Agr Food Chem.* 62, 2906–2912. doi: 10.1021/jf404574c

Zhang, W. S., Li, X., Zheng, J. T., Wang, G. Y., Sun, C. D., Ferguson, I. B., et al. (2008). Bioactive components and antioxidant capacity of Chinese bayberry (*Myrica rubra* Sieb. and Zucc.) fruit in relation to fruit maturity and postharvest storage. *Eur. Food Res. Technol.* 227, 1091–1097. doi: 10.1007/s00217-008-0824-z

Zhang, B., Zhu, Z. Z., Qu, D., Wang, B. C., Hao, N. N., Yang, Y. Z., et al. (2021a). MdBBX21, a B-Box protein, positively regulates light-induced anthocyanin accumulation in apple peel. *Front. Plant Sci.* 12, 774446. doi: 10.3389/fpls.2021.774446

Zhao, L., Sun, J., Cai, Y., Yang, Q., Zhang, Y., Ogutu, C. O., et al. (2022). *PpHYH* is responsible for light-induced anthocyanin accumulation in fruit peel of Prunus persica. *Tree Physiol.* 42, 1662–1677. doi: 10.1093/treephys/tpac025

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 January 2024
DOI 10.3389/fpls.2023.1293599

Check for updates

OPEN ACCESS

EDITED BY
Muhammad Aamir Manzoor,
Shanghai Jiao Tong University, China

REVIEWED BY
Emre Sevindik,
Adnan Menderes University, Türkiye
Nilay Yönet,
Yıldız Technical University, Türkiye

*CORRESPONDENCE
Jianhua Zhang
zhangjianhua@caas.cn
Longyu Huang
huanglongyu1@163.com

RECEIVED 13 September 2023
ACCEPTED 27 December 2023
PUBLISHED 17 January 2024

CITATION
Chen X, Huang L, Fan J, Yan S, Zhou G and Zhang J (2024) KASP-IEva: an intelligent typing evaluation model for KASP primers. *Front. Plant Sci.* 14:1293599. doi: 10.3389/fpls.2023.1293599

COPYRIGHT
© 2024 Chen, Huang, Fan, Yan, Zhou and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

KASP-IEva: an intelligent typing evaluation model for KASP primers

Xiaojing Chen[1,2], Longyu Huang[2,3,4]*, Jingchao Fan[1,2], Shen Yan[1], Guomin Zhou[2,3] and Jianhua Zhang[1,2]*

[1]National Agriculture Science Data Center, Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China, [2]National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya, China, [3]Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, China, [4]Hainan Yazhou Bay Seed Laboratory, Sanya, China

KASP marker technology has been used in molecular marker-assisted breeding because of its high efficiency and flexibility, and an intelligent evaluation model of KASP marker primer typing results is essential to improve the efficiency of marker development on a large scale. To this end, this paper proposes a gene population delineation method based on NTC identification module and data distribution judgment module to improve the accuracy of K-Means clustering, and introduces a decision tree to construct the KASP-IEva primer typing evaluation model. The model firstly designs the NTC identification module and data distribution judgment module to extract four types of data, grouping and categorizing to achieve the improvement of the distinguishability of amplification product signals; secondly, the K-Means algorithm is used to aggregate and classify the data, to visualize the five aggregated clusters and to obtain the morphology location eigenvalues; lastly, the evaluation criteria for the typing effect level are constructed, and the logical decision tree is used to make conditional discrimination on the eigenvalues in order to realize the score prediction. The performance of the model was tested by the KASP marker typing test results of 2519 groups of cotton varieties, and the following conclusions were obtained: the model is able to visualize the aggregation and classification effects of the amplification products of NTC, pure genotypes, heterozygous genotypes, and untyped genotypes, enabling rapid and accurate KASP marker typing evaluation. Comparing and analyzing the model evaluation results with the expert evaluation results, the average accuracy rate of the four grades evaluated by the model was 87%, and the overall evaluation results showed an uneven distribution of the grades with significant differential characteristics. When evaluating 2519 KASP fractal maps, the expert evaluation consumes 15 hours, and the model evaluation only uses 8min27.45s, which makes the model intelligent evaluation significantly better than the expert evaluation from the perspective of time. The establishment of the model will further enhance the application of KASP markers in molecular marker-assisted breeding and provide technical support for the large-scale screening and identification of excellent genotypes.

KEYWORDS

intelligent evaluation, KASP marker, decision tree, genotyping, cotton, molecular marker-assisted selection

1 Introduction

Cotton is an important fiber crop in the world, and is also an important strategic material related to the national economy and people's livelihood of China, with high economic value (Gu et al., 2020; Lu et al., 2023; Chen et al., 2024). A number of important agronomic traits of cotton are characterized by quantitative genetic features, which are easily disturbed by external environmental conditions and are genetically negatively correlated, resulting in a large workload and low selection efficiency in cotton variety selection (Li and Yuan, 2013). The traditional breeding methods of selection of low accuracy, long cycle, poor predictability, and molecular marker-assisted selection (MAS) can make up for these shortcomings. Molecular marker-assisted selection is a direct selection of genotypes for target traits with the help of molecular markers, which greatly shortens the breeding time and reduces the population planting scale, and is of great significance for the rapid breeding of new cotton materials and varieties (Abdelraheem et al., 2021; Gao et al., 2023).

At present, the commonly used molecular marker technologies mainly include RFLP, RAPD, AFLP, SSR, InDel, SNP, etc (Amiteye, 2021; Geng et al., 2021; Al-Khayri et al., 2022; Kumar et al., 2022), among which SNP (single nucleotide polymorphisms) markers have been designated as one of the priority recommended marker methods by the International Union for the Protection of New Plant Variety Rights (UPOV) BMT molecular testing guidelines (Nie et al., 2021), and KASP (kompetitive allele-specific PCR), as a mainstream SNP high-throughput typing technology, is a novel PCR-based homogeneous fluorescent SNP typing method. KASP has high analytical stability and accuracy, and KASP provides great flexibility in terms of the number of SNPs and samples used for the determination and can achieve high-precision dual-allele genotyping (SNPs and InDels) for a small number of target markers in large-scale segregating or natural populations (Yang et al., 2020; Wang et al., 2021; Zhao et al., 2021).

Some scholars at home and abroad are working on kompetitive allele-specific PCR studies in cotton and other crops. Guo et al. (2023) developed the M-1590 KASP marker to classify 90 kinds of cotton materials and used the grid lines of the coordinate axes to assist in visually inspecting the high and low signal values of the allele population, and the results showed that the marker could only classify four kinds of materials. Fan Tao (Fan et al., 2021) and others screened multiple spikes and long-grain wheat resources based on KASP marker technology, using two straight lines parallel to the axes to divide the area composed of the axes into four parts on average, and observing the high and low fluorescence values of the signal points and the size of the angle between the different genotypes, to evaluate whether the primers were able to type the different materials in the population well or not. Xu Biyu (Xu et al., 2023) and others used KASP to identify key variants or genes responsible for stem trichome traits in cotton stalks and confirmed that mutations co-segregated with the stem trichome phenotype by directly observing the relative independence of gene populations in allelic discrimination maps. Byers et al. (2012) and others transformed hundreds of putative allotetraploid cotton SNPs into functional SNPs. In the genotyping determination, the genotyping map of Fluidigm SNP analysis software was divided into tables to determine whether the heterozygote group was distinguished from the homozygote group, so as to evaluate the amplification and separation of the genotype group. Sheng et al. (2022) used the designed markers to genotype 86 kale crop materials at the target loci, and they verified the accuracy and applicability of the markers by outlining and observing the distribution trend lines of the signal points. Li Lihua (Li et al., 2022) and others designed KASP primers based on the results of pre-fiber strength association analysis and genotyped 376 land cotton materials, and they proposed that the population was classified into three categories by the marker, the pure and genotypes were close to the axes, and the heterozygous genotypes were located in the center of the typing diagram of the typing results could prove that the marker had DNA polymorphism in the group. Wang et al. (2022) used 48 KASP markers to genotype 348 grape germplasm for genotyping, observed the separation status of pure and heterozygous populations in the output genotyping map of LGC's KASP detection technology platform, and finally screened out 46 markers with good fluorescence genotyping results. In the traditional KASP primer typing results validation, the evaluation of competitive primer combination status relies heavily on professional knowledge and long-term experience judgment. The expert visual evaluation requires that the evaluators in the relevant fields have a high degree of professionalism and data analysis ability, and the diversity of genotypic amplification signal patterns puts forward a higher demand for the expert's ability to make judgments. The approach also suffers from highly subjective results, high physical effort, low precision and slow validation of large-scale materials.

With the development of high-throughput SNP genotyping technology, new molecular marker technologies have been continuously developed, and KASP genotyping technology has been gradually involved in selection tests to optimize the best technology to be applied in the field of crop breeding. For example, in order to compare the detection differences among TaqMan, KASP and rhAmp, Broccanello et al. (2018) utilized the indicators of distance of genotypic clusters to the NTC, cluster angle segregation and cluster spread to perform analysis of variance to quantitatively assess the typing effect by comparing the differences in the values of the indexes. On this basis, Ayalew et al. (2019) proposed to measure allelic discrimination by using the cluster separation angle, and cluster compactness by calculating the standard deviation of the distances between the data points in the clusters with the mean coordinate value of the clusters to realize the comparison of the three genotyping platforms in hexaploid wheat. Although this method provides a more scientific and statistical idea for evaluating the typing effect, it still has the problems of few evaluation indexes and low degree of intelligence, in addition, the method does not provide accurate evaluation standards and is only applicable to multi-detection technology difference comparison tests.

To summarize, most KASP studies mainly adopt the way of expert interpretation of primer typing diagrams, and some of them have used statistical analyses. There is a scarcity of studies on intelligent evaluation of typing effects, which makes it difficult to evaluate the amplification efficiency and specificity of competitive primers for crops on a large scale. In this study, in strict accordance

with the grade evaluation criteria, based on the K-Means algorithm, we propose the NTC identification module and data distribution judgment module to improve the accuracy of gene population delineation, and design the decision tree for the grade evaluation of typing effect to construct the intelligent typing evaluation model for the KASP-IEva primers. Based on the results of kompetitive allele-specific PCR study, we organized experimental data, applied the KASP-IEva model to type the amplification products of 2519 groups of KASP markers, realized the classification of different genotypes of land cotton materials, constructed the evaluation decision tree according to the grade evaluation criteria, and intelligently and rapidly screened and identified the KASP markers with excellent typing effect, with the aim of providing data support for improving the success rate of KASP marker development and technical support for molecular marker-assisted breeding and other work.

2 Materials and methods

2.1 Experimental data

In this study, the results of the KASP marker test of cotton variety resource materials were used as experimental data, which came from the Cotton Quality Supervision, Inspection and Testing Center of the Ministry of Agriculture and Rural Affairs of the Chinese Academy of Agricultural Sciences (CAAS) Cotton Research Institute, and the KASP marker test materials contained 450 resource varieties, 260 line materials and 1,200 audited varieties and 609 genetically segregated population materials. The content of the experimental data is 2519 groups of amplification product information, each group contains 46 or 94 DNA samples detection data and 2 NTC (negative control reaction without DNA samples added in the PCR assay) detection data, and each detection data includes SNP locus number, sample number, relative value of the sample HEX and FAM fluorescence signal magnitude, daughter plate serial number, mother plate serial number and genotype and other information, and so on. Only the first four items were involved in the experiment, and the specific data are shown in Table 1.

TABLE 1 Information of a set of 48 KASP-labeled amplification products.

Relative value of HEX fluorescence signal magnitude X	Relative value of FAM fluorescence signal magnitude Y	SNP locus number SNPID	Sample Number SubjectID
1.36584	0.40199	Z001	13
1.28488	0.38166	Z001	16
1.35823	0.38800	Z001	24
1.30018	0.38155	Z001	51
1.29644	0.39695	Z001	62
0.33218	1.20284	Z001	63
1.32567	0.40944	Z001	73
1.34497	0.40108	Z001	121
1.30231	0.38416	Z001	123
1.31518	0.40546	Z001	136
1.34153	0.40832	Z001	137
1.58476	0.41924	Z001	138
0.35001	0.35327	Z001	139
0.37357	0.34464	Z001	141
1.35760	0.40401	Z001	151
1.18718	0.35502	Z001	167
1.40429	0.40723	Z001	174
1.67185	0.42256	Z001	203
1.66754	0.42275	Z001	211
1.69116	0.42726	Z001	216
1.74076	0.44449	Z001	229
1.72609	0.42531	Z001	238
0.29392	1.11550	Z001	269
1.60226	0.42337	Z001	277
1.65182	0.40624	Z001	285
1.69974	0.42643	Z001	299
1.63786	0.42332	Z001	308
0.30004	0.46021	Z001	233-2
0.33127	0.31929	Z001	234-2
1.62075	0.43070	Z001	238-1
1.61211	0.41000	Z001	239-2
0.29504	1.40139	Z001	251
0.44045	0.40432	Z001	260
0.27281	1.41842	Z001	267
0.26008	1.22395	Z001	268
0.27529	1.14970	Z001	270

(Continued)

2.2 The KASP-IEva model

2.2.1 Model structure

In general, experts evaluate the typing effect based on the typing map results of the LGC-SNPline platform software by manual judgment based on experience, and there is a lack of unified typing effect evaluation standards and evaluation models. In order to quantify the evaluation criteria and meet the need for the model to be evaluated with higher precision and faster speed, this study constructs the KASP-IEva model by using the intelligent typing module and decision tree on the basis of more detailed evaluation criteria for the typing effect level, so that the factors within the typing diagram undergo a recursive process, and the precise calculation is performed at each intermediate node to classify the

TABLE 1 Continued

Relative value of HEX fluorescence signal magnitude X	Relative value of FAM fluorescence signal magnitude Y	SNP locus number SNPID	Sample Number SubjectID
0.29831	0.32121	Z001	286
1.68704	0.43980	Z001	290
1.57209	0.43311	Z001	294
1.48042	0.41833	Z001	295
1.68928	0.42789	Z001	JM-11
1.64988	0.42502	Z001	ZZM-2
1.67375	0.42881	Z001	277
0.23985	1.54469	Z001	278
1.63963	0.43460	Z001	285
1.65306	0.44516	Z001	Z1
0.26926	0.83101	Z001	NTC
0.79678	0.36945	Z001	NTC

attributes. The research-constructed model uses unlabeled datasets for processing, groups the data based on the similarity of the underlying structure of the given dataset so that the typing results are in line with professional empirical perceptions, and then realizes score prediction through the computation of eigenvalues of genotypic amplification product classifications and combinations, which then realizes the typing evaluation intelligently.

The KASP-IEva primer intelligent typing evaluation model is mainly composed of four modules, namely, the NTC identification module, data distribution judgment module, K-Means clustering module, and typing effect evaluation, in which the typing effect evaluation module adopts the decision tree theory, and the structure of the model is shown in Figure 1. From the viewpoint of data distribution, there are four main types, namely, NTC data, untyped product data, heterozygous genotype data, and pure genotype data. The distribution of NTC data is characterized by no obvious boundaries, random degree of aggregation, and fixed quantity. The non-NTC data had some missing types, which were categorized into three categories: containing four types of data, not containing untyped data, and not containing heterozygous genotype data. The decentralized distribution of non-NTC data resulted in the boundary of different types of data not being easy to distinguish. Firstly, each group of fluorescence signal magnitude relative value data and sample number are read, NTC identification module controls the separation of NTC data from the rest of the data, and two different branches are used to perform clustering operation separately without destroying the connectivity of the final results, in order to improve the classification ability of the model to the target under the data mixing condition. Data distribution judgment module extracts different types of data by discriminating the location characteristics, and further splits the non-NTC data into 3 parts, which reduces the training time of the subsequent algorithms of the model and improves the distinguishability of the amplified signals. The machine learning algorithm in the K-Means module extracts the deep structure eigenvalues between amplification products on the basis of obtaining 5 aggregation clusters of different sizes in order to compute the evaluation indexes. The decision tree for the evaluation of the fractal effect categorizes the eigenvalue dataset through multiple conditional discriminative processes, the Based on the tree structure, the decision-making judgment of amplification efficiency and specificity of competitive primers is carried out, and the evaluation results of typing effect of each group are finally obtained. These four modules are described in detail below.

2.2.2 NTC identification module

Retrieval is performed within each set of data, and the elements at the corresponding positions of the NTC data are transformed to form a two-dimensional array A[M][N] of pairs of fluorescent signal magnitude relative-value tuples, in which each row represents the coordinates of a point.

2.2.3 Data distribution judgment module

There is a complex nonlinear correlation between amplification products, and the relative value of fluorescence signal magnitude of each data point affects each other to different degrees, so this paper tries to add a data distribution judgment module into the model to make the amplification product characteristic information more distinguishable. This module is based on the maximum value, minimum value, and manually adjusted segmentation value of the data set to quantitatively divide the range of the region, and proposes the extraction method of grouped data, i.e., the target point in the experimental data that is in the same region as the partitioning result is inputted into a two-dimensional array as the data of this genotype. The specific process of the setup is as follows:

For the division of the distribution region of heterozygous genotypes, the segmentation of the value along the horizontal and vertical coordinate axes, respectively, is used to obtain the maximum and minimum values of the relative value of the fluorescence signal level in all non-NTC data, and after calculating the appropriate distance that needs to be intercepted within the interval of the coordinate axes, the relative value of the fluorescence signal level of the data whose horizontal and vertical coordinates meet the requirements of the region is put into the two-dimensional array B [M][N], and the calculation process is represented by Equations 1–7.

$$X_{B\max} = \max(X_1, X_2, \ldots, X_N) \tag{1}$$

$$X_{B\min} = \min(X_1, X_2, \ldots, X_N) \tag{2}$$

$$Y_{B\max} = \max(Y_1, Y_2, \ldots, Y_N) \tag{3}$$

$$Y_{B\min} = \min(Y_1, Y_2, \ldots, Y_N) \tag{4}$$

$$X_{BD} = (X_{B\max} - X_{B\min})/4.2 \tag{5}$$

$$Y_{BD} = (Y_{B\max} - Y_{B\min})/4.2 \tag{6}$$

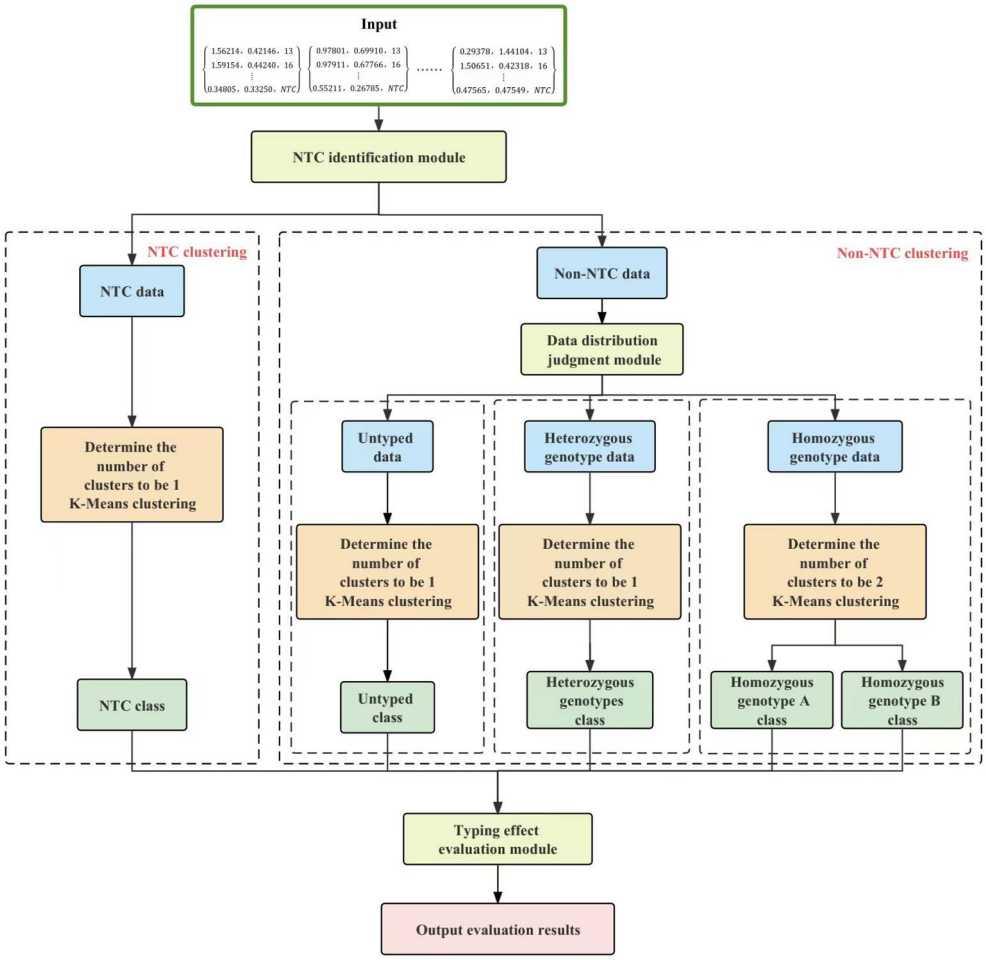


FIGURE 1
Overall structure of the KASP-IEva model.

$$B[I_{Cx}][I_{Cy}] = \begin{cases} \sum[I_{Cx}][I_{Cy}] & \begin{pmatrix} X_{Bmin} + X_{BD} < I_{Bx} < X_{Bmax} - X_{BD,} \\ Y_{Bmin} + Y_{BD} < I_{By} < Y_{Bmax} - Y_{BD} \end{pmatrix} \\ \sum 0 & else \end{cases} \tag{7}$$

Where: X_i and I_{Bx} denote the relative values of HEX fluorescence signal magnitude of experimental data points of non-NTC, Y_i and I_{By} denote the relative values of FAM fluorescence signal magnitude of experimental data points of non-NTC, X_{BD} is the distance of heterozygous genes' distribution region from the ends along the horizontal direction, and Y_{BD} is the distance of heterozygous genes' distribution region from the ends along the vertical direction.

For the division of the distribution region of the untyped product, the same method as described above is used to obtain the maximum and minimum values of the remaining relative values of the fluorescence signal magnitude after the extraction of the NTC data and the heterozygous genotype data, calculate the appropriate interception distance within the range of the coordinate axes, and form a two-dimensional array of the relative values of the fluorescence signal magnitude satisfying the condition of being in the interception region into a two-dimensional array C[M][N], and the remaining data can be expressed directly as D[M][N], and the calculation process is represented by Equations 8–14:

$$X_{C\max} = \max(X_1, X_2, \ldots, X_M) \tag{8}$$

$$X_{C\min} = \min(X_1, X_2, \ldots, X_M) \tag{9}$$

$$Y_{C\max} = \max(Y_1, Y_2, \ldots, Y_M) \tag{10}$$

$$Y_{C\min} = \min(Y_1, Y_2, \ldots, Y_M) \tag{11}$$

$$X_{CD} = (X_{C\max} - X_{C\min})/3.2 \tag{12}$$

$$Y_{CD} = (Y_{C\max} - Y_{C\min})/3.2 \tag{13}$$

$$C[I_{Cx}][I_{Cy}] = \begin{cases} \sum[I_{Cx}][I_{Cy}] & (I_{Cx} < X_{Cmin} + X_{CD}, I_{Cy} < Y_{Cmin} + Y_{CD}) \\ \sum 0 & else \end{cases} \tag{14}$$

Where: X_j and I_{Cx} denote the relative values of HEX fluorescence signal magnitude for unfractionated and pure

amplification product data points, Y_j and I_{Cy} denote the relative values of FAM fluorescence signal magnitude for unfractionated and pure amplification product data points, X_{CD} is the horizontal distance from the distribution area of unfractionated amplification products, and Y_{CD} is the vertical distance from the distribution area of unfractionated amplification products.

2.2.4 K-Means clustering module

Based on the K-Means machine learning algorithm to complete the modeling process of the fractal model, the clustering idea is used to mine the potential correlation of the genotype data, and the data are grouped categorized, and visualized. The K-Means algorithm was proposed by the Lloyd scholars in 1982, and the algorithm is one of the most classical and commonly used unsupervised learning algorithms to solve the clustering problem (Chakraborty et al., 2020; Sinaga and Yang, 2020). It divides the set of samples into K-class clusters and uses Euclidean distance to measure the similarity between the samples, which results in high similarity within clusters of the same class and low similarity between clusters of different classes (Mirzal, 2020). The K-Means algorithm process is as follows:

Randomly select K samples from a set of sample sets as the initial center of mass, calculate the Euclidean distance between each sample point and the K clustering centers, and use this as the basis for assigning all sample points to their nearest clustering centers; in order to reduce the sum of squares of the error of the dataset, calculate the mean vector of the cluster Ci as the new center of mass of the cluster; the smaller the squared error SSE is, the greater the similarity of the samples within the cluster is, and repeat the training of each cluster center, until the cluster center position and the size of SSE value no longer change significantly, to get the final clustering results (Chou, 2016; Yin et al., 2022). The calculation process is represented by equations Equations 15–18:

$$clusterset = \{C_1, C_2, \cdots, C_K\} \tag{15}$$

$$d(x, C_i) = \sqrt{\sum_{j=1}^{m}(x_j - C_{ij})^2}, m = 2 \tag{16}$$

$$\mu i = \frac{1}{|Ci|}\sum_{x \in Ci} x \tag{17}$$

$$SSE = \sum_{i=1}^{k}\sum_{x \in Ci}\left|d(x, C_i)\right|^2 \tag{18}$$

Where: x is the data object, C_i denotes the ith clustering center, k denotes the number of clustering centers, m is the dimension of the data object, x_j is the attribute value of the jth dimension of the data object x, and C_{ij} is the attribute value of the jth dimension of the clustering center C_i.

As can be seen from the above calculation process, before starting the clustering process, the K-Means algorithm will randomly select a certain number of data points as the initial center of mass. The way of selecting the random initial center of mass will directly affect the final results of the clustering, and if the initial selection of the clustering center is not ideal, the approach may fall into an unreasonable local optimal solution, so it is proposed that based on the previously mentioned data distribution judgment module to limit the initial Clustering center selection range, NTC data, heterozygous genotype data, untyped product data and pure genotype data were clustered separately, and the K used for input was 1, 1, 1, and 2 in order. In this study, we inherited the form of Euclidean distances to represent the inter-individual similarity indexes of the amplification product points proposed by the classical K-Means clustering method, and at the end of the clustering process, the order of the clustering results was fixed and different colors of fluorescent light were displayed. NTC data were shown in black, the amplification products with FAM sequence tags were shown in red, those with HEX sequence tags were shown in blue, the untyped amplification products were in pink, the heterozygous genotypes were in green, and the sequence numbers of the clusters for non-NTC data were fixed to 1, 2, 3, and 4 in order.

The typing map contains a large amount of quantifiable positional and morphological feature information, and the center coordinates, center distances, radius eigenvalues, and axes maximum values of K-Means clustering of the experimental data of all groups to form aggregated clusters are obtained to provide data support for the evaluation of the typing effect in the next step.

2.2.5 Typing effect evaluation module

The purpose of the validation of primer typing results is to evaluate KASP markers and screen and identify KASP markers with excellent typing effect. After classifying and partitioning the experimental data, the model in this paper will intelligently evaluate the partitioning results. The evaluation of the KASP primer typing effect is characterized by many evaluation indexes, strong intrinsic correlation, great difficulty in evaluation, etc. In order to overcome the subjectivity of the expert evaluation method which scores by visually inspecting the unmeasurable abstract indexes, we propose the hierarchical evaluation method based on the decision tree to improve the performance of evaluation and perform data statistics and analysis of the characteristic values such as the radius of the clusters, the center position and the center relative distance to ensure the objectivity and science of the hierarchical evaluation.

2.2.5.1 Criteria for evaluating the typing effect of KASP primers

According to a large amount of experimental data, if NTC did not show a significant fluorescence signal, it indicated that the PCR primer amplification was normal and the test results were credible, and the grading of the typing effect was classified as grade 0, grade 1, grade 2, and grade 3 according to the status of the combination of the red competitive primer and the status of the combination of the blue competitive primer in descending order. The grading is independent of the cluster number, and the criteria for the classification of each grade are shown in Table 2. If the NTC showed a significant fluorescence signal, it was evaluated as grade 2 according to its amplification product combination status if it was not classified as grade 3.

TABLE 2 Classification criteria of KASP primer typing effect level.

Hierarchy	Evaluations	Red competitive primer combination status			Blue competitive primer combination status		
		Amplification efficiency	Amplification specificity	Portfolio competitiveness	Amplification efficiency	Amplification specificity	Portfolio competitiveness
0	Excellent	Higher	Higher		Higher	Higher	
1	Good	Higher	Worse		Higher	Higher	
		Higher	Higher	**Stronger**	Lower	Worse	
		Lower	Worse		Higher	Higher	**Stronger**
		Higher	Higher	**Stronger**	Higher	Higher	**Stronger**
2	Fair	Higher	Weaker		Higher	Higher	
		Higher	Higher		Higher	Worse	
		Higher	Higher		Higher	Weaker	
		Higher	Higher	**Stronger**	Higher	Higher	
		Higher	Higher		Higher	Higher	**Stronger**
		Higher	Higher	**Weaker**	Higher	Higher	**Weaker**
		Lower	Higher		Lower	Higher	
		Lower	Lower		Lower	Lower	
3	Poor	Lower	Lower		Lower	Lower	

"Blank" in the table means none.

2.2.5.2 KASP primer typing effect evaluation module

The decision tree classification algorithm has the advantages of fast speed, low computational cost, clear classification rules, and high accuracy (Adibi, 2019; Zhang et al., 2023). Each internal node in the decision tree represents a judgment on an attribute, each branch represents the output of a judgment result, and finally, each leaf node represents a classification result (Che et al., 2011; Charbuty and Abdulazeez, 2021; Huang et al., 2022). In this paper, we use cluster 1 to denote the pure genotype with FAM sequence tag, cluster 2 to denote the pure genotype with HEX sequence tag, cluster 3 to denote the untyped amplification product, and cluster 4 to denote the heterozygous genotype and construct a complete decision tree with a maximum depth of 7 for the morphology location eigenvalues such as radius, center position and relative distance from the center obtained by clustering, and the structure is shown in Figure 2. Layer 1 uses "module start" as the root node of the decision tree, layer 2 mainly judges whether there is cluster 4 data in the typing diagram, and filters the input data set of the subsequent recursive conditions, layer 3 is the evaluation index layer, which is the factor condition to discriminate the state of primer combinations expressed by the eigenvalues, and also the core of the decision tree. The main task of layer 4 is to quantify the formula of the indicators, and the formula will help to transform the evaluation indicators into numerical calculation forms for better comparison and analysis. Layer 5 is mainly used to classify the cases in which the eigenvalues satisfy each indicator item, and two layers of logical decision-making for judging the degree of dispersion among the groups of amplification products have been included in the strategic paths existing in the cluster 4 data, which can further ensure the reasonableness of the classification of the evaluation results. Finally, the classification results of each index item were combined to determine the evaluation grade of typing effect.

3 Results

3.1 Experimental conditions

The operating system environment for this experiment is Windows 11 with a 12th Gen Intel (R) Core (TM) i5-12500 3.00 GHz CPU, 32.0 GB of RAM on board, NVDIA GeForce RTX 3080 GPU, and 10 GB of RAM. The environment is configured for Python 3.8.3.

3.2 Analysis of model typing results

Input the KASP marker test data of cotton variety resource materials in the KASP-IEva model to verify the typing ability of the model. As shown in Figure 3A, the model first read the relative value of signal magnitude and sample number of each group of amplification products, and then extracted the NTC data, heterozygous genotype data, untyped product data, and pure genotype data, which were temporarily stored in a two-dimensional array, and then classified and combined the amplification product data and output the characteristic values. The results showed that the KASP-labeled pure genotype amplification products, heterozygous genotype amplification products, and untyped detection data could be successfully classified and analyzed using this model, and the final typing effect is shown in Figure 3B.

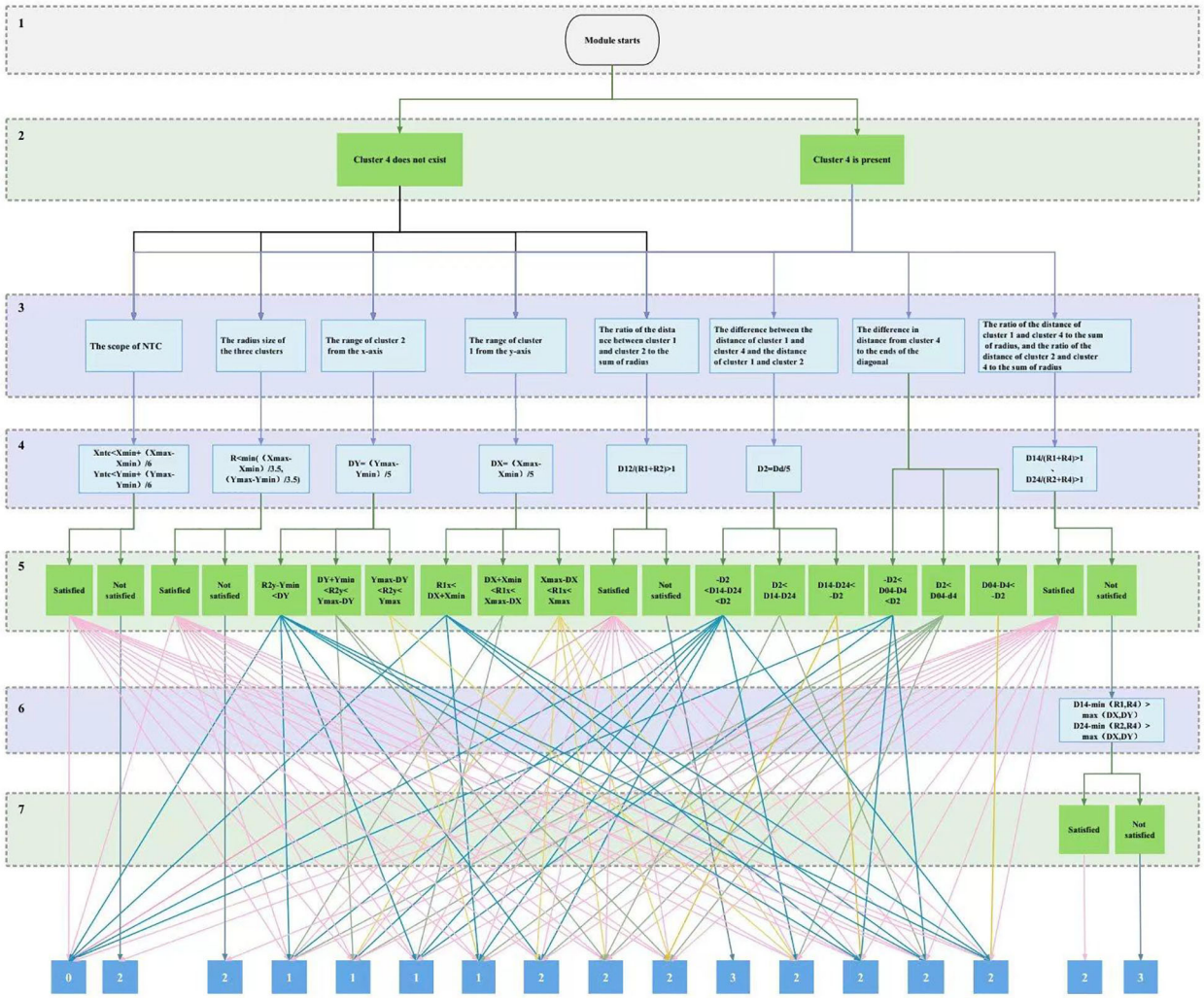


FIGURE 2
Decision tree rank evaluation module structure. Where: X_{ntc} and Y_{ntc} are x and y coordinate values of NTC clustering centers, X_{max} and X_{min} are maximum and minimum values of the X-axis, Y_{max} and Y_{min} are the same as above, R denotes the radius of the three genotypes clusters, R_{2y} denotes the y coordinate value of the center of cluster 2, *D12*, *D14*, *D24*, and *D04* denote the distances of cluster 1 and 2 centers, cluster 1 and 4 centers, cluster 2 and 4 centers and the cluster 4 centers to the origin, respectively, *R1*, *R2*, and *R4* denote the radii of clusters 1, 2, and 4, respectively, and D_d denotes diagonal distance.

The horizontal and vertical coordinates indicate the relative values of HEX and FAM fluorescence signal magnitudes, respectively. From the NTC typing results, the NTC identification module can accurately identify the detection data of negative control samples, and the model is indicated by the use of black dots. From the non-NTC data typing results, the data distribution judgment module can correctly classify the data categories, the pure genotype group is divided into cluster 1 and cluster 2, the HEX signal of the amplification product of the pure allele genotypes close to the horizontal axis is significantly high, and the FAM signal is not obvious, which is shown in red, and the FAM signal of the amplification product of the pure allele genotypes close to the vertical axis is significantly high, and the HEX signal is not obvious, which is shown in blue; The dots in the pink circle with the serial number labeled 3 indicate unsuccessfully typed DNA samples; the green population located in the middle position of the typing diagram can detect both FAM and HEX signals, indicating a heterozygous amplification product containing both alleles, with the serial number labeled 4. In summary, in terms of classification, the KASP-IEva model can aggregate and categorize the amplification product data into the correct distributions that accurately reflect the relationship between homozygous and heterozygous alleles. Morphologically, the KASP-IEva model uses stars and circles to represent the center position and range of the amplification signal clusters of each genotype, which makes the effect of distinguishing and clustering genotypes more obvious, and better expresses the amplification efficiency and specificity of competitive primers for crops. The characteristic values of the aggregation cluster information of each genotype produced by the model are shown in Table 3, which can provide data support for evaluating the typing effect.

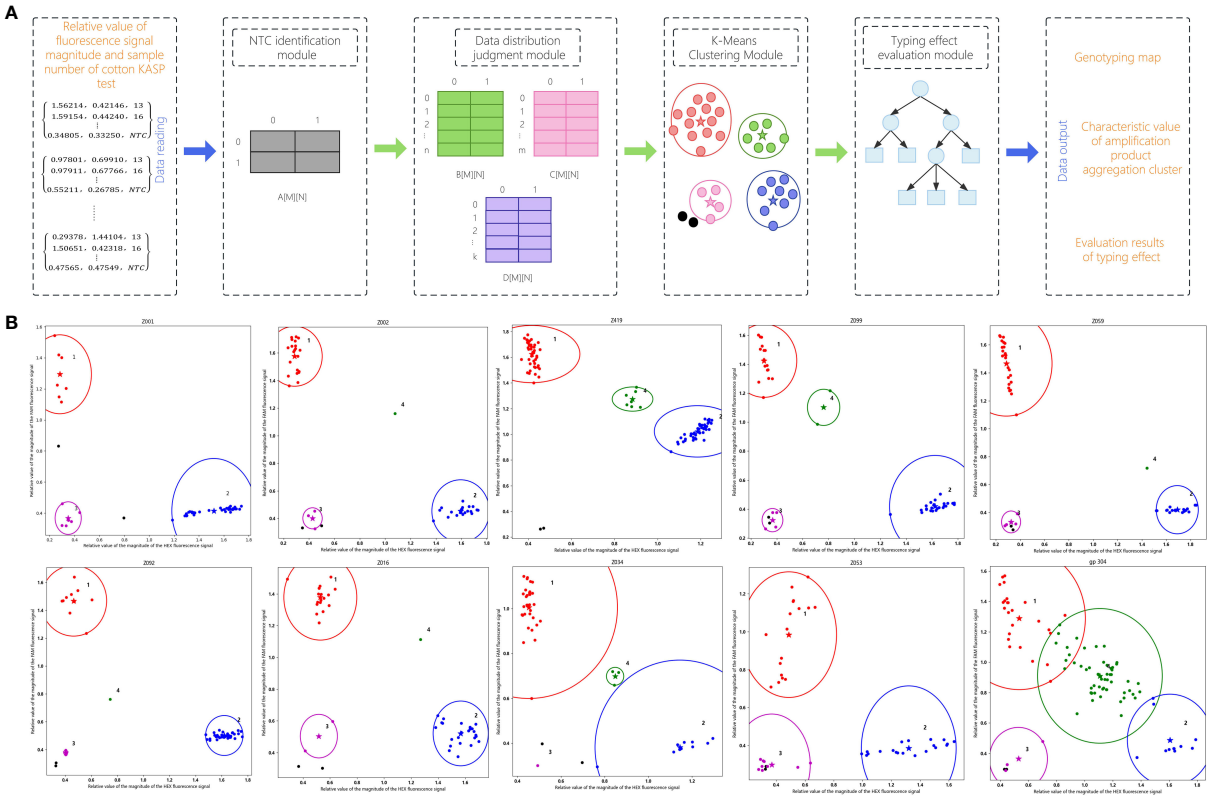


FIGURE 3
KASP-IEva model typing display. **(A)** Model composition schematic. **(B)** Different typing effects.

3.3 Analysis of model evaluation results

The KASP-IEva model evaluates the typing effect by calculating the eigenvalues of each genotypic aggregation cluster information in 2519 cotton KASP marker typing result maps, and some of the evaluation results are shown in Table 4. As shown in the first set of samples in the table, the original data are X=0.43932, 0.43774, 0.84223……0.41968, 0.50936, 0.48568, Y=1.39457, 1.44224, 1.04104……1.63594, 0.28467, 0.29018, and the modeling process yields eigenvalues 0.4503,1.4939, 1.7889……0.3965, 1.6608, 0.2847, using the expert evaluation grade scores as the standard, the model evaluation results are compared and analyzed with the expert evaluation results. It can be seen that the evaluation scores of the model are the same as the expert evaluation scores, indicating that the model has good accuracy and reliability in evaluating the effect of typing.

In order to effectively assess the computational accuracy of the KASP-IEva primer intelligent typing evaluation model for 2519 sets of experimental materials, the accuracy measure of the evaluation results was performed by constructing a confusion matrix. As shown in Figure 4, the rows of the confusion matrix, i.e., the real labels, represent the expert evaluation levels, the columns, i.e., the prediction labels, represent the model evaluation levels, the diagonal elements represent the correct rate of judgment of each level of the KASP-IEva primer intelligent typing evaluation model, and the off-diagonal elements are the proportion of judgment errors, which are calculated by dividing the number of evaluation errors of the level by the total number of samples. It can be seen that a higher diagonal value indicates a higher evaluation accuracy and a better performance of the model for grade evaluation. After analysis, it can be seen that the average accuracy of the four evaluation levels is 87%, and the evaluation accuracy of levels 0, 2, and 3 is 91%, 80%, and 95%, respectively, which indicates that the model judgment ability is better for these three types of typing diagrams, and the evaluation accuracy of level 1 is 72%. The reason for this is, on the one hand, the number of experimental data of this kind of sample is small, on the other hand, considering the influence of human subjective factors, the description of NTC not showing significant fluorescent signals in the class classification standard of primer typing effect is ambiguous, and the reasonable distribution area of NTC is not strictly defined in the process of expert evaluation, which results in the interpretation of the 2-level typing effect as 1-level. In comparison, the evaluation results of the intelligent model based on the principle of the decision tree machine algorithm are more objective and reasonable.

As can be seen from the counting statistics of the intelligent evaluation results of the experimental data in Figure 5, among the 2519 test samples, the proportion of level 2 and 3 fractal effects is significantly higher, reaching 48.55% and 41.23%, respectively, and the proportion of level 0 and 1 fractal maps are both smaller, 6.74% and 3.47%, respectively. In the model evaluation results, 42.39% of the samples were grade 2, 47.98% of the samples were grade 3, and the proportion of grade 1 was the least, only 3.08%. The overall evaluation results of the model showed an uneven distribution of

10.3389/fpls.2023.1293599

TABLE 3 Eigenvalue information corresponding to different typing effects.

Eigenvalue Indicators	Z001	Z002	Z419	Z099	Z059	Z092	Z016	Z034	Z053	gp 304
Transverse coordinates of the center of cluster 1	0.2813	0.2868	0.4169	0.3019	0.2856	0.4647	0.5286	0.4519	0.4882	0.5341
Vertical coordinates of the center of cluster 1	1.2938	1.5752	1.6247	1.4240	1.4658	1.4677	1.3784	1.0054	0.9845	1.2893
Transverse coordinates of the center of cluster 2	1.5193	1.5940	1.1849	1.6237	1.6908	1.6202	1.5670	1.1442	1.3212	1.6053
Vertical coordinates of the center of cluster 2	0.4139	0.4533	1.0230	0.4177	0.4191	0.5012	0.5204	0.3814	0.3847	0.4865
Transverse coordinates of the center of cluster 3	0.3489	0.4281	0.0000	0.3672	0.3249	0.4038	0.5177	0.4926	0.3691	0.5279
Vertical coordinates of the center of cluster 3	0.3672	0.4005	0.0000	0.3229	0.3316	0.3796	0.5032	0.3010	0.2976	0.3641
Transverse coordinates of the center of cluster 4	0.0000	1.0763	0.8839	0.7624	1.4413	0.7445	1.2686	0.8475	0.0000	1.1076
Vertical coordinates of the center of cluster 4	0.0000	1.1612	1.2744	1.1018	0.7176	0.7615	1.1129	0.6980	0.0000	0.9099
Transverse coordinates of the center of the NTC	0.5330	0.4236	0.4637	0.3409	0.3318	0.3281	0.4554	0.6042	0.3265	0.4353
Vertical coordinates of the NTC center	0.6002	0.3402	0.2678	0.3236	0.2902	0.2957	0.3079	0.3555	0.2824	0.2964
The radius of cluster 1	0.2543	0.2174	0.2229	0.2532	0.3761	0.2533	0.2698	0.4064	0.3315	0.4699
The radius of cluster 2	0.3373	0.2264	0.2000	0.3520	0.1727	0.1434	0.2041	0.3887	0.3293	0.3028
The radius of cluster 3	0.1051	0.0764	0.0000	0.0804	0.0784	0.0160	0.1372	0.0000	0.2679	0.2065
The radius of cluster 4	0.0000	0.0000	0.0943	0.1250	0.0000	0.0000	0.0000	0.0397	0.0000	0.4434
Distance between centers of cluster 1 and cluster 2	1.5188	1.7226	0.9756	1.6612	1.7522	1.5065	1.3469	0.9320	1.0265	1.3387
Distance between centers of cluster 1 and cluster 3	0.9291	1.1832	0.0000	1.1031	1.1348	1.0898	0.8752	0.7056	0.6972	0.9252
Distance between centers of cluster 1 and cluster 4	0.0000	0.8914	0.5838	0.5621	1.3768	0.7597	0.7862	0.5010	0.0000	0.6877
Distance between centers of cluster 2 and cluster 3	1.1713	1.1671	0.0000	1.2600	1.3686	1.2225	1.0494	0.6565	0.9560	1.0844
Distance between centers of cluster 2 and cluster 4	0.0000	0.8770	0.3922	1.0998	0.3891	0.9136	0.6634	0.4339	0.0000	0.6534
Distance between centers of cluster 3 and cluster 4	0.0000	0.9995	0.0000	0.8734	1.1813	0.5118	0.9673	0.5325	0.0000	0.7962
Maximum relative value of HEX	1.7408	1.7314	1.2512	1.7714	1.8474	1.7579	1.7009	1.2954	1.6363	1.7885
Minimum relative value of HEX	0.2399	0.2346	0.3686	0.2585	0.2308	0.3275	0.2848	0.4205	0.2779	0.3951
Maximum relative value of FAM	1.5447	1.7182	1.7766	1.6029	1.6682	1.6399	1.5077	1.1442	1.2888	1.5686
Minimum relative value of FAM	0.3193	0.3262	0.2647	0.2645	0.2770	0.2853	0.3027	0.2958	0.2696	0.2885

grades, with significant differential characteristics, which basically reflected the low success rate of KASP marker development in the batch of experimental data, and was in line with the actual grade distribution. The fact that the number of level 3 results obtained from the model evaluation exceeded that of the expert evaluation was due to the fact that the eigenvalue indexes selected in the decision tree algorithm had certain limitations on the representativeness of the clustering pattern (Sagi and Rokach, 2020), which potentially introduced bias in the branching structure for calculating the degree of dispersion among the groups of amplification products.

The speed of the expert and model evaluation methods was tested, and the comparison test of different groups was selected. The results are shown in Table 5. From the table, it can be seen that only for the evaluation of 1 group of KASP amplification product typing effect, the expert evaluation computation time is 3s, the KASP-IEva model computation time is 0.28s, the model is shorter than the time of the expert visual inspection method, the evaluation of the 2519 KASP typing diagrams, the expert evaluation consumes 15hour, the model evaluation is only 8min27.45s, the model intelligent evaluation shows more advantages, on the one hand, because the model front data distribution judgment module shortens the K-Means algorithm training time, the decision tree algorithm computational complexity is not high, and the output is fast, on the other hand, there are many influencing factors in the process of the expert evaluation, and the inefficiency of evaluation in a long period of time may lead to omitted evaluation, wrong evaluation, checking and repeated evaluation consume a lot of time. Therefore, from the perspective of time, the model evaluation in this paper is significantly better than the expert evaluation.

4 Discussion

KASP markers are considered to be an efficient and flexible high-throughput typing technology in the field of molecular marker-assisted breeding. In the traditional validation of KASP primer typing results, due to the dependence of the manual visual inspection method on professional knowledge and long-term experience, high precision and rapid evaluation of large quantities

10.3389/fpls.2023.1293599

TABLE 4 Results of the evaluation of the effect of partial subtyping.

Sample Serial number	Raw data	Eigenvalue	Model evaluation results	Expert evaluation results
1	X=0.43932, 0.43774, 0.84223……0.41968, 0.50936, 0.48568, Y=1.39457, 1.44224, 1.04104……1.63594, 0.28467, 0.29018	0.4503, 1.4939, 1.7889…… 0.3965, 1.6608, 0.2847	2	2
2	X=0.60457, 0.57431, 0.65949……0.63905, 0.71255, 0.70928, Y=0.35818, 0.33858, 0.43917……0.43223, 0.48933, 0.4905	0.5906, 0.4538, 0.6791…… 0.5576, 0.4981, 0.3350	3	3
3	X=0.50657, 0.51646, 0.98894……0.49561, 0.52378, 0.50983, Y=1.01964, 1.08208, 0.7055……1.28604, 0.33016, 0.32921	0.5167, 1.1084, 1.7231…… 0.4666, 1.2860, 0.3292	2	2
4	X=0.45103, 0.42839, 0.91412……0.39512, 0.45118, 0.46644, Y=1.36386, 1.47295, 1.28254……1.56937, 0.76839, 0.71924	0.4235, 1.5148, 1.1915…… 0.3928, 1.6517, 0.7192	1	1
5	X=0.44263, 0.44069, 1.28643……0.44412, 0.53078, 0.57653, Y=1.33613, 1.3216, 0.95744……1.3247, 0.4802, 0.43717	0.4420, 1.3473, 1.6820…… 0.4255, 1.4937, 0.3933	0	0
6	X=0.45865, 0.44352, 0.70477……0.44976, 0.52592, 0.53284, Y=1.07028, 1.06589, 0.95067……1.10195, 0.30809, 0.44488	0.4568, 1.1315, 0.9247…… 0.4430, 1.2614, 0.3081	2	1
7	X=0.457, 0.41995, 0.89926……0.48447, 0.46364, 0.48052, Y=1.55319, 1.54596, 1.49991……1.59789, 0.28759, 0.2887	0.5973, 1.4677, 1.5237…… 0.3979, 1.6069, 0.2876	2	2
8	X=0.45655, 0.43385, 1.27705……0.49476, 0.51694, 0.46114, Y=1.61637, 1.6914, 1.23828……1.54833, 0.28303, 0.26206	0.4545, 1.4756, 1.4532…… 0.3853, 1.6914, 0.2621	1	1
9	X=0.45065, 0.44332, 1.55996……0.37383, 0.44326, 0.44204, Y=1.43939, 1.39367, 0.82486……1.72951, 0.31645, 0.28992	0.4142, 1.5499, 1.8423…… 0.3544, 1.7295, 0.2900	3	2
10	X=0.49894, 0.4636, 1.59811……0.5059, 0.4372, 0.45986, Y=1.46816, 1.49855, 1.03515……1.72295, 0.29714, 0.31563	0.4831, 1.5996, 1.6093…… 0.3829, 1.7360, 0.2931	3	3
11	X=0.69556, 0.62995, 1.47723……0.68204, 1.30596, 0.4447, Y=1.40773, 1.45091, 1.04785……1.60087, 0.33923, 0.24859	0.6793, 1.4922, 1.7165…… 0.4447, 1.6149, 0.2486	2	2
12	X=0.54169, 0.57037, 1.05047……0.4853, 0.45474, 0.47241, Y=1.45685, 1.37958, 1.19831……1.63692, 0.27228, 0.28638	0.5302, 1.5411, 1.4786…… 0.4248, 1.6741, 0.2723	3	3
13	X=0.50719, 0.48245, 1.52441……0.50909, 0.44464, 0.45769, Y=1.46333, 1.50677, 0.89832……1.46611, 0.27971, 0.2722	0.4904, 1.4980, 1.6842…… 0.4219, 1.6333, 0.2722	2	2
14	X=0.42744, 0.38678, 1.31314……0.42961, 0.44257, 0.43698, Y=1.4936, 1.50528, 1.12856……1.50871, 0.26177, 0.25463	0.4083, 1.5368, 1.6830…… 0.3686, 1.6570, 0.2546	0	0
15	X=0.42368, 0.40204, 0.90772……0.41492, 0.44749, 0.43715, Y=1.4702, 1.49327, 1.20338……1.31487, 0.25624, 0.25149	0.4036, 1.4711, 1.1958…… 0.3785, 1.5802, 0.2515	1	1
16	X=0.47723, 0.44637, 1.4121……0.51865, 0.45703, 0.43806, Y=1.4776, 1.52423, 1.15026……1.45825, 0.25368, 0.24879	0.4798, 1.5147, 1.4785…… 0.4030, 1.6015, 0.2488	3	3
⋮ ⋮	⋮ ⋮	⋮ ⋮	⋮ ⋮	⋮ ⋮
2517	X=0.43663, 0.43324, 1.39222……0.40587, 0.42239, 0.44341, Y=1.52676, 1.52671, 0.83752……1.72138, 0.27678, 0.2601	0.4022, 1.5936, 1.7156…… 0.3649, 1.7373, 0.2601	2	2
2518	X=0.46683, 0.4499, 1.21902……0.44592, 0.4437, 0.4663, Y=1.34006, 1.30644, 0.74495……1.60008, 0.26654, 0.26475	0.4511, 1.4711, 1.6680…… 0.4021, 1.6029, 0.2648	2	2
2519	X=0.39699, 0.40567, 1.24456……0.35578, 0.44603, 0.44563, Y=1.3522, 1.29601, 0.49863……1.54919, 0.29102, 0.2879	0.3812, 1.4571, 1.3394…… 0.3502, 1.6360, 0.2879	2	2

of primer typing results are not possible, however, intelligent evaluation of KASP marker primer typing results has not yet been widely discussed in the industry. Therefore, in this study, the NTC identification module and the data distribution judgment module were combined with the K-Means algorithm and decision tree theory to propose an intelligent typing evaluation model for KASP-IEva primers, and the model was tested by using the results of the kompetitive allele-specific PCR test of the 2519 groups of cotton varietal resource materials, and the following conclusions were obtained:

(1) The KASP-IEva Primer Intelligent Typing Evaluation Model limits the initial clustering center selection range based on the data distribution judgment module, which effectively improves the reasonableness of genotype group typing. The model is able to visualize and display the data

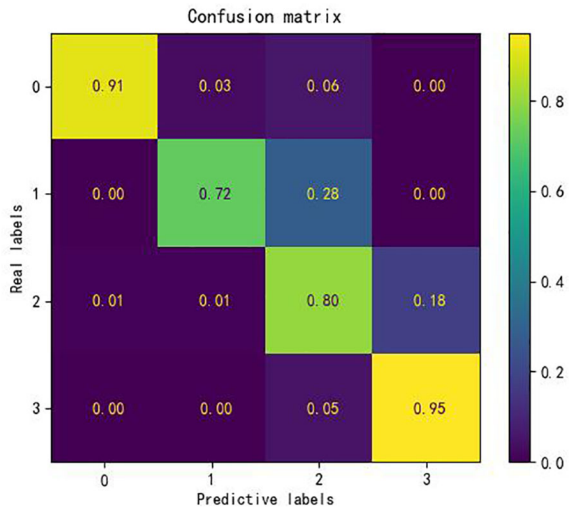


FIGURE 4
Detection effect of KASP-IEva primer intelligent typing evaluation model.

aggregation and classification effects of the amplification products of pure genotypes, heterozygous genotypes, and untyped genotypes, and at the same time, it can extract the positional morphology eigenvalues, which provides powerful data support for the evaluation decision.

(2) The typing effect rating evaluation criteria was constructed, and an intelligent typing effect evaluation model based on a logical decision tree was created on the basis of which the model predicts the scores for the calculation of eigenvalues of genotype classification and aggregation, realizing a fast and accurate typing evaluation of KASP markers, which is used to screen KASP markers with excellent typing effect.

(3) The typing effect evaluation test of 2519 groups of KASP markers was carried out, and the model evaluation results were compared and analyzed with the expert evaluation results. From the results of the comparative analysis, of 2519 groups of KASP markers, the model typing effect evaluation is correct, the average accuracy rate is 87%, and the evaluation results show the uneven distribution of the grades, with significant differential characteristics, basically can reflect the low success rate of the development of the KASP markers in the experimental data of the batch of the low success rate of the KASP markers, in line with the actual distribution of the grades; for the evaluation of the 2519 KASP typing diagrams, the expert evaluation consumes 15 hours, and the model evaluation only uses 8min27.45s, from the perspective of time, the model intelligent evaluation is obviously better than the expert evaluation. Therefore, the KASP-IEva model has good evaluation performance.

(4) The research method in this paper provides an intelligent evaluation idea for the KASP primer typing effect. The model is not only applicable to the results of the cotton KASP test but also can be used to evaluate the KASP typing effect of resource data of varieties including wheat, soybean, and corn.

5 Conclusion

In the study of this paper, we designed an intelligent KASP marker primer typing evaluation model to realize the rapid typing evaluation of KASP markers for 2519 sets of cotton varietal resource materials. First, the separation of NTC data from the rest of the data was controlled by the NTC identification module to enhance the classification ability of the target under data mixing conditions. In addition, the data distribution judgment module further classifies the data types to improve the distinguishability of amplified signals. The K-Means machine learning algorithm is introduced to analyze the potential association between amplification products and extract the deep structure feature values. Finally, the typing effect is evaluated based on the decision tree, which overcomes the subjectivity of the existing expert evaluation method that scores by visual inspection of unmeasurable abstract indicators. The

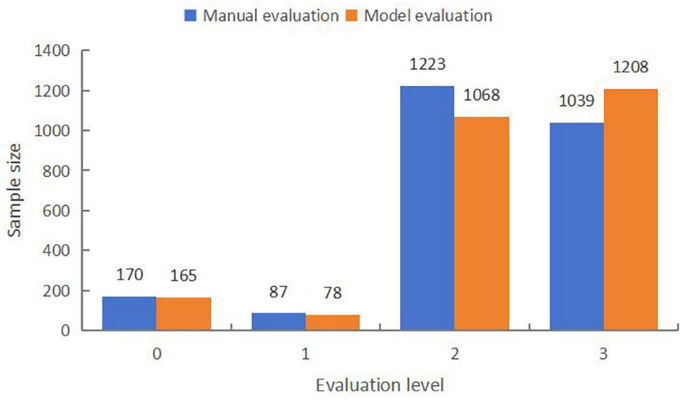


FIGURE 5
Comparative results of typing effect evaluation.

TABLE 5 Comparison of computation time between expert and model evaluation methods.

Data size	Expert evaluation computation time	Model evaluation computation time
Group 1	3s	0.28s
Group 2519	15hour	8min27.45s

average accuracy of the four levels of model evaluation was 87%, and the model evaluation took only 8min27.45s compared to the expert evaluation which took 15hour, all the tests and results show that the model has good performance and enough speed to be used for screening KASP markers with excellent typing results at scale. In the subsequent application research, the introduction of deep learning optimization evaluation method will be considered to further improve the evaluation accuracy, with a view to realizing the application of the KASP-IEva intelligent evaluation model in the field of KASP marker testing.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

XC: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. LH: Writing – review & editing. JF: Writing – review & editing. SY: Writing – review & editing. GZ: Writing – review & editing. JZ: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key R&D Program of China (2022YFF0711805); the National Science Foundation of China (31971792, 32160421); Innovation Project of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2023-AII, ZDXM23011); Special Project for Basic Research Operating Costs of Central Public Welfare Scientific Research Institutes (JBYW-AII-2023- 06, Y2022XK24, Y2022QC17, JBYW-AII-2022-14); Special Project on Southern Prosperity for the National Institute of Southern Prosperity, Chinese Academy of Agricultural Sciences, Sanya (YBXM2312, YDLH01, YDLH07, YBXM10); the Project of Sanya Yazhou Bay Science and Technology City(SCKJ-JYRC-2023-45, SCKJ-JYRC-2022-62).

Acknowledgments

We thank all the participants in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdelraheem, A., Kuraparthy, V., Hinze, L., Stelly, D., Wedegaertner, T., and Zhang, J. (2021). Genome-wide association study for tolerance to drought and salt tolerance and resistance to thrips at the seedling growth stage in US Upland cotton. *Ind. Crops Products* 169, 113645. doi: 10.1016/J.INDCROP.2021.113645

Adibi, M. A. (2019). Single and multiple outputs decision tree classification using bi-level discrete-continues genetic algorithm. *Pattern Recognition Lett.* 128, 190–196. doi: 10.1016/j.patrec.2019.09.001

Al-Khayri, J. M., Mahdy, E. M., Taha, H. S., Eldomiaty, A. S., Abd-Elfattah, M. A., Abdel Latef, A. A. H., et al. (2022). Genetic and morphological diversity assessment of five kalanchoe genotypes by SCoT, ISSR and RAPD-PCR markers. *Plants* 11 (13), 1722. doi: 10.3390/plants11131722

Amiteye, S. (2021). Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon* 7 (10). doi: 10.1016/j.heliyon.2021.e08093

Ayalew, H., Tsang, P. W., Chu, C., Wang, J., Liu, S., Chen, C., et al. (2019). Comparison of TaqMan, KASP and rhAmp SNP genotyping platforms in hexaploid wheat. *PloS One* 14 (5), e0217222. doi: 10.1371/journal.pone.0217222

Broccanello, C., Chiodi, C., Funk, A., McGrath, J. M., Panella, L., and Stevanato, P. (2018). Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods* 14, 1–8. doi: 10.1186/s13007-018-0295-6

Byers, R. L., Harker, D. B., Yourstone, S. M., Maughan, P. J., and Udall, J. A. (2012). Development and mapping of SNP assays in allotetraploid cotton. *Theor. Appl. Genet.* 124, 1201–1214. doi: 10.1007/s00122-011-1780-8

Chakraborty, S., Paul, D., Das, S., and and Xu, J. (2020). Entropy weighted power kmeans clustering, Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, 26–28.

Charbuty, B., and Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* 2 (01), 20–28. doi: 10.38094/jastt20165

Che, D., Liu, Q., Rasheed, K., and Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Adv. Exp. Med. Biol.* 696, 191–199. doi: 10.1007/978-1-4419-7046-6_19

Chen, T., Li, Y., Rong, E., and Wu, Y. (2024). Identification of traits and floral organ transcriptomic analysis of artificial allotetraploid progeny of the genus Cotton. *Acta Agronomica Sin.* 50 (02), 325–339. doi: 10.3724/SP.J.1006.2023.34061

Chou, Z. H. (2016). *Machine learning* (Beijing: Tsinghua University Press).

Fan, T., Li, Z., Jiang, Q., Chen, S., Ou, X., Chen, Y., et al. (2021). Development and effect evaluation of KASP markers closely linked to major QTLs of spike number per unit area and grain length in wheat. *Scientia Agricultura Sin.* 54 (14), 2941–2951. doi: 10.3864/j.issn.0578-1752.2021.14.002

Gao, W., Chen, Q., Fu, J., Jiang, H., Sun, F., Geng, S., et al. (2023). Using association mapping and local interval haplotype association analysis to improve the cotton drought stress response. *Plant Sci.* 335, 111813. doi: 10.1016/j.plantsci.2023.111813

Geng, X., Qu, Y., Jia, Y., He, S., Pan, Z., Wang, L., et al. (2021). Assessment of heterosis based on parental genetic distance estimated with SSR and SNP markers in

upland cotton (*Gossypium hirsutum* L.). *BMC Genomics* 22 (1), 1–11. doi: 10.1186/s12864-021-07431-6

Gu, Q., Ke, H., Liu, Z., Lv, X., Sun, Z., Zhang, M., et al. (2020). A high-density genetic map and multiple environmental tests reveal novel quantitative trait loci and candidate genes for fibre quality and yield in cotton. *Theor. Appl. Genet.* 133, 3395–3408. doi: 10.1007/s00122-020-03676-z

Guo, Y., Chen, Q., Qu, Y., Deng, X., Zheng, K., Wang, N., et al. (2023). Development and identification of molecular markers of *GhHSP70-26* related to heat tolerance in cotton. *Gene* 874, 147486. doi: 10.1016/j.gene.2023.147486

Huang, H., Wang, W., Li, C., Hou, K., Zheng, Y., and Chen, Q. (2022). The maximum Lyapunov exponent and Elman-Decision tree based fault warning diagnosis method. *China Rural Water Hydropower* 2), 168–173.

Kumar, S. J., Susmita, C., Sripathy, K. V., Agarwal, D. K., Pal, G., Singh, A. N., et al. (2022). Molecular characterization and genetic diversity studies of Indian soybean (*Glycine max* (L.) Merr.) cultivars using SSR markers. *Mol. Biol. Rep.* 49, 2129–2140. doi: 10.1007/s11033-021-07030-4

Li, L., Sun, Z., Ke, H., Gu, Q., Wu, L., Zhang, Y., et al. (2022). Development and effect evaluation of KASP markers for fiber strength in Gossypium hirsutum L. *J. Agric. Sci. Technol.*, 1–10. doi: 10.13304/j.nykjdb.2022.0818

Li, F. G., and Yuan, Y. L. (2013). *Cotton molecular breeding* (Beijing: China: Agricultural University Press).

Lu, Z., Huang, S., Zhang, X., Yang, W., Zhu, L., and Huang, C. (2023). Intelligent identification on cotton verticillium wilt based on spectral and image feature fusion. *Plant Methods* 19 (1), 1–16. doi: 10.1186/s13007-023-01056-4

Mirzal, A. (2020). Statistical analysis of microarray data clustering using NMF, spectral clustering, Kmeans, and GMM. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 19 (2), 1173–1192. doi: 10.1109/TCBB.2020.3025486

Nie, X. H., Wang, Z. H., Liu, N. W., Song, L., Yan, B. Q., Xing, Y., et al. (2021). Fingerprinting 146 Chinese chestnut (*Castanea mollissima* Blume) accessions and selecting a core collection using SSR markers. *J. Integr. Agric.* 20 (5), 1277–1286. doi: 10.1016/s2095-3119(20)63400-1

Sagi, O., and Rokach, L. (2020). Explainable decision forest: Transforming a decision forest into an interpretable tree. *Inf. Fusion* 61, 124–138. doi: 10.1016/j.inffus.2020.03.013

Sheng, X., Shen, Y., Yu, H., Wang, J., Zhao, Z., and Gu, H. (2022).). Development and application of KASP marker of *BoCAL* gene related to curd development in cauliflower. *Acta Agriculturae Zhejiangensis* 34 (6), 1183–1192. doi: 10.3969/j.issn.1004-1524.2022.06.09

Sinaga, K. P., and Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access* 8, 80716–80727. doi: 10.1109/ACCESS.2020.2988796

Wang, F. Q., Fan, X. C., Zhang, Y., Lei, S., Liu, C. H., and Jiang, J. F. (2022). Establishment and application of an SNP molecular identification system for grape cultivars. *J. Integr. Agric.* 21 (4), 1044–1057. doi: 10.1016/S2095-3119(21)63654-7

Wang, P., Tian, Z., Kang, C., Li, Y., Wang, H., Yang, C., et al. (2021). Establishment and application of a tomato KASP genotyping system based on five disease resistance genes. *Acta Hortic. Sin.* 48 (11), 2211–2226. doi: 10.16420/j.issn.0513-353x.2020-0913

Xu, B., Zhang, J., Shi, Y., Dai, F., Jiang, T., Xuan, L., et al. (2023). *GoSTR*, a negative modulator of stem trichome formation in cotton. *Plant J.* 116 (2), 389–403. doi: 10.1111/tpj.16379

Yang, S., Yu, W., Wei, X., Wang, Z., Zhao, Y., Zhao, X., et al. (2020). An extended KASP-SNP resource for molecular breeding in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *PloS One* 15 (10), e0240042. doi: 10.1371/journal.pone.0240042

Yin, Z., Chen, J., Shen, Z., Fu, Y., Zheng, Z., and Wang, Y. (2022). Identification algorithm of distribution Network transformer winding material based on Kmeans clustering. *J. North China Electric Power Univ. (Natural Sci. Edition)*, 1–9.

Zhang, Y., Zhang, W., Xu, K., and Li, J. (2023). Phenological phase identification of oilseed rape (Brassica napus L.) Using typical stokes parameters. *Geomatics Inf. Sci. Wuhan Univ.* 48 (8), 1322–1330. doi: 10.13203/j.whugis20210394

Zhao, Y., Chen, W., Cui, Y., Sang, X., Lu, J., Jing, H., et al. (2021). Detection of candidate genes and development of KASP markers for Verticillium wilt resistance by combining genome-wide association study, QTL-seq and transcriptome sequencing in cotton. *Theor. Appl. Genet.* 134, 1063–1081. doi: 10.1007/s00122-020-03752-4

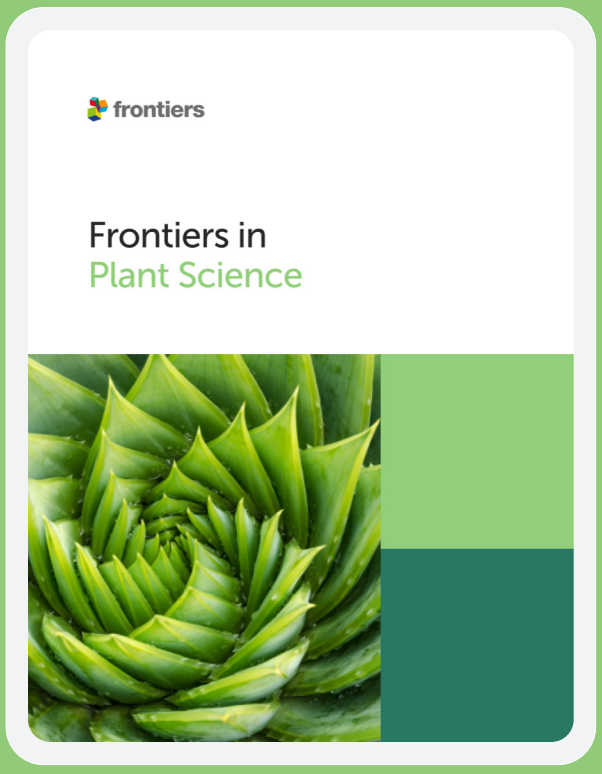
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact