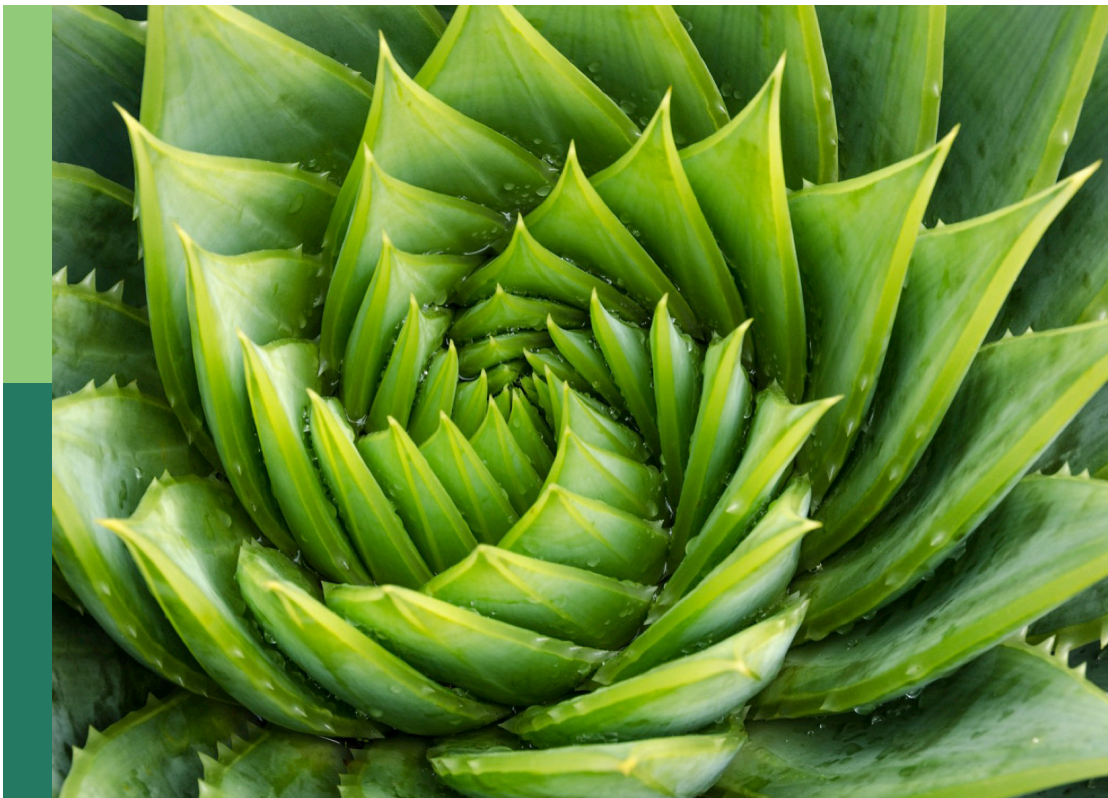
frontiers | Research Topics

Multi-omics, genetic evolution and crop domestication

Edited by
Xuming Li, Hao Tong and Changmian Ji

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-6636-7
DOI 10.3389/978-2-8325-6636-7

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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, genetic evolution and crop domestication

Topic editors

Xuming Li — Department of Scientific Affairs, Hugo Biotechnologies Co., Ltd., China

Hao Tong — Max Planck Institute of Molecular Plant Physiology, Germany

Changmian Ji — Chinese Academy of Tropical Agricultural Sciences, China

Citation

Li, X., Tong, H., Ji, C., eds. (2025). *Multi-omics, genetic evolution and crop domestication*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6636-7

Table of contents

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 April 2024
DOI 10.3389/fpls.2024.1388163

OPEN ACCESS

EDITED BY
Xuming Li,
Department of Scientific Affairs, Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Beimi Cui,
University of Edinburgh, United Kingdom
Meixiang Zhang,
Shaanxi Normal University, China

*CORRESPONDENCE
Ruimin Li
liruimin@gnnu.edu.cn
Guiyan Huang
huangguiyan@gnnu.edu.cn

RECEIVED 19 February 2024
ACCEPTED 27 March 2024
PUBLISHED 10 April 2024

CITATION
Li R, Wang X, Hu Y and Huang G (2024) Analysis of huanglongbing-associated RNA-seq data reveals disturbances in biological processes within *Citrus* spp. triggered by *Candidatus* Liberibacter asiaticus infection. *Front. Plant Sci.* 15:1388163. doi: 10.3389/fpls.2024.1388163

Analysis of huanglongbing-associated RNA-seq data reveals disturbances in biological processes within *Citrus* spp. triggered by *Candidatus* Liberibacter asiaticus infection

Ruimin Li[1,2]*, Xinyou Wang[1], Yanan Hu[1] and Guiyan Huang[1,2]*

[1]College of Life Sciences, Gannan Normal University, Ganzhou, China, [2]China-USA Citrus Huanglongbing Joint Laboratory, National Navel Orange Engineering Research Center, Gannan Normal University, Ganzhou, China

Introduction: Huanglongbing (HLB), a disease that's ubiquitous worldwide, wreaks havoc on the citrus industry. The primary culprit of HLB is the gram-negative bacterium *Candidatus* Liberibacter asiaticus (CLas) that infects the phloem, but its damaging mechanism is yet to be fully understood.

Methods and results: In this study, a multitude of tools including weighted correlation network analysis (WGCNA), protein-protein interaction (PPI) network analysis and gene expression profiling are employed to unravel the intricacies of its pathogenesis. The investigation pinpoints various central genes, such as the ethylene-responsive transcription factor 9 (*ERF9*) and thioredoxin reductase 1 (*TrxR1*), that are associated with CLas invasion and resultant disturbances in numerous biological operations. Additionally, the study uncovers a range of responses through the detection of differential expressed genes (DEGs) across different experiments. The discovery of core DEGs leads to the identification of pivotal genes such as the sieve element occlusion (*SEO*) and the wall-associated receptor kinase-like 15 (*WAKL15*). PPI network analysis highlights potential vital proteins, while GO and KEGG pathway enrichment analysis illustrate a significant impact on multiple defensive and metabolic pathways. Gene set enrichment analysis (GSEA) indicates significant alterations in biological processes such as leaf senescence and response to biotic stimuli.

Discussion: This all-encompassing approach extends valuable understanding into the pathogenesis of CLas, potentially aiding future research and therapeutic strategies for HLB.

KEYWORDS

Citrus sinensis, Candidatus Liberibacter asiaticus, RNA-Seq, WGCNA, biological processes

1 Introduction

The citrus industry is significantly impacted by citrus huanglongbing (HLB), which is a disease that has spread globally and causes severe damage (Wang, 2019). HLB can affect all citrus varieties, leading to symptoms such as yellowing of leaves, stunted growth, and the production of small, misshapen, and bitter fruits (Bové and Barros, 2006; Wang, 2019; Pandey et al., 2022). Ultimately, the disease leads to premature tree death, rendering orchards unproductive, resulting in many orchards being abandoned or replaced with alternative crops (Yuan et al., 2021).

HLB is typically believed to be caused by *Candidatus* Liberibacter asiaticus (CLas), which is a gram-negative bacterium that parasitizes to the phloem (Zheng et al., 2023). The pathogenesis of CLas is unclear during its unculturable *in vitro* (Hu et al., 2021). Nevertheless, there are still numerous research findings attempting to uncover the pathogenesis of CLas (Pitino et al., 2016; Loto et al., 2017; Clark et al., 2018; Pitino et al., 2018; Shi et al., 2019; Liu et al., 2019a; Du et al., 2021; Wang et al., 2023; Shi et al., 2023a, Shi et al., 2023b). Following the release of the CLas genome, researchers found that the bacterium lacks type III or IV secretory systems, but it does have a complete sec-dependent system (Duan et al., 2009). The Sec-dependent effectors (SDEs) of CLas disrupt autophagy, development, and papain-like cysteine proteases, while also suppressing the immune response in citrus (Clark et al., 2018; Wang et al., 2023; Shi et al., 2023a, Shi et al., 2023b). For instance, SDE1 targets citrus proteases and declines defense responses in plants (Clark et al., 2018). Overexpression of SDE1 accelerated senescence related biological processes both in *Arabidopsis thaliana* and *Citrus paradisi* (Clark et al., 2020). Moreover, SDE15 suppresses the immune response to bacterial infection by interacting with citrus CsACD2 (Pang et al., 2020). Interestingly, both SDE3 and SDE4405 interfere with citrus autophagy; SDE3 suppresses it, while SDE4405 stimulates it (Shi et al., 2023a, Shi et al., 2023b). Nevertheless, both SDE3 and SDE4405 exert a negative influence on the citrus defense response (Shi et al., 2023a, Shi et al., 2023b). Although the functions of several SDEs have been analyzed, the pathogenic mechanism of CLas is still unclear. For instance, whether these SDEs work synergistically or independently during the infection process and which effector protein is the key pathogenic factor, remain uncertain. Moreover, the response of citrus to CLas infection and the regulatory mechanisms involved in citrus need to be clarified. Dynamic changes in the signal transduction pathway also require further investigation.

Upon CLas infection, the citrus transcriptome undergoes significant alterations, leading to the up-regulation or down-regulation of numerous genes (Liu et al., 2023). These genes are primarily associated with plant defense responses, photosynthesis, cell wall modification, nutrient transport, and metabolism among others (Hu et al., 2017; Zhao et al., 2019; Liu et al., 2019b; Ribeiro et al., 2022). Several defense-related genes and pathogenesis-related proteins (PRs) are up-regulated in response to CLas infection. For instance, genes encoding PR proteins, such as peroxidase, chitinase, and β-1,3-glucanase, have been observed to increase, indicating an active defense response (Rawat et al., 2015). Additionally, genes related to secondary metabolite biosynthesis, like phytoalexins, are up-regulated, providing further biochemical defense against the pathogen (Fu et al., 2016). CLas infection has been associated with a broad down-regulation of genes related to photosynthesis and carbohydrate metabolism (Martinelli et al., 2012; Liu et al., 2019b, 2023). The down-regulation of these genes likely contributes to the pronounced leaf yellowing observed in CLas-infected trees. Transcriptomic analysis has revealed an up-regulation of genes related to cell-wall modification, such as those encoding pectin methylesterases and pectate lyases (Wang et al., 2016; Liu et al., 2023). These enzymes are involved in the modification and degradation of the cell wall, suggesting an alteration in plant cell wall integrity and structure during CLas infection. Genes involved in nutrient transport, such as those encoding sugar and amino acid transporters, show variable expression patterns depending on the stage of CLas infection (Shahzad et al., 2020). These intricate changes in the citrus transcriptome in response to CLas infection underscore the complexity of the plant-pathogen interaction. Notably, the up-regulation of defense-related genes suggests a robust, yet not fully effective, physiological response to the pathogen, while the down-regulation of photosynthetic genes highlights the detrimental effects of HLB on the primary metabolic processes of citrus. Despite ongoing transcriptomic research on HLB, the conservative response regulation pattern to CLas infection in citrus remains unclear due to the different stages of CLas infection and the variety of citrus species involved. Therefore, in this study, we utilize the previous sequencing data from our research group and publicly available data from National Center for Biotechnology Information (NCBI) SRA database to determine the core regulatory genes, conservative biological response processes, metabolic pathways, and response patterns of citrus.

Weighted gene co-expression network analysis (WGCNA) offers a robust computational approach to decipher complex biological systems at the genetic level (Langfelder and Horvath, 2008). The performance of WGCNA in identifying co-expression modules has become a critical tool in plant transcriptional regulation (Lu et al., 2019; Zhu et al., 2019; Yu et al., 2023). WGCNA is helpful to identify hub genes that control plant growth, development, and responses to environmental stimuli (Yao et al., 2023; Yu et al., 2023). For instance, multiple transcription factors associated with regulation of salt response in rice were identified by WGCNA (Zhu et al., 2019). Moreover, Gene set enrichment analysis (GSEA) is an important computational method in analyzing transcriptomic data, particularly in studies focused on plant-pathogen interactions (Muchero et al., 2018; Rody et al., 2021). Understanding the complex and dynamic interplay between a plant and its pathogen at a molecular level requires tools that can identify and categorize patterns in gene expression (Dong et al., 2015; Jiang et al., 2017). Applications of GSEA in plant-pathogen interaction studies have led to the identification of key gene sets and pathways that are modulated during infection (Bautista et al., 2021). Furthermore, the utilization of Gene ontology (GO) and Kyoto Encyclopedia of genes and genomes (KEGG) enrichment analyses enables identification of the biological processes and metabolic pathways in which differentially expressed genes (DEGs) are involved (Liu et al., 2023).

In order to elucidate the disturbances in biological processes within *Citrus* spp. triggered by CLas infection, we collected HLB

associated RNA-seq data, including the RNA-seq data sequenced in our previous study (Liu et al., 2023) and public RNA-seq data deposited in NCBI SRA database from various research groups, for analysis. Subsequently, we utilized WGCNA to identify gene sets and pinpointed hub genes positive correlated with CLas infection. Moreover, we performed GO and KEGG enrichment analysis, protein-protein interactions (PPI) network analysis and GESA to determine biological processes and metabolic pathways disturbed by CLas infection. Our findings would lay the groundwork for better understanding of CLas pathogenesis, elucidating the complex interaction and molecular mechanisms between citrus and CLas.

2 Materials and methods

2.1 RNA-seq data collection and analysis

To collect HLB associated RNA-seq data, we used '*Citrus*', '*Candidatus* Liberibacter asiaticus', and 'Huanglongbing' as keywords to search related biosamples in SRA database. In addition, we incorporated RNA-seq data from a previous publication by our research group into this study (Liu et al., 2023). We then utilized the 'Kallisto Super GUI Wrapper' within the TBtools software to quantify the expression profiles of genes in *C. sinensis* using the 3.0 version reference genome in the Citrus Pan-genome to Breeding Database (CPBD) (http://citrus.hzau.edu.cn/) (Bray et al., 2016; Liu et al., 2022; Chen et al., 2023). The bias correction parameters were configured with a kmer size of 31.

2.2 Weighted correlation network analysis

The WGCNA was performed with WGCNA-shinyApp (https://github.com/ShawnWx2019/WGCNA-shinyApp) (Langfelder and Horvath, 2008). The raw count values of all genes were normalized using the variance-stabilizing transformation method (Lin et al., 2008), and then the gene sets were filtered twice. First, genes with 90% of samples having a count value less than 10 were removed. Then, genes were further filtered using the 'median absolute deviation' method (Pham-Gia and Hung, 2001). The normalized count values of the remaining genes were utilized to calculate the suggested power value. Then module net was constructed with parameters of 'min Module size = 30' and 'module cuttree height = 0.25'. All CLas infected samples were designated as 'HLB', while CLas free samples were designated as 'MOCK'. The correlation between module and trait data (such as 'HLB') was computed, and significant 'module-trait' pairs were utilized to identify hub genes.

2.3 Hub genes identification

The selection of hub genes was determined by filtering with a cut off value greater than 0.5 for both the 'kME' and 'GS' generated with WGCNA-shinyApp. Subsequently, the edge information for each module was generated and any weight values below 0.8 were eliminated. We then utilized Cytoscape 3.9.0 to arrange the co-expression network (Kohl et al., 2011).

2.4 Differential expressed genes identification

DEGs were generated with R package 'edgeR', with a threshold of absolute log fold change (|log2FC|) greater than 1 and a false discovery rate of less than 0.05 (Robinson et al., 2010). The DEGs from various experiments were combined, and those found in over 2/3 of the experiments were classified as core DEGs (cDEGs), while those found in more than 1/2 of the experiments were classified as soft core DEGs (scDEGs). The Venn diagram was drawn by InteractiVenn (Heberle et al., 2015), and the heatmap was drawn by TBtools (Chen et al., 2023), while the GO IDs were clustered by simplifyEnrichment (Gu and Hübschmann, 2023).

2.5 Protein-protein interaction network analysis

cDEGs were utilized for conducting PPI network analysis with the STRING database (Szklarczyk et al., 2019), and subsequently, the network was visualized using OmicSuite software (Miao et al., 2023). The annotation of proteins in the network was verified through blastp in the NCBI database (Johnson et al., 2008).

2.6 GO and KEGG pathway enrichment analysis

scDEGs were used for GO and KEGG pathway enrichment analysis. The log2FC value for each scDEG was calculated by taking the log 2 of the average count value of all CLas infected samples dividing the average count value of all CLas free samples. The GO terms of all genes in *C. sinensis* were annotated by eggNOG-mapper 2.1.12 (Cantalapiedra et al., 2021), while the KEGG ID of each gene of *C. sinensis* were annotated by KofamKOALA (Aramaki et al., 2020). Enrichment analysis for both GO and KEGG pathways was carried out using gogseasenior and pathwaygseasenior online tools on the omicshare website (www.omicshare.com).

2.7 Gene set enrichment analysis

GSEA was conducted using 'Simple GO GSEA Wrapper' within the TBtools software (Subramanian et al., 2005; Chen et al., 2023). scDEGs were used for gene ranking and gene sets with *P* value less than 0.05 were considered as significantly enriched gene sets.

2.8 qRT-PCR analysis

Quantitative real-time PCR (qRT-PCR) was implemented to corroborate the RNA-seq data. We gathered mature leaves from

both CLas-infected and CLas-free *C. sinensis* cv. "Newhall", from which total RNA was extracted utilizing the *EasyPure*® Plant RNA Kit (Transgen Biotech, Beijing). RNA was subsequently reverse-transcribed to produce first-strand cDNA by means of the *EasyScript*® All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (One-Step gDNA Removal) (Transgen Biotech, Beijing). Six arbitrarily chosen DEGs up-regulated in CLas-infected samples were subjected to qRT-PCR analysis with *TransStart*® Green qPCR SuperMix (Transgen Biotech, Beijing) employing specific primers (Supplementary Table 1). The *glyceraldehyde-3-phosphate dehydrogenase* gene (NCBI Reference Sequence: XM_006468885.2) of *C. sinensis* was utilized as an internal control, and the $2^{-\Delta\Delta CT}$ method was employed to determine the relative expression profiles of the chosen genes (Livak and Schmittgen, 2001). The statistical significance of results from qRT-PCR analysis was assessed using an unpaired two-sided Student's *t*-test through SPSS 25.0.

3 Results

3.1 Six BioProjects with 66 BioSamples were generated in this study

To better understand the disruptions in biological processes of *C. sinensis* caused by CLas infection, we carried out keyword searches within the SRA database in NCBI. A total of 6 BioProjects containing 66 BioSamples were collected and analyzed in this study. The 6 BioProjects included PRJNA951807 (Shahzad et al., 2023), PRJNA599503 (Peng et al., 2021), PRJNA579742 (Wu et al., 2020), PRJNA394061 (Zhao et al., 2019), PRJNA203307 (Martinelli et al., 2012, 2013), and PRJNA953196 (Liu et al., 2023) (Table 1) and detailed accessions of 66 BioSamples were listed in Supplementary Table 2. The 66 BioSamples included tissues from leaf, bark, root, fruit, and calyx abscission zones, divided into two groups: MOCK and HLB (Table 1).

3.2 WGCNA revealed key modules associated with CLas infection

To identify key modules related with CLas infection, we performed co-expression analysis with WGCNA. After normalizing and filtering the raw count values of all genes in *C. sinensis*, a total of 14,587 genes were retained. Due to the power-law distribution of the degree of nodes in a biological network, we calculated the soft threshold and determined that a power of 9 was appropriate for further analysis (Supplementary Figure 1). The remaining genes were clustered into 33 modules (Figure 1A) and the correlation between each module were calculated (Figure 1B). When conducting association analysis on the trait data of various BioSamples that were previously categorized as either 'HLB' or 'MOCK', two modules were identified as significantly associated with CLas infection. These modules were the salmon module and the green module (Figure 2). The salmon module showed a correlation of 0.42 with 'HLB' and a *P* value of 0.00045, while the green module had a correlation of 0.25 with 'HLB' and a *P* value of 0.042 (Figure 2).

3.3 Identification of candidate disease resistant and ROS scavenge gene networks through co-expression analysis

To identify hub genes in the salmon and green module, we plotted gene significance (|GS|) versus module membership (|MM|) in coordinate system and genes with |MM| > 0.5 and |GS| > 0.5 were defined as hub genes (Figure 3). Then weight value of gene connectivity greater than 0.8^9 were remained for co-expression network construction. In salmon module, 7 genes with top ranking were labeled with their gene functional descriptions. These genes included 'MACPF domain-containing protein NSL1' (*Cs_ont_2g004620*), 'Leaf rust 10 disease-resistance locus receptor-like protein kinase-like' (*Cs_ont_6g004280*), 'protein ROH1' (*Cs_ont_3g007360*), 'Uncharacterized protein' (*Cs_ont_7g000040*), 'protein ALP1-like' (*Cs_ont_4g011310*), 'Ethylene-responsive transcription factor 9' (*ERF9*) (*Cs_ont_2g022570*) and 'Ethylene-regulated transcript 2' (*Cs_ont_2g008210*) (Figure 4A; Supplementary Table 3). In green module, 6 genes including 'Thioredoxin reductase 1' (*TrxR1*) (*Cs_ont_1g029680*), 'Ferrochelatase' (*Cs_ont_4g005230*), 'peroxisomal adenine nucleotide carrier 1' (*Cs_ont_3g016200*), 'protein ESMERALDA 1' (*Cs_ont_2g012890*), 'Anamorsin-like' (*Cs_ont_7g018200*), and 'ERV-F (C)1 provirus ancestral Env polyprotein' (*Cs_ont_2g019740*) were identified (Figure 4B; Supplementary Table 4). These hub genes positive correlated with CLas infection would be key candidates for functional validation in further study.

3.4 DEGs varied in different experiments: a case study

We conducted DEGs identification to uncover the biological processes disrupted by CLas infection, resulting in the identification of diverse DEGs in different experiments (Supplementary Tables 5, 6). A maximum of 4,250 DEGs were identified in C7, comprising 1,522 up-regulated genes and 2,728 down-regulated genes. In contrast, the minimum number of DEGs was observed in C1, where 5 genes were found to be down-regulated (Figure 5; Supplementary Table 4). Due to the limited number of DEGs in C1 and C2 from experiment 1, the subsequent analysis did not include these two groups of DEGs. Following that, we proceeded with a thorough analysis of experiment 2, which includes a variety of tissues, such as leaves, bark, and roots. The comparison between HLB and MOCK samples of leaves, bark, and roots resulted in the identification of 145 common up-regulated DEGs and 14 common down-regulated DEGs (Figures 6A, B). The analysis of GO annotation revealed that the commonly down-regulated DEGs are associated with 'Development', 'Morphogenesis', 'Organization', and 'Growth', while the up-regulated DEGs are associated with 'Response', 'Regulation', 'Transport', and 'Homeostasis', as well as other GO keywords like 'Defense', 'Stimulus', 'Stress' and 'Programmed cell death' (Figure 6C). In this study, we pay

10.3389/fpls.2024.1388163

TABLE 1 Details of RNA-seq data of *Citrus sinensis* associated with CLas infection used in this study.

Experiment	Group	Control group	Stress group	Comparison	Accession
Experiment 1	2018M11	Mild-2018M11-R1 Mild-2018M11-R2 Mild-2018M11-R3 Mild-2018M11-R4	Severe-2018M11-R1 Severe-2018M11-R2 Severe-2018M11-R3 Severe-2018M11-R4	C1	PRJNA951807
	2019M2	Mild-2019M2-R1 Mild-2019M2-R2 Mild-2019M2-R3 Mild-2019M2-R4	Severe-2019M2-R1 Severe-2019M2-R2 Severe-2019M2-R3 Severe-2019M2-R4	C2	
Experiment 2	Leaf	CK_L1 CK_L2 CK_L3	HLB_L1 HLB_L2 HLB_L3	C3	PRJNA599503
	Bark	CK_P1 CK_P2 CK_P3	HLB_P1 HLB_P2 HLB_P3	C4	
	Root	CK_R1 CK_R2 CK_R3	HLB_R1 HLB_R2 HLB_R3	C5	
Experiment 3	Leaf	CS-M_1	CS-HLB_1	C6	PRJNA579742
Experiment 4	Calyx Abscission zones	Rh1 Rh2	Rd1 Rd2	C7	PRJNA394061
		Dh1 Dh2	Dd1 Dd2	C8	
Experiment 5	Mature fruit	MF-CO	MF-SY	C9	PRJNA203307
		MF-CO	MF-AS	C10	
		MF-CO	MF-AH	C11	
	Immature fruit	IF-CO	IF-SY	C12	
		IF-CO	IF-AS	C13	
		IF-CO	IF-AH	C14	
	Mature leaves	ML-CO	ML-SY	C15	
		ML-CO	ML-AS	C16	
		ML-CO	ML-AH	C17	
	Young leaves	YL-CO	YL-SY	C18	
		YL-CO	YL-AS	C19	
		YL-CO	YL-AH	C20	
Experiment 6	Young leaves	MOCK1 MOCK2 MOCK3	HLB1 HLB2 HLB3	C21	PRJNA953196

more attention to the GO enrichment results related with defense response, which were not detailed described in previous analysis (Peng et al., 2021).

3.5 PPI network analysis identified key SEO proteins

The count function in Excel was used to obtain cDEGs, resulting in the generation of 114 cDEGs. Following this, the protein sequences of cDEGs were downloaded from the CPBD database and submitted to the String database for analysis of PPI networks. Five separate PPI networks were created (Figure 7; Supplementary Table 7). It is interesting to note that one PPI network consisted of nine nodes, with three of them being sieve element occlusion (SEO) B, one being an HTH type transcriptional regulator, and one being a wall-associated receptor kinase-like 15. Furthermore, there was another PPI network with three nodes, which consisted of very-long-chain aldehyde decarbonylase CER1, salicylic acid-binding protein 2-like, and pathogenesis-related protein PR-1. The other PPI networks also consisted of resistance associated genes like chitinase, abscisic acid receptor PYL4 (Figure 7). Our analysis indicated SEO proteins may be significantly involved in citrus-CLas interaction.

10.3389/fpls.2024.1388163

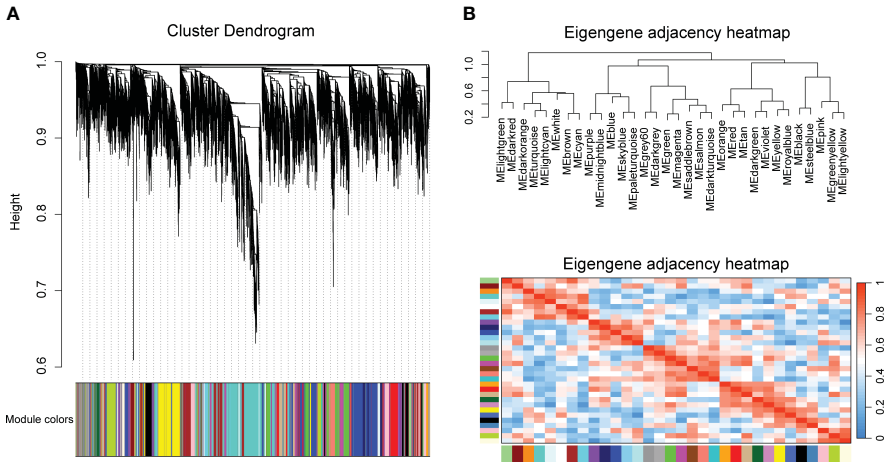


FIGURE 1
Clustering genes of *Citrus sinensis* into different modules base on the expression profiles of CLas infected samples by RNA-seq analysis. **(A)** Cluster dendrogram of all expressed genes. **(B)** Correlation heatmap of different modules.

3.6 Defense responsive GO terms and secondary metabolic pathways were disturbed by CLas infection

A total of 676 scDEGs were filtered and used for GO and KEGG pathway enrichment analysis. The GO enrichment analysis showed a significant influence on GO terms related to the cell wall, particularly GO:0009505 (plant-type cell wall) (Figure 8A; Supplementary Table 8). In addition, several defense related GO terms like GO:0009605 (response to external stimulus), GO:0009607 (response to biotic stimulus), and GO:0006952 (defense response) were identified (Figure 8A). Moreover,

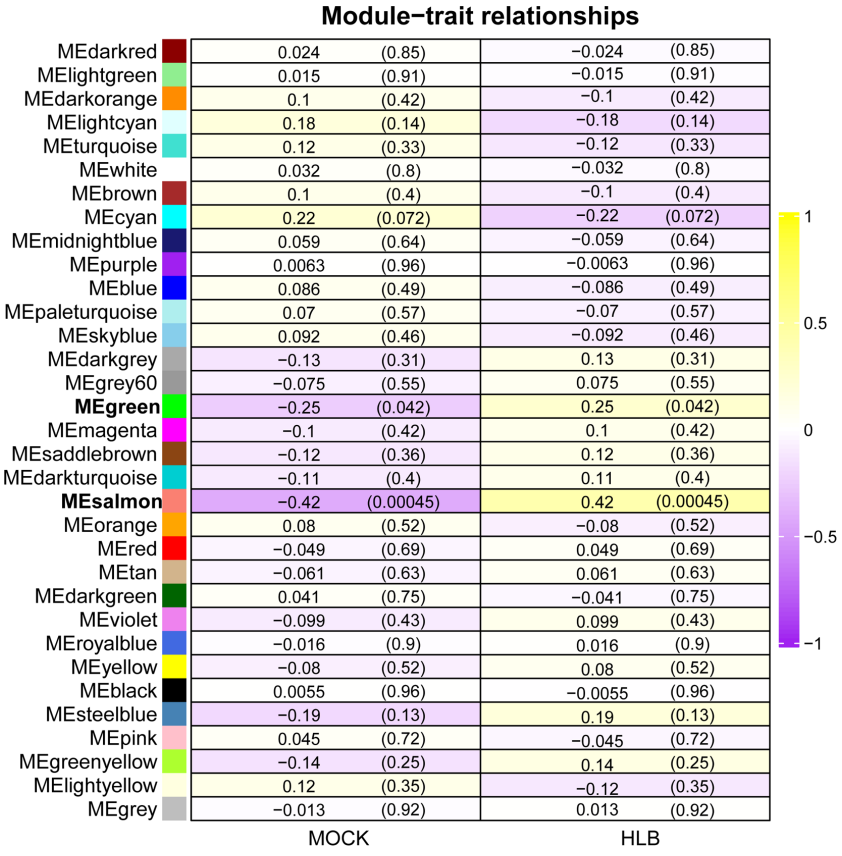


FIGURE 2
Identifying key modules of gene sets from *Citrus sinensis* associated with CLas infection by WGCNA. The numbers in the rectangular columns show the correlation coefficient and *P* value. The bold labels indicate significant modules.

10.3389/fpls.2024.1388163

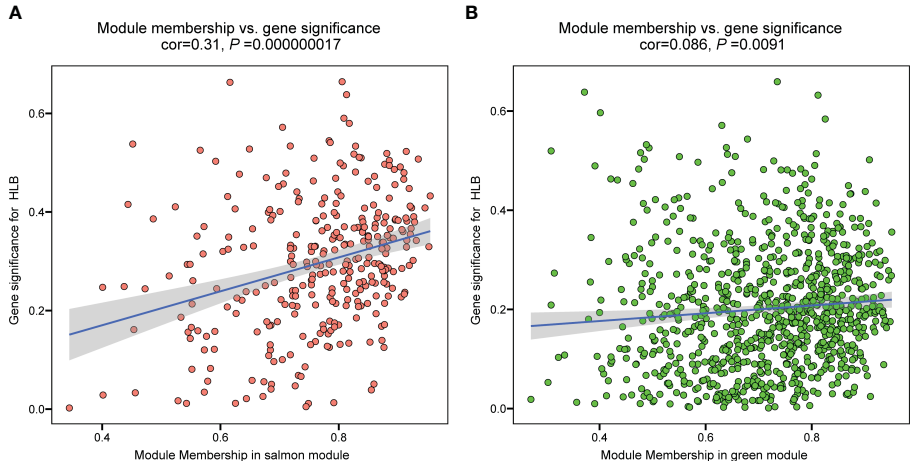


FIGURE 3
The relationship between module membership and gene significance of significant modules in the analysis of the RNA-seq data of *Citrus sinensis* associated with CLas infection. **(A)** Salmon module. **(B)** Green module.

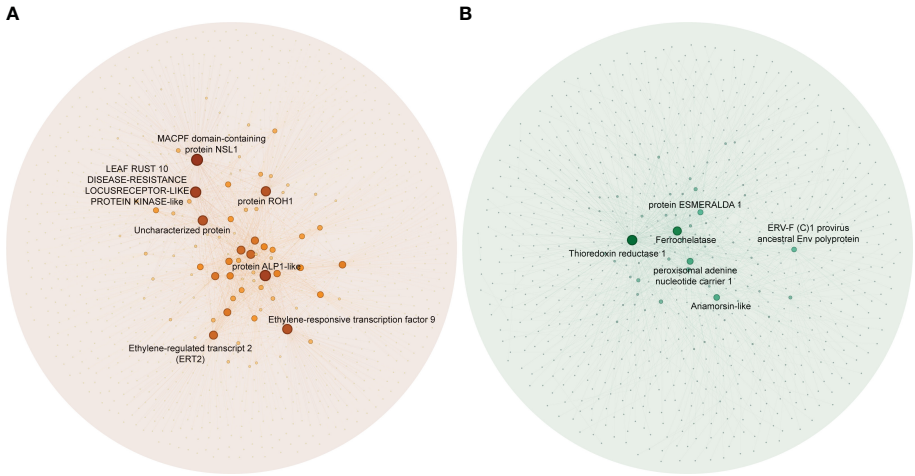


FIGURE 4
Co-expression networks in significant modules in the analysis of the RNA-seq data of *Citrus sinensis* associated with CLas infection. **(A)** Salmon module. **(B)** Green module. Blast analysis was employed to annotate the hub genes, using their descriptions found in the National Center for Biotechnology Information (NCBI) database.

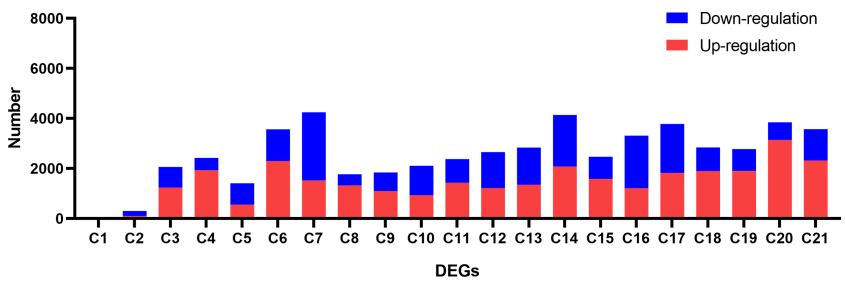


FIGURE 5
Identification of differential expressed genes (DEGs) between CLas infected and control samples in the analysis of the RNA-seq data of *Citrus sinensis*. The details for C1 to C21 are described in Supplementary Table 6.

10.3389/fpls.2024.1388163

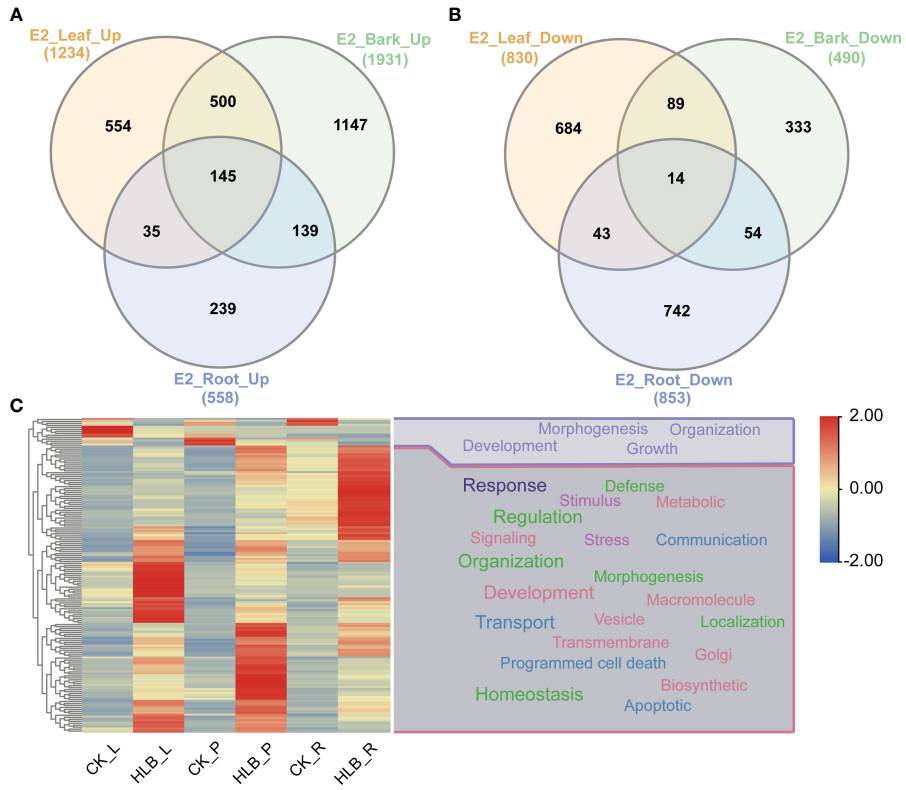


FIGURE 6
Differential expressed genes identification and GO enrichment analysis in experiment 2. **(A)** Common up-regulated genes. **(B)** Common down-regulated genes. **(C)** GO enrichment analysis. E2 signifies experiment 2, the specifics of which are documented in Table 1. The symbols L, P, and R denote leaf, bark, and root samples respectively.

GO:0010150 (leaf senescence) was significantly enriched in this study (Figure 8A). KEGG pathway enrichment indicated various pathways were disturbed during CLas infection (Figure 8B; Supplementary Table 9). It is worth noting that sugar metabolism, including 'Amino sugar and nucleotide sugar metabolism', 'Glucosinolate biosynthesis', 'Starch and sucrose metabolism', and 'pentose and glucuronate interconversions', was affected (Figure 8B). Signaling pathways such as 'MAPK signaling pathway-plant' was altered (Figure 8B). Amount of biosynthesis pathways were impacted like 'Cutin, suberine and wax biosynthesis', 'Terpenoid backbone biosynthesis', 'Sesquiterpenoid and triterpenoid biosynthesis', 'Phenylpropanoid biosynthesis', 'Brassinosteroid biosynthesis', 'Stilbenoid, diarylheptanoid and gingerol biosynthesis', and 'Flavonoid biosynthesis' (Figure 8B). Additionally, there had been a change in the 'Plant-pathogen interaction' (Figure 8B). The expression of genes related to biotic stresses in *Citrus* spp. was found to be significantly influenced according to the GO and KEGG enrichment analysis.

3.7 Various enriched gene sets positively correlated with CLas infection

GSEA was conducted utilizing scDEGs to identify gene sets exhibiting positive correlations with CLas infection. The results of GSEA were in alignment with GO enrichment analysis (Figure 9; Supplementary Table 10). The GSEA revealed a significant number of up-regulated DEGs associated with GO:0010150 (leaf senescence), GO:0009607 (response to biotic stimulus), GO:0009505 (plant-type cell wall), and GO:1901135 (Carbohydrate derivative metabolic process) (Figure 9; Supplementary Table 10). These results indicated CLas infection interferes with multiple biological processes within citrus plants.

3.8 RNA-seq data validation by qRT-PCR analysis

QRT-PCR was employed to verify the accuracy of the RNA-seq data. Six DEGs that were found to be up-regulated in CLas-infected *C. sinensis* were randomly selected for further analysis, including vacuolar amino acid transporter YPQ1 (*Cs_ont_1g026080*), uncharacterized protein (*Cs_ont_2g006750*), glucose-6-phosphate/phosphate translocator 2 (*Cs_ont_2g006950*), kunitz trypsin inhibitor 3 (*Cs_ont_5g028190*), o-methyltransferase 3 (*Cs_ont_6g024150*), and glucose-1-phosphate adenylyl transferase large subunit 3 (*Cs_ont_8g021290*). The qRT-PCR analysis revealed an up-regulation of all chosen genes, consistent with the RNA-seq data (Figure 10). Therefore, this validation confirmed the reliability of the RNA-seq analysis.

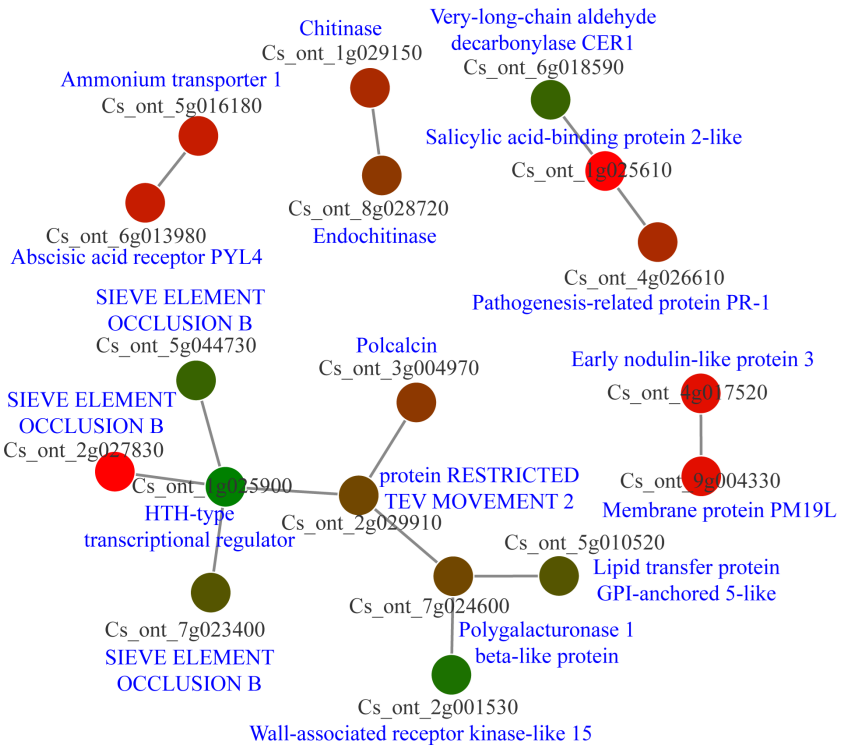


FIGURE 7
Protein-protein interaction (PPI) network analysis of core differential expressed genes in *Citrus sinensis* during CLas infection. PPI networks were derived using the STRING database, with each node arbitrarily assigned a hue from the rainbow spectrum. Detailed descriptions pertaining to each gene were sourced from the National Center for Biotechnology Information (NCBI) database.

4 Discussion

The citrus industry grapples with the destructive impact of HLB, a disease propagated globally causing significant damage (Wang, 2019; Hu et al., 2021). This study has sought to analyze the pathogenesis of CLas, despite the ambiguity surrounding its unculturable nature *in vitro*. In the realm of plant science, WGCNA has proven to be a powerful tool for exploring complex genetic traits and plant responses to various environmental factors (Mutinda et al., 2023). Unveiling these hub genes can lead us to the genetic controls underlying critical physiological traits, stress responses, and adaptation strategies in plants (Yu et al., 2023). Utilizing 289 transcriptome samples of *Dendrobium catenatum*, key genes including *DcCHIL*, *DcFLS*, *DcDFR*, and *DcWRKY3/4* were determined to react to methyl Jasmonate treatment. Further experimental data confirmed the role of *DcWRKY3/4* in regulating the metabolic pathway of flavonoids (Li et al., 2024). In the complex interplay between sorghum and the parasitic plant *Striga hermonthica*, the use of WGCNA facilitated the identification of several key genes playing a role in resistance response (Mutinda et al., 2023). During *Sporisorium scitamineum* infection, the transcriptomic profiling of sugarcane pinpointed 38 pivotal genes by WGCNA, incorporating those encoding chitinase, glutathione S-transferase, and heavy metal-associated isoprenylated plant protein (Wu et al., 2022). In this study, key gene modules were identified in *C. sinensis* during CLas infection. This was achieved through a co-expression analysis with WGCNA, retaining a total of 14,587 genes for analysis. Through our intricate research methodology, we identified two modules, namely 'Salmon' and 'Green', significantly associated with the infection. The hub genes identified within the 'Salmon' and 'Green' modules included *ERF9* (*Cs_ont_2g022570*), 'Leaf rust 10 disease-resistance locus receptor-like protein kinase-like' (*Cs_ont_6g004280*), and *TrxR1* (*Cs_ont_1g029680*). It has been demonstrated that *ERF9* is engaged in plant defense mechanisms against necrotic fungi (Maruyama et al., 2013). The leaf rust 10 disease-resistance locus receptor-like protein kinase-like protein is also implicated in the complex relationship between wheat and *Puccinia triticina*, which triggers leaf rust (Lee et al., 2020). *TrxR1* serves a crucial function in maintaining the equilibrium of redox reactions within plant cells (Gelhaye et al., 2005). The genes within these modules are involved in intricate processes such as disease resistance, reactive oxygen species (ROS) scavenging, and a range of other functional activities. Our study supplements these findings by identifying specific genes that might be involved in these complex processes.

Previous transcriptome analyses linked to HLB have pointed out significant influences on GO terms and KEGG pathways associated with stimulus and metabolic processes (Martinelli et al., 2012; Rawat et al., 2017; Arce-Leal et al., 2020; Liu et al., 2023). For example, during the asymptomatic stage of CLas infection, the most enriched biological process GO terms were associated with stimulus response, energy generation, lipid metabolism, and cellular homeostasis (Arce-Leal et al., 2020). In order to understand the biological processes affected by CLas

10.3389/fpls.2024.1388163

ID	Descrption
GO:0009505	plant−type cell wall
GO:0005618	cell wall
GO:0030312	external encapsulating structure
GO:0071944	cell periphery
GO:0009605	response to external stimulus
GO:0004497	monooxygenase activity
GO:1901700	response to oxygen−containing compound
GO:0071554	cell wall organization or biogenesis
GO:0051707	response to other organism
GO:0043207	response to external biotic stimulus
GO:0009607	response to biotic stimulus
GO:0006952	defense response
GO:0042221	response to chemical
GO:0050896	response to stimulus
GO:0001101	response to acid chemical
GO:0071669	plant−type cell wall organization or biogenesis
GO:0014070	response to organic cyclic compound
GO:0031225	anchored component of membrane
GO:0010150	leaf senescence
GO:0010260	organ senescence

FIGURE 8
GO and KEGG pathway enrichment analysis of soft-core differential expressed genes in *Citrus sinensis* during CLas infection. **(A)** GO enrichment analysis of soft-core DEGs. **(B)** KEGG pathway enrichment analysis of soft-core DEGs. The enrichment GO terms and KEGG pathways were organized according to -log10 (*P* value), and the leading 20 enriched GO terms and the top 25 enriched KEGG pathways were subsequently visualized.

infection, we performed GO enrichment analysis of DEGs. The identification of GO terms related to 'Programmed cell death' and 'Homeostasis' is noteworthy. Considering that HLB stimulates the accumulation of reactive oxygen species (ROS), triggers programmed cell death, and is perceived as an immune-mediated disease (Ma et al., 2022), Our findings further indicate that CLas infection significantly impacts the gene expression patterns related to programmed cell death and redox homeostasis. Furthermore, in the HLB-tolerant variety *Poncirus trifoliata*, there was a significant expansion in the number of genes associated with the MAPK signaling pathway (Bao et al., 2023). In contrast, the MAPK signaling pathway genes were significantly induced in another HLB-tolerant variety, *C. jambhiri*, but not in the HLB-susceptible variety *C. sinensis* (Yu et al., 2017). Nevertheless, our study revealed a slight activation of the MAPK signaling pathway following CLas infection, as determined by KEGG enrichment analysis. Moreover, our study identified activation in the metabolic pathway of brassinosteroid biosynthesis. Prior research has demonstrated that treatment with brassinosteroids notably mitigated the symptoms of CLas (Canales et al., 2016). Our findings further substantiate the

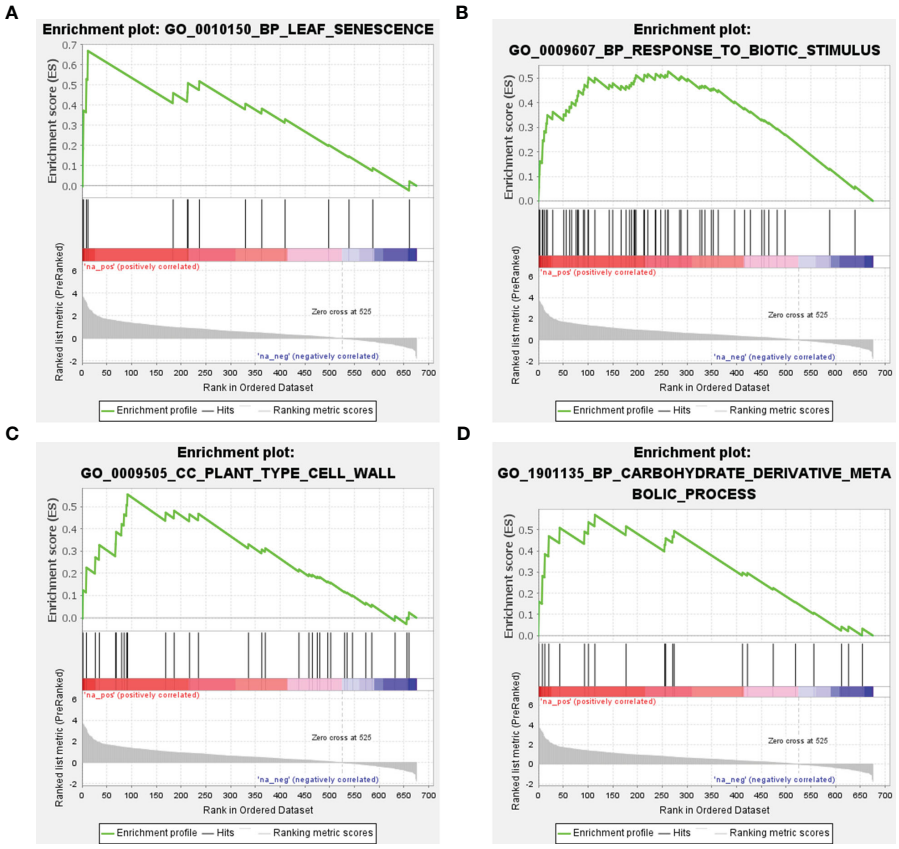


FIGURE 9
Gene set enrichment analysis of soft-core differential expressed genes in *Citrus sinensis* during CLas infection. **(A)** Leaf senescence. **(B)** Response to biotic stimulus. **(C)** Plant-type cell wall. **(D)** Carbohydrate derivative metabolic process. The enrichment score and the ranked list metric for each gene were individually ploted.

understanding that citrus plants response to the CLas infection via conserved biological processes or metabolic pathways.

The PPI network analysis plays a critical role in illuminating the mechanics of transcriptional regulatory networks in plants (Alves et al., 2014; Konishi and Yanagisawa, 2019). The identification of 114 cDEGs and the subsequent PPI network analysis shed light on key responsive elements within the host. The PPI networks underscored the participation of numerous genes associated with resistance, accentuating the reaction of citrus plants to CLas infection. Noteworthy, *SEO* genes are responsible for the production of structural phloem proteins, which play an essential role in the wound-sealing process within the phloem (Ernst et al., 2012). The HTH-type transcriptional regulator has been observed to respond during instances of pathogen infection (Matoušek et al., 2015). The wall-associated receptor kinase serves as a crucial protein, facilitating resistance against pathogens in a variety of plants (Li et al., 2009; Qi et al., 2021a, Qi et al., 2021b). Another intriguing PPI network comprises very-long-chain aldehyde decarbonylase CER1, salicylic acid-binding protein 2-like, and pathogenesis-related protein PR-1. Salicylic acid-binding protein 2-like and pathogenesis-related protein PR-1 are integral components of the salicylic acid signaling pathway (Yang et al., 2022). However, the expression of CER1 was inhibited during salicylic acid treatment (Wang et al., 2021). The interplay of these proteins suggests a complex defense response, possibly involving pathways related to salicylic acid signaling and related defense pathways.

Furthermore, GSEA has emerged as a powerful method for interpreting gene expression data in the context of plant-pathogen interactions (Jiang et al., 2017). In rice-blast fungus interaction. multiple upregulated gene sets related to pathogen recognition, signal transduction, and the production of antimicrobial compounds in the rice cultivars resistant to the blast fungus were identified using GSEA (Wei et al., 2013). An analysis of the transcriptomic response of Flax to *Fusarium oxysporum* revealed a notable enrichment of genes associated with the terpenoid backbone biosynthesis pathway by GSEA (Galindo-González and Deyholos, 2016). Additionally, our GSEA results validated the conclusions obtained from the GO enrichment analysis. Many up-regulated DEGs were involved in leaf senescence, response to biotic stimulus, plant-type cell wall, and carbohydrate derivative metabolic process, strengthening the evidence of the extensive impact of CLas infection at the cellular and molecular levels. There has been a demonstrated correlation between leaf senescence and the susceptibility of plants to pathogens (Jarosch et al., 2005; Häffner et al., 2015). Notably, GO:0010150 (leaf senescence) was significantly enriched, supporting observations of premature aging in HLB-infected citrus trees. In this study, results of enrichment analysis revealing a substantial impact on GO terms

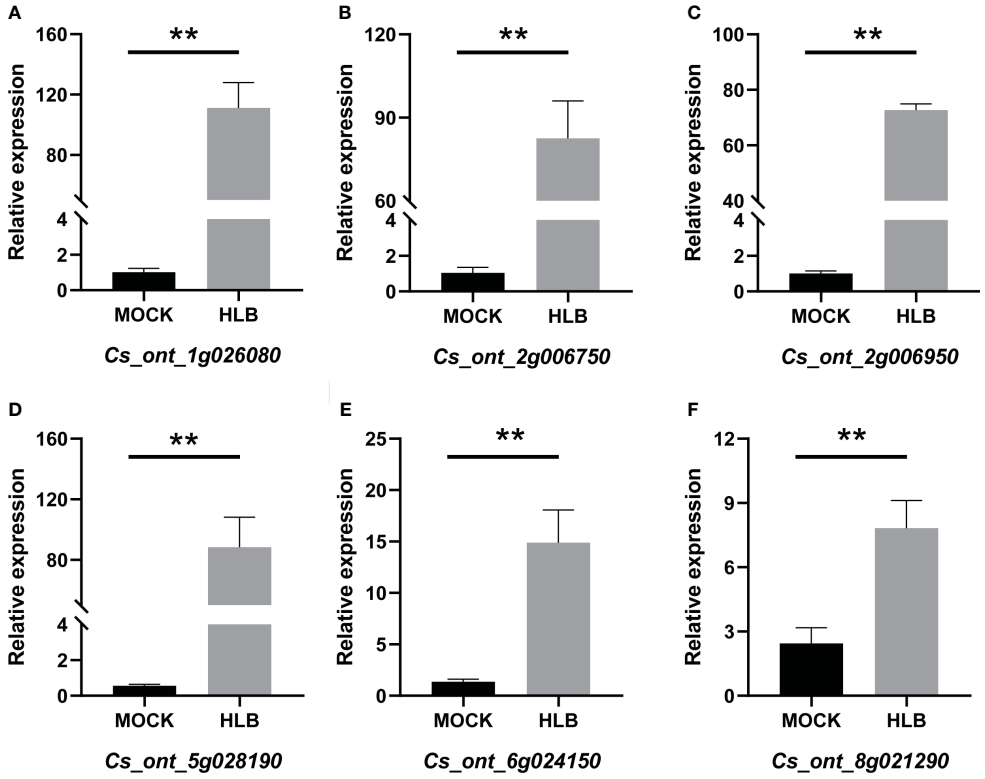


FIGURE 10
QRT-PCR analysis of the expression profiles of six candidate genes from *Citrus sinensis* during CLas infection. Relative expression of **(A)** *Cs_ont_1g026080*, **(B)** *Cs_ont_2g006750*, **(C)** *Cs_ont_2g006950*, **(D)** *Cs_ont_5g028190*, **(E)** *Cs_ont_6g024150*, and **(F)** *Cs_ont_8g021290*. The *C. sinensis GAPDH* gene was used as internal reference, three biological replicates were adopted, and the relative expression values were calculated by $2^{-\Delta\Delta CT}$ method. The symbol '**' denotes a *P* value of less than 0.01, indicating statistical significance.

related to the cell wall and various defense responses. These variations in the host's transcriptome profile underscore the pervasive and disruptive nature of HLB disease, affecting both structural and defense-related components of the host.

5 Conclusion

In summary, a comprehensive analysis of *C. sinensis* RNA-seq data in relation to CLas infection was conducted in this study. Hub genes positive correlated with CLas infection, including 'Leaf rust 10 disease-resistance locus receptor-like protein kinase-like', *ERF9* and *TrxR1*, were identified using WGCNA. Moreover, GO and KEGG enrichment revealed various biological processes and metabolic pathways, including 'Programmed cell death', 'Homeostasis', 'MAPK signaling pathway', and 'Brassinosteroid biosynthesis' were significantly influenced. PPI network analysis identified key responsive proteins such as SEO, HTH type transcriptional regulator, and wall-associated receptor kinase-like 15. Furthermore, GESA indicated a significant number of up-regulated DEGs associated with leaf senescence, biotic stimulus, plant-type cell wall and carbohydrate derivative metabolic process. These findings would broaden our understanding of citrus-CLas interactions and pave the way for future studies on the intricate transcriptional regulations in *C. sinensis* during CLas infection.

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

RL: Funding acquisition, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. XW: Investigation, Software, Writing – original draft. YH: Methodology, Writing – original draft. GH: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China [grant numbers 32160621, 32260659] and the Innovation Program of short-term Leader of Jiangxi Province [grant number jxsq2020102129].

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.2024.1388163/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The scatter and bar diagrams for **(A)** scale independence, **(B)** mean connectivity, **(C)** soft connectivity, and **(D)** check scale free topology generated during progress of WGCNA.

References

Alves, M. S., Dadalto, S. P., Gonçalves, A. B., De Souza, G. B., Barros, V. A., and Fietto, L. G. (2014). Transcription factor functional protein-protein interactions in plant defense responses. *Proteomes* 2, 85–106. doi: 10.3390/proteomes2010085

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859

Arce-Leal, Á.P., Bautista, R., Rodríguez-Negrete, E. A., Manzanilla-Ramírez, M.Á., Velázquez-Monreal, J. J., Santos-Cervantes, M. E., et al. (2020). Gene expression profile of Mexican lime (*Citrus aurantifolia*) trees in response to Huanglongbing disease caused by *Candidatus* Liberibacter asiaticus. *Microorganisms* 8, 528. doi: 10.3390/microorganisms8040528

Bao, Y., Zeng, Z., Yao, W., Chen, X., Jiang, M., Sehrish, A., et al. (2023). A gap-free and haplotype-resolved lemon genome provides insights into flavor synthesis and huanglongbing (HLB) tolerance. *Hortic. Res.* 10, uhad020. doi: 10.1093/hr/uhad020

Bautista, D., Guayazan-Palacios, N., Buitrago, M. C., Cardenas, M., Botero, D., Duitama, J., et al. (2021). Comprehensive Time-Series Analysis of the Gene Expression Profile in a Susceptible Cultivar of Tree Tomato (Solanum betaceum) During the Infection of Phytophthora betacei. *Front. Plant Sci.* 12, 730251. doi: 10.3389/fpls.2021.730251

Bové, J., and Barros, A. (2006). Huanglongbing: A destructive, newly emerging, century-old disease of citrus. *J. Plant Pathol.* 88, 7–37. doi: 10.4454/jpp.v88i1.828

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Canales, E., Coll, Y., Hernández, I., Portieles, R., Rodríguez García, M., López, Y., et al. (2016). '*Candidatus* Liberibacter asiaticus', causal agent of citrus Huanglongbing, is reduced by treatment with Brassinosteroids. *PloS One* 11, e0146223. doi: 10.1371/journal.pone.0146223

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Clark, K., Franco, J. Y., Schwizer, S., Pang, Z., Hawara, E., Liebrand, T. W. H., et al. (2018). An effector from the Huanglongbing-associated pathogen targets citrus proteases. *Nat. Commun.* 9, 1718. doi: 10.1038/s41467-018-04140-9

Clark, K. J., Pang, Z., Trinh, J., Wang, N., and Ma, W. (2020). Sec-delivered effector 1 (SDE1) of 'Candidatus liberibacter asiaticus' Promotes citrus huanglongbing. *Mol. Plant-Microbe Interactions®* 33, 1394–1404. doi: 10.1094/MPMI-05-20-0123-R

Dong, X., Jiang, Z., Peng, Y.-L., and Zhang, Z. (2015). Revealing shared and distinct gene network organization in Arabidopsis immune responses by integrative analysis. *Plant Physiol.* 167, 1186–1203. doi: 10.1104/pp.114.254292

Du, M., Wang, S., Dong, L., Qu, R., Zheng, L., He, Y., et al. (2021). Overexpression of a "Candidatus liberibacter asiaticus" Effector gene caLasSDE115 contributes to early colonization in citrus sinensis. *Front. Microbiol.* 12, 797841. doi: 10.3389/fmicb.2021.797841

Duan, Y., Zhou, L., Hall, D. G., Li, W., Doddapaneni, H., Lin, H., et al. (2009). Complete genome sequence of citrus huanglongbing bacterium, 'Candidatus Liberibacter asiaticus' obtained through metagenomics. *Mol. Plant Microbe Interact.* 22, 1011–1020. doi: 10.1094/MPMI-22-8-1011

Ernst, A. M., Jekat, S. B., Zielonka, S., Müller, B., Neumann, U., Rüping, B., et al. (2012). Sieve element occlusion (SEO) genes encode structural phloem proteins involved in wound sealing of the phloem. *Proc. Natl. Acad. Sci.* 109, E1980–E1989. doi: 10.1073/pnas.1202999109

Fu, S., Shao, J., Zhou, C., and Hartung, J. S. (2016). Transcriptome analysis of sweet orange trees infected with '*Candidatus* Liberibacter asiaticus' and two strains of Citrus Tristeza Virus. *BMC Genomics* 17, 1–18. doi: 10.1186/s12864-016-2663-9

Galindo-González, L., and Deyholos, M. K. (2016). RNA-seq Transcriptome Response of Flax (Linum usitatissimum L.) to the Pathogenic Fungus Fusarium oxysporum f. sp. lini. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01766

Gelhaye, E., Rouhier, N., Navrot, N., and Jacquot, J.-P. (2005). The plant thioredoxin system. *Cell. Mol. Life Sci. CMLS* 62, 24–35. doi: 10.1007/s00018-004-4296-4

Gu, Z., and Hübschmann, D. (2023). simplifyEnrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinf.* 21, 190–202. doi: 10.1016/j.gpb.2022.04.008

Häffner, E., Konietzki, S., and Diederichsen, E. (2015). Keeping control: the role of senescence and development in plant pathogenesis and defense. *Plants* 4, 449–488. doi: 10.3390/plants4030449

Heberle, H., Meirelles, G. V., Da Silva, F. R., Telles, G. P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinf.* 16, 169. doi: 10.1186/s12859-015-0611-3

Hu, B., Rao, M. J., Deng, X., Pandey, S. S., Hendrich, C., Ding, F., et al. (2021). Molecular signatures between citrus and Candidatus Liberibacter asiaticus. *PloS Pathog.* 17, e1010071. doi: 10.1371/journal.ppat.1010071

Hu, Y., Zhong, X., Liu, X., Lou, B., Zhou, C., and Wang, X. (2017). Comparative transcriptome analysis unveils the tolerance mechanisms of *Citrus hystrix* in response to '*Candidatus* Liberibacter asiaticus' infection. *PloS One* 12, e0189229. doi: 10.1371/journal.pone.0189229

Jarosch, B., Collins, N. C., Zellerhoff, N., and Schaffrath, U. (2005). RAR1, ROR1, and the actin cytoskeleton contribute to basal resistance to Magnaporthe grisea in barley. *Mol. Plant Microbe Interact.* 18, 397–404. doi: 10.1094/MPMI-18-0397

Jiang, Z., He, F., and Zhang, Z. (2017). Large-scale transcriptome analysis reveals arabidopsis metabolic pathways are frequently influenced by different pathogens. *Plant Mol. Biol.* 94, 453–467. doi: 10.1007/s11103-017-0617-5

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., Mcginnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201

Kohl, M., Wiese, S., and Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Data Min. Proteomics: Standards to Appl.* 696, 291–303. doi: 10.1007/978-1-60761-987-1_18

Konishi, M., and Yanagisawa, S. (2019). The role of protein-protein interactions mediated by the PB1 domain of NLP transcription factors in nitrate-inducible gene expression. *BMC Plant Biol.* 19, 1–12. doi: 10.1186/s12870-019-1692-3

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 1–13. doi: 10.1186/1471-2105-9-559

Lee, A., Trinh, C. S., Lee, W. J., Kim, M., Lee, H., Pathiraja, D., et al. (2020). Characterization of two leaf rust-resistant Aegilops tauschii accessions for the synthetic wheat development. *Appl. Biol. Chem.* 63, 13. doi: 10.1186/s13765-020-00496-z

Li, C., Gong, Q., Liu, P., Xu, Z., Yu, Q., Dai, H., et al. (2024). Co-expressed network analysis based on 289 transcriptome samples reveals methyl jasmonate-mediated gene regulatory mechanism of flavonoid compounds in Dendrobium catenatum. *Plant Physiol. Biochem.* 206, 108226. doi: 10.1016/j.plaphy.2023.108226

Li, H., Zhou, S.-Y., Zhao, W.-S., Su, S.-C., and Peng, Y.-L. (2009). A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. *Plant Mol. Biol.* 69, 337–346. doi: 10.1007/s11103-008-9430-5

Lin, S. M., Du, P., Huber, W., and Kibbe, W. A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* 36, e11–e11. doi: 10.1093/nar/gkm1075

Liu, C., Chang, X., Li, F., Yan, Y., Zuo, X., Huang, G., et al. (2023). Transcriptome analysis of Citrus sinensis reveals potential responsive events triggered by Candidatus Liberibacter asiaticus. *Protoplasma*. doi: 10.1007/s00709-023-01911-0

Liu, X., Fan, Y., Zhang, C., Dai, M., Wang, X., and Li, W. (2019a). Nuclear Import of a Secreted "Candidatus Liberibacter asiaticus" Protein is Temperature Dependent and Contributes to Pathogenicity in Nicotiana benthamiana. *Front. Microbiol.* 10, 1684. doi: 10.3389/fmicb.2019.01684

Liu, H., Wang, X., Liu, S., Huang, Y., Guo, Y.-X., Xie, W.-Z., et al. (2022). Citrus Pan-Genome to Breeding Database (CPBD): A comprehensive genome database for citrus breeding. *Mol. Plant* 15, 1503–1505. doi: 10.1016/j.molp.2022.08.006

Liu, X., Zheng, Y., Wang-Pruski, G., Gan, Y., Zhang, B., Hu, Q., et al. (2019b). Transcriptome profiling of periwinkle infected with Huanglongbing ('*Candidatus* Liberibacter asiaticus'). *Eur. J. Plant Pathol.* 153, 891–906. doi: 10.1007/s10658-018-01607-9

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

Loto, F., Coyle, J. F., Padgett, K. A., Pagliai, F. A., Gardner, C. L., Lorca, G. L., et al. (2017). Functional characterization of LotP from Liberibacter asiaticus. *Microb. Biotechnol.* 10, 642–656. doi: 10.1111/1751-7915.12706

Lu, C., Pu, Y., Liu, Y., Li, Y., Qu, J., Huang, H., et al. (2019). Comparative transcriptomics and weighted gene co-expression correlation network analysis (WGCNA) reveal potential regulation mechanism of carotenoid accumulation in Chrysanthemum× morifolium. *Plant Physiol. Biochem.* 142, 415–428. doi: 10.1016/j.plaphy.2019.07.023

Ma, W., Pang, Z., Huang, X., Xu, J., Pandey, S. S., Li, J., et al. (2022). Citrus Huanglongbing is a pathogen-triggered immune disease that can be mitigated with antioxidants and gibberellin. *Nat. Commun.* 13, 529. doi: 10.1038/s41467-022-28189-9

Martinelli, F., Reagan, R. L., Uratsu, S. L., Phu, M. L., Albrecht, U., Zhao, W., et al. (2013). Gene regulatory networks elucidating huanglongbing disease mechanisms. *PloS One* 8, e74256. doi: 10.1371/journal.pone.0074256

Martinelli, F., Uratsu, S. L., Albrecht, U., Reagan, R. L., Phu, M. L., Britton, M., et al. (2012). Transcriptome profiling of citrus fruit response to huanglongbing disease. *PloS One* 7, e38039. doi: 10.1371/journal.pone.0038039

Maruyama, Y., Yamoto, N., Suzuki, Y., Chiba, Y., Yamazaki, K.-I., Sato, T., et al. (2013). The Arabidopsis transcriptional repressor ERF9 participates in resistance against necrotrophic fungi. *Plant Sci.* 213, 79–87. doi: 10.1016/j.plantsci.2013.08.008

Matoušek, J., Piernikarczyk, R. J. J., Týcová, A., Duraisamy, G. S., Kocábek, T., and Steger, G. (2015). Expression of SANT/HTH Myb mRNA, a plant morphogenesis-regulating transcription factor, changes due to viroid infection. *J. Plant Physiol.* 183, 85–94. doi: 10.1016/j.jplph.2015.06.001

Miao, B.-B., Dong, W., Gu, Y.-X., Han, Z.-F., Luo, X., Ke, C.-H., et al. (2023). OmicsSuite: a customized and pipelined suite for analysis and visualization of multi-omics big data. *Horticulture Res.* 10, uhad195. doi: 10.1093/hr/uhad195

Muchero, W., Sondreli, K. L., Chen, J.-G., Urbanowicz, B. R., Zhang, J., Singan, V., et al. (2018). Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree. *Proc. Natl. Acad. Sci.* 115, 11573–11578. doi: 10.1073/pnas.1804428115

Mutinda, S., Mobegi, F. M., Hale, B., Dayou, O., Ateka, E., Wijeratne, A., et al. (2023). Resolving intergenotypic Striga resistance in sorghum. *J. Exp. Bot.* 74, 5294–5306. doi: 10.1093/jxb/erad210

Pandey, S. S., Hendrich, C., Andrade, M. O., and Wang, N. (2022). Candidatus Liberibacter: From movement, host responses, to symptom development of citrus Huanglongbing. *Phytopathology®* 112, 55–68. doi: 10.1094/PHYTO-08-21-0354-FI

Pang, Z., Zhang, L., Coaker, G., Ma, W., He, S.-Y., and Wang, N. (2020). Citrus csACD2 is a target of candidatus liberibacter asiaticus in huanglongbing disease. *Plant Physiol.* 184, 792–805. doi: 10.1104/pp.20.00348

Peng, T., Kang, J.-L., Xiong, X.-T., Cheng, F.-T., Zhou, X.-J., Dai, W.-S., et al. (2021). Integrated transcriptomics and metabolomics analyses provide insights into the response of chongyi wild mandarin to candidatus liberibacter asiaticus infection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.748209

Pham-Gia, T., and Hung, T. L. (2001). The mean and median absolute deviations. *Math. Comput. Model.* 34, 921–936. doi: 10.1016/S0895-7177(01)00109-1

Pitino, M., Allen, V., and Duan, Y. (2018). LasΔ5315 Effector Induces Extreme Starch Accumulation and Chlorosis as Ca. Liberibacter asiaticus Infection in Nicotiana benthamiana. *Front. Plant Sci.* 9, 113. doi: 10.3389/fpls.2018.00113

Pitino, M., Armstrong, C. M., Cano, L. M., and Duan, Y. (2016). Transient expression of candidatus liberibacter asiaticus effector induces cell death in nicotiana benthamiana. *Front. Plant Sci.* 7, 982. doi: 10.3389/fpls.2016.00982

Qi, H., Guo, F., Lv, L., Zhu, X., Zhang, L., Yu, J., et al. (2021a). The wheat wall-associated receptor-like kinase TaWAK-6D mediates broad resistance to two fungal pathogens Fusarium pseudograminearum and Rhizoctonia cerealis. *Front. Plant Sci.* 12, 758196. doi: 10.3389/fpls.2021.758196

Qi, H., Zhu, X., Guo, F., Lv, L., and Zhang, Z. (2021b). The wall-associated receptor-like kinase TaWAK7D is required for defense responses to Rhizoctonia cerealis in wheat. *Int. J. Mol. Sci.* 22, 5629. doi: 10.3390/ijms22115629

Rawat, N., Kiran, S. P., Du, D., Gmitter, F. G., and Deng, Z. (2015). Comprehensive meta-analysis, co-expression, and miRNA nested network analysis identifies gene candidates in citrus against Huanglongbing disease. *BMC Plant Biol.* 15, 1–21. doi: 10.1186/s12870-015-0568-4

Rawat, N., Kumar, B., Albrecht, U., Du, D., Huang, M., Yu, Q., et al. (2017). Genome resequencing and transcriptome profiling reveal structural diversity and expression patterns of constitutive disease resistance genes in Huanglongbing-tolerant *Poncirus trifoliata* and its hybrids. *Horticulture Res.* 4, 17064. doi: 10.1038/hortres.2017.64

Ribeiro, C., Xu, J., Hendrich, C., Pandey, S. S., Yu, Q., Gmitter, F. G., et al. (2022). Seasonal transcriptome profiling of susceptible and tolerant citrus cultivars to citrus huanglongbing. *Phytopathology®* 113, 286–298. doi: 10.1094/PHYTO-05-22-0179-R

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rody, H. V. S., Camargo, L. E. A., Creste, S., Van Sluys, M.-A., Rieseberg, L. H., and Monteiro-Vitorello, C. B. (2021). Arabidopsis-based dual-layered biological network analysis elucidates fully modulated pathways related to sugarcane resistance on biotrophic pathogen infection. *Front. Plant Sci.* 12, 707904. doi: 10.3389/fpls.2021.707904

Shahzad, F., Chun, C., Schumann, A., and Vashisth, T. (2020). Nutrient uptake in huanglongbing-affected sweet orange: Transcriptomic and physiological analysis. *J. Am. Soc. Hortic. Sci.* 145, 349–362. doi: 10.21273/JASHS04929-20

Shahzad, F., Tang, L., and Vashisth, T. (2023). Unraveling the mystery of canopy dieback caused by citrus disease Huanglongbing and its link to hypoxia stress. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1119530

Shi, J., Gong, Y., Shi, H., Ma, X., Zhu, Y., Yang, F., et al. (2023b). 'Candidatus Liberibacter asiaticus' secretory protein SDE3 inhibits host autophagy to promote Huanglongbing disease in citrus. *Autophagy* 19, 2558–2574. doi: 10.1080/15548627.2023.2213040

Shi, Q., Pitino, M., Zhang, S., Krystel, J., Cano, L. M., Shatters, R. G. Jr., et al. (2019). Temporal and spatial detection of Candidatus Liberibacter asiaticus putative effector transcripts during interaction with Huanglongbing-susceptible, -tolerant, and -resistant citrus hosts. *BMC Plant Biol.* 19, 122. doi: 10.1186/s12870-019-1703-4

Shi, H., Yang, Z., Huang, J., Wu, H., Fu, S., Li, W., et al. (2023a). An effector of 'Candidatus Liberibacter asiaticus' manipulates autophagy to promote bacterial infection. *J. Exp. Bot.* 74, 4670–4684. doi: 10.1093/jxb/erad176

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Wang, N. (2019). The citrus huanglongbing crisis and potential solutions. *Mol. Plant* 12, 607–609. doi: 10.1016/j.molp.2019.03.008

Wang, S., Du, M., Dong, L., Qu, R., Ran, D., Ma, J., et al. (2023). Function and molecular mechanism analysis of CaLasSDE460 effector involved in the pathogenesis of "Candidatus Liberibacter asiaticus" in citrus. *Mol. Hortic.* 3, 14. doi: 10.1186/s43897-023-00062-3

Wang, D., Ni, Y., Liao, L., Xiao, Y., and Guo, Y. (2021). Poa pratensis ECERIFERUM1 (PpCER1) is involved in wax alkane biosynthesis and plant drought tolerance. *Plant Physiol. Biochem.* 159, 312–321. doi: 10.1016/j.plaphy.2020.12.032

Wang, Y., Zhou, L., Yu, X., Stover, E., Luo, F., and Duan, Y. (2016). Transcriptome profiling of Huanglongbing (HLB) tolerant and susceptible citrus plants reveals the role of basal resistance in HLB tolerance. *Front. Plant Sci.* 7, 933. doi: 10.3389/fpls.2016.00933

Wei, T., Ou, B., Li, J., Zhao, Y., Guo, D., Zhu, Y., et al. (2013). Transcriptional profiling of rice early response to Magnaporthe oryzae identified OsWRKYs as important regulators in rice blast resistance. *PloS One* 8, e59720. doi: 10.1371/journal.pone.0059720

Wu, H., Hu, Y., Fu, S., Zhou, C., and Wang, X. (2020). Coordination of multiple regulation pathways contributes to the tolerance of a wild citrus species (Citrus ichangensis '2586') against Huanglongbing. *Physiol. Mol. Plant Pathol.* 109, 101457. doi: 10.1016/j.pmpp.2019.101457

Wu, Q., Pan, Y.-B., Su, Y., Zou, W., Xu, F., Sun, T., et al. (2022). WGCNA identifies a comprehensive and dynamic gene co-expression network that associates with smut resistance in sugarcane. *Int. J. Mol. Sci.* 23, 10770. doi: 10.3390/ijms231810770

Yang, F., Wu, C., Zhu, G., Yang, Q., Wang, K., and Li, Y. (2022). An integrated transcriptomic and metabolomic analysis for changes in rose plant induced by rose powdery mildew and exogenous salicylic acid. *Genomics* 114, 110516. doi: 10.1016/j.ygeno.2022.110516

Yao, Y., Xiong, E., Qu, X., Li, J., Liu, H., Quan, L., et al. (2023). WGCNA and transcriptome profiling reveal hub genes for key development stage seed size/oil content between wild and cultivated soybean. *BMC Genomics* 24, 494. doi: 10.1186/s12864-023-09617-6

Yu, Q., Chen, C., Du, D., Huang, M., Yao, J., Yu, F., et al. (2017). Reprogramming of a defense signaling pathway in rough lemon and sweet orange is a critical element of the early response to 'Candidatus Liberibacter asiaticus'. *Hortic. Res.* 4, 17063. doi: 10.1038/hortres.2017.63

Yu, T., Zhang, J., Cao, J., Ma, X., Li, W., and Yang, G. (2023). Hub gene mining and co-expression network construction of low-temperature response in maize of seedling by WGCNA. *Genes* 14, 1598. doi: 10.3390/genes14081598

Yuan, X., Chen, C., Bassanezi, R. B., Wu, F., Feng, Z., Shi, D., et al. (2021). Region-wide comprehensive implementation of roguing infected trees, tree replacement, and insecticide applications successfully controls citrus huanglongbing. *Phytopathology®* 111, 1361–1368. doi: 10.1094/PHYTO-09-20-0436-R

Zhao, W., Baldwin, E. A., Bai, J., Plotto, A., and Irey, M. (2019). Comparative analysis of the transcriptomes of the calyx abscission zone of sweet orange insights into the huanglongbing-associated fruit abscission. *Horticulture Res.* 6, 71. doi: 10.1038/s41438-019-0152-4

Zheng, Y., Li, Y., Xu, P., Liu, C., Chen, J., Deng, X., et al. (2023). Genome sequence resource for "Candidatus Liberibacter asiaticus" strain GDCZ from a historical HLB endemic region in China. *BMC Genom Data* 24, 63. doi: 10.1186/s12863-023-01160-3

Zhu, M., Xie, H., Wei, X., Dossa, K., Yu, Y., Hui, S., et al. (2019). WGCNA analysis of salt-responsive core transcriptome identifies novel hub genes in rice. *Genes* 10, 719. doi: 10.3390/genes10090719

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 May 2024
DOI 10.3389/fpls.2024.1375646

OPEN ACCESS

EDITED BY
Hao Tong,
Max Planck Institute of Molecular Plant Physiology, Germany

REVIEWED BY
Junbo Yang,
Chinese Academy of Agricultural Sciences, China
Jindong Liu,
Chinese Academy of Agricultural Sciences, China
Huatao Chen,
Jiangsu Academy of Agricultural Sciences (JAAS), China

*CORRESPONDENCE
Zhangxiong Liu
liuzhangxiong@caas.cn
Xu Zhu
zhu.xu@ruibiotech.com
Ming Yuan
y.m@haas.cn

†These authors have contributed equally to this work

RECEIVED 24 January 2024
ACCEPTED 15 April 2024
PUBLISHED 14 May 2024

CITATION
Han D, Zhao X, Zhang D, Wang Z, Zhu Z, Sun H, Qu Z, Wang L, Liu Z, Zhu X and Yuan M (2024) Genome-wide association studies reveal novel QTLs for agronomic traits in soybean.
Front. Plant Sci. 15:1375646.
doi: 10.3389/fpls.2024.1375646

Genome-wide association studies reveal novel QTLs for agronomic traits in soybean

Dongwei Han[1,2†], Xi Zhao[3†], Di Zhang[1], Zhen Wang[1], Zhijia Zhu[1], Haoyue Sun[1], Zhongcheng Qu[1], Lianxia Wang[1], Zhangxiong Liu[4*], Xu Zhu[5*] and Ming Yuan[1*]

[1]Qiqihar Branch of Heilongjiang Academy of Agricultural Science, Qiqihar, Heilongjiang, China, [2]Heilongjiang Chinese Academy of Sciences Qiuying Zhang Soybean Scientist Studio, Qiqihar, Heilongjiang, China, [3]Biotechnology Institute, Heilongjiang Academy of Agricultural Science, Harbin, Heilongjiang, China, [4]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [5]Department of Research and Development, Ruibiotech Co., Ltd, Beijing, China

Introduction: Soybean, as a globally significant crop, has garnered substantial attention due to its agricultural importance. The utilization of molecular approaches to enhance grain yield in soybean has gained popularity.

Methods: In this study, we conducted a genome-wide association study (GWAS) using 156 Chinese soybean accessions over a two-year period. We employed the general linear model (GLM) and the mixed linear model (MLM) to analyze three agronomic traits: pod number, grain number, and grain weight.

Results: Our findings revealed significant associations between *qgPNpP-98*, *qgGNpP-89* and *qgHGW-85* QTLs and pod number, grain number, and grain weight, respectively. These QTLs were identified on chromosome 16, a region spanning 413171bp exhibited associations with all three traits.

Discussion: These QTL markers identified in this study hold potential for improving yield and agronomic traits through marker-assisted selection and genomic selection in breeding programs.

KEYWORDS

soybean, GWAS, pod number per plant, grain number per plant, 100-grain weight

Introduction

Cultivated soybean (*Glycine max L.*) is a principal oilseed crop that is grown globally for its significant contribution to edible protein and oil production (Zong et al., 2017). Despite the success of the revolution in enhancing the yields of rice, wheat, and maize, comparatively lower progress has been made in improving soybean yields (Liu et al., 2020). The global objective of soybean breeders, particularly in China, is to develop genotypes that possess

increased yield potential. In soybean, pod number, grain number, and grain weight exist a substantial positive association with grain yield. However, when there is an excessive number of pods and seeds, as well as when the seeds become heavier, soybean plants become more vulnerable to lodging. This can have a detrimental effect on the overall crop yield. Therefore, it is important to pursue a balance between these traits and grain yield in order to optimize the productivity of soybean crops (Zhang et al., 2015; Ning et al., 2018). Additionally, it is worth noting that pod number, grain number, and hundred grain weight are controlled by multiple loci/genes and can be influenced by interactions between genotype and environment. Therefore, it is crucial to consider both genetic factors and environmental conditions when aiming to achieve an optimal balance between these traits and grain yield in soybean crops (Lamichhane et al., 2020; Happ et al., 2021). Both conventional breeding and marker-assisted breeding (MAB) approaches have been utilized to enhance yield traits in soybean crops (Raju et al., 2018). Recent studies have demonstrated that MAB approaches are particularly effective for traits that are sensitive to environmental conditions, such as seed germination, seedling development, and maturity (Saminadane et al., 2024; Xu et al., 2024). By utilizing a specially designed marker primer design program, MAB demonstrates significant potential in molecular breeding, offering a promising avenue for the development of improved crop varieties with desired traits (Xia et al., 2023). Hence, understanding the genetic architecture of yield-related traits is crucial in effectively utilizing marker-assisted breeding (MAB) for the development of high-yielding soybean varieties.

Genome-wide association studies (GWAS) have emerged as a highly effective approach for identifying alleles/QTLs linked to specific traits with high resolution (Zhang et al., 2021). With the advancements in sequencing-based genotyping technologies, GWAS has become increasingly popular in crop genetics research (Zhang et al., 2016, 2022). Numerous studies conducted on various plant species, including rice, maize, wheat, soybean, potato, cucumber, and tomato, have demonstrated the ability of GWAS to uncover associations between marker-trait associations (MTAs) and effectively identify the underlying genes (Zhao et al., 2011; Wang et al., 2016; Yan et al., 2019; Torkamaneh et al., 2020; Gao et al., 2021; Zhou et al., 2021; Liang et al., 2022; Luo et al., 2022; Zhao et al., 2023; Sun et al., 2022a; Sun et al., 2022b; Wang et al., 2015, 2018; Jhon et al., 2023; Zhu et al., 2023). In soybean, multiple research studies have specifically investigated the genetic basis of yield (Dong et al., 2021; Zhao et al., 2021; Bhat et al., 2022). Over the past 20 years, more than 3000 quantitative trait loci (QTLs) have been identified through GWAS in soybean (https://www.soybase.org/). These QTLs are spread across the 20 chromosomes of the soybean genome. However, the lack of effective utilization of the three yield-related QTLs (pod number, total seed number and one-hundred grain weight) has been a persistent obstacle in the development of improved soybean varieties with desirable yield traits.

The utilization of whole genome sequencing data allows for the precise mapping of QTLs associated with agronomic traits. In this study, we conducted a GWAS to identify significant MTAs, candidate genes, and QTLs in soybean cultivars selected from major soybean growing regions in China, which holds great potential for the development of enhanced soybean varieties with targeted yield characteristics through a QTL-based breeding approach.

Materials and methods

Experimental materials and cultivate management

A total of 156 soybean genotypes, including 100 accessions selected from germplasm collection by Dr. Lijuan Qiu's laboratory at the Chinese Academy of Agricultural Sciences and 56 cultivars from Qiqihar Branch of Heilongjiang Academy of Agricultural Science. These genotypes were originally from 3 provinces in China (79, 50.6%) and 10 states in the United States (77, 49.4%), representing a wide range of genetic diversity within and outside China. Experiments were carried out in Qiqihar (123.685996°E, 47.274543°N) during the years 2020 and 2021, utilizing a single-row plot system with 3-meter-long rows spaced at 0.5-meter intervals. The field trials followed a randomized complete block design and were conducted across multiple testing environments, with each test environment containing three replicates. Throughout the growing season, field management practices adhered to standard cropping system protocols, including fertilization, weed management, and insecticide fungicide application.

Trait identification and statistical analysis

The assessment of all traits from 10 randomly selected plants in each line was conducted after reaching full maturity. All plants and pods derived from conventionally developed plants were gathered for trait examination. Specifically, three traits related to yield were analyzed, including pod number per plant (PNpP), grain number per plant (GNpP), and 100-grain weight (HGW). The PNpP is calculated by counting all pods on 10 randomly selected soybean plants, including both mature and immature pods. The GNpP is the number of seeds contained in all pods of 10 randomly selected soybean materials, divided by 10 to obtain the average number of grains per plant. The statistical method of HGW is the value obtained by randomly selecting 10 soybean materials and then selecting and weighing 100 seeds from the seeds obtained. All these traits were subsequently measured in the laboratory, or determined through electronic weighing. The broad-sense heritability (H^2) was calculated following a previously reported method (Wyman and Baker, 1991). Statistical analyses of obtained data were calculated by using GraphPad Prism 7.0.

DNA extraction and SNP Genotyping

DNA samples extracted by CTAB method from 156 accessions were genotyped for SNPs using a soybean 200K array. This array named "ZDX1", which is a high-throughput SNP genotyping chip

developed jointly by the Institute of Crop Sciences of Chinese Academy of Agriculture Science and Beijing Compass Biotechnology Co., Ltd., using the Illumina platform for customization (Sun, et al., 2022c). And a total of 158959 high-quality SNPs were used for association mapping.

Linkage disequilibrium analysis and population structure

We conducted genome-wide LD analysis using Plink 2 (Chang et al, 2015) with $R^2 < 0.2$ as the threshold to identify 3026 unlinked loci. The population structure analysis was conducted using software FastStructure (Raj et al., 2014). Then analysis the kinship applying the software TASSEL version 5.0 (Bradbury et al., 2007).

Genome-wide association analysis

Two models, mixed linear model (MLM) and general linear model (GLM), were employed to investigate associations between genotypic and phenotypic data in TASSEL version 5.0. Briefly, the MLM approach was used to account for both population structure and kinship matrix, which are jointly incorporated via the Q+K approach for enhanced statistical power (Yu et al., 2006). On the other hand, GLM was utilized to analyze individual location datasets using a least square fixed effect model with Q acting as a covariate with flexible assumptions.

Results

Phenotypic analysis of three yield-related traits

To explore loci associated with agronomic traits, a total of 156 soybean landraces were cultivated over two consecutive years, and an investigation of three yield-related traits was conducted (Figure 1A). The frequency distribution of the three yield-related traits across the two years showed a continuous distribution in the GWAS panel of 156 soybean landraces, with a wide range of variation (Figure 1B). GNpP exhibited 7-fold variation, ranging from 41 to 307, with an average of 112.9 ± 20.1. HGW and PNpP showed approximately 4-fold and 7-fold differences, ranging from 6.8 g to 28.2 g and 18 to 128g, respectively. The frequency distribution of the three yield-related traits displayed an approximately normal distribution except for a few materials that showed large deviation. The broad-sense heritability (H^2) was determined for each of the three traits. All traits presented an H^2 above 54%, suggesting that genetic effects play a predominant role in the phenotype variation of these traits (Supplementary Table 1). The results of the phenotypic correlation analysis revealed strong correlations between PNpP and GNpP, while other traits exhibited no significant correlations. This suggests that the genetic-associated loci are likely to be similar between PNpP and GNpP, while the other traits are not influenced in a correlated manner (Figure 1C).

SNP profile

A total of 80163 high-quality SNPs (MAF > 0.05, missing rate < 20%) were used for a GWAS of the three traits, with an average marker density of 11.47 kb/SNP at the genome wide scale. The lowest marker density (17.10 kb/SNP) was found on chromosome 1, and the highest marker density (8.55 kb/SNP) was found on chromosome 16. Furthermore, SNPs exhibit a predominant enrichment in sub telomeric regions, particularly in areas distal to the centromeres (Figure 2A).

Population structure of 156 soybean materials

To elucidate the most promising genetic variations and contribute to a comprehensive understanding of the genetic underpinnings of the trait, we conducted a structure analysis. The 156 genotypic materials were stratified into two subpopulations based on the results obtained from the K value (Figure 2B). The kinship analysis revealed that the soybean materials utilized in this investigation originated from two primary lineages, further substantiating the existence of dual ancestors for the 156 soybean materials (Figure 2C). Examination of the linkage disequilibrium (LD) decay rates with the high-quality SNPs demonstrated that the decay curves of LD exhibited a discernible pattern dependent on distance, revealing steeper decay at longer distances. Furthermore, beyond a marker distance of 5 Mb, the r^2 value generally remains below 0.1. Precisely, at a distance of 100 kb, the LD exhibited decay with an r^2 value of 0.4, signifying a notably robust LD correlation among proximate variants. At 200 kb, the LD decayed with an r^2 value of 0.35, denoting a moderate level of LD correlation between adjacent variants. Lastly, at 300 kb, the LD decayed with an r^2 value of 0.2, indicating a diminished level of LD correlation between nearby variants. These outcomes correspond to a physical distance of approximately 200 Kb. Consequently, we performed QTL anchoring using a 200 Kb window, specifically focusing on the region extending 100 kb upstream and downstream of the MTA (Figure 2D). To discern the degree of SNP variation among the materials, a cluster analysis was executed through principal component analysis (PCA). The results indicated that the 156 materials could be classified into two subgroups based on PC1 (Figure 2E), constituting 9.99% of the total variance. This suggests that the 156 soybean materials can be delineated into two subgroups, reflecting a composite of two ancestral populations. The geographical origins of the 156 soybean landraces were the foreign region and domestic region.

Genome wide association analysis of four yield-related traits

To identify significant SNPs associated with target traits, we employed two Genome-Wide Association Study (GWAS) models, namely the General Linear Model (GLM) and the Mixed Linear Model (MLM), and FDR correction, for the analysis of high-quality

10.3389/fpls.2024.1375646

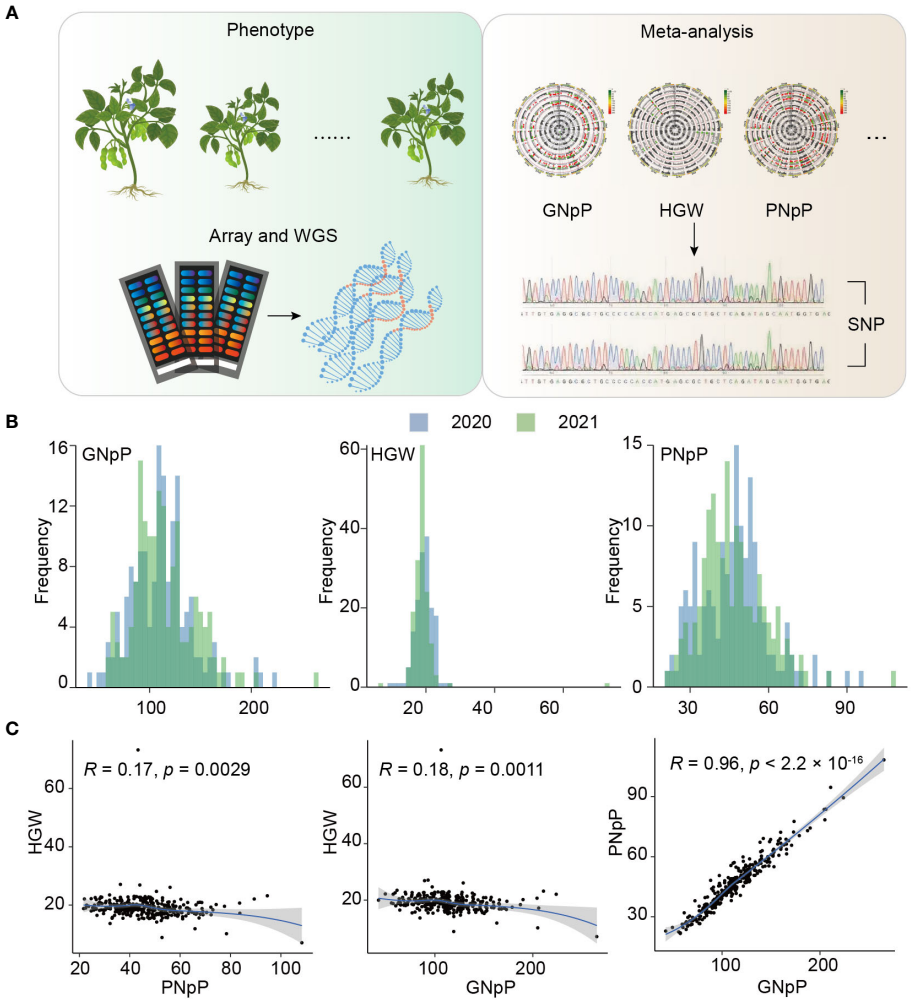


FIGURE 1
Schematic and statistics of the Experiment. **(A)** Workflow of genotyping by sequencing. **(B)** Frequency distribution of three yield-related traits over two years. **(C)** Phenotypic correlations between three traits GNpP, HGW and PNpP.

SNPs within a dataset comprising 156 soybean germplasms. The significant threshold value for the association between SNP and traits were determined by -log10 (P) >4.78, which is equivalent to P <0.5/3026, for GLM, MLM and FDR correction. Quantile–quantile plots are employed to evaluate whether the distribution of observed p-values from statistical tests deviates from the expected distribution. Deviations from the diagonal suggest potential deviations from the null hypothesis, indicating possible associations between genetic variants and the trait (Supplementary Figure 1). The Manhattan plots result showed that a few sets of significant SNPs associated with GNpP, HGW and PNpP was detected in 2020 by the two models. The data from 2021 found a few numbers of SNPs with false positives. Based on all the result, SNPs, that were significantly tested at least two times for each trait, were regarded as MTAs. Specifically, in the GLM, a total of 628 QTLs were identified, consisting of 205 QTLs related to GNpP, 206 QTLs related to HGW and 217 QTLs related to PNpP. In contrast, the MLM approach detected a total of 52 QTLs, with 15 QTLs related to GNpP, 9 QTLs related to HGW and 28 QTLs related to PNpP. The method of MLM eliminated lots of false positive SNPs by comparison. Among these QTLs, two or three QTLs can form a colocalization region, and thirteen colocalization regions in all were found. These colocalization regions distributed on 9 of 20 soybean chromosomes, and most of them associated with both PNpP and GNpP. Only the region on chromosome 6 associated with PNpP, GNpP and HGW (Figure 3).

Prediction of candidate genes

There are three QTLs named *qgGNpP-89*, *qgHGW-85*, and *qgPNpP-98*, associated with GNpP, HGW and PNpP individually, which are in a specific region spanning from 3030820 to 3424009 base pairs on chromosome 16 was identified by GLM. This region contained ten marker sites carrying different alleles, and showed a significant positive correlation between them. More notably, seven

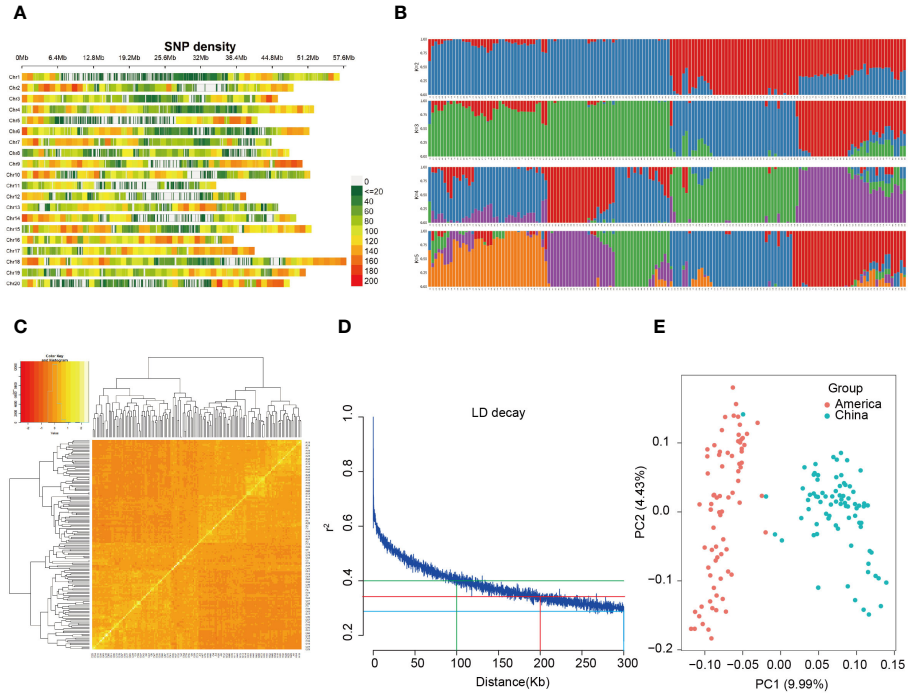


FIGURE 2
SNP distribution and population structure of 156 soybean landraces. **(A)** Distribution of 80,163 SNP markers across 20 soybean chromosomes for GWAS. A density distribution map of SNPs across the entire genome with bins of 1 Mb. **(B)** Population structure analysis of 156 materials. **(C)** Heatmap of the kinship matrix of the 156 soybean accessions. **(D)** Scatter plot of pairwise SNPs showing genome-wide linkage disequilibrium (LD) decay. The red curve line represents the smoothing spline regression model fitted to LD decay. The vertical red line indicates the genetic distance (200 Kb) at which the LD half-decay (r^2 = 0.35, the horizontal blue line) intersect with the LD decay curve. **(E)** Principal component analysis (PCA) of the population structure. Distribution of the accessions in the association panel under PC1 and PC2.

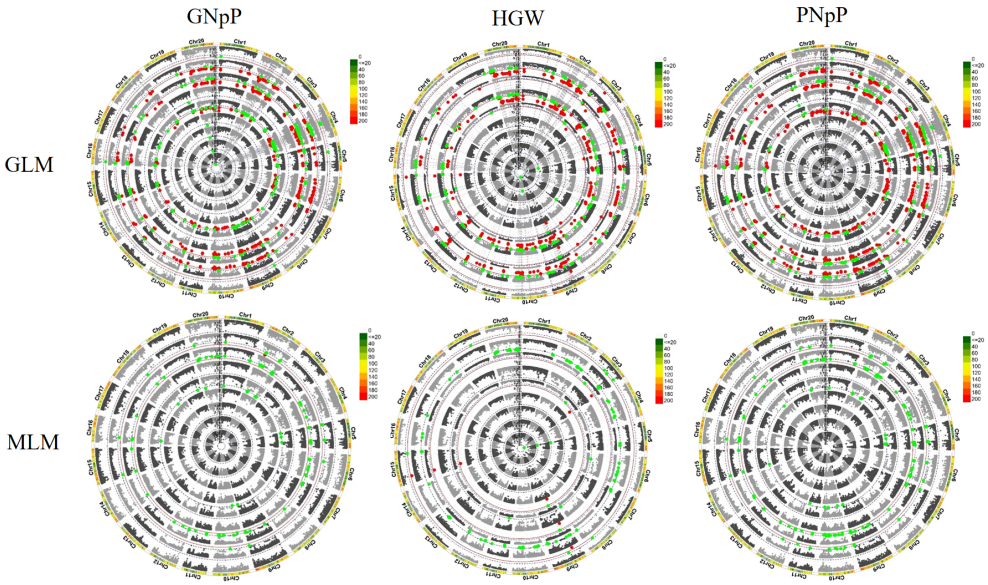


FIGURE 3
Manhattan plot displays the SNPs significantly associated with agronomic traits detected in GWAS analysis. The data consists of the testing results of population materials in 2020 and 2021, and analysis was conducted using the GLM and MLM modules. The soybean materials are subjected to three repeated experiments every year, and the average of the three repeated experiments each year is taken (denoted as "m"). The circular Manhattan plot represents the values from the center to the outer 1-9 circles as follows: 2020 (m), 2020 (1), 2020 (2), 2020 (3), 2021 (m), 2021 (1), 2021 (2), 2021 (3), and BLUE value of two years. Each colored dot represents a SNP.

10.3389/fpls.2024.1375646

TABLE 1 Traits co-location and associated markers.

Traits	QTLs	chromo-some	RegionS-tart	RegionE-nd	LOD	R^2	Markers
GNpP	*qgGNpP-89*	16	3010838	3424009	4.81 - 6.69	12.19 - 16.86	Gm16_3110838; Gm16_3130820; Gm16_3169888; Gm16_3195848; Gm16_3203580; Gm16_3260654; Gm16_3271292; Gm16_3283717; Gm16_3304635; Gm16_3324009
HGW	*qgHGW-85*	16	3030820	3424009	4.99 - 5.90	11.78 - 13.32	**Gm16_3130820**; Gm16_3169888; Gm16_3195848; **Gm16_3203580**; **Gm16_3260654**; **Gm16_3271292**; **Gm16_3283717**; **Gm16_3304635**; **Gm16_3324009**
PNpP	*qgPNpP-98*	16	3030820	3424009	4.91 - 6.82	12.05 - 17.06	Gm16_3130820; Gm16_3203580; Gm16_3260654; Gm16_3271292; Gm16_3283717; Gm16_3304635; Gm16_3324009

The seven marker loci in bold: Gm16_3130820, Gm16_3203580, Gm16_3260654, Gm16_3271292, Gm16_3283717, Gm16_3304635, and Gm16_3324009, were related to GNpP, HGW and PNpP at the same time.

of these marker loci were related to GNpP, HGW and PNpP at the same time (Table 1). We were particularly interested in the markers with large effects, such as marker Gm16_3130820 and Gm16_3195848 on chromosome 16. Compared with the alternative alleles, the GNpP and PNpP of the materials carrying the favorable allele (TT) at Gm16_3130820 was higher than the materials carrying the unfavorable allele (CC), the HGW of the materials carrying the favorable allele (CC) at Gm16_3130820 was higher than the materials carrying the unfavorable allele (TT). Otherwise, the GNpP and PNpP of the materials carrying the favorable allele (TT) at Gm16_3195848 was higher than the materials carrying the unfavorable allele (GG), the HGW of the materials carrying the favorable allele (GG) at Gm16_3195848 was higher than the materials carrying the unfavorable allele (TT) (Figure 4). It's also worth mentioning that three of these seven markers, GM16_3130820, GM16_3195848, and GM16_3260654, were found to be located on three candidate genes, *Glyma.16G033100*, *Glyma.16G034100*, and *Glyma.16G034600*, respectively. Candidate genes *Glyma.16G033100* is responsible for encoding the S-adenosylmethionine-binding subunit, which plays a crucial role in various biochemical processes. It associated with dividing tissues, particularly reproductive organs (Zhong et al., 2008) such as floral organ number and size (Bodi et al., 2012), tiller bud formation (Yu et al., 2021), and shoot meristems. It

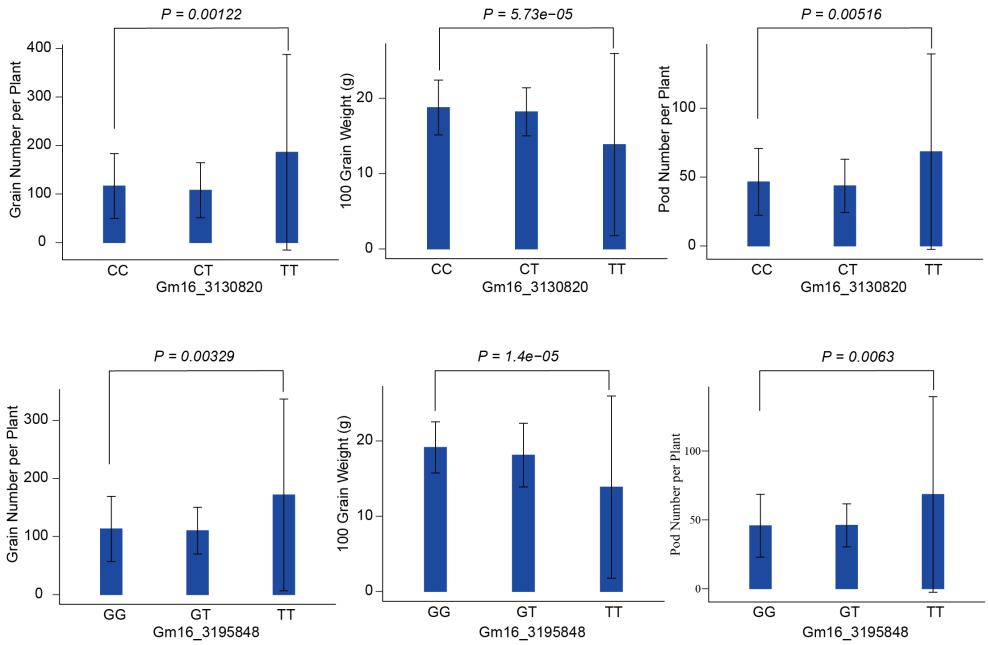


FIGURE 4
Phenotypic differences between accessions carrying different alleles. They are the allele effects for the marker Gm16_3130820 and Gm16_3195848 of GNpP, HGW and PNpP in soybean. GNpP means grain number per plant, HGW means 100 grain weight, and PNpP means pod number perplant.

10.3389/fpls.2024.1375646

TABLE 2 Candidate gene information.

Functional annotation	Gene ID	RegionStart	RegionEnd
protein kinase	*Glyma.16g032700*	3094789	3099121
glycosyl hydrolase family 28	*Glyma.16g033000*	3116063	3122058
S-adenosylmethionine-binding subunit	*Glyma.16g033100*	3127716	3133331
protein tyrosine kinase	*Glyma.16g034600*	3254425	3263715
ubiquitin-conjugating enzyme	*Glyma.16g035000*	3311992	3314979
cyclin-like domain	*Glyma.16g035400*	3348514	3362406
PPR repeat	*Glyma.16g031800*	3017265	3019791
zinc finger and C3HC4 type (ring finger)	*Glyma.16g034100*	3195056	3196022

also influenced seed development with protein and the starch synthesis-related pathway enriched in the later stages (Li et al., 2022). *Glyma.16G034100* codes for a zinc finger and C3HC4 type (ring finger), a protein domain known for its ability to bind zinc ions and regulate gene expression. In rice, research showed that *FRRP1* probably regulates flowering time and yield potential by affecting histone H2B monoubiquitylation (Du et al., 2016). The *AtYY1* gene is a negative regulator of the *Arabidopsis* ABA response network in *Arabidopsis* (Li et al., 2016). And these proteins also can regulate seed weight (Barmukh et al., 2021), tiller number (Zhou et al., 2017), and seed size and plant height in crop plants (Du et al., 2016). Lastly, *Glyma.16G034600* is associated with the protein tyrosine kinase, a protein plays a fundamental role in regulation of most cellular activities. They have been shown to impact the seed and yield of crops (*Arabidopsis*, rice, sunflower, potato and so on) through various pathways, such as regulation of carbon supply (Zheng et al., 2010), development of embryo (Thakur and Bhatla, 2014; Mazin et al., 2019), and cell proliferation (Zhang et al., 2020), BR signaling and control (Tian et al., 2021).

In the region of co-localization, a total of ten candidate genes were found, encoding five kinases (protein kinase, glycosyl hydrolase, m^6A methyl-transferase, protein tyrosine kinase and ubiquitin-conjugating enzyme), two domains (cyclin-like domain and zinc finger and ring finger) and one repeat (PPR repeat) Table 2. Most of the candidate genes associated with kinases have relatively long sequences, ranging from approximately 3000 to 10000 base pairs. The candidate gene sequences related to PPR repeat are shorter, falling within the range of 2000 to 3000 bps. On the other hand, the candidate gene sequences associated with the cyclin-like domain are relatively longer, approaching 14000 bps. Lastly, the candidate gene sequences related to the zinc finger protein have lengths less than 1000 bps. Otherwise, some of the candidate genes were closely related to each other, such as candidate genes *Glyma.16G033000* and *Glyma.16G033100* are 5658 bps apart, encoding glycosyl hydrolase family and S-adenosylmethionine-binding subunit, respectively (Supplementary Tables 3-8).

Discussion

The quest for increasing soybean yield is a paramount goal for breeders, and the use of GWAS has proven to be an effective method to uncover the genetic components related to soybean yield. Several studies have successfully identified various SNP loci, QTLs, and candidate genes (Hao et al., 2012; Liu et al., 2017; Jing et al., 2018; Hu et al., 2020; Mohsen et al., 2021; Chanditha et al., 2022) associated with soybean yield and its components. In the current study, the focus was on three essential yield-related traits: PNpP, GNpP, and HGW. Using GWAS, a substantial number of SNPs associated with these traits were identified within the soybean population. Most significantly, a co-localization interval on chromosome 16 was discovered, associating with all three traits. This interval contained eight candidate genes involved in crop growth and development, which had not been reported previously.

These candidate genes have diverse functions, such as transcriptional initiation of protein-coding genes (e.g., *Glyma.16g035400*), regulation of genes involved in DNA repair, cell cycle, and apoptosis (e.g., *Glyma.16g031800*), and modulation of alternative RNA splicing (e.g., *Glyma.16g033100*). Furthermore, two candidate genes (*Glyma.16g032700* and *Glyma.16g034600*) were associated with protein kinases, while another (*Glyma.16g033000*) was linked to cell wall metabolism. The ubiquitin-proteasome system (UPS) was also represented in the candidate genes, with *Glyma.16g035000* acting as a ubiquitin-conjugating enzyme (E2), and *Glyma.16g034100* likely functioning as E3 ubiquitin ligases. In summary, the study employed GWAS on both Chinese and foreign soybean varieties, revealing a co-localization locus associated with PNpP, GNpP, and HGW. Within this region, ten candidate genes and seven common markers were identified, with three of the markers mapping to three candidate genes. The proximity of these candidate genes related to yield traits underscores the importance of this locus. Further investigations into the expression patterns of these candidate genes at specific growth stages will help in improving soybean varieties and overcoming yield limitations.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: FigShare, https://figshare.com/articles/dataset/_b_Genome-wide_association_studies_b_b_reveal_novel_b_b_QTLs_b_b_for_b_b_agronomic_traits_b_b_in_soybean_b_/25690689.

Author contributions

DH: Writing – original draft, Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. XZ: Visualization, Supervision, Resources, Project administration, Methodology,

Conceptualization, Writing – review & editing, Writing – original draft, Validation, Software, Investigation, Formal analysis, Data curation. DZ: Writing – review & editing, Project administration. ZW: Writing – review & editing, Project administration. ZZ: Writing – review & editing, Project administration. HS: Writing – review & editing, Project administration. ZQ: Writing – review & editing, Project administration. LW: Writing – review & editing, Project administration. ZL: Writing – review & editing, Conceptualization. XZ: Writing – review & editing, Software, Formal analysis, Conceptualization. MY: Writing – review & editing, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Innovation Project of Heilongjiang Academy of Agricultural Sciences (CX23ZD03); Heilongjiang Chinese Academy of Sciences Qiuying Zhang soybean scientist studio; China Agriculture Research System of MOF and MARA (CARS-04); Evaluation, Innovation, and Excellent Gene Excavation for Elite Soybean Cultivars between China and Europe (2019YFE0105900).

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.

The reviewer JL declared a shared affiliation with the author ZL to the handling editor at the time of review.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1375646/full#supplementary-material

References

Barmukh, R., Soren, K. R., Madugula, P., Singh, N. P., Roorkiwal, M., Varshney, R. K., et al. (2021). Construction of a high-density genetic map and QTL analysis for yield, yield components and agronomic traits in chickpea (*Cicer arietinum* L.). *PloS One* 16, e0251669. doi: 10.1371/journal.pone.0251669

Bhat, J. A., Adeboye, K. A., Ganie, S. A., Barmukh, R., Varshney, R. K., Yu, D. Y., et al. (2022). Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.). *Front. Genet.* 13. doi: 10.3389/fgene.2022.953833

Bodi, Z., Zhong, S. L., Mehra, S., Song, J., Graham, N., Fray, R. G., et al. (2012). Adenosine methylation in *Arabidopsis* mRNA is associated with the 3' End and reduced levels cause developmental defects. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00048

Bradbury, P. J., Zhang, Z. W., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping ofcomplex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Chanditha, P., Davoud, T., and Istvan, R. (2022). Genome-wide association study of soybean germplasm derived from Canadian × Chinese crosses to mine for novel alleles to improve seed yield and seed quality traits. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.866300

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 4, 7. doi: 10.1186/s13742-015-0047-8

Dong, L. D., Fang, C., Weller, J. L., Lu, S. J., Kong, F. J., Liu, B. H., et al. (2021). Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* 12, 5445. doi: 10.1038/s41467-021-25800-3

Du, Y. W., He, W., Deng, C. W., Chen, X., Zhang, J. F., Tao, W., et al. (2016). Flowering-related RING Protein 1 (*FRRP1*) regulates flowering time and Yield Potential by affecting histone H2B monoubiquitination in rice (*Oryza sativa*). *PloS One* 11, e0150458. doi: 10.1371/journal.pone.0150458

Gao, J. S., Yang, S. X., Tang, K. Q., Li, G., Gao, X., Feng, X. Z., et al. (2021). GmCCD4 controls carotenoid content in soybeans. *Plant Biotechnol. J.* 19, 801–813. doi: 10.1111/pbi.13506

Hao, D., Cheng, H., Yin, Z., Cui, S., Zhang, D., Yu, D., et al. (2012). Identification of single nucleotide polymorphisms and haplotypes associated with yield and yi eld components in soybean (Glycine max) landraces across multiple environments. *Theor. Appl. Genet.* 124, 447–458. doi: 10.1007/s00122-011-1719-0

Happ, M. M., Graef, G. L., Wang, H., Howard, R., Posadas, L., and Hyten, D. L. (2021). Comparing a mixed model approach to traditional stability estimators for mapping genotype by environment interactions and yield stability in soybean [*Glycine max* (L.) Merr.]. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.630175

Hu, D., Zhang, H., Du, Q., Hu, Z., Yang, Z., Kan, G., et al. (2020). Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (Glycine soja Sieb. and Zucc.). *Planta.* 251, 39. doi: 10.1007/s00425-019-03329-6

Jhon, A. B.-C., María, D. S. C-L., and Roxana, Y. (2023). Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1007104

Jing, Y., Zhao, X., Wang, J., Qiu, L., Han, Y., Li, W., et al. (2018). Identification of the genomic region underlying seed weight per plant in soybean (*Glycine max* L. Merr.) via high-throughput single-nucleotide polymorphisms and a genome-wide association study. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01392

Lamichhane, J. R., Aubertot, J.-N., Champolivier, L., Debaeke, P., and Maury, P. (2020). Combining experimental and modeling approaches to understand genotype x sowing date x environment interaction effects on emergence rates and grain yield of soybean. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.558855

Li, T., Wu, X. Y., Li, H., Song, J. H., and Liu, J. Y. (2016). A dual-function transcription factor, *AtYY1*, is a novel negative regulator of the *Arabidopsis* ABA response network. *Mol. Plant* 9, 650–661. doi: 10.1016/j.molp.2016.02.010

Li, W. X., Yu, Y., Chen, X. R., Fang, Q., Yang, A. Q., Chen, X. Y., et al. (2022). N6-Methyladenosine dynamic changes and differential methylation in wheat grain development. *Planta.* 255, 125. doi: 10.1007/s00425-022-03893-4

Liang, Q. J., Chen, L. Y., Yang, X., Yang, H., Kong, F. J., Tian, Z. X., et al. (2022). Natural variation of Dt2 determines branching in soybean. *Nat. Commun.* 13, 6429. doi: 10.1038/s41467-022-34153-4

Liu, Z., Li, H., Fan, X., Huang, W., Wang, S., Qiu, L., et al. (2017). Phenotypic characterization and genetic dissection of nine agronomic traits in Tokachi nagaha and its derived cultivars in soybean (Glycine max (L.) Merr.). *Plant Sci.* 256, 72–86. doi: 10.1016/j.plantsci.2016.11.009

Liu, S. L., Zhang, M., Feng, F., and Tian, Z. X. (2020). Toward a "green revolution" for soybean. *Mol. Plant* 13, 688–697. doi: 10.1016/j.molp.2020.03.002

Luo, Y., Zhang, M. L., Liu, Y., Liu, J., Li, W. Q., Yan, J. B., et al. (2022). Genetic variation in YIGE1 contributes to ear length and grain yield in maize. *New Phytol.* 234, 513–526. doi: 10.1111/nph.17882

Mazin, B. D., Joly, V., and Matton, D. P. (2019). The ScFRK2 mitogen-activated protein kinase kinase kinase (MAP3K) is involved in early embryo sac development in Solanum chacoense. *Plant Signal Behav.* 14, 1620059. doi: 10.1080/15592324.2019.1620059

Mohsen, Y.-N., Sepideh, T., Dan, T., Istvan, R., and Milad, E. (2021). Genome-wide association studies of soybean yield-related hyperspectral reflectance bands using machine learning-mediated data integration methods. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.777028

Ning, H. L., Yuan, J. Q., Dong, Q. Z., Li, W. B., Xue, H., Li, W. X., et al. (2018). Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [Glycine max (L.) Merri]. *PloS One* 13, e0195830. doi: 10.1371/journal.pone.0195830

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics.* 197, 573–589. doi: 10.1534/genetics.114.164350

Raju, S. K. K., Shao, M. R., Sanchez, R., Xu, Y. Z., Sandhu, A., Mackenzie, S., et al. (2018). An epigenetic breeding system in soybean for increased yield and stability. *Plant Biotechnol. J.* 16, 1836–1847. doi: 10.1111/pbi.12919

Saminadane, T., Geddam, S., Krishnaswamy, P., Jothiganapathy, K., Tamilselvan, A., Ramadoss, B. R., et al. (2024). Development of early maturing salt-tolerant rice variety KKL(R) 3 using a combination of conventional and molecular breeding approaches. *Front. Genet.* 2. doi: 10.3389/fgene.2023.1332691

Sun, M., Li, Y., Zheng, J., Wu, D., Zhao, X., Li, Y. G., et al. (2022a). A nuclear factor Y-B transcription factor, *GmNFYB17*, regulates resistance to drought stress in soybean. *Int. J. Mol. Sci.* 23, 7242. doi: 10.3390/ijms23137242

Sun, M., Na, C., Jing, Y., Cui, Z., Zhao, X., Han, Y., et al. (2022b). Genome-wide association analysis and gene mining of resistance to China race 1 of frogeye leaf spot in soybean. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.867713

Sun, R. J., Sun, B. C., Tian, Y., Su, S. S., Li, Y. H., and Qiu, L. J. (2022c). Dissection of the practical soybean breeding pipeline by developing ZDX1, a high-throughput functional array. *Theor. Appl. Genet.* 135, 1413–1427. doi: 10.1007/s00122-022-04043-w

Thakur, A., and Bhatla, S. C. (2014). A probable crosstalk between Ca^{+2}, reactive oxygen species accumulation and scavenging mechanisms and modulation of protein kinase C activity during seed development in sunflower. *Plant Signal Behav.* 9, e27900. doi: 10.4161/psb.27900

Tian, X. J., He, M. L., Mei, E. Y., Zhang, B. W., Tang, J. Q., Bu, Q. Y., et al. (2021). WRKY53 integrates classic brassinosteroid signaling and the mitogen-activated protein kinase pathway to regulate rice architecture and seed size. *Plant Cell.* 33, 2753–2775. doi: 10.1093/plcell/koab137

Torkamaneh, D., Chalifour, F.-P., Beauchamp, C. J., Agrama, H., Boahen, S., Belzile, F., et al. (2020). Genome-wide association analyses reveal the genetic basis of biomass accumulation under symbiotic nitrogen fxation in African soybean. *Theor. Appl. Genet.* 133, 665–676. doi: 10.1007/s00122-019-03499-7

Wang, X. L., Wang, H. W., Liu, S. X., Ferjani, A., Yang, X. H., Qin, F., et al. (2016). Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48, 1233–1241. doi: 10.1038/ng.3636

Wang, S., Yang, X., Xu, M., Lin, X., Lin, T., Huang, S., et al. (2015). A rare SNP identified a TCP transcription factor essential for tendril development in cucumber. *Mol. Plant* 8, 1795–1808. doi: 10.1016/j.molp.2015.10.005

Wang, Y., Zhou, Q., Zhu, G., Wang, S., Zhang, Z., Gu, X., et al. (2018). Genetic analysis and identification of a candidate gene associated with in *vitro* regeneration ability of cucumber. *Theor. Appl. Genet.* 131, 2663–2675. doi: 10.1007/s00122-018-3182-7

Wyman, E. N., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313

Xia, H., Zhang, Z., Luo, C., Yang, J. B., Guan, Y. L., David, C. L., et al. (2023). Multiprime: a reliable and efficient tool for targeted next-generation sequencing. *iMeta.* 1, e1. doi: 10.1002/imt2.143

Xu, G. W., Cheng, Y. J., Wang, X. Q., Dai, Z. G., Xie, D. W., Sun, J., et al. (2024). Identification of single nucleotide polymorphic loci and candidate genes for seed germination percentage in okra under salt and no-salt stresses by genome-wide association study. *Plants (Basel).* 13, 588. doi: 10.3390/plants13050588

Yan, X. F., Zhao, L., Ren, Y., Dong, Z. D., Cui, D. Q., and Chen, F. (2019). Genome-wide association study revealed that the TaGW8 gene was associated with kernel size in Chinese bread wheat. *Sci. Rep.* 9, 2702. doi: 10.1038/s41598-019-38570-2

Yu, J., Pressoir, G., Briggs, W. H., Bi, I., Vroh,, and Yamasaki, M. (2006). A unifified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Yu, Q., Liu, S., Yu, L., Song, B. A., He, C., Jia, G. F., et al. (2021). RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* 39, 1581–1588. doi: 10.1038/s41587-021-00982-9

Zhang, J., Song, Q., Cregan, P. B., and Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycinemax*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x

Zhang, W., Xu, W. J., Zhang, H. M., Liu, X. Q., and Chen, H. T. (2021). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi: 10.1007/s00122-021-03774-6

Zhang, Y., He, J., Wang, Y., Xing, G., Zhao, J., Li, Y., et al. (2015). Establishment of a 100-seed weight quantitative trait locus-allele matrix of the germplasm population for optimal recombination design in soybean breeding programmes. *J. Exp. Bot.* 66, 6311–6325. doi: 10.1093/jxb/erv342

Zhang, H. M., Zhang, G. W., Zhang, W., Wang, Q., Xu, W. J., Chen, H. T., et al. (2022). Identification of loci governing soybean seed protein content via genome-wide association study and selective signature analyses. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1045953

Zhang, Y. Y., Yao, W. J. S., Wang, F., Su, Y. H., Zhang, D. J., Zhang, X. S., et al. (2020). AGC protein kinase AGC1-4 mediates seed size in *Arabidopsis*. *Plant Cell Rep.* 39, 825–837. doi: 10.1007/s00299-020-02533-z

Zhao, K. Y., Tung, C.-W., Eizenga, G. C., Wright, M. H., Bustamante, C. D., McCouch, S. R., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2, 467. doi: 10.1038/ncomms1467

Zhao, Q. S., Shi, X. L., Yan, L., Yang, C. Y., Zhang, M. C., Yang, Y. Q., et al. (2021). Characterization of the common genetic basis underlying seed hilum size, yield, and quality traits in soybean. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.610214

Zhao, L., Zheng, Y. T., Wang, Y., Wang, S. S., Wang, T. Z., Chen, F., et al. (2023). A HST1-like gene controls tiller angle through regulating endogenous auxin in common wheat. *Plant Biotechnol. J.* 21, 122–135. doi: 10.1111/pbi.13930

Zheng, Z., Xu, X., Crosley, R. A., Greenwalt, S. A., Sun, Y., Blakeslee, B., et al. (2010). The protein kinase SnRK2.6 mediates the regulation of sucrose metabolism and plant growth in *Arabidopsis*. *Plant Physiol.* 153, 99–113. doi: 10.1104/pp.109.150789

Zhong, S. L., Li, H. Y., Bodi, Z., Button, J., Vespa, L., Fray, R. G., et al. (2008). MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell.* 20, 1278–1288. doi: 10.1105/tpc.108.058883

Zhou, B., Lin, J. Z., Peng, D., Yang, Y. Z., Guo, M., Tang, D. Y., et al. (2017). Plant architecture and grain yield are regulated by the novel DHHC-type zinc finger protein genes in rice (*Oryza sativa* L.). *Plant Sci.* 254, 12–21. doi: 10.1016/j.plantsci.2016.08.015

Zhou, H., Xia, D., Li, P. B., Ao, Y. T., Xu, X. D., He, Y. Q., et al. (2021). Genetic architecture and key genes controlling the diversity of oil composition in rice grains. *Mol. Plant* 14, 456–469. doi: 10.1016/j.molp.2020.12.001

Zhu, Y., Zhu, G., Xu, R., Jiao, Z., Chong, L., Zhu, J., et al. (2023). A natural promoter variation of *SlBBX31* confers enhanced cold tolerance during tomato domestication. *Plant Biotechnol. J.* 21, 1033–1043. doi: 10.1111/pbi.14016

Zong, Y., Yao, S., Crawford, G. W., Fang, H., Chen, X., Jiang, H., et al. (2017). Selection for oil content during soybean domestication revealed by X-Ray tomography of ancient beans. *Sci. Rep.* 7, 43595. doi: 10.1038/srep43595

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/fpls.2024.1412953

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Yiyong Zhao,
Harvard University, United States
Sang Shi Fei,
Henan Normal University, China
Junbo Yang,
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Wenqiu Pan
wenqiu_pan@nwafu.edu.cn
Yihan Li
liyihan@jxau.edu.cn

†These authors have contributed equally to this work

RECEIVED 06 April 2024
ACCEPTED 08 May 2024
PUBLISHED 22 May 2024

CITATION
Li T, Cai S, Cai Z, Fu Y, Liu W, Zhu X, Lai C, Cui L, Pan W and Li Y (2024) TriticeaeSSRdb: a comprehensive database of simple sequence repeats in *Triticeae*. *Front. Plant Sci.* 15:1412953. doi: 10.3389/fpls.2024.1412953

TriticeaeSSRdb: a comprehensive database of simple sequence repeats in *Triticeae*

Tingting Li[1,2†], Shaoshuai Cai[1†], Zhibo Cai[2], Yi Fu[1], Wenqiang Liu[1], Xiangdong Zhu[1], Chongde Lai[1,3], Licao Cui[1], Wenqiu Pan[2*] and Yihan Li[1*]

[1]College of Bioscience and Engineering, Jiangxi Agricultural University, Nanchang, Jiangxi, China, [2]State Key Laboratory for Crop Stress Resistance and High-Efficiency Production, Northwest A&F University, Yangling, Shaanxi, China, [3]The Public Instrument Platform of Jiangxi Agricultural University, Jiangxi Agricultural University, Nanchang, Jiangxi, China

Microsatellites, known as simple sequence repeats (SSRs), are short tandem repeats of 1 to 6 nucleotide motifs found in all genomes, particularly eukaryotes. They are widely used as co-dominant markers in genetic analyses and molecular breeding. *Triticeae*, a tribe of grasses, includes major cereal crops such as bread wheat, barley, and rye, as well as abundant forage and lawn grasses, playing a crucial role in global food production and agriculture. To enhance genetic work and expedite the improvement of *Triticeae* crops, we have developed TriticeaeSSRdb, an integrated and user-friendly database. It contains 3,891,705 SSRs from 21 species and offers browsing options based on genomic regions, chromosomes, motif types, and repeat motif sequences. Advanced search functions allow personalized searches based on chromosome location and length of SSR. Users can also explore the genes associated with SSRs, design customized primer pairs for PCR validation, and utilize practical tools for whole-genome browsing, sequence alignment, and *in silico* SSR prediction from local sequences. We continually update TriticeaeSSRdb with additional species and practical utilities. We anticipate that this database will greatly facilitate trait genetic analyses and enhance molecular breeding strategies for *Triticeae* crops. Researchers can freely access the database at http://triticeaessrdb.com/.

KEYWORDS

genome, microsatellite, SSR, database, molecular breeding

1 Introduction

The *Triticeae* tribe, a member of the *Poaceae* grass family's subfamily *Pooideae*, encompasses a total of 503 diploid and polyploid species distributed among 27 genera (Soreng et al., 2015). This tribe includes economically significant cereal crops, such as bread wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and rye (*Secale cereale*), along with

their wild relatives (Sato et al., 2021). Since the Neolithic Agricultural Revolution, these crops have played essential roles in human sustenance and global food security (Li et al., 2023).

Recent advancements in novel genome assembly algorithms and strategies have progressively improved the quality of genome assemblies in *Triticeae*, resulting in a rapid increase in the availability of assemblies (Adamski et al., 2020). These high-quality assemblies enable the decoding of chromosome and gene evolutionary history through clade-specific comparative genomic studies (Chen et al., 2020). Deciphering the genomes of *Triticeae* species has revealed their characteristic features of large, complex, and highly repetitive genomes. For instance, diploid *Triticeae* species like *H. vulgare* and *S. cereale* have genome sizes ranging from 4.3 Gb to 7.9 Gb. The tetraploid *T. dicoccoides* genome comprises 10.5 Gb of genomic sequence, while the hexaploid *T. aestivum* Chinese Spring genome spans 14.5 Gb, with approximately 80% of these genomes consisting of repetitive elements (Wang et al., 2022). In addition to the assembly of high-quality reference genomes, recent *Triticeae* pan-genome studies have commenced, capturing the genomic diversities of these species and enhancing our understanding of the genetic basis of domestication and environmental adaptation in *Triticeae* crops (Gao et al., 2023). However, these genomic complexities present challenges for genetic studies and the identification of agronomic traits in *Triticeae* crops.

The majority of *Triticeae* breeding programs heavily rely on conventional breeding selection, which involves replicated and time-consuming field trials. However, Marker-assisted selection (MAS) has emerged as an increasingly popular and valuable tool for tracking alleles associated with key agricultural traits (Hasan et al., 2021). Recent advancements in the availability of reference sequences and high-throughput genotyping platforms for major *Triticeae* species have facilitated comprehensive genome analysis. This has provided valuable insights into marker development and laid the foundation for identifying DNA markers closely linked to target traits. In various crops, microsatellites, or simple sequence repeat (SSRs) have become the preferred marker type for MAS, further enhancing its effectiveness and application (Duhan and Kaundal, 2021).

SSRs are short nucleotide motifs that are tandemly repeated and range in length from 1 to 6 base pairs (Tóth et al., 2000). The high polymorphism of microsatellites arises from the variation in the number of repeat units among different alleles. SSR markers offer several advantages, including easy detection through polymerase chain reaction (PCR), unambiguous scoring, and high reproducibility (Powell et al., 1996). Consequently, they have been extensively employed in genetic studies, encompassing genetic diversity assessment, phylogenetic analysis, genetic linkage mapping, and quantitative trait locus (QTL) mapping (Kalia et al., 2011). The identification and development of SSR markers at a genome-wide scale have been greatly facilitated by the availability of reference sequences and high-throughput genotyping platforms. Microsatellites serve significant roles in genome organization and are widely recognized as popular neutral genetic markers (Schlötterer, 2000). The characterization and application of SSR markers hold practical implications for breeding programs, providing valuable assistance to breeders in genotype association and selection processes. Despite the impact of the rise of high-throughput sequencing technologies on traditional molecular markers, SSR technology, which relies on PCR systems, can be conducted in any molecular biology laboratory, and has an absolute advantage, particularly in the rapid preliminary screening of materials. In contrast, high-throughput sequencing technologies are limited by the cost of instruments and are not widely applicable in laboratories. As a result, successful whole-genome identification and development of SSR markers have been achieved in various species (Deng et al., 2016; Guan et al., 2019; Lu et al., 2019). To facilitate data sharing, user-friendly and accessible databases have been established, such as LegumeSSRdb (Duhan and Kaundal, 2021), citSATdb (Duhan et al., 2020), PSSRD (Song et al., 2021), and MSDB (Avvaru et al., 2017, 2020).

Given the various applications of SSRs in genetic and genomic research, such as genetic mapping, diversity analysis, marker-assisted selection, and population genetics, it is evident that there is a lack of a comprehensive database specifically catering to *Triticeae* species. Therefore, in addition to the existing databases, the development of a dedicated database that focuses on the genomic data of the *Triticeae* tribe and incorporates user-friendly features would be a valuable genomic resource for improving and characterizing *Triticeae* crops. In this study, we introduce TriticeaeSSRdb, an integrated web resource that offers a wide range of services, including JBrowse visualization, Basic Local Alignment Search Tool (BLAST), *in silico* SSR prediction, primer design for genotyping, and other enriching features. The species included in our database have undergone complete genome sequencing, ensuring a more comprehensive representation of SSR information. With the establishment of TriticeaeSSRdb, we aim to not only provide comprehensive and representative SSR information but also enable large-scale systematic and comparative analyses of SSRs in *Triticeae* species.

2 Materials and methods

2.1 Data collection

A total of 21 species from the *Triticeae* tribe were included in our study (Supplementary Table S1). Sequence information was downloaded in FASTA format, and the genomic annotations were obtained in GFF or GTF format. If multiple versions were available for the same species, we selected the most recent or favored assembly for download. The genomic information of six species, namely *Aegilops tauschii*, *Hordeum vulgare*, *T. aestivum*, *T. dicoccoides*, *T. spelta*, and *T durum* was obtained from the Ensembl Plants database (https://plants.ensembl.org/). The genomes of nine species, including *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, *Ae. sharonensis*, *Ae. speltoides*, *Ae. umbellulata*, *T. aestivum* ssp. *tibetanum* Shao, *T. monococcum*, and *T. urartu*, were downloaded from the National Center for Biotechnology Information (NCBI) Database (https://www.ncbi.nlm.nih.gov/). *Dasypyrum villosum*, *H. marinum*, *Leymus chinensis*, and *Thinopyrum elongatum* (4 species) were collected from the China National Center for Bioinformation - National Genomics Data Center (CNCB-NGDC) (https://ngdc.cncb.ac.cn/). *Hordeum vulgare* var. *nudum* was obtained from the WheatOmics database (http://wheatomics.sdau.edu.cn/jbrowse.html). *H. spontaneum* was retrieved from the IPK database (https://barley-pangenome.ipk-gatersleben.de/).

2.2 SSR identification and primer design

The genome sequences of all 21 species were analyzed using Pytrf v1.3.0, a Python package specifically designed for detecting both exact or perfect SSRs in genomic sequences (https://github.com/lmdu/pytrf/blob/master/docs/index.rst) (Du et al., 2018). Microsatellites were identified based on specific criteria for six different categories: mono-nucleotide repeats (MNRs) with a minimum repeat count of 15, di-nucleotide repeats (DNRs) with a minimum repeat count of 10, tri-nucleotide (TNRs) repeats with a minimum repeat count of 8, and tetra- (TtNRs), penta- (PNRs), and hexa-nucleotide repeats (HNRs) with a minimum repeat count of 5. To facilitate primer design, an in-house Python script was developed to exclude a total of 500 bp flanking sequences upstream (250 bp) and downstream (250 bp) of the identified SSR regions. The forward and reverse primers flanking the microsatellite repeat motifs were designed in batches using the primer3_core program v2.6.1 (Untergasser et al., 2012). Two Perl scripts, p3_in.pl and p3_out.pl, were implemented to enable seamless data interchange between Pytrf and the primer modeling software Primer v3.0. The primer design parameters were set as follows: primer size ranging from 20 to 22 base pairs, GC content between 40% and 60%, melting temperature between 55°C and 65°C, and expected product size ranging from 100 to 500 base pairs.

2.3 Functional annotation of SSR markers

To determine the precise genomic locations of the identified SSRs, we developed a custom Python script. Using the available genomic annotation file, we categorized the SSRs into intergenic and genic regions. The genic regions were further subdivided into intron, untranslated region (UTR), and coding sequence (CDS). The genomic distribution of SSRs was visualized using the RIdeogram v0.2.2 R package (Hao et al., 2020). To facilitate data visualization, we integrated comprehensive gene function annotations for all 21 *Triticeae* species into the Jbrowse genome browser v1.16.11 within the TriticeaeSSRdb database (Buels et al., 2016). This genome browser allows for the visualization of markers in relation to the reference sequence, gene coordinates, and detailed structural and functional information. This integration enhances the accessibility and understanding of the SSR data within the context of gene features and functions. Gene ontology (GO) enrichment analyses were conducted using KOBAS v3.0 (http://bioinfo.org/kobas).

2.4 Database implementation and web interface

The main website of the database is hosted on a Tencent Cloud server with specifications of an 8-core CPU, 64GB RAM, 500GB storage, and a 10Mbps bandwidth (https://cloud.tencent.com/). The domain name (http://triticeaessrdb.com/) was registered and linked to the server's IP address. The server operates on the Linux CentOS v7.9 operating system (http://www.centos.org). To facilitate platform development and web page creation, a Nginx v1.25.0 reverse proxy server architecture was configured.

The front-end web interface was developed using HTML5 (https://www.w3.org/html/), JavaScript (https://www.javascript.com/), CSS3 (http://www.w3.org), and Bootstrap5 (https://getbootstrap.com/). On the server side, the back-end was implemented using PHP (https://www.php.net/). MySQL v5.7.22 (https://www.mysql.com/) was utilized for storing, maintaining, and operating the microsatellite information. Custom PHP code was written to facilitate data retrieval from MySQL, which was then transferred to the front end for display. Data visualization charts were implemented using the Chart.js (https://www.chartjs.org/) component. Real-time primer designing functionality was achieved by integrating Primer3 into TriticeaeSSRdb. Furthermore, we employed a back-end Perl script called Microsatellite Identification Tool (MISA) (Thiel et al., 2003) and a Python script named Pytrf to facilitate the *in silico* prediction of microsatellites within uploaded sequences. For sequence homology searches, the BLAST was implemented using ViroBlast, a standalone BLAST web server. To enable online BLAST searches, ViroBLAST constructed a standalone database specifically for the 21 *Triticeae* species (Deng et al., 2007). Local scripts were also modified to enhance search services. The whole workflow of TriticeaeSSRdb is illustrated in Figure 1.

3 Results

3.1 SSR abundance and distribution in *Triticeae* species

In the current study, we characterized and compared the distribution of SSRs in 21 *Triticeae* genomes. A total of 3,891,705 microsatellites were identified, with an overall frequency of 27.21 SSRs/Mb (Supplementary Table S1). Considering their hexaploid genome characteristics, *T. aestivum* ssp. *tibetanu* Shao (321,257), *T. spelta* (314,128), and *T. aestivum* (269,000) exhibited a relatively higher total number of SSRs. At the diploid level, *H. vulgare* had the highest number of SSRs (304,626), while *Ae. tauschii* (101,264) and *H. marinum* (71,983) had the fewest SSRs. The maximum microsatellite density was observed in species *H. vulgare* (72.09 SSRs/Mb), whereas the minimum was observed in species *T. aestivum* (18.49 SSRs/Mb). Generally, species with larger genomes exhibited lower SSR density. Previous studies have reported a significantly negative correlation between SSR frequency and genome sizes in *Poaceae* species (Deng et al., 2016). However, our analysis of the 21 *Triticeae* species did not yield similar results. Although differences in genome size may contribute to the level of microsatellite repetition, we found no evidence of a relationship between SSR density and genome size (Portis et al., 2018).

3.2 Characterization of SSR motifs

The predicted SSR loci were classified into six categories: MNR, DNR, TNR, TtNR, PNR, and HNR. Across all species, MNR, DNR,

10.3389/fpls.2024.1412953

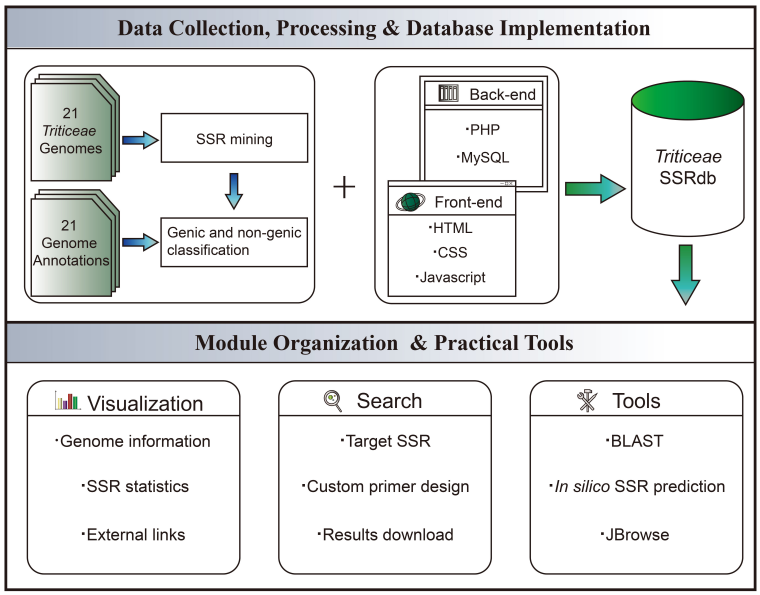


FIGURE 1
Schematic Workflow of TriticeaeSSRdb.

and TNR accounted for 90.06% of the total SSR. DNR was the most abundant category (49.53%), followed by TNR and MNR (Supplementary Table S2). Among the mono-nucleotide repeats, *Th. elongatum* had the highest percentage, while *T. dicoccoides* and *T. durum* had the lowest. Dimeric repeats had the highest occurrence in most species. Regarding trimeric repeats, three species from the *Hordeum* genus (*H. vulgare* var. *nudum*, *H. vulgare*, and *H. spontaneum*) and *Th. elongatum* exhibited higher percentages compared to other *Triticeae* species, suggesting their distinct evolutionary history. In contrast, SSRs with longer repeat motifs demonstrated significantly lower frequency, with average ratios of 7.71%, 0.85%, and 1.38% for tetrameric, pentameric, and hexameric repeats, respectively.

The distribution of microsatellites in *Triticeae* species was examined in terms of repeat unit length. Our analysis revealed a negative correlation between repeat length and frequency, indicating that longer repeats were less common (Supplementary Figure S1; Supplementary Table S3). Moreover, the abundance of SSRs decreased as the repeat number increased. It is worth noting that the decline rate of mono-, di-, tri-, and tetra-nucleotide repeats was noticeable, while the decline trend of penta- and hexa-nucleotide repeats was slightly less steep.

3.3 Analysis of SSR motif sequence composition

A total of 3424 distinct SSR motifs were identified, encompassing all possible base combinations of mono-nucleotide (4 types), di-nucleotide (12 types), tri-nucleotide (60 types), and tetra-nucleotide (240 types) repeats. Additionally, variations of penta- and hexa-nucleotide repeats contributed 955 and 2153 additional motifs, respectively (Supplementary Table S4). We further conducted a comprehensive analysis of individual repetitive motifs for each type of SSR (Supplementary Table S5). Considering the relatively low GC content observed in *Triticeae* species, the base composition of repeat motifs exhibited a notable bias towards A and T nucleotides. Remarkably, the motif (AT)n ranked among the top five most frequently occurring motifs in the 20 *Triticeae* genomes. Notably, AT repeats emerged as the most frequent motif overall, accounting for 10.11% of all SSRs in *Triticeae*. In contrast, GC repeats were exceptionally rare, with (G)n accounting for 0.157%-18.079%, (C)n accounting for 0.158%-18.245%, (GC)n accounting for 0.003%-0.050%, and (CG)n accounting for 0.001%-0.030%. The most prevalent motifs in *Triticeae* species encompassed mono- to hexanucleotide repeats, including C (40.63% of mono-nucleotide repeats), AT (20.40% of di-nucleotide repeats), AAG (7.44% of tri-nucleotide repeats), TTAA (10.14% of tetra-nucleotide repeats), CCTCT (3.54% of penta-nucleotide repeats), and TTTTCA (6.17% of hexa-nucleotide repeats).

3.4 Functional annotation of SSRs

The identified SSR loci were categorized based on their distribution across the pseudomolecules. The distribution of SSR loci was found to be relatively uneven among the different chromosomes (Supplementary Table S6). In *H. vulgare*, for example, the density of SSRs varied from 90.91 SSRs/Mb (chr4H) to 53.42 SSRs/Mb (chr5H), with an average density of 68.99 SSRs/Mb. Among the chromosomes, the longest linkage group, chr2H in *H. vulgare*, contained a total of 39,071 SSRs, while the fourth longest chromosome, chr4H, had the highest number of SSRs (55,487).

Furthermore, the analysis of chromosome location revealed a tendency for the identified SSRs to be located on the distal chromosome arm rather than in the pericentromeric region (Supplementary Figure S2). This may be attributed to the fact that the terminal regions of chromosomes are often gene-rich and contain other functional elements, which are relatively stable and subject to less selective pressure and recombination, thus providing more opportunities for the accumulation of microsatellite repeat sequences.

Based on the corresponding annotation file, we successfully categorized the identified SSRs as genic and intergenic SSRs. Genic SSRs were further classified into intron, UTR, and CDS regions. The majority of SSRs were found in the intergenic region, ranging from 83.8% (*H. spontaneum*) to 58.5% (*Ae. tauschii*), with an average ratio of 69.9% (Supplementary Table S7). Noncoding regions exhibited a higher frequency of SSRs compared to CDS, possibly due to negative selection against frameshift mutations in coding regions. Genic SSR markers, in contrast to noncoding microsatellites, are highly portable among related species, making them useful as anchor markers for comparative genetics purposes (Varshney et al., 2005). In terms of the genic region, 19.66% and 8.56% of SSRs were located in the intron and CDS regions, respectively. It is important to note that the available released genomes employ different annotation workflows, and a proportion of genomes lack UTR annotations. With the exception of *Ae. tauschii* and *T. dicoccoides*, where 10.5% and 7.2% of SSRs were predicted to be in the UTR, respectively, the remaining genomes showed an annotation of 1.1% to 5.3% of SSRs in this region.

Genic SSRs serve as valuable functional markers due to their involvement in gene expression and function in plants (Li et al., 2004). To investigate the potential biological functions of SSR-associated genes, we conducted a GO enrichment analysis using *H. vulgare* as a model. The genic SSRs showed significant enrichment in 232 GO terms (corrected p-value < 0.05), encompassing 127 biological processes, 37 cellular components, and 68 molecular functions (Supplementary Table S8). The top three enriched terms in the molecular function category were protein binding (GO:0005515), kinase activity (GO:0016301), and protein serine/threonine kinase activity (GO:0004674), while the most common terms in the cellular components category were plasma membrane (GO:0005886), cytosol (GO:0005829), and nucleus (GO:0005634). These genes were found to be implicated in various biological functions (e.g., GO:0006468, protein phosphorylation; GO:0006952, defense response; and GO:0042742, defense response to bacterium), and demonstrated involvement in response to diverse stresses (e.g., GO:0006979, response to oxidative stress; GO:0009651, response to salt stress; and GO:0071472, cellular response to salt stress) within the biological process category. These results indicate the extensive involvement of SSR tightly linked genes in diverse signal transduction pathways and their ability to adapt to various stresses. Therefore, the development of new SSR markers can contribute to the improvement of genetic analysis, QTL mapping, and molecular breeding in *Triticeae* crops.

3.5 Database implementation and practical tools

To facilitate research on *Triticeae* crop breeding and improvement, we developed the *Triticeae* SSR database (TriticeaeSSRdb) accessible at http://triticeaessrdb.com/. This user-friendly web platform offers a range of features and charts for data visualization, browsing, and downloading. The web interface comprises five primary sections: Home, Species, Tools, Help, and Contact (Figure 2). Below, we provide a detailed description and instructions for utilizing the interactive pages within TriticeaeSSRdb.

3.5.1 Home

The homepage of TriticeaeSSRdb provides an overview of the database and offers convenient access to plots and microsatellite data for *Triticeae* species. The homepage consists of two main sections. The first section displays a gallery of images containing SSR information for the 21 *Triticeae* species. Users can search for species using either their scientific or common names through the search bar or filter species of interest using the species panel. This comprehensive collection includes SSR details for all species, and the images are interactive, allowing users to click on specific species to access the second-level website interface dedicated to that species.

Additionally, the homepage features informative charts that summarize essential information about the database. These charts include the distribution of SSRs across species, the distribution of SSR motif types, the total number of SSRs, and the counts of SSRs located in genic and intergenic regions. These graphical representations provide a concise and visual representation of the database's key characteristics.

3.5.2 Species

The species selection panel, accessible from the top navigation bar, allows users to add or change species. This panel connects to the secondary website, which is the same as the species panel on the homepage. TriticeaeSSRdb provides species information, including a brief introduction, genome size, and total number of SSRs. A bar chart is available to visualize the distribution of SSR motif types and identify the most abundant motif category. Moreover, citation information for the reference genome is provided, enabling users to access relevant publications by clicking on the links.

The database offers versatile search options for retrieving SSRs based on various criteria, such as the genic or intergenic region, chromosome or scaffold-wise ID, motif type, repeat type, length, and genomic location. Notably, multiple chromosomes can be selected. In the advanced query, users can further refine their search by specifying parameters like the chromosome location (start and end coordinates) and size of SSR (length range). The result page dynamically displays tables with real-time visualization of the search results. The output includes comprehensive

10.3389/fpls.2024.1412953

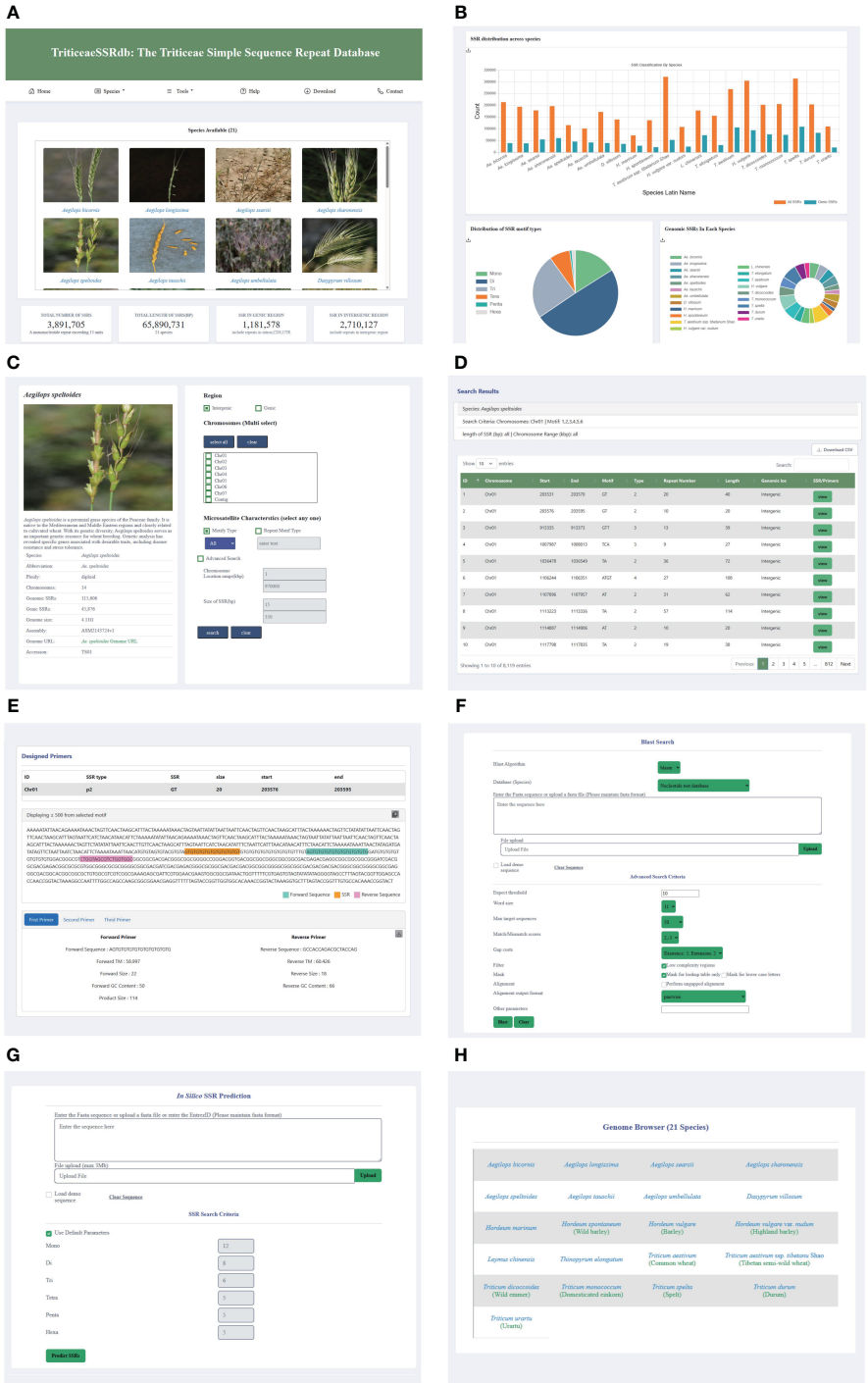


FIGURE 2
Main Functions and Modules of TriticeaeSSRdb. **(A)** List of species on the home page; **(B)** SSR statistics on the homepage; **(C)** Species page (*e.g.*, *Aegilops speltoides*); **(D)** Results page showing the desired search results; **(E)** Results page displaying SSR and corresponding primers; **(F)** Sequence BLAST tools; **(G)** *In silico* SSR prediction tools; **(H)** JBrowse framework.

information that can be sorted by chromosome number, chromosome location, motif type, repeat number, and SSR size.

The search results provide a visualization of the repeat and designed primers in the flanking region, which is extracted from a sequence spanning 500 base pairs upstream (250 bp) and downstream (250 bp) of the repeat. Additionally, users can access three sets of designed primers for each SSR by clicking the view button. The primer information includes primer sequence, size, GC content, the melting temperature of forward and reverse primers, and product size. The primer page showcases the designed primers in the flanking region and allows users to download the information in text format using the download button.

3.5.3 Tools

To enhance user experience, three practical tools have been integrated into TriticeaeSSRdb. The JBrowse framework, known for its versatility and customizability, is widely used for interpreting, and visualizing genomic features. For each species, specific gene annotation tracks have been established, including mRNA, CDS, UTR, and SSR loci, among other key data. This visualization plugin significantly improves the accessibility of genomic data, providing researchers with a powerful tool to access the desired genomic loci quickly and accurately, aiding in gene discovery and genetic breeding applications.

When users encounter sequence fragments without explicit IDs, TriticeaeSSRdb provides an online BLAST service that allows queries within the database collection of different *Triticeae* species. Users can submit amino acid or nucleotide sequences in Fasta format directly into the search box or upload text files. Depending on the query sequence and the database, users can select the appropriate BLAST algorithm to identify putative homologous sequences. In the advanced settings, users have the flexibility to set parameters such as expected threshold, Word size, and Max target sequences. After the alignment, the candidate hits are ranked in descending order of the e-value and displayed side by side in the result window.

We developed custom scripts to implement the MISA and Pytrf tools for the identification of SSRs in user input sequences. Users have the option to enter query sequences in the provided dialog box, and for larger file sizes, bulk identification can be performed by uploading files. Customization options allow users to specify the repetitive motifs for SSR detection. Once the search is initiated, the tool identifies microsatellites in the uploaded sequences and employs the primer3 program in the background to design primers for the target regions.

3.5.4 Help

The Help page provides users with an introduction to the database and a comprehensive tutorial on how to effectively utilize its features.

3.5.5 Download

To facilitate personalized local services for users, all raw data of this database are included under the Download interface. Users can select specific species according to their needs and download information such as whole-genome SSR identification results, SSR genomic distributions, and SSR annotations. Additionally, the statistical results in the supplementary tables can also be downloaded through this interface to enhance the user experience.

3.5.6 Contact

The Contact page includes contact information for researchers to facilitate communication and provides quick access to generic external links to genomic resources. Additionally, the website offers researcher information and other relevant contact details.

4 Discussion

4.1 Characterization of SSR in *Triticeae* species

The application of SSR markers has demonstrated great value in genetic analysis, linkage mapping, breeding programs across different crop species, and germplasm characterization (Zhao et al., 2017). To further enhance the understanding of complex traits and facilitate molecular breeding, it is crucial to develop novel genome-wide markers and construct a physical map with uniformly distributed markers (Lu et al., 2019). Despite the growing popularity of single nucleotide polymorphism markers in *Triticeae* crop research, SSR markers continue to be widely utilized by research institutions due to their cost-effectiveness, simplicity, and rapidity (Li et al., 2015).

The advancement of improved sequencing technology and bioinformatics has resulted in the release of novel plant genome assemblies, particularly for cereal crops (Michael and VanBuren, 2020). This availability of genomic sequences for *Triticeae* species presents an opportunity to investigate and compare the distribution of microsatellites across these genomes. Computational approaches for SSR identification offer advantages such as high throughput, low cost, and minimal workload (Sharma et al., 2007). Various tools emerged from the late 20th century to the early 21st century to explore microsatellites within genome sequences. These tools include the Repeat Pattern Toolkit (RPT) (Agarwal, 1994), Tandem Repeats Finder (TRF) (Benson, 1999), Microsatellite Identification Tool (MISA) (Thiel et al., 2003), and Tandem Repeat Occurrence Locator (TROLL) (Castelo et al., 2002). However, the preference has shifted towards the utilization of microsatellite software that offers faster processing speeds, identification of SSR polymorphisms, and broader applicability. Notable examples of such tools include Krait (Pytrf) (Du et al., 2018) and CandiSSR (Xia et al., 2016).

In a previous study, we conducted a screening of genomic SSR loci in eight *Triticeae* species (Deng et al., 2016). However, existing genome databases have become outdated due to advances in sequencing technology and the release of new species genomes. Therefore, there is a need for comprehensive databases that collect, store, and maintain genomics data for further study of biological functions and molecular mechanisms. While some web resources provide information related to *Triticeae* species, they lack microsatellite markers, SSR prediction, PCR primer design, visualization, and other related information.

To address these limitations, we first characterized the distribution of SSRs in 21 *Triticeae* species using the most up-to-date genomes. In total, 3,891,705 microsatellites were identified. The distribution and frequency of SSRs vary among different species genomes. Our findings suggest that the relationship between SSR density and genome size is not evident in *Triticeae* species. Although differences in genome size may contribute to the level of microsatellite repetition, the density of SSRs has been found to be unrelated to genome size, consistent with previous studies (Behura and Severson, 2015; Portis et al., 2016).

Subsequently, we conducted a detailed evaluation of individual repeat motifs for each type of SSR in the *Triticeae* species. The base composition of SSR motifs in *Triticeae* species shows a strong bias toward A and T, consistent with their low GC content characteristics (Wang et al., 2022). Dimeric repeats were found to be the most abundant SSR motifs in *Triticeae* species. Previous studies have indicated that mono-nucleotide repeats are the most prevalent in *Poaceae* species. However, discrepancies in these findings may be attributed to the inherent limitations of next-generation sequencing methods used for data generation. In this study, we utilized the newly released reference genome, particularly integrating third-generation sequencing technologies, which greatly improved the accuracy of repeat sequence identification. Furthermore, we observed a significant increase in microsatellite abundance as the motif repeat number decreased, possibly due to higher mutation rates and increased instability in longer repeats.

4.2 TriticeaeSSRdb architecture, features, and advantages

The TriticeaeSSRdb, accessible at http://triticeaessrdb.com/, is a publicly available resource. It encompasses a comprehensive collection of 3,891,705 SSRs from 21 *Triticeae* species, with an overall frequency of 27.21 SSRs per Mb. These high-density SSRs serve as a valuable resource for genetic and genomic studies, including genetic diversity analysis, genetic mapping, marker-trait association, and molecular breeding. By facilitating the mining of microsatellite loci and the design of primers for genic and intergenic SSRs on a chromosome-wise basis, the database enables efficient genotyping with specific search criteria. Additionally, TriticeaeSSRdb incorporates additional tools such as genomic browsing, *in silico* SSR prediction for user-provided data, and BLAST-based similarity searches. TriticeaeSSRdb not only offers the largest collection of microsatellites but also provides a user-friendly web interface for data analysis and visualization. Through its statistical and plotting functions, users can easily grasp global microsatellite trends within a genome, eliminating the need for manual analysis of millions of data points. The information is presented through interactive plots and a filterable table view, ensuring that relevant microsatellite information is easily accessible to researchers.

TriticeaeSSRdb possesses specific advantages and features:

(i) TriticeaeSSRdb represents the pioneering effort in identifying and characterizing *Triticeae* SSRs, crucial for genetic and genomics research. It provides critical information on genetic variation, kinship relationships, gene mapping, and breeding. With a comprehensive collection of 3,891,705 SSRs across 21 *Triticeae* species, TriticeaeSSRdb stands as the most extensive and inclusive database for *Triticeae* SSR research.

(ii) TriticeaeSSRdb has been successfully developed as a one-stop, user-friendly platform for *Triticeae* species SSR research. *In silico* predicted markers can be easily searched based on criteria such as genic or genomic origin, chromosome-wise location, motif type, repeat type, length, and genomic position. TriticeaeSSRdb offers advanced search capabilities, including frequency-based searches and filtering based on motif type, repeat type, length, and genomic location. The search results provide visual representations of repeat motifs and flanking primers, with 500 base pairs upstream and downstream of the SSR. All results can be downloaded in clipboard, CSV, and Excel formats, facilitating further analysis. This genomic resource holds immense value for the global research community and can expedite chromosome-wise microsatellite locus mining and primer design for both genic and intergenic SSRs, enabling rapid genotyping.

(iii) TriticeaeSSRdb offers a range of practical tools, making it easy and efficient for users to browse, search, and download specific areas of interest. The database provides statistical analyses, including SSR distribution, motif categories, and repeat sequence composition. The platform supports data visualization in various formats. Additionally, TriticeaeSSRdb integrates external generic databases such as IPK, NCBI, e! DAL, and Ensembl Plants, enabling quick cross-database searches.

4.3 Prospects

TriticeaeSSRdb is continuously improving, and these resources contribute to understanding genomic research, species identification, and evolution in *Triticeae* species. In the future, we plan to enhance TriticeaeSSRdb with additional features and utilities:

(i) Continuous updates: As more *Triticeae* species datasets become available in the public domain, we will regularly update the database with newly released genomes, expanding beyond the *Triticeae* tribe and potentially encompassing the entire *Poaceae* family or even larger scales. With its scalable framework, TriticeaeSSRdb aims for at least one update per year.

(ii) Pan-genome integration: Recognizing that a single reference genome may not capture the full genetic information of a species, future research will shift towards pan-genomics, encompassing different subspecies, varieties, or strains with significant genetic variations. TriticeaeSSRdb intends to incorporate the pan-genomes of *T. aestivum*, *H. vulgare*, and other related species to enrich the genetic information available.

(iii) Integration of new analysis tools: TriticeaeSSRdb plans to integrate new analysis tools, especially those focused on pan-genomics and population genomics data, enabling cross-genome e-PCR and cross-validation to rapidly identify highly reliable and polymorphic SSR loci.

(iv) Integration of QTL information: We aim to enhance the database by incorporating QTL information derived from our studies as well as publicly available sources in our

ongoing efforts. This integration will facilitate more accurate gene mapping and enable its application in breeding programs.

In summary, our approach maximizes the utility of TriticeaeSSRdb for studying various aspects of *Triticeae* crop development. It enables researchers, including those from small labs without extensive computing resources, to analyze genetic diversity and distinguish populations within *Triticeae* 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 authors.

Author contributions

TL: Formal analysis, Visualization, Writing – original draft. SC: Data curation, Writing – original draft. ZC: Formal analysis, Visualization, Writing – original draft. YF: Resources, Writing – original draft. WL: Resources, Visualization, Writing – original draft. XZ: Supervision, Writing – review & editing. CL: Writing – review & editing. LC: Writing – original draft. WP: Writing – review & editing. YL: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (Grant No. 32360157 and 32160016), the Natural Science Foundation of Jiangxi Province (Grant No. 20232BAB205012), and the Open Project Program of State Key Laboratory for Crop Stress Resistance and High-Efficiency Production (Grant No. SKLCSRHPKF11). The funders played no role in the study design, data collection, analysis, decision to publish, or manuscript preparation.

Acknowledgments

We would like to express our gratitude to the research groups that made their reference genomes publicly available. Due to space constraints, we apologize for not being able to cite all relevant papers in this manuscript. We are also grateful to Prof. Xiaojun Nie for his valuable comments and to the High-Performance Computing platform at Northwest A&F University for their assistance with data processing. Furthermore, we extend our appreciation to the diligent researchers involved in the investigation of the *Triticeae* tribe, whose previous studies and specifically captured photographs have significantly aided in the dynamic representation of our non-profit research database. We emphasize our commitment to promptly replace any images that may infringe upon copyright, ensuring adherence to legal guidelines.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1412953/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of SSR motif types with varying repeat numbers.

SUPPLEMENTARY FIGURE 2
Distribution overview of high-density SSR on chromosomes (using *T. aestivum* as a representative example). The bar depicts the number of SSR markers within a 1-Mb window.

References

Adamski, N. M., Borrill, P., Brinton, J., Harrington, S. A., Marchal, C., Bentley, A. R., et al. (2020). A roadmap for gene functional characterisation in crops with large genomes: lessons from polyploid wheat. *Elife* 9, e55646. doi: 10.7554/eLife.55646

Agarwal, P. (1994). "The Repeat Pattern Toolkit (RPT): analyzing the structure and evolution of the C. elegans genome," in *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, ed. States, D. J. (Stanford, CA, USA) 2, 1–9.

Avvaru, A. K., Saxena, S., Sowpati, D. T., and Mishra, R. K. (2017). MSDB: a comprehensive database of simple sequence repeats. *Genome Biol. Evol.* 9, 1797–1802. doi: 10.1093/gbe/evx132

Avvaru, A. K., Sharma, D., Verma, A., Mishra, R. K., and Sowpati, D. T. (2020). MSDB: a comprehensive, annotated database of microsatellites. *Nucleic Acids Res.* 48, D155–D159. doi: 10.1093/nar/gkz886

Behura, S. K., and Severson, D. W. (2015). Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Res.* 22, 29–38. doi: 10.1093/dnares/dsu036

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

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17, 1–12. doi: 10.1186/s13059-016-0924-1

Castelo, A. T., Martins, W., and Gao, G. R. (2002). TROLL—tandem repeat occurrence locator. *Bioinformatics* 18, 634–636. doi: 10.1093/bioinformatics/18.4.634

Chen, Y., Song, W., Xie, X., Wang, Z., Guan, P., Peng, H., et al. (2020). A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the triticeae tribe as a pilot practice in the plant pangenomic era. *Mol. Plant* 13, 1694–1708. doi: 10.1016/j.molp.2020.09.019

Deng, P., Wang, M., Feng, K., Cui, L., Tong, W., Song, W., et al. (2016). Genome-wide characterization of microsatellites in Triticeae species: abundance, distribution and evolution. *Sci. Rep.* 6, 32224. doi: 10.1038/srep32224

Deng, W., Nickle, D. C., Learn, G. H., Maust, B., and Mullins, J. I. (2007). ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 23, 2334–2336. doi: 10.1093/bioinformatics/btm331

Du, L., Zhang, C., Liu, Q., Zhang, X., and Yue, B. (2018). Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics* 34, 681–683. doi: 10.1093/bioinformatics/btx665

Duhan, N., and Kaundal, R. (2021). LegumeSSRdb: A comprehensive microsatellite marker database of legumes for germplasm characterization and crop improvement. *Int. J. Mol. Sci.* 22, 11350. doi: 10.3390/ijms222111350

Duhan, N., Meshram, M., Loaiza, C. D., and Kaundal, R. (2020). citSATdb: genome-wide simple sequence repeat (SSR) marker database of Citrus species for germplasm characterization and crop improvement. *Genes* 11, 1486. doi: 10.3390/genes11121486

Gao, Z., Bian, J., Lu, F., Jiao, Y., and He, H. (2023). Triticeae crop genome biology: an endless frontier. *Front. Plant Sci.* 14, 1222681. doi: 10.3389/fpls.2023.1222681

Guan, L., Cao, K., Li, Y., Guo, J., Xu, Q., and Wang, L. (2019). Detection and application of genome-wide variations in peach for association and genetic relationship analysis. *BMC Genet.* 20, 1–13. doi: 10.1186/s12863-019-0799-8

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* 6, e251. doi: 10.7717/peerj-cs.251

Hasan, N., Choudhary, S., Naaz, N., Sharma, N., and Laskar, R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J. Genet. Eng. Biotechnol.* 19, 128. doi: 10.1186/s43141-021-00231-1

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177, 309–334. doi: 10.1007/s10681-010-0286-9

Li, C., Bai, G., Carver, B. F., Chao, S., and Wang, Z. (2015). Single nucleotide polymorphism markers linked to QTL for wheat yield traits. *Euphytica* 206, 89–101. doi: 10.1007/s10681-015-1475-3

Li, Y.-C., Korol, A. B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007. doi: 10.1093/molbev/msh073

Li, M., Wang, H., Tian, S., Zhu, Y., and Zhang, Y. (2023). Triticeae-BGC: a web-based platform for detecting, annotating and evolutionary analysis of biosynthetic gene clusters in Triticeae. *J. Genet. Genomics* 50, 921–923. doi: 10.1016/j.jgg.2023.09.014

Lu, Q., Hong, Y., Li, S., Liu, H., Li, H., Zhang, J., et al. (2019). Genome-wide identification of microsatellite markers from cultivated peanut (Arachis hypogaea L.). *BMC Genomics* 20, 1–9. doi: 10.1186/s12864-019-6148-5

Michael, T. P., and VanBuren, R. (2020). Building near-complete plant genomes. *Curr. Opin. Plant Biol.* 54, 26–33. doi: 10.1016/j.pbi.2019.12.009

Portis, E., Lanteri, S., Barchi, L., Portis, F., Valente, L., Toppino, L., et al. (2018). Comprehensive characterization of simple sequence repeats in eggplant (Solanum melongena L.) genome and construction of a web resource. *Front. Plant Sci.* 9, 401. doi: 10.3389/fpls.2018.00401

Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S., et al. (2016). A genome-wide survey of the microsatellite content of the globe artichoke genome and the development of a web-based database. *PloS One* 11, e0162841. doi: 10.1371/journal.pone.0162841

Powell, W., Machray, G. C., and Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1, 215–222. doi: 10.1016/1360-1385(96)86898-1

Sato, K., Yoshida, K., and Takumi, S. (2021). RNA-Seq-based DNA marker analysis of the genetics and molecular evolution of Triticeae species. *Funct. Integr. Genomics* 21, 535–542. doi: 10.1007/s10142-021-00799-4

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109, 365–371. doi: 10.1007/s004120000089

Sharma, P. C., Grover, A., and Kahl, G. (2007). Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 25, 490–498. doi: 10.1016/j.tibtech.2007.07.013

Song, X., Yang, Q., Bai, Y., Gong, K., Wu, T., Yu, T., et al. (2021). Comprehensive analysis of SSRs and database construction using all complete gene-coding sequences in major horticultural and representative plants. *Horticulture Res.* 8, 122. doi: 10.1038/s41438-021-00562-7

Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., et al. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Systematics Evol.* 53, 117–137. doi: 10.1111/jse.12150

Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Tóth, G., Gáspári, Z., and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10, 967–981. doi: 10.1101/gr.10.7.967

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115–e115. doi: 10.1093/nar/gks596

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005

Wang, X., Yan, X., Hu, Y., Qin, L., Wang, D., Jia, J., et al. (2022). A recent burst of gene duplications in Triticeae. *Plant Commun.* 3, 100268. doi: 10.1016/j.xplc.2021.100268

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, 1171. doi: 10.3389/fpls.2015.01171

Zhao, C., Qiu, J., Agarwal, G., Wang, J., Ren, X., Xia, H., et al. (2017). Genome-wide discovery of microsatellite markers from diploid progenitor species, Arachis duranensis and A. ipaensis, and their application in cultivated peanut (A. hypogaea). *Front. Plant Sci.* 8, 1209. doi: 10.3389/fpls.2017.01209

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 May 2024
DOI 10.3389/fpls.2024.1403220

OPEN ACCESS

EDITED BY
Ertugrul Filiz,
Duzce University, Türkiye

REVIEWED BY
Xiaomei Zhang,
Chinese Academy of Agricultural Sciences, China
Neelam Mishra,
St Joseph's College (Autonomous), India
Yi Shang,
Yunnan Normal University, China

*CORRESPONDENCE
Weichao Ren
lzyrenweichao@126.com
Qifeng Fu
269702613@qq.com
Wei Ma
mawei@hljucm.edu.cn

RECEIVED 19 March 2024
ACCEPTED 13 May 2024
PUBLISHED 28 May 2024

CITATION
Wang Z, Wang P, Cao H, Liu M, Kong L, Wang H, Ren W, Fu Q and Ma W (2024) Genome-wide identification of bZIP transcription factors and their expression analysis in *Platycodon grandiflorus* under abiotic stress.
Front. Plant Sci. 15:1403220.
doi: 10.3389/fpls.2024.1403220

Genome-wide identification of bZIP transcription factors and their expression analysis in *Platycodon grandiflorus* under abiotic stress

Zhen Wang[1], Panpan Wang[1], Huiyan Cao[1], Meiqi Liu[1], Lingyang Kong[1], Honggang Wang[2], Weichao Ren[1*], Qifeng Fu[3*] and Wei Ma[1,3*]

[1]Pharmacy of College, Heilongjiang University of Chinese Medicine, Harbin, China, [2]Research Office of Development and Utilization of Medicinal Plants, Heilongjiang Academy of Forestry, Yichun, China, [3]Experimental Teaching and Practical Training Center, Heilongjiang University of Chinese Medicine, Harbin, China

The Basic Leucine Zipper (bZIP) transcription factors (TFs) family is among of the largest and most diverse gene families found in plant species, and members of the bZIP TFs family perform important functions in plant developmental processes and stress response. To date, *bZIP* genes in *Platycodon grandiflorus* have not been characterized. In this work, a number of 47 *PgbZIP* genes were identified from the genome of *P. grandiflorus*, divided into 11 subfamilies. The distribution of these *PgbZIP* genes on the chromosome and gene replication events were analyzed. The motif, gene structure, *cis*-elements, and collinearity relationships of the *PgbZIP* genes were simultaneously analyzed. In addition, gene expression pattern analysis identified ten candidate genes involved in the developmental process of different tissue parts of *P. grandiflorus*. Among them, Four genes (*PgbZIP5*, *PgbZIP21*, *PgbZIP25* and *PgbZIP28*) responded to drought and salt stress, which may have potential biological roles in *P. grandiflorus* development under salt and drought stress. Four hub genes (*PgbZIP13*, *PgbZIP30*, *PgbZIP32* and *PgbZIP45*) mined in correlation network analysis, suggesting that these *PgbZIP* genes may form a regulatory network with other transcription factors to participate in regulating the growth and development of *P. grandiflorus*. This study provides new insights regarding the understanding of the comprehensive characterization of the PgbZIP TFs for further exploration of the functions of growth and developmental regulation in *P. grandiflorus* and the mechanisms for coping with abiotic stress response.

KEYWORDS

***Platycodon grandiflorus*, bZIP transcription factor, evolutionary analyses, expression profiling, abiotic stress**

1 Introduction

Platycodon grandiflorus (Jacq.) A. DC. is a perennial herbal in the *Platycodon* genus of the Campanulaceae family (Ma et al., 2016). Mainly from the northeastern, northern, central and eastern provinces of China, it is also found in the Russian Far East and the Korean Peninsula (Zhang F. et al., 2015; Zhang L. et al., 2015). *P. grandiflors* has high medicinal and edible value, and is a homologous category of medicine and food. At the same time, it also has high ornamental value (Ji et al., 2020). *P. grandiflorus* is rich in natural chemical products. In the past many years, at least 100 compounds have been isolated, including steroidal saponins, sterols, flavonoids and phenolic acids. Modern pharmacological studies have shown that the pharmacological components of *P. grandiflorus* are triterpenoid saponins, and Platycodon D possesses several medicinal activities such as anti-obesity, anti-inflammatory, antioxidant, antitumor and immune regulatory activities (Zhang H. et al., 2022; Zhang S. et al., 2022). The growth and development of *P. grandiflorus* is a delicate process, and the study of molecular mechanisms of resistance to abiotic stress is an important production guide in the face of increasingly challenging natural environments. Transcription factors (TFs) are key regulatory proteins with DNA binding domains that can inhibit or activate gene expression (Tian et al., 2023). The YABBY (Kong et al., 2023), WRKY (Jing et al., 2022) and Trihelix (Liu et al., 2023) TF families have been proven to play a role in the response of *P. grandiflorus* to abiotic stress. However, this is only a drop in the ocean of a complex transcriptional regulatory network.

The Basic Leucine Zipper (bZIP) TFs are among of the most widely families of TFs in eukaryotes (Zhao et al., 2016b). The plant bZIP protein has two highly conserved domains composed of 60–80 amino acids: the basic region and leucine zipper region (Nijhawan et al., 2008). bZIP transcription factors mainly regulate tissue and organ development in plants, including seed germination and maturation, embryonic development, flowering, and light morphogenesis (Gangappa and Botto, 2016). Earlier research has indicated that the *bZIP* gene family *ABF1* gene in *Arabidopsis thaliana* regulate seed dormancy and germination, affecting winter *A. thaliana* growth (Sharma et al., 2011). *Litchi chinensis LcbZIP1*, *LcbZIP4*, *LcbZIP7*, *LcbZIP21* and *LcbZIP28* may be involved in the regulation of senescence during postharvest storage of fruit (Hou et al., 2022). Meanwhile, the *bZIP* family also performs essential functions in responding to abiotic stress and regulating secondary metabolites. Overexpression of *Capsicum annuum CabZIP25* in *A. thaliana* can improve tolerance to salt stress (Gai et al., 2020). Similarly, *TabZIP15* can also improve salt tolerance in *Triticum aestivum* (Bi et al., 2021). Overexpression of *Phyllostachys edulis PhebZIP47* in *A. thaliana* and *Oryza sativa* can increase the drought resistance of plants at the adult stages (Lan et al., 2023). In transgenic *Artemisia annua* overexpressing *AabZIP9*, the biosynthetic accumulation of artemisinin, dihydroartemisinic acid, and artemisinic acid was significantly increased (Shen et al., 2019). Reports on *bZIP* genes have now been carried out in a wide variety plants, However, it is unclear whether *bZIP* genes are participating in the growth process of *P. grandiflorus* and their role in responding to abiotic stress. At present, the chromosome-scale genome of *P. grandiflorus* has been released (Lee et al., 2023). Therefore, explore the possible role of bZIP TFs in the growth, development, and abiotic stress of *P. grandiflorus* on the basis of genome-wide and transcriptome data is of great significance.

In this study, we carried out a genome-wide identification and characterization of the *bZIP* gene family of *P. grandiflorus* and comprehensively analyzed the physicochemical properties, phylogeny, synteny relationship, gene structure and *cis*-elements. In addition, by mine the expression patterns of *PgbZIP* genes in eight tissues and constructing correlation networks, it was determined that they may interact with different genes and participate in multiple biological processes together. More importantly, the expression of *PgbZIP* genes during abiotic stress was analyzed by RT-qPCR, which indicated that some *PgbZIP* genes were responsive to salt and drought stress. The present study was conducted to provide a reference for the screening of *bZIP* candidate genes involved in regulating the developmental process of *P. grandiflorus* and in resistance to drought and salinity stress.

2 Materials and methods

2.1 Plant materials

The plant material *P. grandiflorus* was cultivated from the Medicinal Botanical Garden of Heilongjiang University of Chinese Medicine (HLJUCM), Harbin, China. There are two types of plant materials used in this study, one of which was the roots, stems, leaves and flowers of 1-year-old *P. grandiflorus* in the normal growth state without treatment. The other was a three-month-old *P. grandiflorus* seedling grown in a simulated abiotic stress environment. *P. grandiflorus* seedlings were watered with 200 mmol/L NaCl and 20% PEG6000 (Coolaber Science & Technology, Beijing, China) to simulate drought and salt stress, respectively, and roots were collected at 48 h. Following the sample process, all plant components were instantly frozen in liquid nitrogen immediately after sampling and kept in a medical refrigerator at -80°C for additional analysis.

2.2 Identification of putative *bZIP* genes in *P. grandiflorus*

The complete genome sequence and annotated file of *P. grandiflorus* can be downloaded from the Figshare database (https://doi.org/10.6084/m9.figshare.21511020). TBtools software was used to identify every member of the bZIP TFs in the genome of *P. grandiflorus* (Chen et al., 2020). TBtools software was used to extract all genes sequences in the *P. grandiflorus* genome. The hidden Markov model file in the *bZIP* domain was used as a template to identify each gene and obtain the candidate PgbZIP protein sequence. For validation, the candidate sequences were uploaded to the NCBI CD-search tools and the plant transcription factor database. After removing redundant and structurally incomplete domain sequences, the final PgbZIP protein was obtained. Use the web tool ExPASy ProtParam

(https://web.expasy.org/protparam/) to search the molecular weight (MW), isoelectric point, grand average of hydropathicity (GRAVY) of PgbZIP proteins.

2.3 Chromosomal location and gene duplication analysis of *PgbZIP* genes

The genome annotation file of *P. grandiflorus* was analyzed and the chromosome location information of *PgbZIP* gene was obtained, and its corresponding chromosomal physical location was mapped using Tbtools software. The genomic information of *Malus domestica* (GCF_002114115.1), *Cannabis sativa* (GCA_900626175.1), *Vitis vinifera* (GCF_000003745.3), *Solanum lycopersicum* (GCF_000188115.4) and *Oryza_sativa* (GCA_001433935.1) were downloaded from the NCBI database. MCScanX software was used to analyze the gene duplication and collinearity of *PgbZIP* gene family (Wang et al., 2012). The *PgbZIP* gene replication events were visualized using the TBtools software's "Advanced Circos" function, and the collinearity relationship between *P. grandiflorus* and other species was shown using the "Dual Systeny Plot" function.

2.4 Phylogenetic analyses of the PgbZIP and AtbZIP proteins

The TAIR database (https://www.arabidopsis.org/) provided the sequence of the *A. thaliana* bZIP protein. MEGA X was used to perform multiple sequence alignment of AtbZIP, IibZIP (Jiang et al., 2022) and PgbZIP protein, respectively (Kumar et al., 2018). The phylogenetic tree was constructed by neighbor- joining (NJ) method, and the bootstrap was repeated 1000 times. PgbZIP proteins are divided into different subfamilies based on *A. thaliana* bZIP proteins (Jakoby et al., 2002).

2.5 Analyses of gene structure and cis-element compositions of *PgbZIP*

The *PgbZIP* gene family was subjected to motif analysis using the Multiple Em for Motif Elicitation website. The parameters were set to Classic mode, and ten motifs were discovered (Bailey et al., 2009). Extract the upstream 2 kb sequence of the *PgbZIP* family gene using the "GXF Sequences Extract" tool in TBtools software, and then identify and retrieve the *cis*-elements of the gene family through the PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for analysis and comparison (Lescot et al., 2002).

2.6 Analysis of the expression profile of the *PgbZIP* gene family in different tissues

The NCBI Sequence Read Archive (SRA) database has the raw transcriptome data for the eight *P. grandiflorus* tissues that were used in this investigation. These accession numbers are as follows: SER912510-SER912517 (Kim et al., 2020). The expression of every *PgbZIP* gene used to be quantified in fragments per kilobase of exon model per million mapped fragments (FPKM). All FPKM values have been processed using row scale transformation, and heatmaps were generated using TBtools.

2.7 RNA extraction and evaluation of *bZIP* gene expression patterns in *P. grandiflorus*

Total RNA was extracted from leaves, roots and stems of 1-year-old *P. grandiflorus*, drought, and salt stress treated *P. grandiflorus* seedlings, respectively, using the RNAprep Pure Plant Kit (TIANGEN Biotech Co.,Ltd, Beijing, China). Then, The first strand cDNA was synthesized using MS 1st Strand cDNA Synthesis SuperMix for qPCR (+gDNA wiper) kit (Msunflowers Biotech Co.,Ltd, Beijing, China) reverse transcribed RNA as a template, and finally, qRT-PCR experiments were performed using 2×SYBR Green qPCR Mix (With ROX) kit (Shandong SparkJade Biotechnology Co., Ltd, Shandong, China) kit. qRT-PCR experiments had been carried out the use of an AriaMx real-time PCR system (Agilent Technologies, USA). The *PgGAPDH* had been used as a reference gene (Ma et al., 2016), and three technical replicates had been set up for each experiment. The relative expression of genes was computed by the $2^{-\Delta\Delta CT}$ Method (Livak and Schmittgen, 2001). The Primer3web (version 4.1.0) website (https://primer3.ut.ee/) was used to layout unique primers (Untergasser et al., 2012). GraphPad Prism (v8.0.2) software program was once used to draw the histogram of relative expression of genes, which was once analyzed by *t*-test.

2.8 Statistical analysis

PlantTFDB was used to identify TFs for the whole genome proteins of *P. grandiflorus*, and the online database of Pfam was used to annotate the conserved domains of genes. The correlation between transcriptomes was calculated using a Python script primarily based on the Pearson correlation coefficient (PCC) (Zhang et al., 2017). Visualization of correlation network diagrams using Cytoscape (v3.7.1) software (Shannon et al., 2003). The GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis using the R language package clusterProfiler (Zhang et al., 2017).

3 Results

3.1 Identification and physicochemical properties of the *bZIP* genes in *P.grandiflorus*

After removing redundant and incomplete domain sequences, we finally identified 47 *bZIP* genes in *P.grandiflorus* (Supplementary Table 1). In subsequent analyses, we named the genes *PgbZIP1* to *PgbZIP47* based on their location on the chromosome or contig (from Pg_chr01_03120T to

Pg_contig01856_00120T). Then, we analyzed the physicochemical parameters of the PgbZIP protein in detail. The results confirmed that the amino acid (AA) size of the PgbZIP proteins ranged from 132 AA (PgbZIP40) to 704 AA (PgbZIP13). The molecular weight (MW) of the PgbZIP proteins ranged from 14840.52 Da to 75818.79 Da, with an theoretical isoelectric point (pI) of 4.66 (PgbZIP28) to 9.77 (PgbZIP27). The GRAVY values of all PgbZIP proteins were less than 0, suggesting that all PgbZIP proteins may be hydrophilic.

3.2 Chromosome location and replication events of *bZIP* genes in *P.grandiflorus*

The distribution of *PgbZIP* genes on the chromosomes of *P.grandiflorus* did not show any obvious pattern, with 47 *PgbZIP* genes unevenly distributed on 9 chromosomes and 2 contigs (Figure 1). Chromosome 3 contained the most *bZIP* genes (9), and chromosome 9, Pg_contig00116 and Pg_contig01856 contained the fewest *bZIP* genes (only a single gene each). The *bZIP* genes were distributed on all chromosomes, but none of the genes showed a preferential distribution on a particular chromosome.

Investigations into the gene replication events in *P. grandiflorus* were conducted to understand the expansion of the *PgbZIP* genes. Gene duplication within the same chromosome or on different chromosomes, but without one following the other, is considered a segmental duplication event. In the *PgbZIP* gene family, four *PgbZIP* gene pairs were generated by segmental duplication. It is worth noting that 1 pair of segmental duplication events (*PgbZIP5* and *PgbZIP46*) is not shown in the Figure 1B, because the *PgbZIP46* gene was not mapped to the chromosome and was distributed on the contig.

3.3 Phylogenetic analysis of the *PgbZIP* gene family

To explore the evolutionary relationships and classification of the *PgbZIP* gene family, we constructed NJ phylogenetic trees using MEGA X software (Supplementary Table 2). Based on the earlier classification results of the *A. thaliana* and *I. indigotica* bZIP gene family, 47 PgbZIP proteins were divided into 11 subfamilies (Figure 2), comprising A (10), B (1), C (2), D (6), E (4), F (1), G (5), H (2), I (6), K (1) and S (9). The outcome of phylogenetic tree evaluation confirmed that the *bZIP* genes of *P. grandiflorus*, *A. thaliana* and *I. indigotica* were highly homologous in each cluster.

3.4 Gene structure, motif and *cis*-element analysis of *PgbZIP*

To explore the structural similarity of PgbZIP proteins, 10 conserved motifs have been recognized the use of the MEME website (Figure 3A). These motifs vary in size from 21–50 AA, with at least 1 to 5 motifs distributed across 47 PgbZIP proteins, and all sequences have motif 1. Members of gene families with close

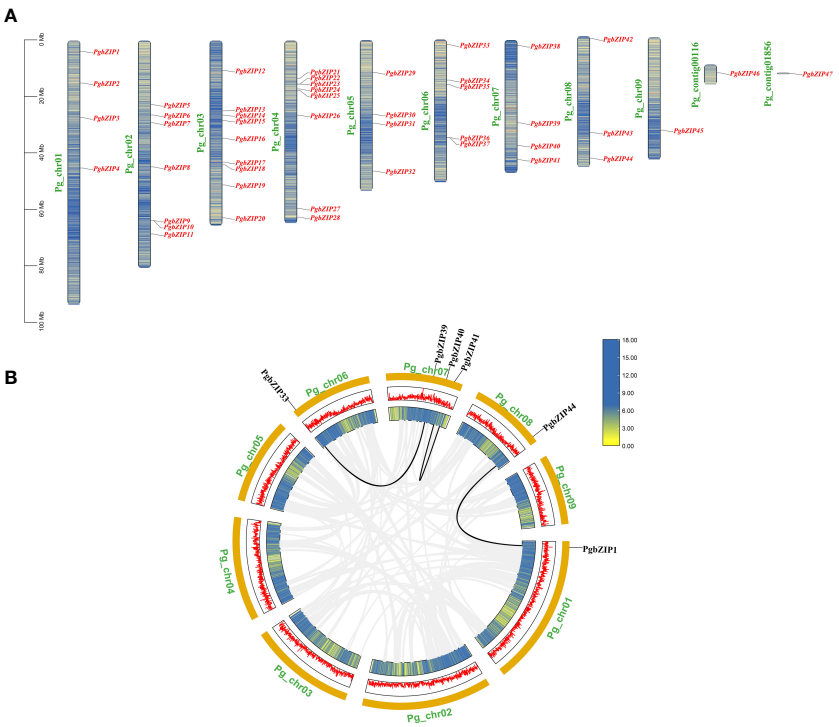


FIGURE 1
The distribution information of *PgbZIP* genes on chromosomes and gene replication events in the *P.grandiflorus*. **(A)** Information on the location of the *PgbZIP* genes on the *P.grandiflorus*. **(B)** Gene duplication events of *bZIP* genes in *P.grandiflorus*. Black lines indicate gene segmental duplication events in the *PgbZIP* gene, and gray lines show that collinear pairs of all *P.grandiflorus* genes. The yellow lines and red bar indicate the genes density in each chromosome.

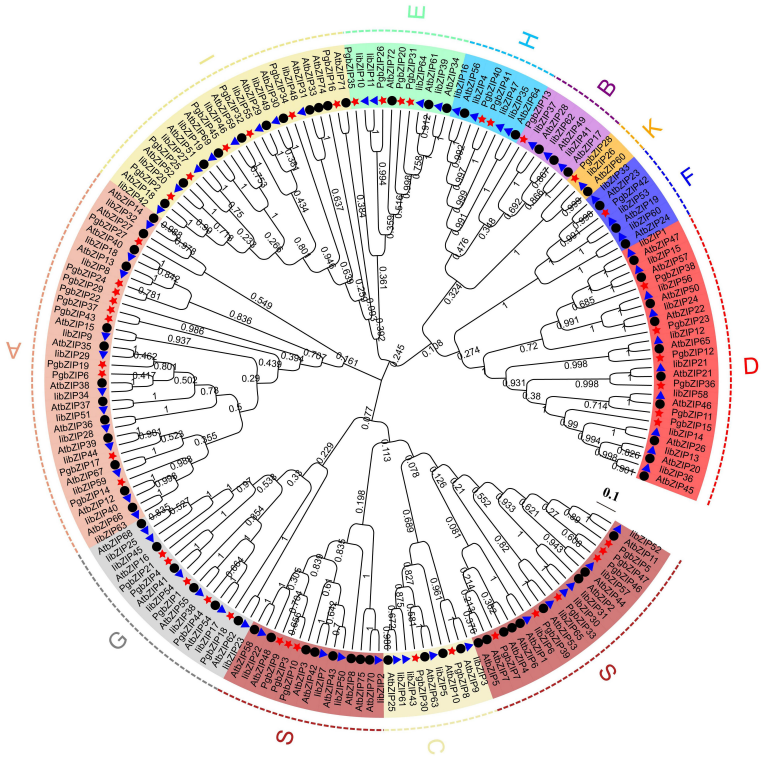


FIGURE 2
The phylogenetic relationship and classification of the of the bZIP proteins in *P. grandiflorus*, *I. indigotica* and *A. thaliana*. The bootstrap value were set to 1000 replications. Subfamilies are marked with distinct colors. The red star, blue triangle and black dots point out the bZIP proteins of *P. grandiflorus*, *I. indigotica and A. thaliana*, respectively.

evolutionary relationships have similar motif compositions; for example, G subfamily members all have motifs 1 and 3. Furthermore, all PgbZIP proteins contain conserved domains of the *bZIP* genes, proving that the outcome of gene identification had been dependable (Figure 3B).

To research the potential transcriptional regulatory role of *PgbZIP* genes, in this study, *cis*-elements were extracted from the PlantCARE database using 2K bp upstream of the *PgbZIP* genes as the promoter region. A total of 929 valuable *cis*-elements were recognized in the promoter regions of 47 *PgbZIP* genes, these *cis*-

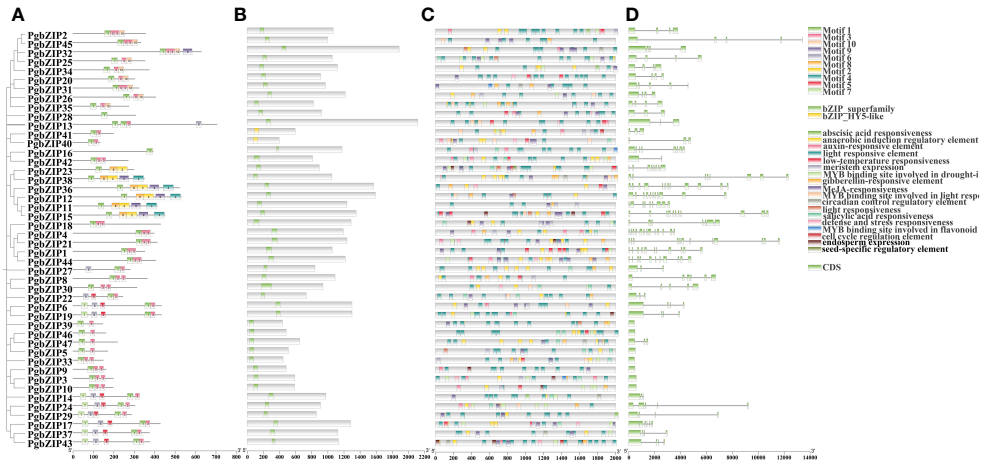


FIGURE 3
Phylogenetic analysis motifs, conserved domains and *cis*-element analysis of the *PgbZIP* genes. **(A)** Distribution of motifs within each PgbZIP protein. **(B)** Conserved domains of PgbZIP proteins. **(C)** Distribution of different types of *cis*-elements in the promoter region of the *PgbZIP* genes. **(D)** *PgbZIP* gene structures, lines indicate introns. **(D)** Distribution of different types of *cis*-elements in the promoter region of the *PgbZIP* genes.

10.3389/fpls.2024.1403220

elements could be broadly categorized into three groups (Figure 3C) (Supplementary Table 3). Relevant to plant growth include light responsive elements, meristem expression, endosperm expression, anaerobically induced regulatory elements, seed-specific regulatory response elements, cell cycle regulatory elements, and circadian regulatory elements. Hormone response related elements include MeJA responsive elements, auxin responsive elements, abscisic acid responsive elements, and gibberellin responsive elements. Involved in abiotic stress related including stress responsive and defense elements and low-temperature responsive elements. Interestingly, the *PgbZIP* gene has *cis*-elements that bind to the *MYB* gene, participating in photoresponse, drought response, and flavonoid synthesis. This suggests that the *PgMYB* gene may regulate *PgbZIP*, forming regulatory networks that exercise different biological functions.

In order to further clarify the evolution of the *PgbZIP* genes, we compared the *PgbZIP* gene sequences and analyzed the coding regions and introns (Figure 3D). Except for seven members of the S family who do not have introns, the number of introns in other members is distributed between 1 and 12. As anticipated, members of the same subfamily have a relatively conserved number of introns, such as the H subfamily, which has 2 introns, and the E subfamily, which has 3 to 4 introns.

3.5 Syntenic analysis of *PgbZIP* genes with other species

For investigating further the relationship of *bZIP* genes evolution among various plants, we compared interspecific synteny of *P. grandiflorus* with that of five dicotyledonous plants and one monocotyledonous plant (*O. sativa*) (Figure 4). Dicotyledonous plants include three fruit (*M. domestica*, *V. vinifera* and *S. lycopersicum*), and two medicinal plants (*C. sativa* and *C. lanceolata*). The results showed that the collinearity relationship between *PgbZIP* genes and *MdbZIP* genes was the closest, with 79 pairs of genes, 40 pairs of *PgbZIP* and *VvbZIP* genes, 32 pairs of *PgbZIP* and *SlbZIP* genes, 23 pairs of *PgbZIP* and *ClbZIP* genes, and the same pairs of *CsbZIP* genes. *PgbZIP* and *OsbZIP* had the worst collinearity relationship, with only 7 pairs of genes. These results indicate that most of these homologous *bZIP* genes occur after the differentiation of dicotyledons and monocotyledons. It is

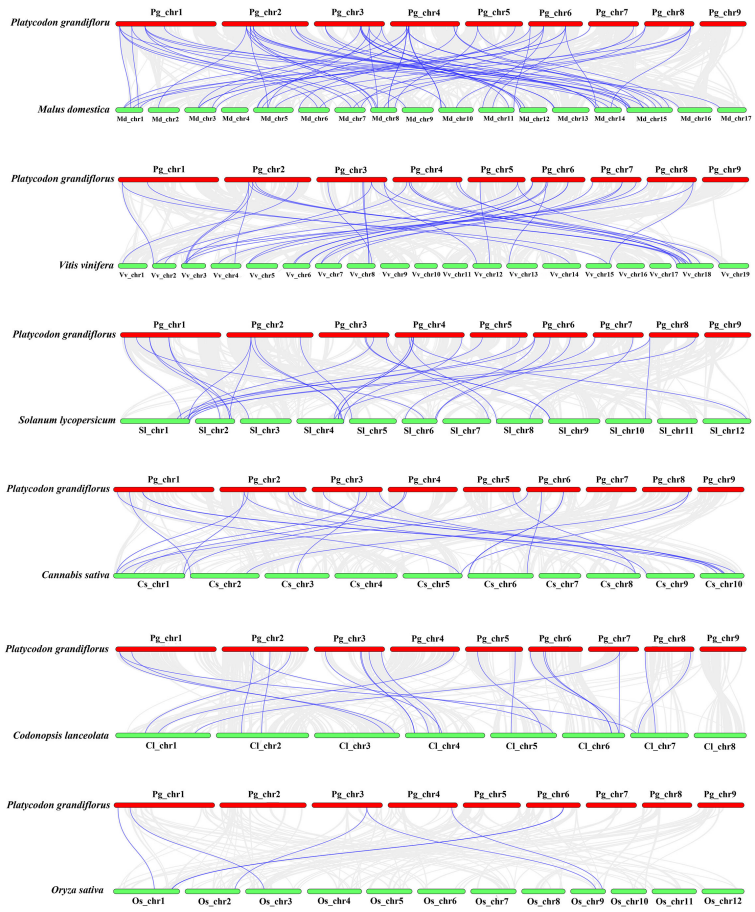


FIGURE 4
Synteny analysis of *bZIP* genes between *P.grandiflorus* and *M. domestica*, *V. vinifera*, *S. lycopersicum*, *C. sativa*, *C. lanceolata*, and *O. sativa*. Gray lines symbolize the colinear blocks within *P.grandiflorus and* other genomes, Purple lines represent the syntenic *bZIP* gene pairs.

worth noting that *PgbZIP1* (*Pg_chr01_03120T*) and *PgbZIP19* (*Pg_chr03_29870T*) have syntenic gene pairs with 6 other species, which may have existed before species divergence and participated in the evolution of these plants.

3.6 Divergent expression of the *bZIP* gene in *P. grandiflorus* tissues and verification of candidate genes by quantitative real-time PCR

To investigate the expression profile of *bZIP* genes in *P. grandiflorus* tissues, the gene relative expression of 47 *PgbZIP* genes were analyzed based on RNA-seq data of *P. grandiflorus* root, leaf, seed, petal, stem, stamen, pistil, and sepal (Figure 5) (Supplementary Table 5). The findings indicated that 20 genes were expressed in leaf, 28 genes in petal, 34 genes in pistil, 32 genes in root, 38 genes in seed, 37 genes in sepal, 37 genes in stamen and 37 genes in stem (FPKM>0.5). A total of 10 genes (*PgbZIP5*, *PgbZIP13*, *PgbZIP14*, *PgbZIP21*, *PgbZIP25*, *PgbZIP28*, *PgbZIP30*, *PgbZIP33*, *PgbZIP42* and *PgbZIP45*) were highly expressed in all tissues (FPKM>10), and these genes probably participating in the whole development processes of *P. grandiflorus*. *PgbZIP16* and *PgbZIP17* were not expressed in all tissues and may be pseudogenes or require specific conditions to activate expression. Interestingly, *PgbZIP* genes are also tissue-specific. For example, *PgbZIP26* and *PgbZIP34* are expressed only in stamens. *PgbZIP12*, *PgbZIP35*, *PgbZIP37* and *PgbZIP43* are only expressed in seeds. These genes involved in specific tissue expression may only be involved in the biological process of this tissue.

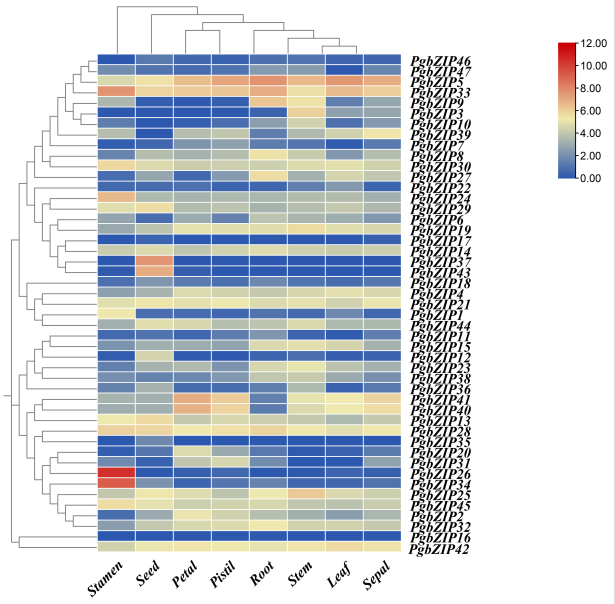


FIGURE 5
Heatmap of *bZIP* gene expression profiles in eight tissues of *P. grandifloru*. Data represent FPKM values from transcriptome data come from 8 tissues, including seed, leaf, petal, pistil, root, sepal, stamen and stem. Gene expression using Log_2(FPKM+1) logarithmic transformation treatment.

As mentioned above, subfamilies A (*PgbZIP14*), B (*PgbZIP13*), C (*PgbZIP30*), F (*PgbZIP42*), G (*PgbZIP21*), I (*PgbZIP25/45*), K (*PgbZIP28*) and S (*PgbZIP5/33*) have been expressed at very high levels in various tissue parts of *P. grandiflorus*, and are probably participated in the regulation of developmental process in all tissues. Therefore, those 10 *PgbZIP* genes were selected as candidate genes for qRT-PCR experiments to validate in this research (Supplementary Table 6). The relative expression of the 10 candidate genes were basically consistent to the expression trends obtained from the RNA-seq data (Figure 6). Notably, *PgbZIP28* showed higher expression in petals, and *PgbZIP30* and *PgbZIP33* were expressed at higher levels in sepals than stems. All of the above suggest that these 10 *PgbZIP* genes possibly are closely involved to the developmental process of *P. grandiflorus*.

3.7 Analysis of expression profile of *PgbZIP* genes family under drought and salt stress

The root of *P. grandiflorus* is the main medicinal part. For studying the potential role of *PgbZIP* gene family under abiotic stress, the qRT-PCR experiments analysis was conducted using the roots of *P. grandiflorus* seedlings under drought and salt stress as templates (Figure 7). Compared with the normally growth group (CK), the expression levels of 4 genes (*PgbZIP5*, *PgbZIP21*, *PgbZIP25* and *PgbZIP28*) were increased under both drought and salt stress. *PgbZIP33* was only highly expressed under salt stress, but its expression was reduced under drought stress. These *PgbZIP* genes with increased expression levels under drought and salt stress could help *P. grandifloru* to resist abiotic stress.

3.8 Correlation network and enrichment analysis of *PgbZIP* genes

The *bZIP* genes often form an interaction networks with various TFs to participate in plant developmental processes. In order to mine the transcriptional regulatory network of *PgbZIP* in different tissues, this study characterized TFs in the RNA-seq data of 8 different tissues of *P. grandiflorus* and constructed a correlation network. A total of 1567 TFs were identified and categorized into 58 gene families, among which the top 5 were ERF (136), bHLH (118), C2H2 (115), MYB (105) and NAC (90). The expression profiles of all genes in the RNA-seq data of 8 different tissues of *P.grandiflorus* were analyzed using Python script to explore the TFs co-expressed with the *PgbZIP* gene. In the correlation network, a total of 19 *PgbZIPs* were co-expressed with 149 TFs, and 168 nodes with 361 network pairs were found. The positive ($r > 0.7$) ($P < 0.05$) and negative ($r < -0.7$) ($P < 0.05$) correlation network pairs are 232 and 129, respectively (Figure 8). *PgbZIP45*, *PgbZIP30*, *PgbZIP13* and *PgbZIP32* were presented as hub genes (degree≧30). The highest correlations (degree≧7) among TFs were Pg_chr05_09390T (GATA), Pg_chr08_31480T (C3H), Pg_chr02_03150T (MYB-related), Pg_chr01_14460T (AP2), Pg_chr06_07740T (bHLH), and Pg_contig26001_00320T (NF-YA) (Supplementary Table 7). These TFs has been shown to function in regulating plant growth

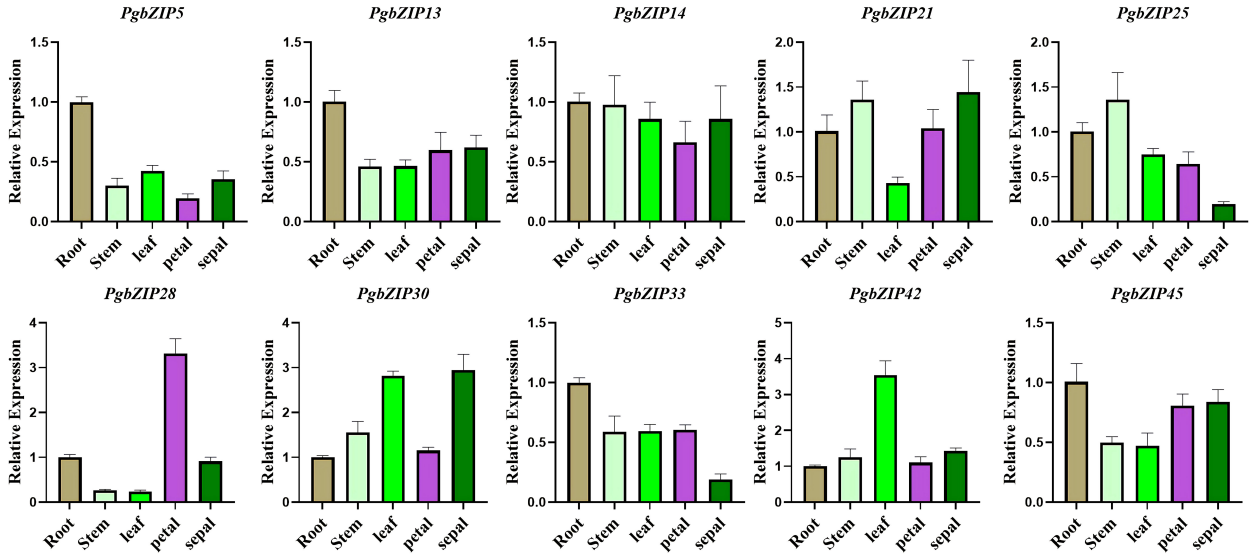


FIGURE 6
qRT–PCR analysis of *PgbZIP* genes in five different tissues of *P. grandiflorus*. Different colors represent different tissues. Use of the *PgGAPDH* gene as a reference gene.

and development in a large number of studies. Therefore, *PgbZIP* genes could be interacting with these TFs to form a regulatory network and participate in the development process of *P.grandiflorus*.

In order to further clarify the functions of TFs in the network in biological processes and the pathways involved, we performed GO annotation, GO and KEGG enrichment analyses of all TFs in the correlation network (Figure 9). The results of the analysis of the GO annotation findings suggests that those genes function in the molecular function classification for recruitment of TFs and activation of transcriptional activity (Figure 9A). In the cellular component classification, these genes were found to be predominantly distributed in the telomeric region, chromosome, nuclear chromosome and spindle microtubule. In biological processes, these genes are mainly involved in response to

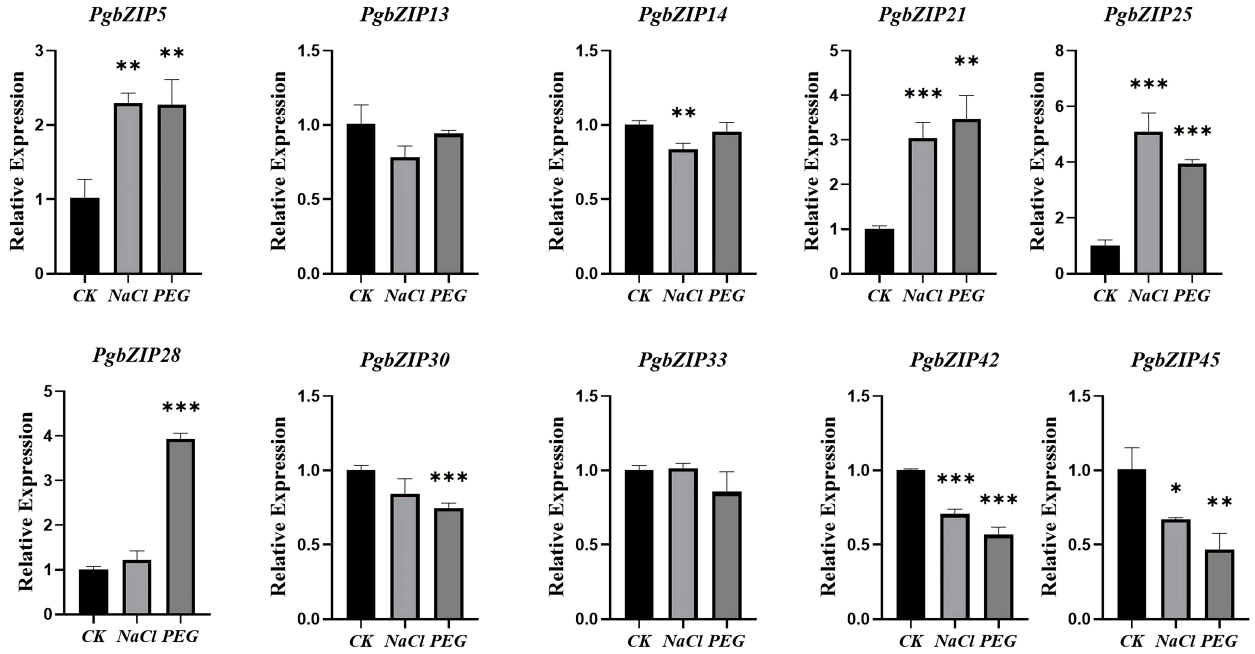


FIGURE 7
qRT-PCR analysis of *PgbZIP* genes under drought and salt stress. CK: normal growth group; Salt stress and drought stress simulated using NaCl and PEG. The *PgGAPDH* gene was used as a reference gene. The significance analysis was carried out the use of a t-test. *, **, *** indicate significant difference in $p < 0.05$, $p < 0.01$, and $p < 0.001$.

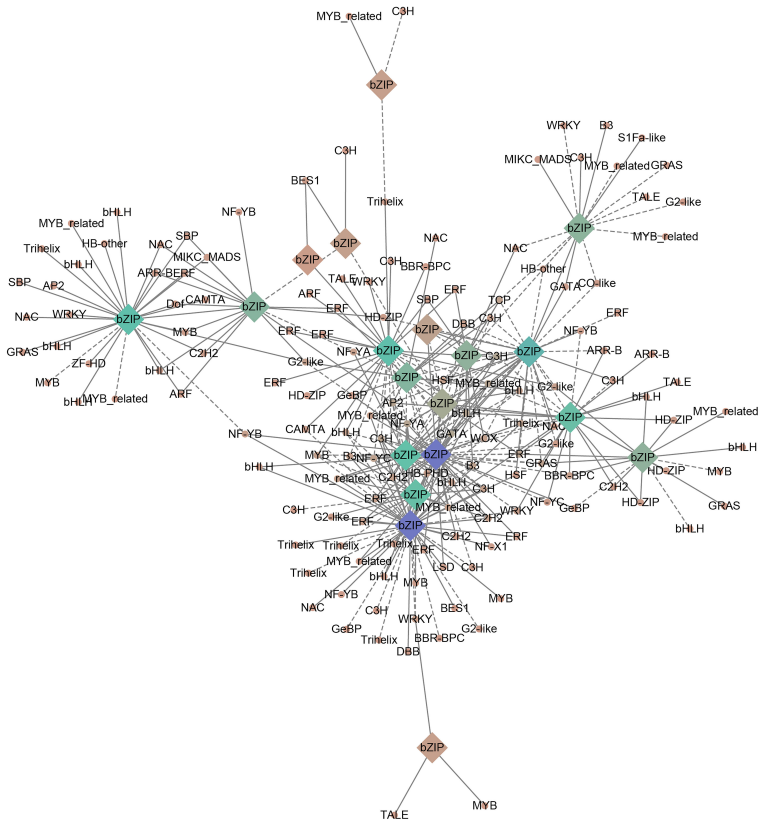


FIGURE 8
PgbZIP genes correlation network established using RNA-seq data from 8 tissues of *P.grandiflorus*. Circles represent co-expressed transcription factors, and diamonds represent the *PgbZIP* genes. Positive and negative correlations are indicated by solid and dashed lines, respectively. The shade of the color indicates the amount of genes correlated.

ethylene, negative regulation of macromolecular biosynthesis processes, negative regulation of RNA metabolic processes and ethylene response (Supplementary Table 8).

Similarly, we performed GO and KEGG enrichment analysis of TFs in the correlation network. The outcome of GO enrichment analysis indicated that the main functions of these genes are transcription factor binding transcription factor activity, cellular response to ethylene stimulus, meristem maintenance, response to ethylene, xylem development, phloem development, floral organ morphogenesis, meristem maintenance, and so on (Figure 9B). The outcome of KEGG enrichment indicated that these TFs were primarily enriched in circadian rhythm, phytohormone signal transduction, nucleotide excision repair, plant-pathogen interaction, pyrimidine metabolism, purine metabolism and so on (Figure 9C). These outcomes indicate that the TFs in the correlation network are participating in most bioprocesses in plants and play a certain role.

4 Discussion

The bZIP transcription factors are among the most widely distributed and conserved families of eukaryotic TFs, and have been more thoroughly researched in the plant field (Li et al., 2016). The bZIP TFs are widely involved in plant developmental process, biotic and abiotic stress response, and regulation of plant secondary metabolite synthesis such as terpenoids (Zhong et al., 2021). *P. grandiflorus* is a species of both medicine and food. Its young shoots and roots are often used to make pickles, cold soups and sauces. Similarly, the dried root of *P. grandiflorus* is a traditional Chinese medicine (TCM) called PLATYCODONIS RADIX, which has a variety of pharmacological effects and is a commonly used bulk of TCM (Ma et al., 2021). In the context of the post-epidemic era, the concept of "preventing disease" has gradually taken root in people's hearts, and the concept of health care based on the same source of medicine and food will be the focus of attention at home and abroad in the future. As one of the representative medicinal materials of the same source of medicine and food, the international requirement for its raw materials is increasing, but the quality of different production areas is different. Therefore, studying the molecular regulatory mechanisms of the developmental process of *P. grandiflorus* to increase its productivity and high quality has emerged as a research hotspot. The bZIP TFs were extensively studied in various plants, but not in *P. grandiflorus*. In this present work, 47 *PgbZIP* genes were recognized by bioinformatics methods, which was similar to most previous reports, such as 49 in *Solanum tuberosum* (Wang Q. et al., 2021; Wang S. et al., 2021), 54 in *Litchi chinensis* (Hou et al., 2022), 59 in *Castanea mollissima* (Zhang et al., 2023), 65 in *Lagenaria siceraria* (Wang et al., 2022) and 65 in *I. tinctoria*. A deeper analysis revealed

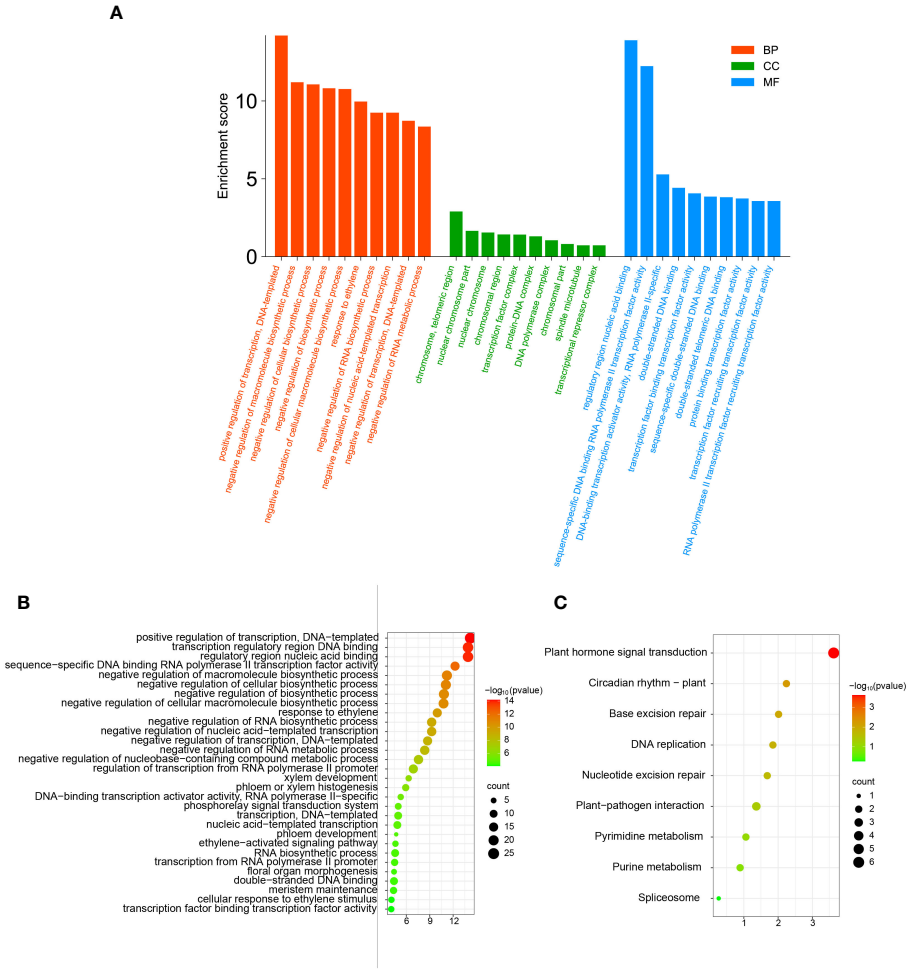


FIGURE 9
The GO annotation and KEGG pathway enrichment analysis of TFs in correlation networks. **(A)** The GO annotation of TFs in correlation networks. **(B)** The GO enrichment analysis of TFs in correlation networks. **(C)** The KEGG pathway enrichment analysis of TFs in correlation networks.

that all *PgbZIP* genes included the bZIP conserved domains, suggesting that the identification results were reliable and accurate. Similarly, motif1 and motif3 comprise a leucine zipper region and base region of the *bZIP* genes, consistent with the identification of bZIP TFs in *Nicotiana tabacum* (Duan et al., 2022).

Phylogenetic analysis can identify homologous genes in different species for the purpose of predicting the function of unknown genes. In this study, a total of 47 PgbZIP protein and 75 A. *thaliana* bZIP protein have been employed to build a phylogenetic trees, which was divided into 11 subfamilies and showed good clustering results. In subfamily A, the PgbZIP6 and PgbZIP19 proteins were homologous with AtbZIP36, and the PgbZIP17 protein was homologous with AtbZIP36, indicating that the three PgbZIP proteins probably have important functions in ABA induction and stress treatment (Choi et al., 2000). Similarly, in subfamily D, AtbZIP46 is participating in the progression of *A. thaliana* flower and plays a role in regulating the size of meristems or floral organ number (Chuang et al., 1999), and its homologous gene *PgbZIP15* probably has the same functionality. Notably, the H subfamily have two genes, namely, *HY5* (*AtbZIP56*) and *HYH* (*AtbZIP64*), while the H subfamily in the *PgbZIP* gene family contains only two *HY5* genes (*PgbZIP40*, *PgbZIP41*) but no *HYH* gene, suggests that there could be possible gene expansions and deletions in the *PgbZIP* gene family that may have occurred during the evolutionary process.

The function of transcription factors is highly correlated with their expression patterns (Wang et al., 2023). Earlier research has indicated that the *bZIP* gens were participating in the developmental processes of various tissues and organs in plants. Coexpression of AtbZIP10/25 with ABI3 significantly increases the activation capacity of the At2S1 promoter to form a regulatory complex for seed-specific expression (Lara et al., 2003). The *AtbZIP9* gene is involved in regulating leaves and vascular bundle development (Silveira et al., 2007). Overexpression of *OsbZIP49* in *O. sativa* reduced internode length and plant height in transgenic rice, which exhibited a tiller-spreading phenotype (Ding et al., 2021). Overexpression of the *Capsicum annuum CabZIP1* gene in *A. thaliana* can slow plant growth and decrease the amount of petals (Lee et al., 2006). To identify the regulatory functions of *PgbZIP* genes in *P. grandiflorus* development, 10 *PgbZIP* genes (*PgbZIP5, PgbZIP13, PgbZIP14, PgbZIP21, PgbZIP25, PgbZIP28, PgbZIP30, PgbZIP33, PgbZIP42* and *PgbZIP45*) that were highly expressed in

the transcriptome data of 8 tissues were selected as potential genes, the relative expression of those genes in stems, leaves, roots, petals and sepals have been analyzed by qRT-PCR. The relative expression of candidate genes were in accordance with the trend of RNA-seq data, suggesting these 10 candidate genes were participated in the development of various tissues of *P. grandiflorus*. Similarly, the *PgbZIP* genes also showed tissue-specific expression in different organs. Transcriptome data showed that *PgbZIP12*, *PgbZIP35*, *PgbZIP37* and *PgbZIP43* were only expressed in seed species, which predicted that they could be participated in seed developmental processes. A comparable tissue-specific expression patterns have been found for the bZIP TFs in species such as *M. domestica* (Wang Q. et al., 2021; Wang S. et al., 2021), *Musa nana* (Hu et al., 2016), and *Citrullus lanatus* (Yang et al., 2019).

In recent years, the biosynthesis of natural products has become increasingly popular. However, the molecular mechanisms by which TFs are involved in regulating the biosynthesis of plant secondary metabolism are complex. The triterpenoid saponins of *P. grandiflorus* are very important secondary metabolites, that are mainly enriched in roots (Zhang F. et al., 2015; Zhang L. et al., 2015). Root development is a complex process regulated by the expression of multiple genes and influenced by endogenous hormone levels and natural resources (Siqueira et al., 2022). During energy deprivation in *A. thaliana*, *AtbZIP63* activates *AtARF19* expression in response to basal lateral root initiation (Muralidhara et al., 2021). Under stress and normal conditions, different levels of ABA (abscisic acid) regulate the development of plant root architecture, including the initiation and elongation of main, adventitious, adventitious roots, and root hairs, as well as root system hydrophilicity and geotropism (Teng et al., 2023). Similarly, bZIP TFs perform important functions in the control of terpenoids; for example, the bZIP TF AaTGA6 in *A. annua* is involved in the regulation of salicylic acid (SA) in the synthesis of artemisinin (Lv et al., 2019), and AabZIP1 is involved in ABA signaling, which in turn regulates artemisinin biosynthesis (Zhang H. et al., 2022; Zhang S. et al., 2022). In this study, *cis*-elements for physiological control, hormonal response and stress response were detected in the promoter region of the *PgbZIP* gene, suggesting that the *PgbZIP* gene with these elements performs an essential function in the regulation of root developmental processes and terpene biosynthesis in *P. grandifloru*.

With the fast speed of modern industrial development and climate change, plants suffer from abiotic stress increasingly frequently in the manner of increase and development, which conduct to a reduction in yield, quality damage, and even plant death (Zhang et al., 2022). To cope with the stress caused by adverse environments, plants have evolved various mechanisms, including the regulation of gene expression through various transcription factors, so that plants can adapt to or escape the effects of stress (Badis et al., 2009). Among these transcription factors, the bZIP TFs has been widely reported to enhance the ability of response to biotic and abiotic stresses. For example, *OsbZIP71* gene in *O. sativa* enhances the tolerance to drought and salinity by activating the expression of the OsNHX1 protein and COR413-TM1 protein (Liu et al., 2014). In *V. vinifera*, the *VvbZIP23* expression is induced by a number of abiotic stresses, including cold, salinity and drought stress (Tak and Mhatre, 2013). The *Hordeum vulgare* bZIP TF HvABI5 is participating in ABA-dependent regulation of resistance to drought stress (Collin et al., 2020). In this study, *PgbZIP5*, *PgbZIP21*, *PgbZIP25* and *PgbZIP28* genes with significantly considerably elevated expression levels beneath drought and salinity stress were verified through qRT-PCR experiments. Additionally, *PgbZIP33* was only highly expressed under salt stress. Among them, *PgbZIP5* is the homologous gene of *AtbZIP11*, which may have the same features in resistance to stress (Weltmeier et al., 2009). *PgbZIP21* and *PgbZIP28* have *cis*-elements participating in defense and stress responses, and *PgbZIP21*, *PgbZIP25*, and *PgbZIP33* have binding sites for drought regulation with *PgMYB* genes. *PgbZIP28* and *PgbZIP33* also contain low-temperature responsiveness elements. These 5 genes also have co-expression links with *bHLH, GATA, MYB* of the stress resistance related genes. It is hypothesized that those TFs might perform key functions in supporting *P. grandiflorus* to resist abiotic stress. However, the specific biological functions of these genes need to be further verified.

5 Conclusion

In this work, a number of 47 *PgbZIP* genes were characterized based on genome-wide analysis, and their chromosomal distribution, phylogenetic, motifs, gene structure, *cis*-element prediction, synteny and expression profiles have been comprehensively analyzed. Ten candidate *PgbZIP* genes that are expressed at high levels in all tissues indicate their crucial role in various physiological and biological processes, and the co-express network results also provide evidence for further research. The expression pattern of candidate genes under drought and salt stress provide valuable information for the expression of *PgbZIP* genes under salt and drought stress. These results provide clues to investigate the functions of the *PgbZIP* genes in the developmental processes of different plant organs and abiotic stress responses.

Data availability statement

This study analyzed publicly available data from the NCBI database (https://www.ncbi.nlm.nih.gov/). Accession numbers: SRR8712510-SRR8712517.

Author contributions

ZW: Investigation, Software, Writing – original draft, Writing – review & editing. PW: Data curation, Methodology, Writing – original draft. HC: Formal analysis, Software, Writing – original draft. ML: Investigation, Software, Writing – original draft. LK: Investigation, Methodology, Writing – original draft. HW: Resources, Writing – original draft. WR: Resources, Software, Writing – original draft. QF: Formal analysis, Supervision, Validation, Writing – original draft. WM: Funding acquisition, Resources, Visualization, Writing – original draft, 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 Mechanism of WRKY transcription factor regulating Platycodin Biosynthesis (CZKYF2022-F04). National Key Research and development Project, research and demonstration of collection, screening and breeding technology of ginseng and other genuine medicinal materials (2021YFD1600901); Talent training project supported by the central government for the reform and development of local colleges and Universities (ZYRCB2021008); Heilongjiang Touyan Innovation Team Program (HLJTYTP2019001).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1403220/full#supplementary-material

References

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, A. S., et al. (2009). Diversity and complexity in dna recognition by transcription factors. *Science* 324, 1720–1723. doi: 10.1126/science.1162327

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, W202–W208. doi: 10.1093/nar/gkp335

Bi, C., Yu, Y., Dong, C., Yang, Y., Zhai, Y., Du, F., et al. (2021). The bzip transcription factor tabzip15 improves salt stress tolerance in wheat. *Plant Biotechnol. J.* 19, 209–211. doi: 10.1111/pbi.13453

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

Choi, H., Hong, J., Ha, J., Kang, J., and Kim, S. Y. (2000). Abfs, a family of aba-responsive element binding factors. *J. Biol. Chem.* 275, 1723–1730. doi: 10.1074/jbc.275.3.1723

Chuang, C. F., Running, M. P., Williams, R. W., and Meyerowitz, E. M. (1999). The perianthia gene encodes a bzip protein involved in the determination of floral organ number in arabidopsis thaliana. *Genes Dev.* 13, 334–344. doi: 10.1101/gad.13.3.334

Collin, A., Daszkowska-Golec, A., Kurowska, M., and Szarejko, I. (2020). Barley abi5 (abscisic acid insensitive 5) is involved in abscisic acid-dependent drought response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01138

Ding, C., Lin, X., Zuo, Y., Yu, Z., Baerson, S. R., Pan, Z., et al. (2021). Transcription factor osbzip49 controls tiller angle and plant architecture through the induction of indole-3-acetic acid-amido synthetases in rice. *Plant J.* 108, 1346–1364. doi: 10.1111/tpj.15515

Duan, L., Mo, Z., Fan, Y., Li, K., Yang, M., Li, D., et al. (2022). Genome-wide identification and expression analysis of the bzip transcription factor family genes in response to abiotic stress in nicotiana tabacum l. *BMC Genomics* 23, 318. doi: 10.1186/s12864-022-08547-z

Gai, W., Ma, X., Qiao, Y., Shi, B., Ul Haq, S., Li, Q., et al. (2020). Characterization of the bzip transcription factor family in pepper (capsicum annuum l.): Cabzip25 positively modulates the salt tolerance. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00139

Gangappa, S. N., and Botto, J. F. (2016). The multifaceted roles of hy5 in plant growth and development. *Mol. Plant* 9, 1353–1365. doi: 10.1016/j.molp.2016.07.002

Hou, H., Kong, X., Zhou, Y., Yin, C., Jiang, Y., Qu, H., et al. (2022). Genome-wide identification and characterization of bzip transcription factors in relation to litchi (litchi chinensis sonn.) Fruit ripening and postharvest storage. *Int. J. Biol. Macromol* 222, 2176–2189. doi: 10.1016/j.ijbiomac.2022.09.292

Hu, W., Wang, L., Tie, W., Yan, Y., Ding, Z., Liu, J., et al. (2016). Genome-wide analyses of the bzip family reveal their involvement in the development, ripening and abiotic stress response in banana. *Sci. Rep.* 6, 30203. doi: 10.1038/srep30203

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). Bzip transcription factors in arabidopsis. *Trends Plant Sci.* 7, 106–111. doi: 10.1016/S1360-1385(01)02223-3

Ji, M., Bo, A., Yang, M., Xu, J., Jiang, L., Zhou, B., et al. (2020). The pharmacological effects and health benefits of platycodon grandiflorus—a medicine food homology species. *Foods* 9, 142. doi: 10.3390/foods9020142

Jiang, M., Wang, Z., Ren, W., Yan, S., Xing, N., Zhang, Z., et al. (2022). Identification of the bzip gene family and regulation of metabolites under salt stress in isatis indigotica. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1011616

Jing, L., Hanwen, Y., Mengli, L., Bowen, C., Nan, D., Chang, X., et al. (2022). Transcriptome-wide identification of wrky transcription factors and their expression profiles in response to methyl jasmonate in platycodon grandiflorus. *Plant Signaling Behav.* 17, 2089473. doi: 10.1080/15592324.2022.2089473

Kim, J., Kang, S. H., Park, S. G., Yang, T. J., Lee, Y., Kim, O. T., et al. (2020). Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in platycodon grandiflorus, a medicinal plant. *Hortic. Res.* 7, 112. doi: 10.1038/s41438-020-0329-x

Kong, L., Sun, J., Jiang, Z., Ren, W., Wang, Z., Zhang, M., et al. (2023). Identification and expression analysis of yabby family genes in platycodon grandiflorus. *Plant Signal Behav.* 18, 2163069. doi: 10.1080/15592324.2022.2163069

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, 1547–1549. doi: 10.1093/molbev/msy096

Lan, Y., Pan, F., Zhang, K., Wang, L., Liu, H., Jiang, C., et al. (2023). Phebzip47, a bzip transcription factor from moso bamboo (phyllostachys edulis), positively regulates the drought tolerance of transgenic plants. *Ind. Crops Products* 197. doi: 10.1016/J.INDCROP.2023.116538

Lara, P., Oñate-Sánchez, L., Abraham, Z., Ferrándiz, C., Díaz, I., Carbonero, P., et al. (2003). Synergistic activation of seed storage protein gene expression in arabidopsis by abi3 and two bzips related to opaque2. *J. Biol. Chem.* 278, 21003–21011. doi: 10.1074/jbc.M210538200

Lee, D. J., Choi, J. W., Kang, J. N., Lee, S. M., Park, G. H., and Kim, C. K. (2023). Chromosome-scale genome assembly and triterpenoid saponin biosynthesis in korean bellflower (platycodon grandiflorum). *Int. J. Mol. Sci.* 24, 6534. doi: 10.3390/ijms24076534

Lee, S. C., Choi, H. W., Hwang, I. S., Choi, D. S., and Hwang, B. K. (2006). Functional roles of the pepper pathogen-induced bzip transcription factor, cabzip1, in enhanced resistance to pathogen infection and environmental stresses. *Planta* 224, 1209–1225. doi: 10.1007/s00425-006-0302-4

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., 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, 325–327. doi: 10.1093/nar/30.1.325

Li, Y., Meng, D., Li, M., and Cheng, L. (2016). Genome-wide identification and expression analysis of the bzip gene family in apple (malus domestica). *Tree Genet. Genomes* 12, 82. doi: 10.1007/s11295-016-1043-6

Liu, M., Liu, T., Liu, W., Wang, Z., Kong, L., Lu, J., et al. (2023). Genome-wide identification and expression profiling analysis of the trihelix gene family and response of pggt1 under abiotic stresses in platycodon grandiflorus. *Gene* 869, 147398. doi: 10.1016/j.gene.2023.147398

Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., et al. (2014). Osbzip71, a bzip transcription factor, confers salinity and drought tolerance in rice. *Plant Mol. Biol.* 84, 19–36. doi: 10.1007/s11103-013-0115-3

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

Lv, Z., Guo, Z., Zhang, L., Zhang, F., Jiang, W., Shen, Q., et al. (2019). Interaction of bzip transcription factor tga6 with salicylic acid signaling modulates artemisinin biosynthesis in artemisia annua. *J. Exp. Bot.* 70, 3969–3979. doi: 10.1093/jxb/erz166

Ma, C., Gao, Z., Zhang, J., Zhang, W., Shao, J., Hai, M., et al. (2016). Candidate genes involved in the biosynthesis of triterpenoid saponins in platycodon grandiflorum identified by transcriptome analysis. *Front. Plant Sci.* 7, 673. doi: 10.3389/fpls.2016.00673

Ma, X., Shao, S., Xiao, F., Zhang, H., and Yan, M. (2021). Platycodon grandiflorum extract: chemical composition and whitening, antioxidant, and anti-inflammatory effects. *Rsc Adv.* 11, 10814–10826. doi: 10.1039/D0RA09443A

Muralidhara, P., Weiste, C., Collani, S., Krischke, M., Kreisz, P., Draken, J., et al. (2021). Perturbations in plant energy homeostasis prime lateral root initiation via snrk1-bzip63-arf19 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2106961118. doi: 10.1073/pnas.2106961118

Nijhawan, A., Jain, M., Tyagi, A. K., and Khurana, J. P. (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* 146, 333–350. doi: 10.1104/pp.107.112821

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, 2498–2504. doi: 10.1101/gr.1239303

Sharma, P. D., Singh, N., Ahuja, P. S., and Reddy, T. V. (2011). Abscisic acid response element binding factor 1 is required for establishment of arabidopsis seedlings during winter. *Mol. Biol. Rep.* 38, 5147–5159. doi: 10.1007/s11033-010-0664-3

Shen, Q., Huang, H., Zhao, Y., Xie, L., He, Q., Zhong, Y., et al. (2019). The transcription factor aabzip9 positively regulates the biosynthesis of artemisinin in artemisia annua. *Front. Plant Sci.* 10, 1294. doi: 10.3389/fpls.2019.01294

Silveira, A. B., Gauer, L., Tomaz, J. P., Cardoso, P. R., Carmello-Guerreiro, S., and Vincentz, M. (2007). The arabidopsis atbzip9 protein fused to the vp16 transcriptional activation domain alters leaf and vascular development. *Plant Ence* 172, 1148–1156. doi: 10.1016/j.plantsci.2007.03.003

Siqueira, J. A., Otoni, W. C., and Araújo, W. L. (2022). The hidden half comes into the spotlight: peeking inside the black box of root developmental phases. *Plant Commun.* 3, 100246. doi: 10.1016/j.xplc.2021.100246

Tak, H., and Mhatre, M. (2013). Cloning and molecular characterization of a putative bzip transcription factor vvbzip23 from vitis vinifera. *Protoplasma* 250, 333–345. doi: 10.1007/s00709-012-0417-3

Teng, Z., Lv, J., Chen, Y., Zhang, J., and Ye, N. (2023). Effects of stress-induced aba on root architecture development: positive and negative actions. *Crop J.* 11, 1072–1079. doi: 10.1016/j.cj.2023.06.007

Tian, J. H., Shu, J. X., Xu, L., Ya, H., Yang, L., Liu, S., et al. (2023). Genome-wide identification of r2r3-myb transcription factor family in *docynia delavayi* (franch.) Schneid and its expression analysis during the fruit development. *Food Bioscience* 54, 102878. doi: 10.1016/j.fbio.2023.102878

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remmet, M., et al. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. doi: 10.1093/nar/gks596

Wang, Q., Guo, C., Li, Z., and Guo, Y. (2021). Identification and analysis of bzip family genes in potato and their potential roles in stress responses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.637343

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, e49. doi: 10.1093/nar/gkr1293

Wang, J., Wang, Y., Wu, X., Wang, B., Lu, Z., Zhong, L., et al. (2022). Insight into the bzip gene family in lagenaria siceraria: genome and transcriptome analysis to understand gene diversification in cucurbitaceae and the roles of lsbzip gene expression and function under cold stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1128007

Wang, Z., Zhang, Z., Wang, P., Qin, C., He, L., Kong, L., et al. (2023). Genome-wide identification of the nac transcription factors family and regulation of metabolites under salt stress in isatis indigotica. *Int. J. Biol. Macromol* 240, 124436. doi: 10.1016/j.ijbiomac.2023.124436

Wang, S., Zhang, R., Zhang, Z., Zhao, T., Zhang, D., Sofkova, S., et al. (2021). Genome-wide analysis of the bzip gene lineage in apple and functional analysis of mhabf in malus halliana. *Planta* 254, 78. doi: 10.1007/s00425-021-03724-y

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schütze, K., Wamg, X., et al. (2009). Expression patterns within the arabidopsis c/s1 bzip transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol. Biol.* 69, 107–119. doi: 10.1007/s11103-008-9410-9

Yang, Y., Li, J., Li, H., Yang, Y., Guang, Y., Zhou, Y., et al. (2019). The bzip gene family in watermelon: genome-wide identification and expression analysis under cold stress and root-knot nematode infection. *Peerj* 7, e7878. doi: 10.7717/peerj.7878

Zhang, S., Chai, X., Hou, G., Zhao, F., and Meng, Q. (2022). Platycodon grandiflorum (jacq.) A. Dc.: A review of phytochemistry, pharmacology, toxicology and traditional use. *Phytomedicine* 106, 154422. doi: 10.1016/j.phymed.2022.154422

Zhang, F., Fu, X., Lv, Z., Lu, X., Shen, Q., Zhang, M., et al. (2015). A basic leucine zipper transcription factor, aabzip1, connects abscisic acid signaling with artemisinin biosynthesis in artemisia annua. *Mol. Plant* 8, 163–175. doi: 10.1016/j.molp.2014.12.004

Zhang, P., Liu, J., Jia, N., Wang, M., Lu, Y., Wang, D., et al. (2023). Genome-wide identification and characterization of the bzip gene family and their function in starch accumulation in chinese chestnut (castanea mollissima blume). *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1166717

Zhang, T., Song, C., Song, L., Shang, Z., Yang, S., Wang, D., et al. (2017). Rna sequencing and coexpression analysis reveal key genes involved in α-linolenic acid biosynthesis in perilla frutescens seed. *Int. J. Mol. Sci.* 18, 2433. doi: 10.3390/ijms18112433

Zhang, L., Wang, Y., Yang, D., Zhang, C., Zhang, N., Li, M., et al. (2015). Platycodon grandiflorus - an ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol* 164, 147–161. doi: 10.1016/j.jep.2015.01.052

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

Zhao, J., Guo, R., Guo, C., Hou, H., Wang, X., and Gao, H. (2016b). Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00376

Zhong, X. J., Feng, X., Li, Y., Guzmán, C., and Jiang, Q. (2021). Genome-wide identification of the bzip transcription factor genes related to starch synthesis in barley (hordeum vulgare l.). *Genome* 12, 1067–1080. doi: 10.1139/gen-2020-0195

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 June 2024
DOI 10.3389/fpls.2024.1407700

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Wei-Min Tian,
Chinese Academy of Tropical Agricultural Sciences, China
Lei Wang,
Huazhong Agricultural University, China

*CORRESPONDENCE
Hao-Fu Dai
daihaofu@itbb.org.cn
Shi-Qing Peng
shqpeng@163.com

†These authors have contributed equally to this work

RECEIVED 27 March 2024
ACCEPTED 30 May 2024
PUBLISHED 24 June 2024

CITATION
Li H-L, Wang Y, Guo D, Zhu J-H, Wang Y, Dai H-F and Peng S-Q (2024) Reprogramming of DNA methylation and changes of gene expression in grafted *Hevea brasiliensis*. *Front. Plant Sci.* 15:1407700. doi: 10.3389/fpls.2024.1407700

Reprogramming of DNA methylation and changes of gene expression in grafted *Hevea brasiliensis*

Hui-Liang Li[1,2,3†], Ying Wang[1†], Dong Guo[1,2], Jia-Hong Zhu[1,2], Yu Wang[1,2,3], Hao-Fu Dai[1*] and Shi-Qing Peng[1*]

[1]Key Laboratory of Biology and Genetic Resources of Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences and Key Laboratory for Biology and Genetic Resources of Hainan Province, Hainan Academy of Tropical Agricultural Resource, Haikou, Hainan, China, [2]National Key Laboratory for Tropical Crop Breeding and Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya, Hainan, China, [3]Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

Rubber tree (*Hevea brasiliensis*) is reproduced by bud grafting for commercial planting, but significant intraclonal variations exist in bud-grafted clones. DNA methylation changes related to grafting may be partly responsible for intraclonal variations. In the current study, whole-genome DNA methylation profiles of grafted rubber tree plants (GPs) and their donor plants (DPs) were evaluated by whole-genome bisulfite sequencing. Data showed that DNA methylation was downregulated and DNA methylations in CG, CHG, and CHH sequences were reprogrammed in GPs, suggesting that grafting induced the reprogramming of DNA methylation. A total of 5,939 differentially methylated genes (DMGs) were identified by comparing fractional methylation levels between GPs and DPs. Transcriptional analysis revealed that there were 9,798 differentially expressed genes (DEGs) in the DP and GP comparison. A total of 1,698 overlapping genes between DEGs and DMGs were identified. These overlapping genes were markedly enriched in the metabolic pathway and biosynthesis of secondary metabolites by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Global DNA methylation and transcriptional analyses revealed that reprogramming of DNA methylation is correlated with gene expression in grafted rubber trees. The study provides a whole-genome methylome of rubber trees and an insight into the molecular mechanisms underlying the intraclonal variations existing in the commercial planting of grafted rubber trees.

KEYWORDS

***Hevea brasiliensis*, grafting, whole genome bisulfite sequencing, DNA methylation, gene expression**

1 Introduction

Rubber tree (*Hevea brasiliensis* Muell. Arg.) is the most unique commercial rubber-producing plant in the world. The propagation of rubber trees for commercial planting and conservation is being performed via artificial grafting scions (axillary buds) of elite clones onto rootstocks (unselected seedlings) to maintain the traits of interest (Hua et al., 2010). However, significant intraclonal variations exist in the growth and rubber yield of the bud-grafted clones, which is mainly attributed to the genetic heterogeneity of the stocks used for propagation (Chandrashekar et al., 1997; Zhou and Lin, 2012; Zhou et al., 2014). Although a great deal of effort has been invested in understanding the stock–scion interactions, most of the disparities reported were in terms of biochemical and phenotypic parameters (Huang and Lin, 2003; Ding et al., 2010; Zhou et al., 2014). The molecular mechanism of rootstock affecting scion growth and characteristics has not been fully understood in rubber trees.

DNA methylation, occurring in CG, CHG, and CHH contexts (where H represents C, T, or A), is one of the main epigenetic modifications. The establishment and maintenance of DNA methylation need several DNA methyltransferases, such as methyltransferase 1, DNA methyltransferase, and chromomethylases 2 and 3 (Law and Jacobsen, 2010; Zhang et al., 2018; Gallego-Bartolomé, 2020). The methylation of DNA plays an important role in plant cells, such as maintaining plant genome stability, chromosome interactions, gene expression, circRNA biogenesis, and mRNA stability and splicing (Law and Jacobsen, 2010; Zhang et al., 2018; Gallego-Bartolomé, 2020). DNA methylation affects plant physiology and development and regulates multiple biological processes, such as fertilization, floral pigmentation, floral scent, and photosynthesis as well as biotic or abiotic stress (Zhang et al., 2018; Gallego-Bartolomé, 2020; Mattei et al., 2022).

Grafting is a traditional method of plant asexual propagation by connecting scion and rootstock. In recent years, some studies have proposed the existence of DNA methylation changes in grafting reactions. The graft-induced DNA methylation in the scion has been connected with variations in grafted plants (Wu et al., 2013; Perrin et al., 2020; Santos et al., 2020; Cerruti et al., 2021; Kapazoglou et al., 2021; Han et al., 2022; Liu et al., 2023). Although grafting caused methylation variations in rubber trees, which were detected using methylation-sensitive amplified polymorphism (MSAP; Uthup et al., 2018), whole-genome DNA methylation by grafting in rubber trees remains largely unknown. Whole-genome bisulfite sequencing (WGBS) is a high-throughput and precise method for DNA methylation analysis, which can identify DNA methylation patterns genome-wide at single-base resolution (Smallwood et al., 2014; Li et al., 2018). Here, the genome-wide DNA methylation patterns between the grafted rubber tree plant (GP) and its donor plant (DP) were investigated using WGBS technology. Our findings reveal that grafting induced reprogramming of DNA methylation in GPs, with evidence suggesting a correlation between DNA methylation and gene expression. The study indicates that a potential molecular mechanism influences intraclonal variations existing in the commercial planting of grafted rubber trees.

2 Materials and methods

2.1 Plant material

Ten-year-old plants of tissue-cultured *H. brasiliensis* Muell. Arg. cultivar (CATA7–33-97) were planted in the National Rubber Tree Varieties Resource Garden of the Chinese Academy of Tropical Agriculture Sciences, Danzhou, Hainan, China. The seeds of the rubber trees (CATA7–33-97) were collected and germinated to produce seedlings, which were subsequently utilized as rootstocks after 1 year. The axillary buds of tissue-cultured rubber trees (CATA7–33-97) were taken and grafted onto the aforementioned rootstocks of the CATA7–33-97 variety. Mature leaf samples were collected from the 15 GPs when they grew to a stable phenology of the third canopy leaf and were pooled together to form one sample for DNA and RNA extraction. Concurrently, mature leaf samples were collected from the 15 DPs of tissue-cultured rubber trees (CATA7–33-97) and processed for DNA and RNA extraction to serve as a comparative control sample.

2.2 Whole-genome bisulfite sequencing

The isolation of genomic DNA was carried out as per the instruction of the Plant Genomic DNA Extraction Kit (TaKaRa, Dalian, China). Genomic DNA was checked on agarose gels. The purity and concentration of DNA were monitored using a NanoPhotometer® spectrophotometer (IMPLEN, Westlake Village, CA, USA) and Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

For library construction, 200–300-bp DNA fragments were obtained by genomic DNA sonication, and then DNA fragments were combined with end-repair mix and adenylation. Cytosine-methylated adapters were ligated to the repaired DNAs, and subsequently, the connections were purified. The purified samples were treated with bisulfite as per the instruction of the Methylation Gold Kit (Zymo Research, Irvine, CA, USA). The treated DNA fragments were amplified to construct a DNA library. The library was qualified by the Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and real-time PCR system (Bio-Rad, Hercules, CA, USA). The qualified libraries were sequenced on an Illumina NovaSeq platform (Novogene, Beijing, China).

2.3 Differentially methylated region analysis

The adaptor sequences in WGBS reads were removed using Trimmomatic (Bolger et al., 2014). The trimmed clean reads were aligned against the rubber tree reference genome (Liu et al., 2020)

using Bismark (Krueger and Andrews, 2011). Methylated cytosines were extracted from aligned reads using the Bismark methylation extractor with default parameters. Considering inefficiencies in the bisulfite conversion reaction and sequencing errors, a binomial test was used to determine whether the observed methylation frequency was above the background [false discovery rate (FDR) <0.05]. The methylation level was assessed according to the previously described method (Lister et al., 2009). The different DNA methylation profiles in different genomic regions were identified according to Lister et al. (2013). Differentially methylated regions (DMRs) between GPs and DPs were determined using the R-package DSS-single (Feng et al., 2014; Wu et al., 2015). The CallDMR function was used to identify DMRs with parameters of twofold change and $p \leq 0.05$ in methylation levels.

2.4 Transcriptome sequencing

Total RNA was extracted according to Tang et al. (2010). Total RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). Total RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Paired-end Illumina cDNA Sequencing libraries were generated following the manufacturer's instructions for mRNASeq sample preparation (NEB, Ipswich, MA, USA). The library quality was assessed using the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The libraries were deep sequenced using the Illumina NovaSeq platform (Novogene, Beijing, China), and 150-bp paired-end reads were generated. Raw data (raw reads) of fastq format were first processed through in-house perl scripts. Clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N, and low-quality reads from raw data. At the same time, Q20, Q30, and GC content of the clean data were calculated. The clean reads were aligned against the rubber tree reference genome (Liu et al., 2020) using HISAT2 v2.0.5 (Kim et al., 2019), then transcripts were assembled, and gene expression levels were calculated using StringTie v1.3.3 (Pertea et al., 2016). The transcript abundance was calculated based on the ratio of fragments per kilobase of transcript per million mapped reads (FPKM) values, and the FDR (control method for the p-value threshold in multiple tests) was used for testing the significance of the differences. DESeq2 R package (v1.16.1) was used to identify differentially expressed genes (DEGs) with cutoff values of FDR ≤ 0.05 and |fold change| ≥ 2 (Love et al., 2014).

2.5 Integrated analysis of DNA methylation and the transcriptome

Pearson's correlation analysis of DNA methylation and gene expression in DPs and GPs was completed as previously described (Yang et al., 2014). DMRs (<5% false discovery rate) are associated with significant changes in gene expression and enriched for an expected inverse relationship between methylation and expression ($p < 2.2 \times 10^{-16}$).

2.6 Functional enrichment analysis of overlapping genes between DEGs and DMGs

Gene Ontology (GO) enrichment analysis of overlapping genes between DEGs and differentially methylated genes (DMGs) was carried out using the GOseq R package at $p \leq 0.05$ (Young et al., 2010). The statistical enrichment of overlapping genes between DEGs and DMGs in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed using KOBAS software with Fisher's exact test selecting $p \leq 0.05$ (Mao et al., 2005).

2.7 Quantitative PCR

The total RNA of rubber tree leaves was extracted according to Tang et al. (2010). Quantitative PCR (qPCR) was performed using *HbACT7* as an internal reference gene (Li et al., 2014) according to the instruction of SYBR®Premix Ex Taq™ II Kit (TaKaRa, Dalian, China). The sequence of primers used in this study is listed in Supplementary Table S1. qPCR conditions were as follows: 95°C for 3 min; 35 cycles of 95°C for 5 s, 60°C for 30 s, 72°C for 20 s. The gene expression level was calculated using the 2(−Delta Delta C (T)) method (Livak and Schmittgen, 2001).

3 Results

3.1 Grafting induced DNA methylation change in *H. brasiliensis*

To clarify the possible effects of rubber trees with grafting on DNA methylation levels, WGBS was performed to investigate whole-genome methylation in GPs and DPs. In total, 204,849,774 clean reads (56.64 G) and 220,533,308 clean reads (60.71 G) were obtained from GPs and DPs. More than 99.23% of DPs and 99.55% of GPs of cytosines were converted by bisulfite treatment, which showed that the conversion was satisfactory with WGBS. In total, 123,873,186 clean reads in DPs and 119,464,270 clean reads in GPs were mapped to the rubber tree reference genome (Table 1). The genome-wide methylation for chromosomes and distinct genomic elements were characterized in DPs and GPs (Figures 1A, B). The proportions of methylated cytosines (mC) decreased from 26.2% in DPs to 16.48% in GPs (Table 1). The mean methylation levels decreased from 77% in DPs to 64.86% in GPs in the CG context, from 72.02% in DPs to 59.83% in GPs in the CHG, and from 11.8% in DPs to 4.36% in GPs in the CHH (Figure 1C). Relative proportions of mC in the CG, CHG, and CHH contexts were respectively 16%, 30.19%, and 53.81% in DPs. In GPs, 21.09%, 40.33%, and 38.75% of mC in CG, CHG, and CHH, respectively, were detected (Figure 1D). CHH methylation significantly decreased from 53.81% in DPs to 38.75% in GPs, while CG and CHG methylation increased from 16% and 30.19% in DPs to 21.09% and 40.33% in GPs, respectively. The above data suggested that grafting induced reprogramming of DNA methylation in grafted rubber trees.

10.3389/fpls.2024.1407700

TABLE 1 Data summary of whole-genome bisulfite sequencing for the two rubber tree samples.

Sample	Clean reads	Clean bases	BS conversion rate	Mapped reads	Mean C mC/C (%)	Mean CG mCG/CG (%)	Mean CHG mCHG/CHG (%)	Mean CHH mCHH/CHH (%)	mCG percent (%) mCG/mC	mCHG percent (%) mCHG/mC	mCHH percent (%) mCHH/mC
DP	204,849,774	56.64 G	99.23%	123,873,186	26.2%	77%	72.03%	11.8%	16%	30.19%	53.81%
GP	220,533,308	60.71 G	99.55%	119,464,270	16.48%	64.86%	59.83%	4.36%	21.09%	40.33%	38.57%

DP, donor plant; GP, grafted rubber tree plant.

3.2 Grafting changed methylation levels in gene regions

The analysis of DNA methylation levels in different gene regions indicated that DNA methylation levels in various gene regions had significant differences between GPs and DPs (Figure 2A). The DNA methylation levels in three contexts of gene regions except for 3′UTR in DPs were more than those in GPs (Figure 2B). Meanwhile, DNA methylation levels in all contexts of gene-body regions, upstream 2K, and downstream 2K were lower in the GPs than in the DPs (Figure 2C). Taken together, grafting changed methylation levels in gene regions.

3.3 Analysis of DMGs between GPs and DPs

DMRs and DMGs were identified by comparing fractional methylation levels between GPs and DPs. A total of 7,570 (819 hyper- and 6,751 hypomethylated) DMRs were identified. According to the location of DMRs, DMGs were categorized into DMR-associated promoter genes and DMR-associated genes. A total of 4,894 DMR-associated genes and 2,676 DMR-associated promoter genes were observed between GPs and DPs (Figures 3A, B). For mCG, 2,135 DMR-associated genes and 596 DMR-associated promoter genes were obtained in the comparison. For mCHG, there were 2,129 DMR-associated genes and 719 DMR-associated promoter genes between GPs and DPs. For mCHH, there were 1,487 DMR-associated genes and 1,548 DMR-associated promoter genes between GPs and DPs, respectively. In DMR-associated genes, the number of DMGs in mCG was the highest in all contexts, and the number of DMGs in mCHH was the lowest in all contexts. In DMR-associated promoter genes, the number of DMGs in mCHH was the highest in all contexts, and the number of DMGs in mCG was the lowest in all contexts. A heat map was also used to visualize the DMGs and their DNA methylation levels in three contexts between GPs and DPs (Figure 4A). Additionally, the methylation levels in the three contexts of the DMGs were obviously lower in the GPs than in the DPs (Figure 4B). The DMR distribution in three contexts in the genic regions varied in GPs and DPs. For the CG and CHG contexts, there were considerably more hypomethylated DMRs than hypermethylated DMRs in all gene regions in GPs and DPs. For the CHH context, there were few hypomethylated DMRs in all gene regions in GPs and DPs (Figure 4C). Thus, the DMR distribution and DNA methylation levels in genic regions may be regulated by grafting. Moreover, the IGV snapshots of the DMGs in GPs and DPs are presented in Figure 5. These results suggested that there were many differentially methylated genes between GPs and DPs.

3.4 Grafting induced changes of gene expression in GPs

To study the correlation of gene expression and overlapping DMGs, the transcriptome analysis of GPs and DPs was accomplished. In total, 45,078,466 clean reads (6.76 G) and 44,625,118 clean reads (6.69 G) were acquired from DPs and GPs

10.3389/fpls.2024.1407700

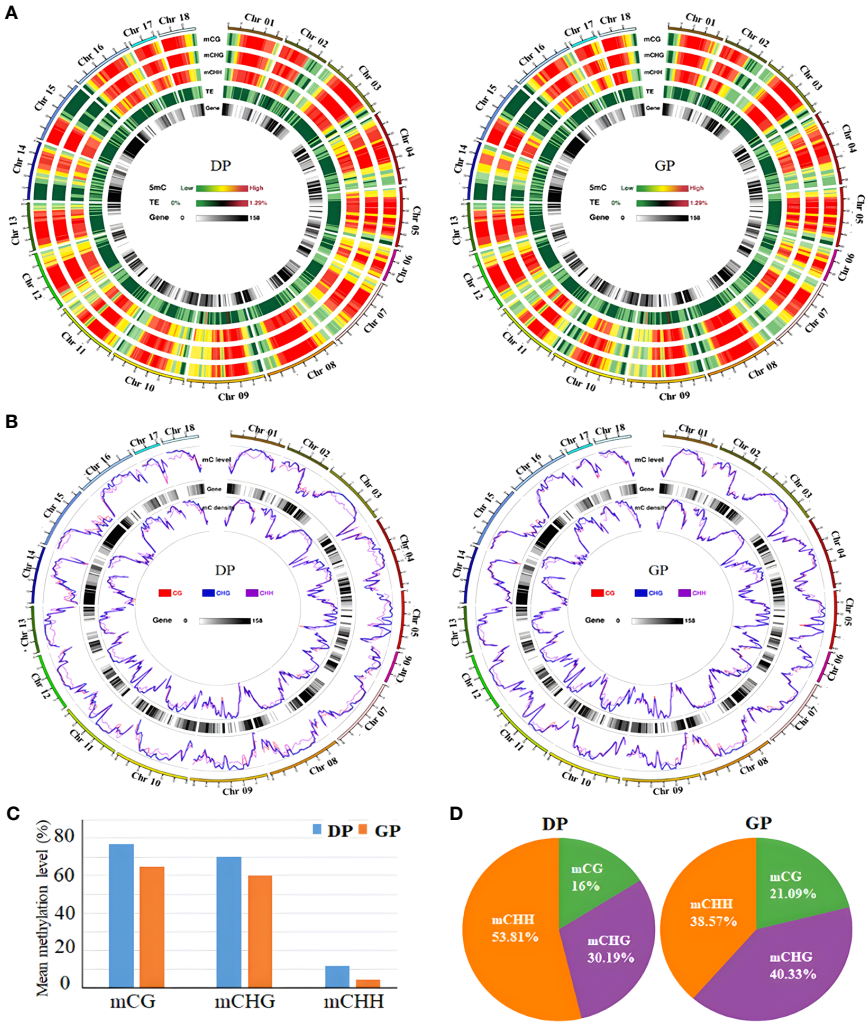


FIGURE 1
Grafting induced DNA methylation change in *Hevea brasiliensis*. **(A)** Chromosome-level methylation features in DPs and GPs. Track from the outside to the inside, as follows: mCG, mCHG, mCHH, TEs, and gene. **(B)** Density plot of 5mC in three contexts in the gene bodies on each chromosome in DPs and GPs. **(C)** The mean methylation levels in three contexts between DPs and GPs. **(D)** The levels of mC in CG, CHG, and CHH in DPs and GPs. DPs, donor plants; GPs, grafted rubber tree plants; TEs, transposable elements.

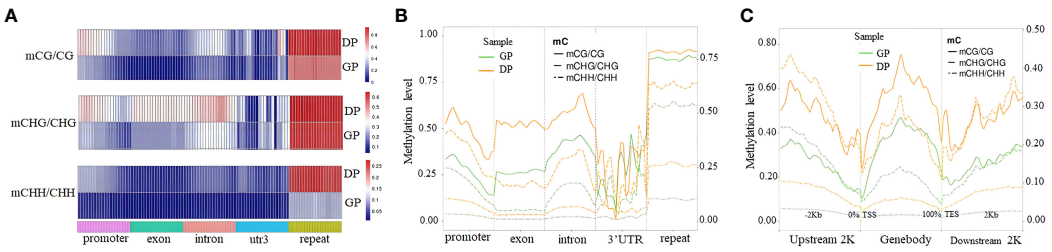


FIGURE 2
DNA methylation profiles in different elements of genes. **(A)** Heat map of methylation levels in three contexts of different elements of genes. **(B)** Distribution of DNA methylation levels in three contexts among elements of genes. **(C)** Methylation levels in three contexts in 2-kb upstream region, gene body, and downstream region in DPs and GPs. DPs, donor plants; GPs, grafted rubber tree plants.

10.3389/fpls.2024.1407700

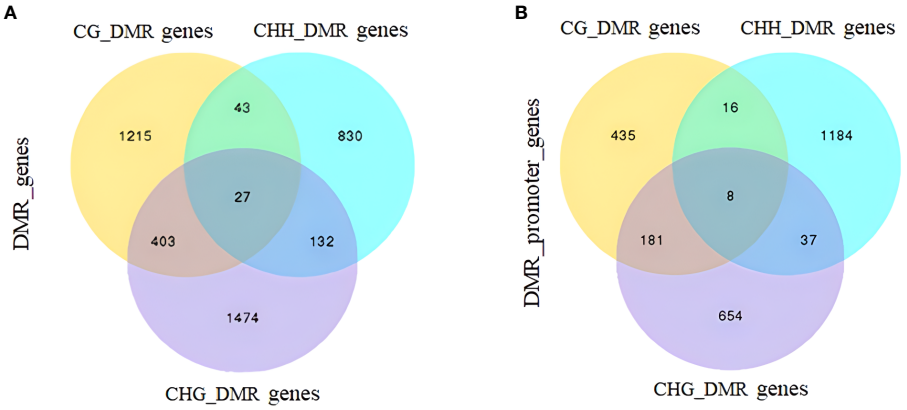


FIGURE 3
Venn analysis of DMGs between DPs and GPs. **(A)** Gene-body regions (DMR_genes). **(B)** The promoter regions (DMR_promoter_genes). DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants.

(Supplementary Table S2), respectively. A total of 9,798 DEGs were identified in the DP and GP comparison. Among DEGs, 5,976 DEGs were upregulated and 3,822 DEGs were downregulated (Figure 6, Supplementary Table S3).

3.5 DMRs correlated with gene expression levels in GPs and DPs

To investigate the relationship between methylation patterns and gene expression levels in rubber tree, genes were classified into the following quartiles based on their FPKM values: no expression (FPKM < 1), low expression (1 < FPKM < the lower quartile), medium expression (the lower quartile < FPKM < the upper quartile), and high expression (FPKM > the upper quartile). Highly expressed genes exhibited lower methylation levels in the CHG and CHH contexts in the gene-body and downstream regions but higher methylation levels in the CHH context within the upstream regions. The non-expressed genes displayed high methylation in CHG within the gene-body, upstream, and downstream regions and low methylation in the CG context within the gene-body regions. Furthermore, non-expressed genes showed lower methylation in CG in the gene-body regions but high methylation in CHG in the gene-body regions. Additionally, the non-expressed genes were highly methylated in all three contexts in

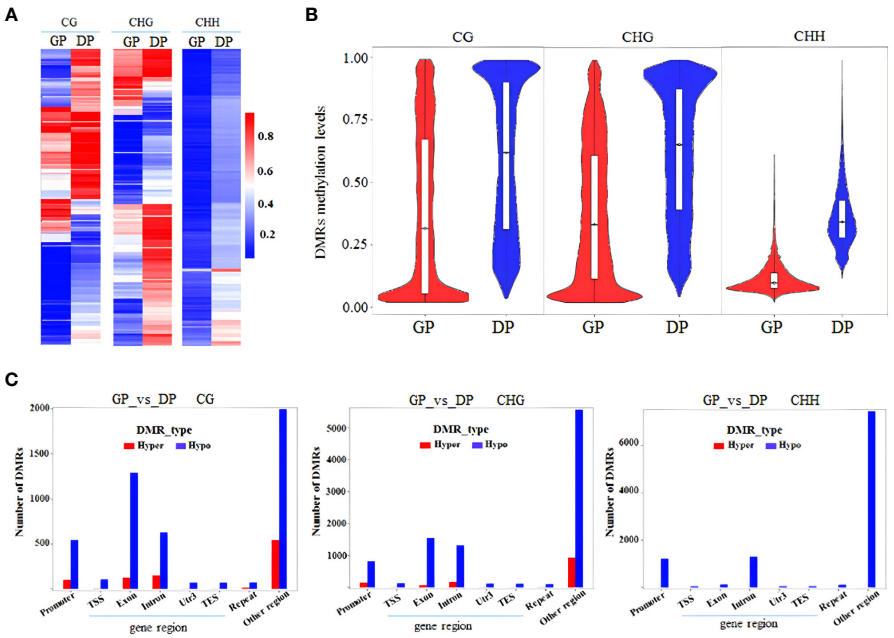


FIGURE 4
Comparative analysis of DMGs between DPs and GPs. **(A)** Heat map of methylation of DMGs in three contexts. **(B)** DMR methylation levels in three contexts. **(C)** Number of DMRs (hyper-/hypomethylated) in three contexts in the different gene regions. DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants; DMR, differentially methylated region.

10.3389/fpls.2024.1407700

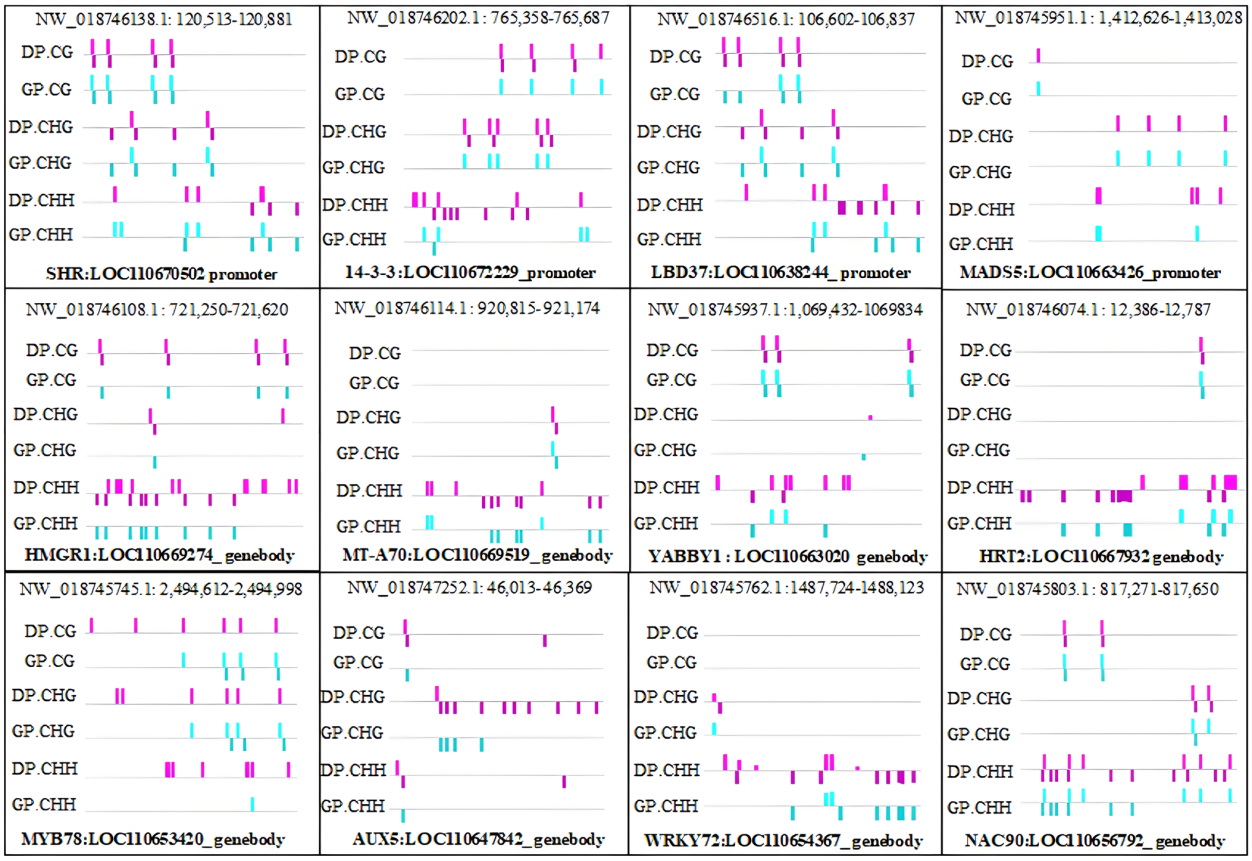


FIGURE 5
IGV snapshots of the representative DMGs in DPs and GPs. DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants.

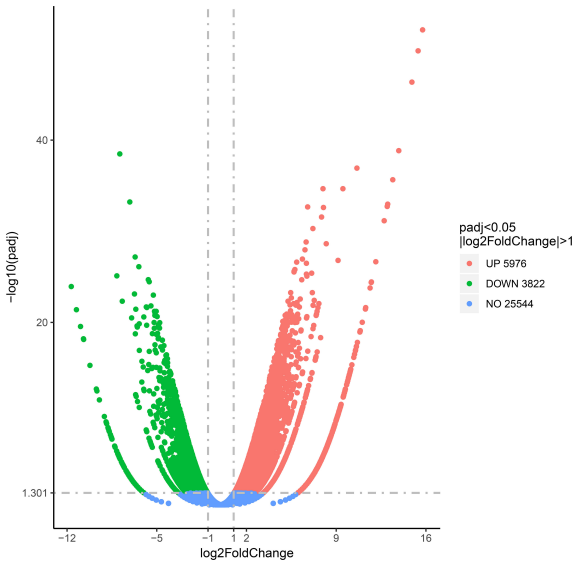


FIGURE 6
Volcano plot of differentially expressed genes (DEGs) in the DP and GP comparison. DP, donor plant; GP, grafted rubber tree plant.

downstream regions. In contrast, medium-expressed genes showed high methylation levels in the CG context in upstream regions and gene-body regions (Figure 7A).

Meanwhile, based on the methylation level in the promoter and gene-body region, methylated genes were classified into five groups: group 1 (methylation level < 20%), group 2 (20%–40%), group 3 (40%–60%), group 4 (60%–80%), and group 5 (>80%). Genes in group 5 within the gene-body region, with the lowest methylation levels, exhibited the lowest expression levels, while genes in group 1 with the lowest methylation level in the promoter region showed the lowest expression levels (Figure 7B), indicating a correlation between methylation status and gene expression levels.

To determine the correlation between gene methylation and gene expression levels, a scatter plot was employed to visualize the methylation level of DMRs along with their corresponding transcriptome differential gene expression levels. The results showed a positive correlation between hypomethylated DMRs and elevated gene expression, and hypermethylated DMRs correlated with decreased gene expression (red dots in Figure 8). However, in the CG, CHG, and CHH contexts, 47.87%, 49.18%, and 54.18% methylation changes did not demonstrate the anticipated associations with gene expression changes, respectively (blue dots in Figure 8).

10.3389/fpls.2024.1407700

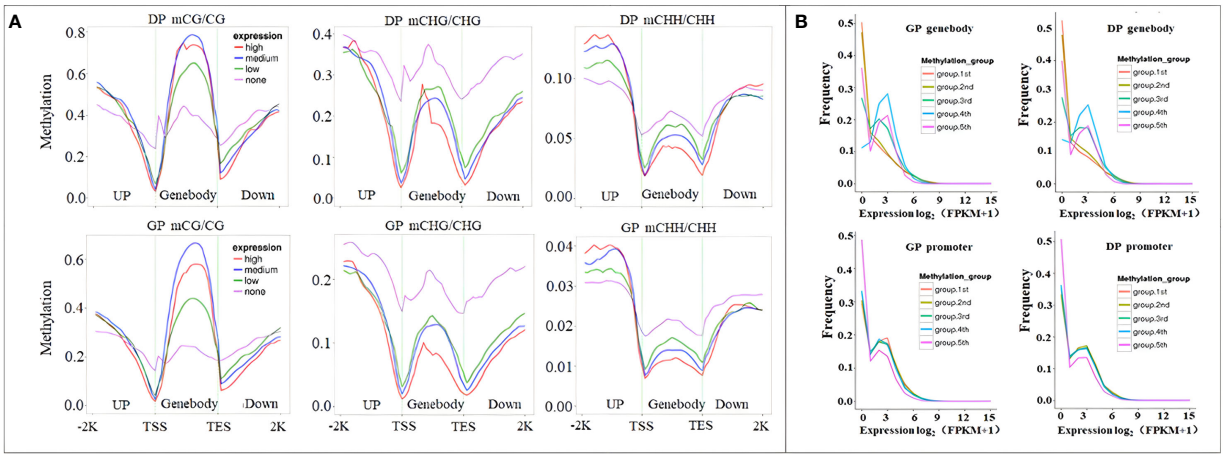


FIGURE 7
Correlation between DNA methylation and gene expression. **(A)** Distribution of methylation levels within gene bodies partitioned by different expression levels. **(B)** Comparison of the expression profiles of methylated and unmethylated genes.

3.6 Functional enrichment analysis of overlapping genes between DEGs and DMGs

Among these obtained DEGs and DMGs, 1,698 overlapping genes between DEGs and DMGs were identified. In detail, 1,122 genes overlapped with DMR_genes, and 576 overlapped with DMR_promoter_genes (Figures 9A, B). A total of 39 and 417 downregulated DEGs overlapped with hypermethylated and hypomethylated DMR_genes, while 70 and 625 upregulated DEGs overlapped with hypermethylated and hypomethylated DMR_genes. In addition, 68 and 304 upregulated DEGs overlapped with hypermethylated and hypomethylated DMR_promoter genes, and 25 and 214 downregulated DEGs overlapped with hypermethylated and hypomethylated DMR_promoter genes (Figures 9C, D).

The functional enrichment analysis of overlapping genes in three contexts was carried out by GO and KEGG analyses. The analysis of GO annotation enrichment is shown in Figure 6. The DMGs in CHH were significantly less in the term of cellular component (CC) than in biological process (BP) and molecular function (MF) than in CG and CHG. In the term of BP, the DMGs were mostly joined in the metabolic process, followed by the cellular process, organic substance metabolic process, primary metabolic process, and cellular metabolic process. In the term of MF, DMGs mostly participated in binding, catalytic activity, heterocyclic compound binding, organic cyclic compound binding, and ion binding. In the term of CC, DMGs were mainly correlated with cell, cell part, and membrane (Figure 10).

KEGG pathway enrichment analysis of DMGs showed that biosynthesis of secondary metabolites and metabolic pathway were significantly enriched pathways in three methylation contexts (Figure 11).

3.7 Analysis of DMGs in a qPCR assay

To demonstrate that the expression of DMGs is related to DNA methylation, 12 DMGs were selected to perform the analysis of gene expression by qPCR. A total of 10 of the 12 DMGs had higher methylation levels in DPs than in GPs and showed lower gene expression in DPs than in GPs, suggesting a significantly negative

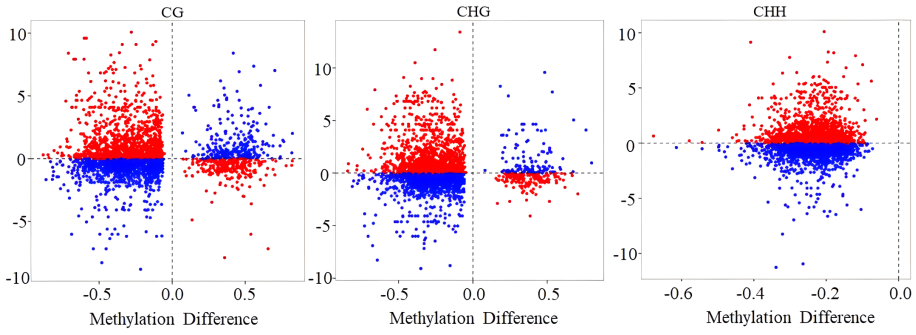


FIGURE 8
The scatter plot of the relationship between methylation level and gene expression level. Red and blue dots represent the gene–differentially methylated region (DMR) pairs exhibiting either inverse or equivalent relationships, respectively.

10.3389/fpls.2024.1407700

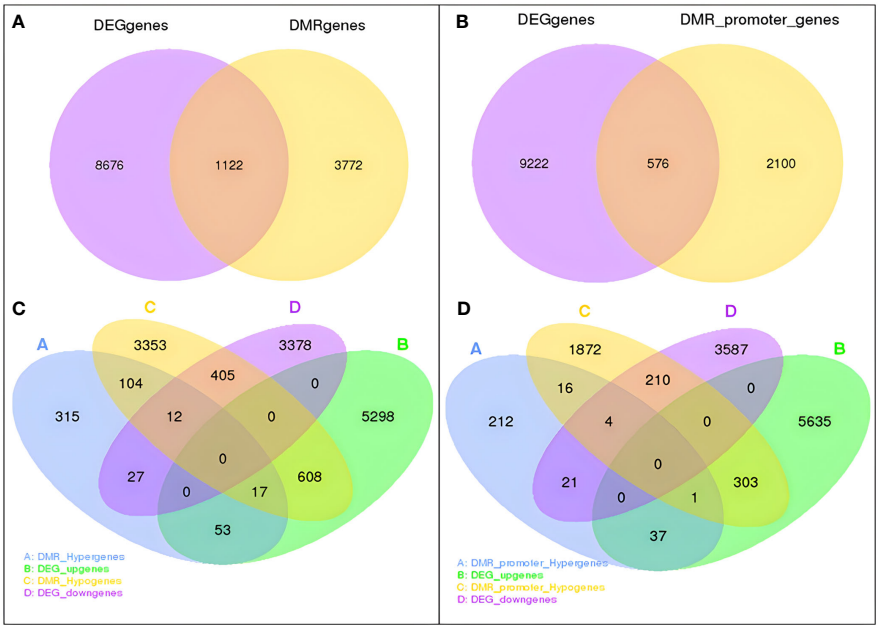


FIGURE 9
Association of DMRs with DEGs in the DP vs. GP comparisons. **(A)** Venn analysis of overlapping genes between DMR_genes and DEGs. **(B)** Number of overlapping genes between DMR_promoter genes and DEGs. **(C)** Venn analysis of upregulated and downregulated DEGs overlapped with hypermethylated and hypomethylated DMR_genes. **(D)** Venn analysis of upregulated and downregulated DEGs overlapped with hypermethylated and hypomethylated DMR_promoter genes. DMRs, differentially methylated regions; DEGs, differentially expressed genes; DP, donor plant; GP, grafted rubber tree plant.

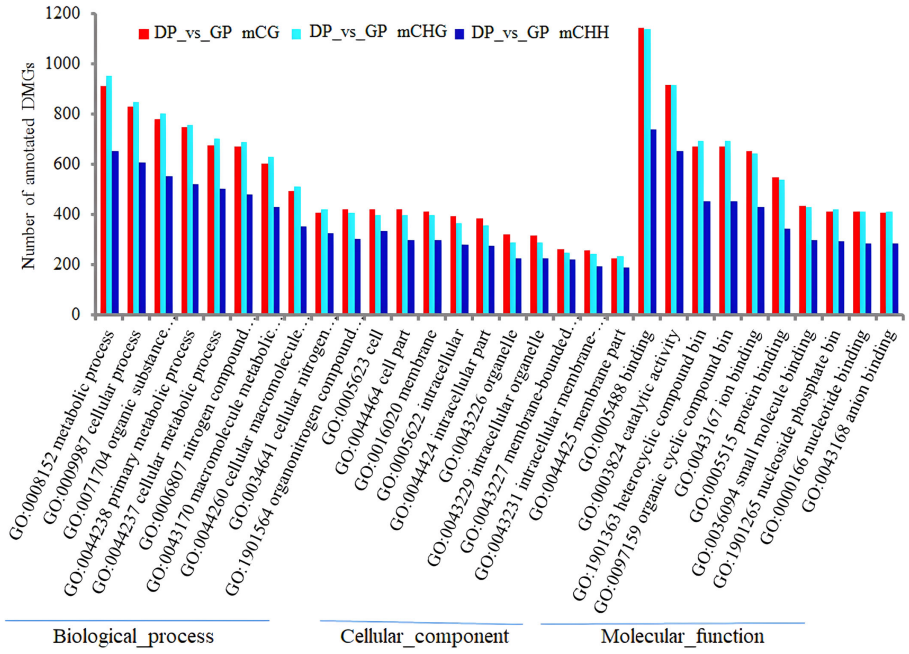


FIGURE 10
The enrichment analysis of GO terms (top 10) of DMGs in three contexts between DPs and GPs. GO, Gene Ontology; DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants.

relationship between gene expression and DNA methylation. However, there is a higher methylation level of DMG (LOC110670502) in GPs than in DPs, and its expression level was higher in GPs than in DPs, while there is a lower methylation level of DMG (LOC110669519) in GPs than in DPs and its expression was lower in GPs than in DPs (Figure 12). These data indicated that the expression of DMGs was correlated with DNA methylation.

4 Discussion

Grafting is a traditional agricultural propagation technology widely employed to reform crop yield, quality, and resistance to environmental stresses. Rubber tree is one of the successful commercially grafted woody plants. Previous research in *Arabidopsis* has demonstrated that grafting can modulate DNA methylation patterns within the genome of the scion cells, leading to physiological alterations (Molnar et al., 2010; Jeynes-Cupper and Catoni, 2023). Additionally, studies conducted in a Solanaceae interspecies grafting system have shown graft-induced modifications in DNA methylation (Wu et al., 2013). In woody species, graft-induced DNA methylation alterations were also reported in rubber trees and apple trees (Uthup et al., 2018; Perrin et al., 2020). Collectively, the above studies suggest that grafting can induce changes in DNA methylation, potentially affecting agronomical traits in crops.

Despite the application of MSAP analysis to investigate DNA methylation profiles in rubber tree heterografts (Uthup et al., 2018), studies regarding rubber tree whole-genome DNA methylation are not sufficient. In this study, we performed whole-genome DNA methylation analysis of GPs and DPs by WGBS. Our findings revealed that grafting triggers the reprogramming of DNA methylation, with a noteworthy decrease in genome-wide CHH methylation observed in GPs. In eggplant, grafting was able to induce grafted-plant vigor, and the enhanced plant vigor was associated with genome-wide CHH hypomethylation in the scions (Cerruti et al., 2021). Likewise, in *Arabidopsis*, a genome-wide CHH methylation decrease was also correlated with hybrid vigor (Groszmann et al., 2011; Greaves et al., 2012). Despite the highly stochastic nature of plant CHH methylation (Harris and Zemach, 2020), it is pertinent to note that DMRs were almost exclusively hypomethylated (Cerruti et al., 2021). It is possible that gene expression changes linked to enhanced vigor were associated with CHH methylation alterations. In addition, CHH methylation is

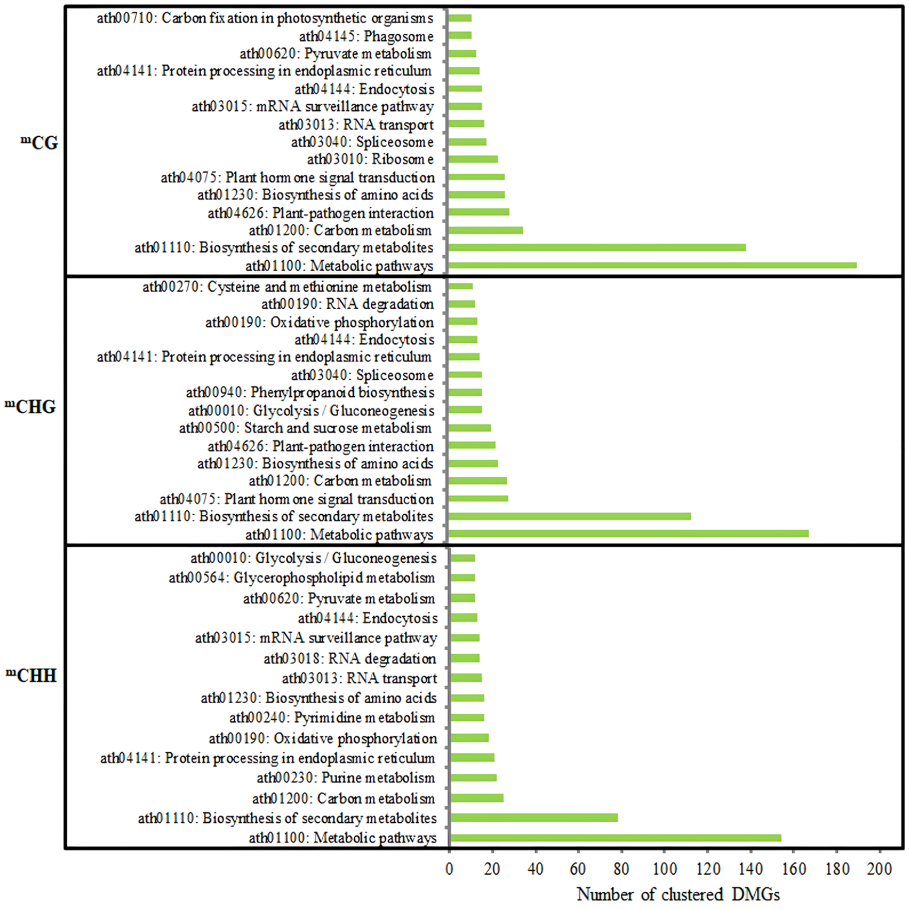


FIGURE 11
Analysis of the KEGG pathway of DMGs (top 15) in three contexts between DPs and GPs. KEGG, Kyoto Encyclopedia of Genes and Genomes; DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants.

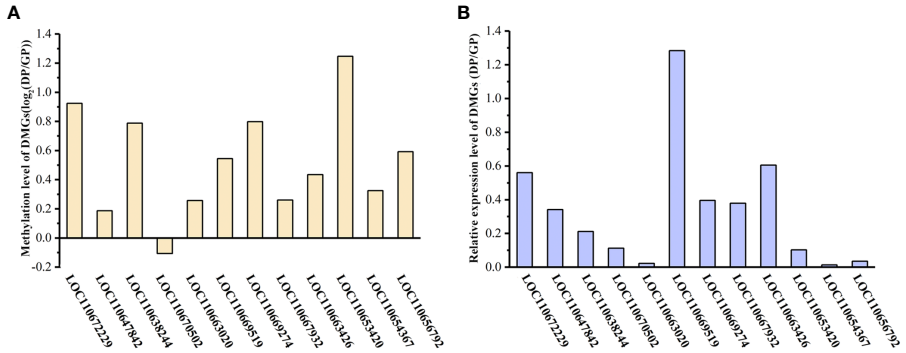


FIGURE 12
Gene expression level of DMGs. **(A)** Differences in the methylation of DMGs between DPs and GPs. **(B)** The expression level of DMGs by qPCR analysis. DMGs, differentially methylated genes; DPs, donor plants; GPs, grafted rubber tree plants.

inversely associated with the expression of transposable elements (TEs; Cerruti et al., 2021), which can regulate their own transcription and impact proximal gene expression through DNA methylation (Liu et al., 2022). Given that rubber trees are cross-pollinated crops and that rootstocks in grafting are typically unselected seedlings, variability in DNA methylation patterns may arise among grafted trees. Therefore, genome-wide CHH methylation decrease in GPs may be associated with intraclonal variations in the grafted rubber trees.

DNA methylation is a common epigenetic regulator of gene expression. DMR represents a notable mark of epigenetic variation and may be related to regulating the impact of DMGs on biological processes (Li et al., 2017). In total, 4,756 DMR-associated genes and 1,183 DMR-associated promoter genes were identified between GPs and DPs, respectively. Transcriptional analysis revealed that there were 9,798 DEGs in the DP and GP comparison. Furthermore, 1,698 overlapping genes were identified between DEGs and DMGs. GO annotation enrichment analysis showed that these overlapping genes were associated with various biological processes and were markedly enriched in metabolic pathway and biosynthesis of secondary metabolites. These findings suggest that graft-induced changes in DNA methylation and gene expression may affect the growth and rubber yield of the bud-grafted clones, potentially contributing to observed intraclonal variations in grafted rubber trees.

5 Conclusions

Whole-genome DNA methylation analyses were performed using WGBS technology to compare genomic methylation patterns between GPs and DPs. The study revealed downregulation of DNA methylation and demonstrated that grafting induced a reprogramming of DNA methylation in GPs. A total of 5,939 DMGs and 9,798 DEGs were identified between GPs and DPs, with 1,698 genes showing overlap between DEGs and DMGs. These overlapping genes were markedly enriched in the metabolic pathway and biosynthesis of secondary metabolites, as determined by KEGG pathway analysis. Global DNA methylation and transcriptional analyses further revealed a correlation between DNA methylation reprogramming and gene expression in grafted rubber trees. The study provides a whole-genome view of the methylome in rubber trees and an insight into the molecular mechanisms underlying the intraclonal variations existing in commercial planting grafted rubber trees.

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

H-LL: Writing – original draft, Writing – review & editing. YiW: Writing – original draft, Writing – review & editing. DG: Writing – original draft, Writing – review & editing. J-HZ: Writing – original draft, Writing – review & editing. YuW: Writing – original draft, Writing – review & editing. H-FD: Writing – original draft, Writing – review & editing. S-QP: Writing – original draft, 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 Hainan Provincial Natural Science Foundation of China (323RC557), National Natural Science Foundation of China (32171827; 31770722), and Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630052022009).

Acknowledgments

We would like to extend our deep gratitude to Weiguo Li and Xiao Huang (Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences) for their collaboration in treating the plant materials.

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.2024.1407700/full#supplementary-material

References

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cerruti, E., Gisbert, C., Drost, H. G., Valentino, D., Portis, E., Barchi, L., et al. (2021). Grafting vigour is associated with DNA de-methylation in eggplant. *Hortic. Res.* 8, 241. doi: 10.1038/s41438-021-00660-6

Chandrashekar, T. R., Mydin, K. K., Alice, J., Varghese, Y. A., and Saraswathyamma, C. K. (1997). Intraclonal variability for yield in rubber (*Hevea brasiliensis*). *Ind. J. Nat.Rubb. Res.* 10, 43–47.

Ding, X., Yuan, K., Yang, L., Cao, J., Xu, Z., He, Z., et al. (2010). The preliminary proteome study on the interaction between rootstock and scion of budgrafted rubber tree. *Chin. J. Trop. Crops.* 31, 1514–1518. doi: 10.3969/j.issn.1000–2561.2010.09.015

Feng, H., Conneely, K. N., and Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* 42, e69. doi: 10.1093/nar/gku154

Gallego-Bartolomé, J. (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol.* 227, 38–44. doi: 10.1111/nph.16529

Greaves, I. K., Groszmann, M., Ying, H., Taylor, J. M., Peacock, W. P., Dennis, E. S., et al. (2012). Trans chromosomal methylation in arabidopsis hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3570–3575. doi: 10.1073/pnas.1201043109

Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J., and Dennis, E. S. (2011). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2617–2622. doi: 10.1073/pnas.1019217108

Han, Q., Song, H., Yang, C., Zhang, S., Korpelainen, H., and Li, C. (2022). Integrated DNA methylation, transcriptome and physiological analyses reveal new insights into superiority of poplars formed by interspecific grafting. *Tree Physiol.* 42, 1481–1500. doi: 10.1093/treephys/tpac013

Harris, K. D., and Zemach, A. (2020). Contiguous and stochastic CHH methylation patterns of plant DRM2 and CMT2 revealed by single-read methylome analysis. *Genome Biol.* 21, 194. doi: 10.1186/s13059-020-02099-9

Hua, Y. W., Huang, T. D., and Huang, H. S. (2010). Micropropagation of self-rooting juvenile clones by secondary somatic embryogenesis in *Hevea brasiliensis*. *Plant Breeding.* 129, 202–207. doi: 10.1111/j.1439-0523.2009.01663.x

Huang, G., and Lin, W. (2003). Biochemical interaction between rootstock and scion of budgrafted *Hevea Brasiliensis*. *Chin. J. Trop. Crops.* 24, 7–11.

Jeynes-Cupper, K., and Catoni, M. (2023). Long distance signalling and epigenetic changes in crop grafting. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1121704

Kapazoglou, A., Tani, E., Avramidou, E. V., Abraham, E. M., Gerakari, M., Megariti, S., et al. (2021). Epigenetic changes and transcriptional reprogramming upon woody plant grafting for crop sustainability in a changing environment. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.613004

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, 907–915. doi: 10.1038/s41587-019-0201-4

Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719

Li, H. L., Guo, D., Yang, Z. P., Tang, X., and Peng, S. Q. (2014). Genome-wide identification and characterization of WRKY gene family in *Hevea brasiliensis*. *Genomics* 104, 14–23. doi: 10.1016/j.ygeno.2014.04.004

Li, Q., Hermanson, P. J., and Springer, N. M. (2018). Detection of DNA methylation by whole-genome bisulfite sequencing. *Methods Mol. Biol.* 1676, 185–196. doi: 10.1007/978–1-4939–7315-6_11

Li, Y., Ding, X., Wang, X., He, T., Zhang, H., Yang, L., et al. (2017). Genome-wide comparative analysis of DNA methylation between soybean cytoplasmic male-sterile line NJCMS5A and its maintainer NJCMS5B. *BMC Genomics* 18, 596. doi: 10.1186/s12864-017-3962-5

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341, 1237905. doi: 10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514

Liu, P., Cuerda-Gil, D., Shahid, S., and Slotkin, R. K. (2022). The Epigenetic control of the transposable element life cycle in plant genomes and beyond. *Annu. Rev. Genet.* 56, 63–87. doi: 10.1146/annurev-genet-072920-015534

Liu, J., Shi, C., Shi, C. C., Li, W., Zhang, Q. J., Zhang, Y., et al. (2020). The chromosome-based rubber tree genome provides new insights into spurge genome evolution and rubber biosynthesis. *Mol. Plant* 13, 336–350. doi: 10.1016/j.molp.2019.10.017

Liu, K., Wang, T., Xiao, D., Liu, B., Yang, Y., Xu, K., et al. (2023). The role of DNA methylation in the maintenance of phenotypic variation induced by grafting chimerism in *Brassica*. *Hortic. Res.* 10, uhad008. doi: 10.1093/hr/uhad008

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, 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, 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

Mattei, A. L., Bailly, N., and Meissner, A. (2022). DNA methylation: a historical perspective. *Trends Genet.* 38, 676–707. doi: 10.1016/j.tig.2022.03.010

Molnar, A., Melnyk, C. W., Bassett, A., Hardcastle, T. J., Dunn, R., and Baulcombe, D. C. (2010). Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* 328, 872–875. doi: 10.1126/science.1187959

Perrin, A., Daccord, N., Roquis, D., Celton, J. M., Vergne, E., and Bucher, E. (2020). Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees. *Epigenomes* 4, 4. doi: 10.3390/epigenomes4010004

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

Santos, A. S., Neves, D. M., Santana-Vieira, D. D. S., Almeida, L. A. H., Costa, M. G. C., Soares, F., et al. (2020). Citrus scion and rootstock combinations show changes in DNA methylation profiles and ABA insensitivity under recurrent drought conditions. *Sci. Hortic.* 267. doi: 10.3389/fpls.2016.01475

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820. doi: 10.1038/nmeth.3035

Tang, C. R., Huang, D. B., Yang, J. H., Liu, S. J., Sakr, S., Li, H. P., et al. (2010). The sucrose transporter HbSUT3 plays an active role in sucrose loading to laticifer and rubber productivity in exploited trees of *Hevea brasiliensis* (para rubber tree). *Plant Cell Environ.* 33, 1708–1720. doi: 10.1111/j.1365-3040.2010.02175.x

Uthup, T. K., Karumamkandathil, R., Ravindran, M., and Saha, T. (2018). Heterografting induced DNA methylation polymorphisms in *Hevea brasiliensis*. *Planta* 248, 579–589. doi: 10.1007/s00425-018-2918-6

Wu, R., Wang, X., Lin, Y., Ma, Y., Liu, G., Yu, X., et al. (2013). Inter-species grafting caused extensive and heritable alterations of DNA methylation in *Solanaceae* plants. *PloS One* 8, e61995. doi: 10.1371/journal.pone.0061995

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., et al. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res.* 43, e141. doi: 10.1093/nar/gkv715

Yang, I. V., Pedersen, B. S., Rabinovich, E., Hennessy, C. E., Davidson, E. J., Murphy, E., et al. (2014). Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 190, 1263–1272. doi: 10.1164/rccm.201408-1452OC

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14. doi: 10.1186/gb-2010-11-2-r14

Zhang, H. M., Lang, Z. B., and Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z

Zhou, L., and Lin, W. (2012). Interaction between scion and rootstock in rubber tree (*Hevea Brasiliensis*). *Chin. J. Trop. Crops.* 33, 1337–1341. doi: 10.1007/s11783-011-0280-z

Zhou, L., Wang, J., Xie, G., An, F., Zeng, X., Chen, X., et al. (2014). Characteristics of latex regeneration and latex flow of rubber tree on different rootstocks. *Chin. J. Trop. Crops* 35, 2161–2167. doi: 10.3969/j.issn.1000–2561.2014.11.011

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 August 2024
DOI 10.3389/fpls.2024.1416913

OPEN ACCESS

EDITED BY
Changmian Ji,
Chinese Academy of Tropical Agricultural Sciences, China

REVIEWED BY
Luming Yang,
Henan Agricultural University, China
Lijun Ou,
Hunan Agricultural University, China
Zhigang Hao,
China Agricultural University, China

*CORRESPONDENCE
Nannan Qin
nnqnews@163.com

†These authors have contributed equally to this work

RECEIVED 13 April 2024
ACCEPTED 24 July 2024
PUBLISHED 12 August 2024

CITATION
Qin N, Yang S, Wang Y, Cheng H, Gao Y, Cheng X and Li S (2024) The *de novo* assembly and characterization of the complete mitochondrial genome of bottle gourd (*Lagenaria siceraria*) reveals the presence of homologous conformations produced by repeat-mediated recombination. *Front. Plant Sci.* 15:1416913. doi: 10.3389/fpls.2024.1416913

The *de novo* assembly and characterization of the complete mitochondrial genome of bottle gourd (*Lagenaria siceraria*) reveals the presence of homologous conformations produced by repeat-mediated recombination

Nannan Qin[1,2*†], Shanjie Yang[1†], Yunan Wang[3], Hui Cheng[3], Yang Gao[1], Xiaojing Cheng[1] and Sen Li[1]

[1]College of Horticulture, Shanxi Agricultural University, Jinzhong, China, [2]Department of Development Planning & Cooperation, Shanxi Agricultural University, Taiyuan, China, [3]Department of Scientific Research Management, Shanxi Agricultural University, Taiyuan, China

Introduction: Bottle gourd is an annual herbaceous plant that not only has high nutritional value and many medicinal applications but is also used as a rootstock for the grafting of cucurbit crops such as watermelon, cucumber and melon. Organellar genomes provide valuable resources for genetic breeding.

Methods: A hybrid strategy with Illumina and Oxford Nanopore Technology sequencing data was used to assemble bottle gourd mitochondrial and chloroplast genomes.

Results: The length of the bottle gourd mitochondrial genome was 357547 bp, and that of the chloroplast genome was 157121 bp. These genomes had 27 homologous fragments, accounting for 6.50% of the total length of the bottle gourd mitochondrial genome. In the mitochondrial genome, 101 simple sequence repeats (SSRs) and 10 tandem repeats were identified. Moreover, 1 pair of repeats was shown to mediate homologous recombination into 1 major conformation and 1 minor conformation. The existence of these conformations was verified via PCR amplification and Sanger sequencing. Evolutionary analysis revealed that the mitochondrial genome sequence of bottle gourd was highly conserved. Furthermore, collinearity analysis revealed many rearrangements between the homologous fragments of *Cucurbita* and its relatives. The Ka/Ks values for most genes were between 0.3~0.9, which means that most of the genes in the bottle gourd mitochondrial genome are under purifying selection. We also identified a total of 589 potential RNA editing sites on 38 mitochondrial protein-coding genes (PCGs) on the basis of long noncoding RNA (lncRNA)-seq data. The RNA editing sites of *nad1-2, nad4L-2, atp6-718, atp9-223 and rps10-391* were successfully verified via PCR amplification and Sanger sequencing.

10.3389/fpls.2024.1416913

Conclusion: In conclusion, we assembled and annotated bottle gourd mitochondrial and chloroplast genomes to provide a theoretical basis for similar organelle genomic studies.

KEYWORDS

Lagenaria siceraria, mitochondrial DNA sequencing, tandem repeats, phylogenetic analysis, RNA editing site

Background

Bottle gourd (*Lagenaria siceraria*) is an annual herbaceous plant in the Cucurbitaceae family that is native to Africa and cultivated worldwide (Konan et al., 2020). In addition to being an important agricultural crop, bottle gourd also has a wide range of applications in traditional medicine and crafts (Zahoor et al., 2021). Bottle gourd has high nutritional value and medicinal properties and is often used as an important rootstock for the grafting of cucurbit crops such as watermelon, cucumber and melon (Yang et al., 2013; Liang et al., 2022). Moreover, this gourd species has good low-temperature tolerance, a high affinity for grafting, and resilience in the quality of the melons it produces; furthermore, there are few obstacles to its continuous cropping, and it is highly disease- and pest-resistant (Xu et al., 2023).

As cellular organelles, mitochondria are essential for maintaining the cellular energy supply and for many biological functions (Liberatore et al., 2016; Møller et al., 2021). Mitochondria have their own genetic material, known as mitochondrial DNA (mtDNA), and their genomes contain key information that encodes proteins within the mitochondria (Huang et al., 2020). The plant mitochondrial genome is the largest known organellar genome after the nuclear genome and has unique characteristics, such as the lowest known rate of synonymous substitutions, relatively rich rearrangements, and high levels of inversion and recombination (Gualberto et al., 2014; Garcia et al., 2021; Yu et al., 2022). The mitochondrial genome harbors genes encoding key proteins in the mitochondria that are involved in the function of the cellular respiratory chain (Cho et al., 2023; Li et al., 2023). Several plant organelle RNA recognition (PORR) members, including WTF1, WTF9 and LEFKOTHEA, play a role in the splicing of introns in angiosperm organelles. *Arabidopsis thaliana* root primordium defective 1 (*rpd1*) plays a role in the splicing of introns located in the coding regions of various complex I (CI) subunits (i.e., *nad2*, *nad4*, *nad5* and *nad7*), as well as in the maturation of ribosomal *rps3* in mitochondria. Changes in the growth and development phenotypes of *rpd1* mutants and changes in respiratory activity are closely related to defects in the respiratory chain caused by CI (Edris et al., 2024). Additionally, genes in the mitochondrial genome are considered important controllers of other biological processes, including apoptosis and cell signaling (Yang H. et al., 2023). The Bcl-2 protein family regulates apoptosis by controlling mitochondrial permeability. The antiapoptotic proteins Bcl-2 and Bcl-xL reside in the outer mitochondrial membrane and inhibit cytochrome c release. Notably, the construction of recombinant fluorescent protein markers confirmed that *cox-8* is closely related to mitochondrial apoptotic cells (Shao et al., 2021).

Several studies conducted in recent years have suggested that the mitochondrial and chloroplast genomes are similar in terms of their phylogeny, biological classification, and kinship. In particular, a small number of species transmit mitochondrial DNA through paternal lines, which provides a unique explanation for the phylogenetic relationships among species from the perspective of paternal inheritance (Del Valle-Echevarria et al., 2016; Mandel et al., 2020). The paternal inheritance of mitochondria in plants was first discovered in green algae, and then, via techniques such as cytology and molecular genetics, it was found that mitochondria in cucumber and melon also exhibit paternal inheritance (Boynton et al., 1987). Mitochondria are also paternally inherited in a few other higher plant species, such as banana and kiwifruit (Faure et al., 1994; Chat et al., 2004). However, owing to the small number of seed plant species that exhibit this type of paternal inheritance, there are very few studies on the mechanism of mitochondrial paternal inheritance. In Cucurbitaceae crops, only a Psm locus on the nuclear genome has been found to control the paternal inheritance of cucumber, and after the publication of the cucumber genome, this locus was found to be located on chromosome 3 and is considered a major QTL (Al-Faifi et al., 2008). Patchy greening or necrosis of cucumber leaves and fruits is a paternal phenotype that is accompanied by a decrease in plant vigor and fertility and promotes the expression of resistance-related genes, and research has shown that this trait is regulated by the mitochondrial genes *nad5* and *atp4* (Del Valle-Echevarria et al., 2015). In addition, the phenotypes of F_1 plants produced via positive and negative crosses between cold-resistant (CH1) and non-cold-resistant (CH4) cucumber cultivars were consistent, and the F_2 generations from CH4×CH1 were more cold resistant than the F_2 generations from the CH1×CH4 cross, suggesting that the mitochondrial genome may be involved in the regulation of this trait (Ali et al., 2014).

To date, the mitochondrial genomes of eight Cucurbitaceae species, including cucumber, watermelon, bitter bottle gourd and zucchini, have been sequenced, and these data lay the foundation for further elucidation of various scientific topics, such as the evolution and inheritance of the mitochondrial genome of Cucurbitaceae plants (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al.,

2023). Cucumber, watermelon and *Cucurbita pepo* all had 37 protein-coding genes, but the protein-coding genes among cucumber, watermelon and *Cucurbita pepo* were not identical; for example, *rps19* is found in watermelon and *Cucurbita pepo* but not in cucumber. However, *rpl10* is found in cucumber but not in watermelon or *Cucurbita pepo*. Melon has only 36 genes encoding proteins and has lost *rpl10* and *rps19* (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al., 2023). Assembling the bottle gourd mitochondrial genome is extremely important for obtaining a thorough understanding of the structure, function, and evolution of bottle gourd mitochondria, but pertinent research on the mitochondrial genome of this gourd has not been performed. Thus, in this study, the bottle gourd mitochondrial genome was sequenced, assembled, and annotated, and its traits, including structural traits, were investigated. PCR assays were subsequently conducted to test the mitochondrial genomic substructure. The bottle gourd chloroplast genes were assembled via the same dataset, and homologous fragments of the bottle gourd mitochondrial and chloroplast genomes were revealed. Mitochondrial RNA editing events and their evolutionary relationships were also analyzed. In conclusion, we thoroughly analyzed the bottle gourd mitochondrial genome to provide a more complete understanding of the genome and crucial data for studying the evolution and phylogeny of bottle gourd and for informing practical applications of these data and further genetic advancements.

Materials and methods

Plant material

The experiment was performed in Zhuangziying village, Xugou town, Qingxu County, Taiyuan city, Shanxi Province, China. We used the bottle gourd cultivar Xiao Shounian (the leaves are small, the branching is strong, the number of fruits per plant is generally 80-100, and the fruits are approximately 3-4 cm tall and divided into upper and lower chambers) as test materials; we selected mature, plump and complete bottle gourd seeds soaked in a 55°C warm water solution and sowed them in a 32-well dish prior to cultivating them under natural light in a solar greenhouse. The whole growth period was managed according to field conventions, and no species of the same family were cultivated around the planting area. On June 13, 2023, the 10th–12th young leaves were removed from the plants by cutting them along the base of the petiole with alcohol-disinfected scissors; the leaves were rapidly placed into a precooled sampling tube, stored in liquid nitrogen, and subsequently stored in a -80°C freezer.

DNA library construction and sequencing

The bottle gourd DNA was extracted via a plant DNA extraction kit (Tiangen, Beijing, China), and its purity, concentration, and integrity were checked via a Nanodrop instrument, Qubit quantification, and 0.35% agarose gel electrophoresis. If the DNA sample was of sufficient quality, it was transported on dry ice to Beijing Biomics Biotech Co., Ltd. (Beijing, China) for sequencing. The DNA was randomly sheared via a Covaris ultrasonic crusher, and terminal repair, A-tail addition, adapter addition, fragment screening, PCR amplification, and purification were then performed to generate the final DNA library (Meyer and Kircher, 2010). The Illumina NovaSeq 6000 high-throughput sequencing platform was used for next-generation sequencing. The raw data output was 6.63 Gb, and the filtered clean data output was 6.61 Gb. gTUBE was used to break the genomic DNA to an average length of approximately 8 kb, and DNA damage repair, end repair, linkage and Qubit library quantification were conducted. Oxford Nanopore platform sequencing was used for third-generation sequencing, and a total of 7.23 Gb of raw data was measured (Lu et al., 2016). After the connectors, short fragments (less than 500 bp in length) and low-quality data were filtered out, a total of 7.14 Gb of clean data was obtained.

Bottle gourd mitochondrial gene splicing, assembly and annotation

First, the bottle gourd mitochondrial genome was assembled on the basis of Nanopore long-read data. The long-read sequencing data were directly assembled via Flye software with the default parameters to generate graphical assembly results in GFA format (Kolmogorov et al., 2019). For all assembled FASTA-format contigs, we constructed a library via makeblastdb and subsequently utilized the BLASTn program to identify contig fragments containing mitochondrial genomes via conserved plant mitochondrial genes in *A. thaliana* as query sequences (-evalue 1e-5 -outfmt 6-max_hsps 10-word_size 7-task blastn-short) (Wick et al., 2015). The GFA files were visualized via Bandage software (version 0.8.1), and the mitochondrial contigs were screened against the BLASTn results to generate a sketch of the bottle gourd mitochondrial genome. The long-read and short-read data were compared to the mitochondrial contigs via BWA software (version 0.7.17) and exported by filtering the mitochondrial reads in the alignment (Wick et al., 2017). A Unicycler with the default parameters was used for hybrid assembly, and the bottle gourd mitochondrial genome was ultimately assembled and then visualized via Bandage software (version 0.8.1) (Tillich et al., 2017).

The PCGs of the mitochondrial genome were selected as reference genes for *A. thaliana* (NC_037304) and watermelon (NC_014043.1), and the sequence of the mitochondrial genome was annotated via GeSeq software (version 2.03). The tRNAs of the mitochondrial genome were annotated via tRNAscan-SE software (version 2.0.11). The rRNAs of the mitochondrial genome were annotated via BLASTN software (version 2.13.0) (Chen et al., 2015). Each mitochondrial genome annotation error was manually corrected via Apollo software (version 1.11.8) (Lewis et al., 2002).

Identification and validation of repeated-mediated recombination events

MISA (version 2.1), TRF (version 4.09) and the REPuter web server (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) were used

to identify repeats, including microsatellite sequence repeats, tandem repeats, and scattered repeats (Kurtz et al., 2001; Beier et al., 2017). The results were visualized with Excel (version 2021) and the Circos package (version 0.69-9) (Zhang et al., 2013). On the basis of the bottle gourd mitochondrial genomic data, sequences 500 bp upstream and downstream of the repeat sequences were extracted, and primers were designed using PrimerBLAST (Supplementary Table S1). The PCR mixture consisted of 1 μl of DNA, 2 μl of 10 μM forward and reverse primers, respectively, 25 μl of 2× Phanta Max Master Mix, and 20 μl of ddH_2O. PCR was performed under the following conditions: 94°C for 3 min; 37 cycles of 94°C for 15 s, 55°C for 15 s and 70°C for 15 s; and 72°C for 5 min. Further sequencing of the PCR products of the expected size was performed via the Sanger method. The PCR products of the correct size in the electrophoresis gel were subsequently sent to Sangon Bioengineering Co., Ltd. (Shanghai) for sequencing. SeqMan was used to visualize the sequencing results.

Sequence transfer and codon preference analysis

The chloroplast genome was assembled via GetOrganelle software (version 1.7.7.0), the chloroplast genome was annotated via CPGAVAS2 software, and the chloroplast genome annotation results were corrected via CPGView software (Shi et al., 2019; Jin et al., 2020). The results were visualized via the Circos package (version 0.69-9) by analyzing the homologous fragments via BLASTN software (version 2.13.0) (Zhang et al., 2013). The protein-coding sequences of the genome were extracted via PhyloSuite software (version 1.1.16) (Zhang et al., 2020). Codon preference analysis of the mitochondrial genome PCGs was performed with Mega software (version 7.0), and relative synonymous codon usage (RSCU) values were calculated (Kumar et al., 2016).

Phylogenetic analysis

The complete mitochondrial genomes of 36 species from four angiosperm orders (Cucurbitales, Rosales, Fagales, and Fabales) related to bottle gourd were downloaded from the NCBI database (see Supplementary Table S2 for details on the new information). Common genes were then extracted using PhyloSuite (version 1.2.2) software, multiplex sequence alignment analysis was performed with MAFFT software (version 7.505), and IQ-TREE software (version 1.6.12) was used for phylogenetic analysis. ITOL software (version 6) was subsequently used to visualize the results from the phylogenetic analysis (Katoh and Standley, 2013; Nguyen et al., 2015). The corresponding nucleotide sequences were aligned and concatenated via Maft (version 7.450) and Phylosuite, respectively. The Ka/Ks value for each gene was calculated via the KaKs_calculator. In the results, the Ka/Ks "NA", which appears when Ks = 0 (in cases with no substitutions in the alignment or 100% match), was replaced with 0.

Collinearity analysis

On the basis of BLASTN (version 2.10.1, parameter: -word_size 7, E value 1e-5), pairwise comparisons of the cucumber species and six proximal taxa (*Cucurbita pepo, Cucurbita maxima, Lagenaria siceraria, Citrullus lanatus, Luffa acutangula* and *Herpetospermum*) were conducted for *Herpetospermum pedunculosum* and *Momordica charantia*, and homologous sequences greater than 500 bp in length were retained as conserved collinear blocks and plotted via a multiple synteny plot.

RNA editing events and collinearity analysis

Total RNA was extracted using an RNA Plant Extraction Kit (Chinese Tiangen). The RNA integrity and concentration were measured with a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and an Agilent 2100 bioanalyzer (Agilent Technologies, USA), respectively. Sequencing libraries were generated after the removal of rRNA (NEB E7420) via the NEBNext Illumina Superdirected RNA Library Preparation Kit. The 5' end of each library was phosphorylated and cyclized, and circular amplification was then performed to generate DNA nanospheres. These DNA nanospheres were subsequently loaded onto the DNBSEQ-T7 sequencing system for long noncoding RNA (lncRNA) sequencing. On the basis of the lncRNA sequencing data, transcripts from the mitochondrial genomes were filtered and mapped onto mitochondrial DNA sequences, and differences in DNA and RNA sequences were further assessed via BEDTools software (version 2.30.0) to identify the most read-supported RNA editing events (Quinlan and Hall, 2010). On the basis of the BLAST procedure, comparisons of individual mitochondrial genomes were performed, and homologous sequences longer than 500 bp were preserved as conserved collinear blocks to generate multiple synteny plots. Multiple synteny plots of bottle gourds and related species were created via the MCScanX source program (Wang et al., 2012).

Results

Assembly and annotation of the *Lagenaria siceraria* mitochondrial genome

The bottle gourd mitochondrial genome was assembled using a combination of long- and short-read and long noncoding RNA (lncRNA) sequencing. We generated a total of 7.23 Gb of raw data from three generations of Nanopore platform sequencers. After filtering out connectors, short fragments and low-quality data, a total of 7.14 GB of clean data with an N50 of 20153 bp was generated, and the average length of the reads in the clean data was 12157 bp (Supplementary Table S3). A total of 6.63 Gb of raw data were generated via a second-generation Illumina platform sequencer, and after filtering out connectors, short fragments and low-quality reads, a total of 6.61 Gb of clean data with a Q20 of 95.89%, a Q30 of 89.58%,

and a GC content of 34.96% were generated (Supplementary Table S3). A total of 15.25 Gb of raw data were generated by the lncRNA sequencer, and after the exclusion of connectors, short fragments and low-quality reads, a total of 13.54 Gb of clean data with a Q20 of 96.92%, a Q30 of 92.41%, and a GC content of 41.42% were generated (Supplementary Table S3). The second-generation Illumina sequencing data were assembled *de novo* into a unit map, and the continuous-repeat-continuous region was decomposed by third-generation Nanopore long-read sequencing. We then identified the bottle gourd mitochondrial genome as a single circular molecule of 357547 bp with a GC content of 45.03% (Figure 1).

Annotation of the bottle gourd mitochondrial genome (Figure 1, Table 1) revealed 38 unique protein-coding genes (PCGs), including 24 unique mitochondrial core genes, 14 noncore genes, 3 ribosomal RNA (rRNA) genes and 19 transfer RNA (tRNA) genes. The core bottle gourd genes included 5 ATP synthase genes, 9 NADH dehydrogenase genes, 1 cytochrome b gene, 4 cytochrome c biogenesis genes, 3 cytochrome c oxidase genes, 1 maturase gene and 1 protein transport subunit. Fourteen noncore genes, namely, 4 ribosomal protein large subunit genes, 8 ribosomal protein small subunit genes, and 2 succinate dehydrogenase genes, were identified. Among the three rRNA genes, rrn5 was a double-copy gene, and rrn18 and rrn26 were single-copy genes. Among the 19 tRNA genes, trnC-GCA and trnW-CCA were double-copy genes, trnP-UGG was a three-copy gene, and the remaining genes were single-copy genes. We submitted the annotated complete bottle gourd mitochondrial genome to GenBank under accession number OR680814.

Structure and codon preferences of the mitochondrial genome of bottle gourd

For the 38 PCGs in bottle gourd mitochondria, we conducted a codon preference analysis. Supplementary Table S4 displays the data on codon utilization by certain amino acids. Codons with RSCU values greater than 1 were considered preferable for certain amino acids. In addition to the starting codons AUG and tryptophan (UGG), which both had RSCU values of 1, the PCGs in bottle gourd mitochondria presented other common preferences for codon use (Figure 2). For example, the stop codon showed a greater preference for UAA, with the highest RSCU occurring at 1.63, followed by alanine (Ala), with a preference for GCU with an RSCU value of 1.59. Notably, phenylalanine (Phe) has an RSCU of 1.11 for UUU and thus does not show a strong codon preference.

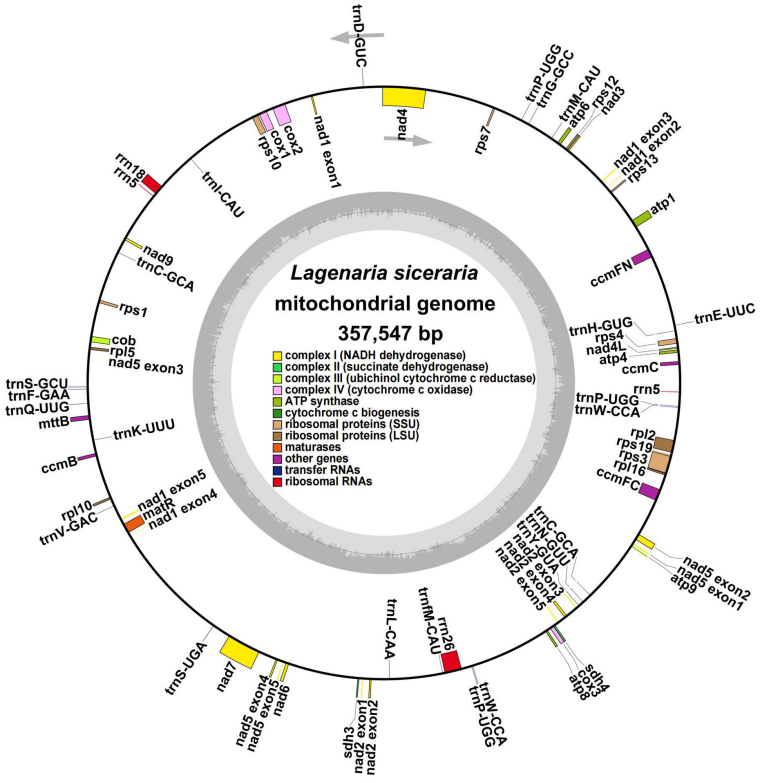


FIGURE 1
Circular map of the whole *Lagenaria siceraria* mitochondrial genome. Each square represents a different mitochondrial gene, and the same gene is represented by the same color. The exons are denoted as 'exons', such as *nad2* exon 1 and *nad2* exon 2, and introns are found between the exons. The arrows indicate the forward and reverse chains.

10.3389/fpls.2024.1416913

TABLE 1 Gene contents of the mitogenome of *Lagenaria siceraria*.

	Gene groups	Names of genes
Core genes	ATP synthase	atp1, atp4, atp6, atp8, atp9
	NADH dehydrogenase	nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9
	Cytochrome b	cob
	Cytochrome c biogenesis	ccmB, ccmC, ccmFC, ccmFN
	Cytochrome c oxidase	cox1, cox2, cox3
	Maturase	matR
	Protein transport subunit	mttB
Noncore genes	Ribosomal protein large subunit	rpl2, rpl5, rpl10, rpl16
	Ribosomal protein small subunit	rps1, rps3, rps4, rps7, rps10, rps12, rps13, rps19
	Succinate dehydrogenase	sdh3, sdh4
rRNA genes	Ribosome RNA	rrn5(×2), rrn18, rrn26
tRNA genes	Transfer RNA	trnC-GCA(×2), trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnH-GUG, trnI-CAU, trnK-UUU, trnL-CAA, trnM-CAU, trnN-GUU, trnP-UGG(×3), trnQ-UUG, trnS-GCU, trnS-UGA, trnV-GAC, trnW-CCA(×2), trnY-GUA

Mitochondrial genome repeat analysis

A total of 101 simple sequence repeats (SSRs) were found in the bottle gourd mitochondrial genome, and monomeric and dimeric SSRs accounted for 56.44% of the total SSRs (Figure 3A). Adenine monomeric repeats accounted for 46.88% (15) of the 32 monomeric SSRs. No hexameric SSRs were detected in this mitochondrial genome. A total of 10 tandem repeats with a matching degree greater than 80% and a length between 12 and 39 bp were detected in the bottle gourd mitochondrial genome. Moreover, the analysis revealed 474 pairs of repeats with lengths greater than or equal to 30 bases; 255 and 219 of these were palindromic and forward repeats, respectively, and no reverse repeats or complementary repeats were detected (Figure 3B). The longest palindromic repeats had a length of 2349 bp, and the longest forward repeats had a length of 1685 bp.

Recombination mediated by repeat sequences

We compared long reads to repeat sequences to determine whether the long reads spanned the repeat region and then derived the most likely structure of the mitochondrial genome. The final results revealed that the bottle gourd mitochondrial genome contained 3 nodes (Figure 4A). On the basis of the long-read data, one cyclic contig sequence (Supplementary Figure S1) was generated after the branching node caused by the repeat sequence (i.e., ctg3) was resolved, and the repeat sequence (node ctg3) was hypothesized to mediate genome recombination and form a structure consisting of two small rings (Supplementary Figure S1). To this end, we designed two pairs of PCR primers (F1/R1 and F2/R2) to verify the presence of the identified recombinant products. The ctg3 repeat sequence amplified bands of approximately the expected size before and after the exchange of reverse primers (Figure 4B). Sanger sequencing revealed that the PCR products were consistent with the template sequence, suggesting that ctg3-mediated recombination can breakdown the bottle-gourd mitochondrial genome into one large circular molecule and one small circular molecule (Figure 4C; Supplementary Figure S2). The repeat sequence ctg3 in the bottle gourd mitochondrial genome can mediate chromosomal recombination into two interconverting conformations (GenBank accession number OR680814).

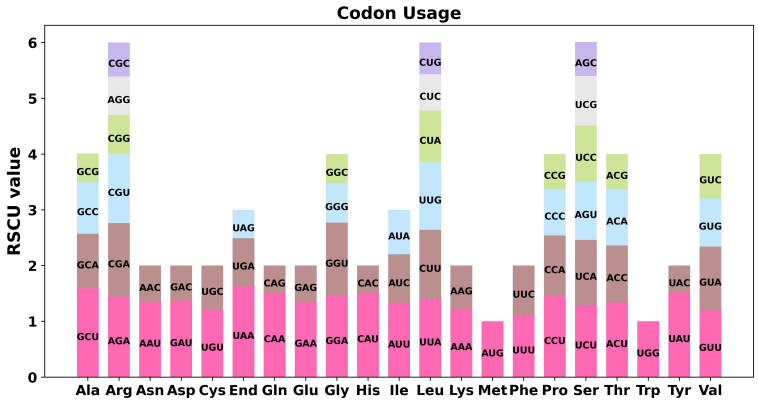


FIGURE 2
Codon preference analysis of the bottle gourd mitochondrial genome. Different colors represent the RSCU values for different codons for the same amino acid. The abscissa represents the different amino acids, and the ordinate represents the RSCU values for the different codons for each amino acid.

10.3389/fpls.2024.1416913

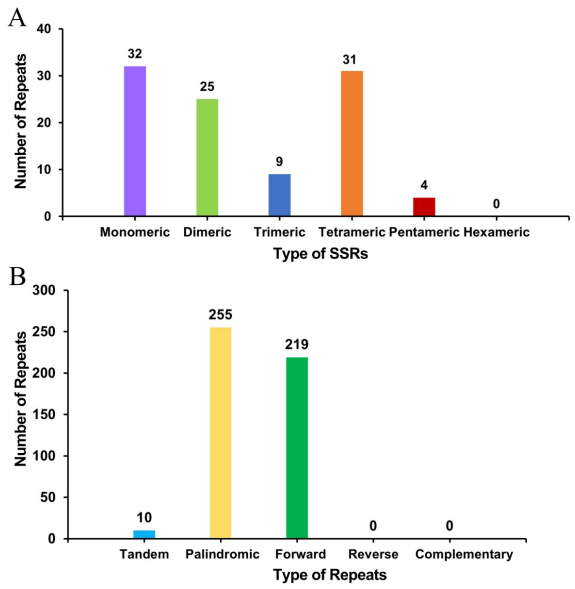


FIGURE 3
Histogram of the results from the repeat sequence analysis. **(A)** The abscissa represents the type of SSR, and the ordinate represents the number of repeating fragments. **(B)** The abscissa represents the type of repeat sequence, and the ordinate represents the number of repeated fragments.

Intergenomic sequence migration

Some chloroplast DNA fragments were found to have moved into the mitochondrial DNA during mitochondrial evolution. We also assembled the bottle gourd chloroplast genome, which is composed of a single-loop double-stranded DNA molecule and has a total sequence length of 157121 bp and a GC content of approximately 50.10% (Supplementary Figure S3). Sequence similarity analysis revealed that a total of 27 segments, which had a combined length of 23226 bp and represented 6.50% of the length of the mitochondrial genome, were homologous sections of the bottle gourd chloroplast and mitochondrial genomes (Figure 5). Mitochondrial plastid DNA (MTPT) 17 was the longest homologous sequence, with a length of 6,184 bp. By annotating the sequences of these homologs, 19 complete genes were found in 27 homologous fragments: 10 PCGs and 9 tRNA genes (Supplementary Table S5). The presence of MTPT sequence ligation sites in the mitochondrial genome of bottle gourd was verified, and both end ligation sites of MTPT6, MTPT7, MTPT8, MTPT9, MTPT10, MTPT17 and MTPT18 could amplify bands consistent with the expected size. Sanger sequencing revealed that the PCR product was consistent with the template sequence, confirming that MTPT6, MTPT7, MTPT8, MTPT9, MTPT10, MTPT17 and MTPT18 was present in the mitochondrial genome

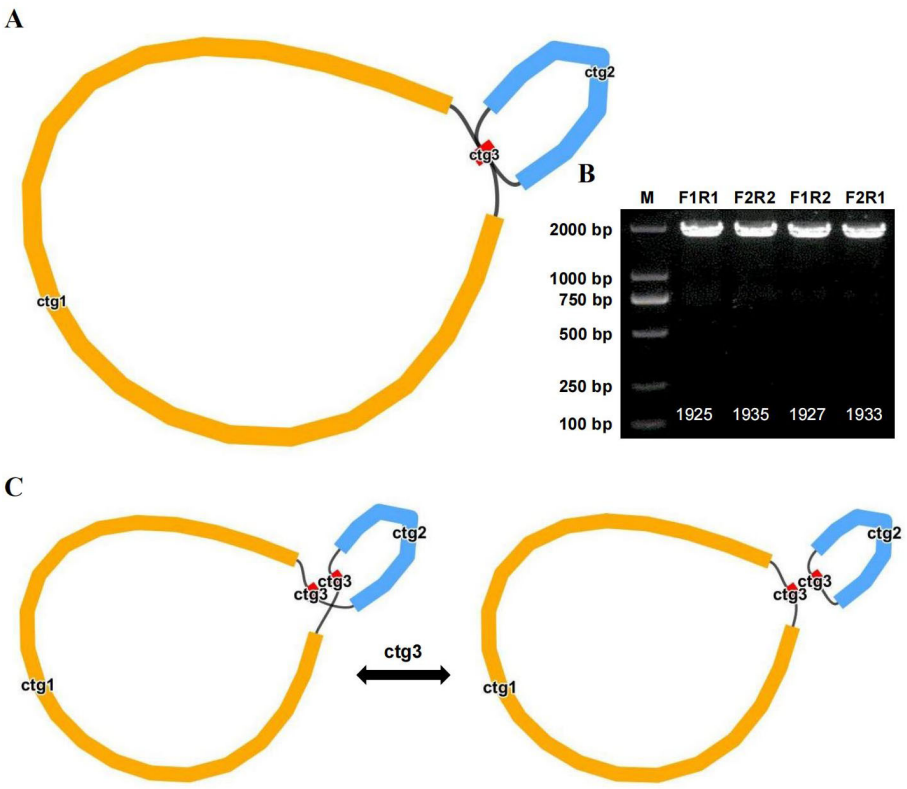


FIGURE 4
Potential isomers of the bottle gourd mitogenome inferred from PCR experiments and long reads. **(A)** Bottle gourd mitochondrial genome conformation. **(B)** Bottle gourd mitochondrial genome repeat verification. Gel electrophoresis results of PCR products amplified with primers for F1, F2, R1, and R2 using the ctg3 sequence described in **(A)**. The sequences shown in the assembled genome were used to generate PCR products using primers for F1/R1 and F2/R2. The presence of recombinant sequences resulted in the generation of PCR products for F1/F2 and R1/R2 using primers. **(C)** Repeat-mediated (ctg3) conformational recombination of the bottle gourd mitochondrial genome.

10.3389/fpls.2024.1416913

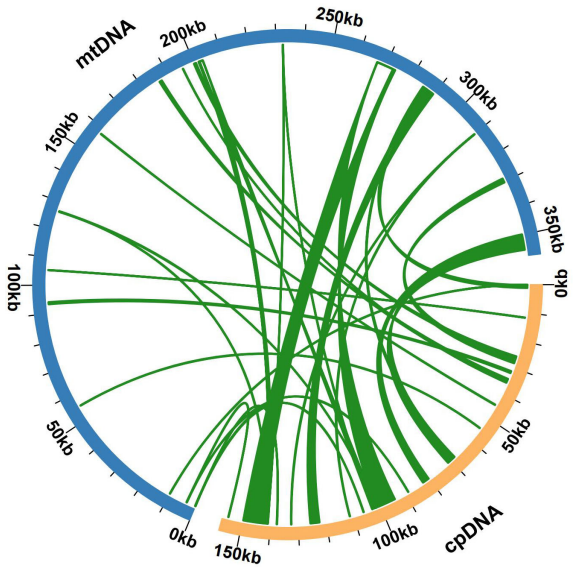


FIGURE 5
Gene exchange in the chloroplast genome and the bottle gourd mitochondrial genome. The blue arc represents the mitochondrial genome, the yellow arc represents the chloroplast genome, and the genome fragment corresponding to the green connection between the arcs is a homologous fragment. The thickness of the green lines represents the size of the cognate fragments, and the number of lines represents the number of homologous fragments.

of bottle gourd (Supplementary Figure S4). Small fragments that were partially derived from chloroplasts were also discovered to be subsets of larger sequence fragments or to occur several times in the mitochondrial genome.

Phylogenetic analysis

The plant mitochondrial genome size and structure are highly variable, but the genes involved are highly conserved. The base replacement rate of PCGs is the most highly conserved among the nuclear, plastid and mitochondrial genomes of plants, but the degree of differentiation is insufficient; therefore, the whole mitochondrial genome is generally not selected as a molecular marker for species phylogenetic analysis and genetic diversity analysis of germplasm resources. However, some protein-coding sequences, such as those of the *COX3* and *CCB203* genes, are often used in such studies. On the basis of the DNA sequences of 26 conserved mitochondrial PCGs, phylogenetic tree reconstruction was performed using 36 species from four orders of angiosperms (Cucurbitales, Rosales, Fagales, and Fabales) (Supplementary Table S2). The two mitochondrial genomes of the legume order were used as outgroups. *Cucumis sativus* and *Cucumis hystrix* were located on the most recently diverged branch, and *Cucumis melo* and *Cucurbita pepo* were most recently related to their origin. The bottle gourd belongs to the Cucurbitaceae family of the Cucurbitales order, is grouped on the same branch as *Citrullus lanatus*, and is closely related to *Cucurbita maxima* (Figure 6). The results of our study are consistent with the theory that plant mitochondrial genes are highly conserved and that plant mitochondrial genomes are not suitable for phylogenetic studies.

Synonymous and nonsynonymous nucleotide substitution patterns are very important markers in gene evolution studies. A Ka/Ks ratio (Ka) with synonymous (Ks) substitution rates < 1 indicates purifying selection; Ka/Ks > 1 indicates probable positive selection, whereas Ka/Ks values close to 1 indicate neutral evolution or relaxed selection. We calculated the Ka/Ks ratios of the bottle gourd mitochondrial genome versus those of 35 species from four orders of angiosperms (Cucurbitales, Rosales, Fagales, and Fabales) (Figure 7). The Ka/Ks values for most genes were between 0.3~0.9, which means that most of the genes in the bottle gourd mitochondrial genome are under purifying selection. The most conserved genes with average Ka/Ks values between 0 and 0.3, indicating very strong purifying selection pressure, are *atp1, ccmC, cox1, nad9, rps4* and *rps12*. Ectopic overexpression of *Ntatp1* in tobacco can result in cytoplasmic male sterility and seed abortion. Cotton mitochondrial *Ghatp1* can affect ATPase production of ATP and promote epidermal hair and fiber elongation. The *nad3* gene can affect the respiration efficiency and ripening of tomato fruits. However, the Ka/Ks values of most genes suggest that the bottle gourd mitochondrial genome may be affected by purifying selection during the selection and domestication process, and some studies have shown that this selection effect is more obvious in the nuclear genome; therefore, these genomic changes may greatly improve the fruit quality of the bottle gourd. The model averaging method in the KaKs calculator yielded average Ka/Ks values > 1 for *ccmB*, *nad4L* and *rps3*. A Ka/Ks > 1, which is indicative of positive selection, could reflect selection pressures specific to the bottle gourd mitochondrial genome. Alternatively, they could be a sign of increased variability in particular proteins within a broader group of species.

Collinearity analysis

Red arcs indicate areas where inversion occurred, and gray areas indicate areas with good homology. To better present the results, collinear blocks with lengths of less than 0.5 kb were not retained. Cucurbita species exhibited homologous collinear blocks with 7 closely related species (*C. pepo, C. maxima, L. siceraria, C. lanatus, L. acutangula, H. pedunculosum* and *M. charantia*), but these collinear blocks were shorter. Additionally, sequences that are specific to bottle gourd species and lack similarity with the blank sections of other species were discovered (Figure 8). The results revealed that the orders of collinear block arrangement between the mitochondrial genomes of these seven species were inconsistent, and many genome rearrangements were detected between the mitochondrial genomes of bottle gourd species and related species. More collinear gene clusters were found between the bottle gourd mitochondria and the mitochondria of *C. maxima* and *C. lanatus*, which further indicates that bottle gourd is more closely related to *C. lanatus* and *C. maxima* than to the other species (Figure 8).

10.3389/fpls.2024.1416913

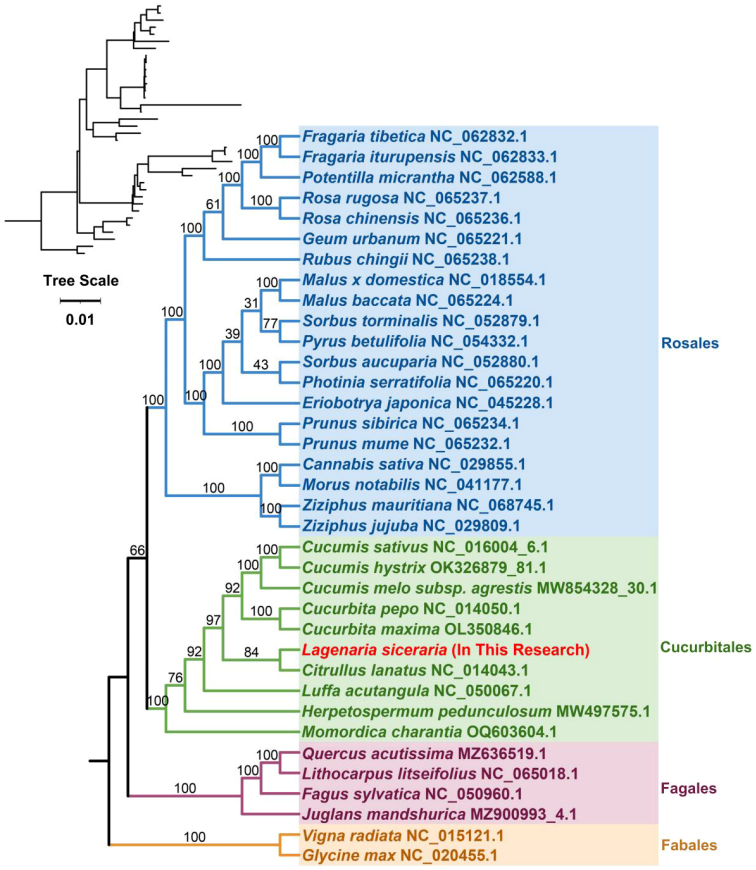


FIGURE 6
In the phylogenetic analysis of bottle gourd, different colors represent different orders, and red represents the bottle gourd mitochondrial genome. The number on a branch indicates the reliability of the branch, and a higher number indicates increased reliability. The upper left corner shows the evolutionary phylogram diagram; in this diagram, the length of the branch represents the evolutionary distance, and a greater distance indicates a greater change from the state of the ancestor.

Bottle gourd mitochondrial RNA editing events

The differences in DNA and RNA sequences were further assessed via BEDTools software (version 2.30.0) to predict mitochondrial genomic RNA editing events in bottle gourd. Thirty-eight distinct PCGs in the bottle gourd mitochondria were used to identify RNA editing events. The 38 mitochondrial PCGs included a total of 589 possible RNA editing events, all of which were C-to-U base edits. Among them, the nad4 gene contained the

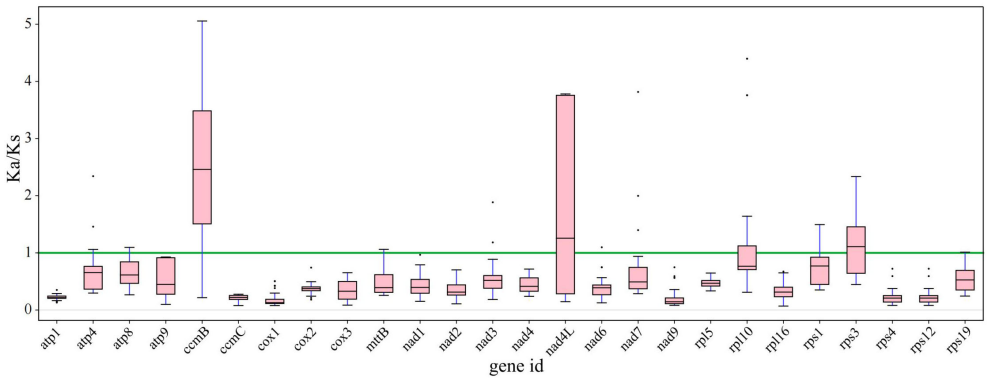


FIGURE 7
The Ka/Ks ratios of 26 conserved mitochondrial PCGs of 36 species from four orders of angiosperms (Cucurbitales, Rosales, Fagales, and Fabales) were compared with those of bottle gourd.

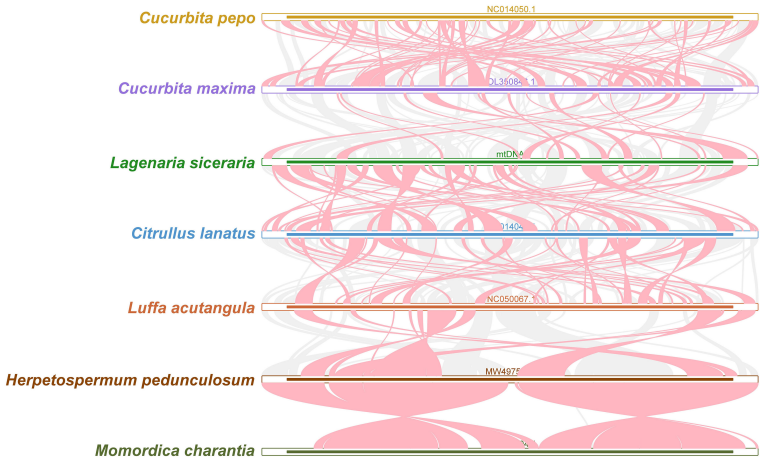


FIGURE 8
Collinear relationships among the mitochondrial genomes of *C. pepo, C. maxima, L. siceraria, C. lanatus, L. acutangula, H. pedunculosum* and *M. charantia*. The red arcs indicate areas where the inversion occurred, and the gray areas indicate regions with good homology. The different species are represented by different colors, the thickness of the lines represents the size of the homologous fragments, and the number of lines represents the number of homologous fragments.

greatest number (45) of identified RNA editing events, accounting for 7.64% of the total editing events (Figures 9A, B). This gene was followed by the ccmB gene, which had 42 RNA editing events, accounting for 7.13% of the total events (Figures 9A, B). The *rpl2* gene had only 1 RNA editing event, which was the lowest among all the bottle-gourd mitochondrial genes. The predicted start codon and stop codon editing sites were verified by PCR and Sanger sequencing (Figure 9C). The stop codon RNA editing phenomenon was successfully verified at five loci, namely, *nad1*-2, *nad4L*-2, *atp6*-718, *atp9*-223 *and rps10*-391. Among them, the four sites of *nad1*-2, *nad4L*-2, *atp9*-223 and *rps10*-391 have high editing efficiency, and the C base is completely replaced by with T base.

Discussion

The energy that plants require for growth and development is provided by their mitochondria, which serve as their power plants (Liu et al., 2023). The assembly of plant mitochondrial genomes is incredibly difficult, especially when next-generation sequencing assembly methodologies are utilized, because plant mitochondrial genomes exhibit vast differences in size, sequence order, and repeat content (Attia et al., 2021; Kang et al., 2021; Yang et al., 2021). In recent years, research on the bottle gourd nuclear genome has greatly promoted the study of the evolution and function of bottle gourd genes (Ma et al., 2022). However, bottle gourd mtDNA has not yet been studied; thus, the lack of representative and genetically clear bottle gourd mtDNA represents a major obstacle in the field (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al., 2023). Despite notable variances in mitochondrial genome size and structure among various Cucurbitaceae species, previous research has indicated that the mitochondrial genomes of Cucurbitaceae species are circular (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al., 2023). For example, the cucumber mitochondrial genome has an especially large main ring and a relatively small subcyclic structure that is approximately 1685 kb in length (Niu et al., 2021). Watermelons and zucchini have single rings of 379 kb and 983 kb, respectively (Cui et al., 2021). The mitochondrial gene of bitter melon is a single loop of 331 kb (Niu et al., 2023). With a size of 2740 kb, the mitochondrial genome of melon is the largest in the Cucurbitaceae family (Chakraborty et al., 2018). However, the cucumber mitochondrial gene has three circular chromosomes that have been mapped and have lengths of 1110 kb, 110 kb, and 92 kb (Kim et al., 2006). Therefore, a thorough study of plant mitochondrial genomes will lead to new information on mitochondrial genome evolution and molecular function. We used Illumina and Oxford Nanopore Technology sequencing data to sequence and assemble the 357-kb circular bottle gourd mitochondrial genome, which is similar in size to the mitochondrial genomes of watermelon and bitter melon (Figure 1). This genome provides important information for the molecular breeding of bottle gourds and can also serve as a reference genome for other species in the Cucurbitaceae family.

Consistent with the findings of previous reports, we found that the mitochondrial genome of Cucurbitaceae crops contains complex I~V genes (nad, sdh cob and cox), cytochrome C biosynthesis genes (ccm), ribosomal protein genes (rpl), matR genes, mttB genes, 3 rRNA genes, and 13~24 tRNA genes (bottle gourd contains 19 genes) (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al., 2023). The protein-encoding genes of different crop species in the Cucurbitaceae family were relatively conserved, and the number of genes was similar. *rpl10* is absent in dicots such as *Arabidopsis thaliana*, *Brassica napus* and *Beta vulgaris*; thus, its function needs to be exercised by nuclear genes (Kubo and Arimura, 2010). The loss of *rps19* in cucumber and melon during the evolution of Cucurbitaceae may be due to its transfer into the nuclear genome or its function as a pseudogene, but no loss of rpl genes was found in bottle gourd. The proportion of genetic coding sequences and intron sequences in the

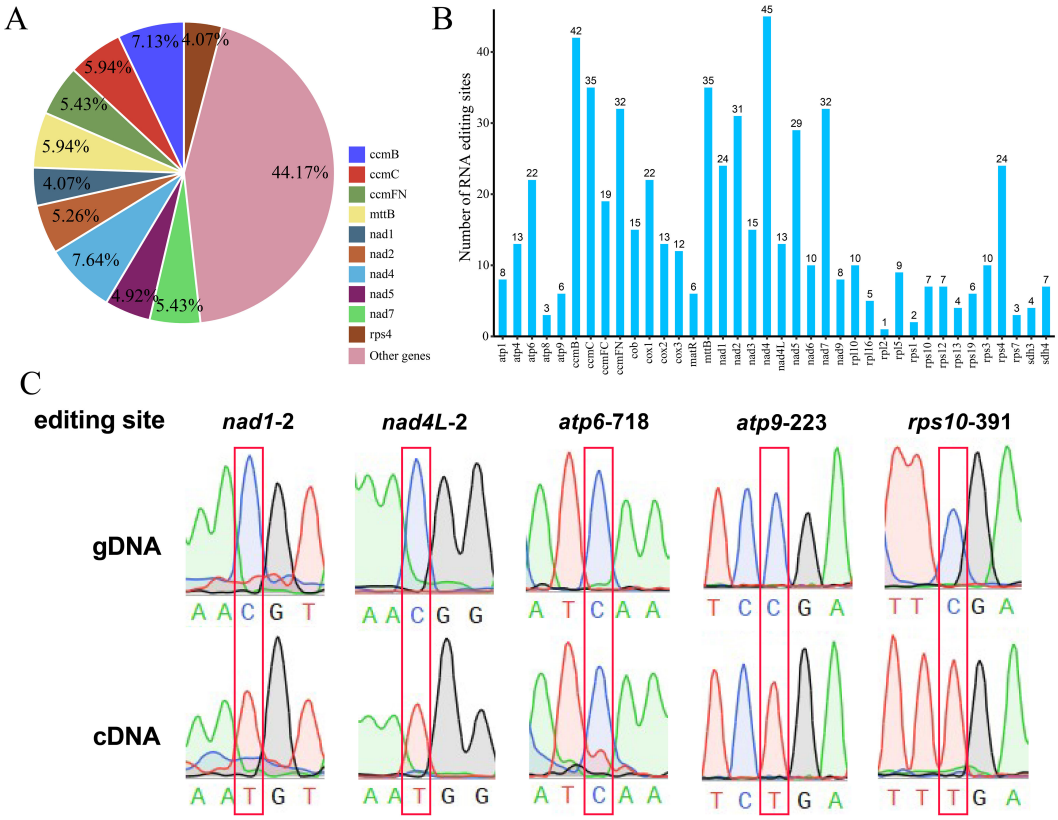


FIGURE 9
Characteristics of the RNA editing sites identified in the PCGs of bottle gourd. **(A)** Pie chart of RNA editing sites predicted by individual PCGs in the mitochondria of cucurbits. Different RNA editing sites are indicated by different colors, and the percentages are shown in the figure. **(B)** Histogram of RNA editing sites predicted by individual PCGs in the mitochondria of bottle gourd. The horizontal axis represents the different RNA editing sites, and the vertical axis represents the number of each RNA editing site. **(C)** Sequencing peak diagram of mitochondrial RNA editing sites in bottle gourd.

mitochondrial genome of Cucurbitaceae species, such as bottle gourd, cucumber, melon, and *Cucurbita pepo*, is inversely proportional to the mitochondrial genome size (Hansen et al., 2007; Cui et al., 2021; Niu et al., 2023; Zhou et al., 2023). This may be because the coding sequences and intron sequences in the Cucurbitaceae mitochondrial genome are relatively conserved and the changes in length are small, whereas the changes in length between genes in the mitochondrial genome are large. The differences in tRNA in the mitochondrial genome of Cucurbitaceae crops were significant, and the tRNA genes in the melon mitochondrial genome (21) were the most common among Cucurbitaceae crops, whereas the tRNA genes in *Cucurbita pepo* were the least common (13). *trnH-GUG*, *trnL-CAA* and *trnN-GUU* are tRNA genes unique to melon and bottle gourd, and trnD-GUC, trnM-CAU, and trnW-CCA, three tRNAs, are present only in bottle gourd, cucumber and melon; these genes may have been obtained from the bottle gourd, cucumber and melon nuclear or chloroplast genomes during evolution. *trnK-UUU* is a tRNA gene that is unique to bottle gourd, watermelon and *Cucurbita pepo*, and both melon and cucumber may have lost the *trnK-UUU* gene during evolution.

The intergenic region of the mitochondrial genome is composed mainly of homologous mitochondrial genome sequences, repeat sequences, homologous chloroplast genome sequences, and homologous nuclear genome sequences (Attia et al., 2021; Kang et al., 2021; Yang et al., 2021). Although only 33% of the sequences in the mitochondrial genomes of the melon genus share a homologous relationship with those of other plants, more than 50% of the melon mitochondrial genome sequences are homologous to those of other plants (Cui et al., 2021). The mechanism of sequence migration between organelle genomes and the patterns of gene expression transmitted by migrating fragments in new genes remain mostly unclear, but it is speculated that further improvement of the bottle gourd genome-wide project will yield beneficial innovations for solving these problems. Repeat amplification is frequently thought to be one of the causes of plant mitochondrial genome expansion. The bottle gourd mitochondrial genome contained a total of 101 SSRs and 474 pairs of repeats longer than or equal to 30 bp, and the longest palindromic repeat and the longest forward repeat measured 2349 bp and 1685 bp, respectively (Figure 3). Only approximately 23 kb of the bottle gourd mitochondrial genome sequence was homologous to the chloroplast genome sequence, whereas approximately 113 kb of the chloroplast homologous sequence was observed in *C. pepo*, suggesting that chloroplast homologous sequences were not the cause of the expansion of the bottle gourd mitochondrial genome (Ma et al., 2022). Therefore, on the basis of the results from this study, it can be preliminarily speculated that the expansion of the bottle gourd mitochondrial genome was due to

increases in homologous sequences, repeat sequences and nuclear genome homologous sequences.

Short-sequence duplication is thought to be an important mediator of chloroplast genome recombination in some algae and lower plants (Ogihara et al., 1988). Although repeat-mediated recombination yields a dizzying array of genomic conformations, in the mitotic genomes of *Salvia* and *Scutellaria skullcap*, repeat sequences can mediate homologous recombination, leading to genomic conformational alterations (Li et al., 2021; Yang J. X. et al., 2023). We resolved repeats (i.e., ctg3) on the basis of long-read data, which may mediate the recombination of the bottle gourd mitochondrial genome and lead to the formation of two small loops. Sanger sequencing revealed that the PCR products were consistent with the template sequence, suggesting that ctg3-mediated recombination can break the bottle-gourd mitochondrial genome into one large circular molecule and one small circular molecule (Figure 4). The results demonstrated that the repeat sequence ctg3 in the bottle gourd mitochondrial genome can mediate chromosomal recombination into two interconverting conformations. This finding also suggested that repeat-mediated recombination is the driving force for the structural diversification of plant mitotic genomes.

The inconsistent order of the colinear arrangement suggested that the mitochondrial genome of Cucurbita species underwent extensive rearrangement, which may have contributed to its evolution and diversification (Lee and Park, 2022). After transcription, RNA editing events alter information, particularly at transcription sites, and terrestrial plants frequently undergo mitochondrial RNA editing events (Wu and Chaw, 2022). Additionally, the angiosperm mitochondrial genome has certain conserved RNA editing sites (Grienenberger, 2009). A growing body of evidence suggests that RNA editing by mitochondria is associated with important cultivation traits in plants. For example, RNA editing at the C1292 and C1415 loci of the cotton mitochondrial *Ghatp1* gene can affect ATPase production of ATP and promote epidermal hair and fiber elongation, whereas decreased RNA editing of the *nad3* and *sdh4* genes in tomato can disrupt mitochondrial biological functions, reduce fruit respiration efficiency, and ultimately inhibit tomato fruit ripening (He et al., 2018; Yu et al., 2019). RNA editing is a posttranscriptional regulatory process that occurs mainly in the chloroplasts and mitochondria of plants and is an important means of maintaining the normal biological functions of chloroplasts and mitochondria. The main type of mitochondrial gene RNA editing is cytosine to uracil (C-to-U), which occurs in protein-coding regions, rarely occurs in noncoding regions, and acts on the first and second sites of codons (Alverson et al., 2010). The number of editing sites varies greatly among species; for example, 11 RNA editing sites are found in the mitochondrial genome of *Physcomitrium patens*, while 692 RNA editing sites are found in the mitochondria of cotton. The role of RNA editing is usually to maintain the conservation of amino acid sequences of important functional proteins in organelles, but sometimes, it can also affect the translation of important genes, such as those that encode the production of new start codons or terminators, thereby altering the length and normal function of proteins. The RNA editing times for different genes differ among Cucurbitaceae crops, and ribosomal protein genes (*rpl2, rps1* and *rps7*) undergo less RNA editing than other genes do, whereas the *mttB*, *ccmB* and *ccmFN* genes have greater RNA editing times, which is consistent with our findings (Alverson et al., 2010). For example, most of the nonsynonymous edits in the *ccmFC*, *cob*, *matR* and *mttB* genes were fully edited in *Cucurbita pepo* and partially edited in watermelon and bottle gourd. Therefore, understanding RNA editing in plant mitochondria is crucial for elucidating the function of plant mitochondrial genomes. However, because RNA editing plays a complex and unclear role in plant mitochondrial genomes, additional research is needed to fully understand its regulatory mechanism. The 589 mitochondrial RNA editing sites identified in this study can inform future research related to RNA editing in bottle gourd.

Conclusions

In this study, the bottle gourd mitochondrial genome was assembled and annotated, revealing a genome size of 357547 bp and 38 PCGs. One pair of repeats was confirmed by PCR amplification and Sanger sequencing to facilitate homologous recombination into one main conformation and one small conformation in the mitochondrial genome. Evolutionary analysis revealed that the bottle gourd mitochondrial genome was conserved, and collinearity analysis revealed that it underwent many rearrangements with the genomes of related species. Moreover, 27 homologous portions of the bottle gourd chloroplast and mitochondrial genomes, which totaled 23226 bp or 6.50% of the length of the mitochondrial genome, were identified. Thirty-eight mitochondrial PCGs were found to include a total of 589 possible RNA editing sites, all of which were altered from base C to U. In conclusion, we assembled and annotated a Cucurbitaceae mitochondrial genome, which advances our understanding of the Cucurbitaceae genome, lays a foundation for elucidating the phylogenetic relationships of Cucurbitaceae and emphasizes the need for further sequencing of the chloroplast, mitochondrial and nuclear genomes of additional Cucurbitaceae and closely related wild species.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number PRJNA1026380.

Author contributions

NQ: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. SY: Formal analysis, Methodology, Software, Visualization, Writing – review & editing. YW: Data curation, Formal analysis, Visualization, Writing – review & editing. HC: Data curation, Formal analysis, Visualization, Writing

– review & editing. YG: Data curation, Formal analysis, Writing – review & editing. XC: Formal analysis, Visualization, Writing – review & editing. SL: Methodology, Software, 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 Shanxi Basic Research Program Youth Project (202203021212432).

Acknowledgments

We would like to thank Professor Guoming Xing and Associate Professor Xiuping Kang for their valuable advice on the experimental design and thesis writing.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1416913/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Lagenaria siceraria mitochondrial genome assembly results and recombinant structure.

SUPPLEMENTARY FIGURE 2
Sanger sequencing results of *Lagenaria siceraria* mitochondrial genome repeats.

SUPPLEMENTARY FIGURE 3
Circular map of the whole *Lagenaria siceraria* chloroplast genome.

SUPPLEMENTARY FIGURE 4
Lagenaria siceraria particle MTPT verification.

References

Al-Faifi, S., Meyer, J. D., Garcia-Mas, J., Monforte, A. J., and Havey, M. J. (2008). Exploiting synteny in Cucumis for mapping of Psm: a unique locus controlling paternal mitochondrial sorting. *Theor. Appl. Genet.* 117, 523–529. doi: 10.1007/s00122-008-0796-1

Ali, A., Bang, S. W., Yang, E. M., Chung, S. M., and Staub, J. E. (2014). Putative paternal factors controlling chilling tolerance in Korean market-type cucumber (Cucumis sativus L.). *Scientia. Hortic.* 167, 145–148. doi: 10.1016/j.scienta.2014.01.004

Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K., and Palmer, J. D. (2010). Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). *Mol. Biol. Evol.* 27, 1436–1448. doi: 10.1093/molbev/msq029

Attia, Z., Pogoda, C., Vergara, D., and Kane, N. C. (2021). Mitochondrial genomes do not appear to regulate flowering pattern/reproductive strategy in Cannabis sativa. *AoB. Plants* 14, plab068. doi: 10.1093/aobpla/plab068

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

Boynton, J. E., Harris, E. H., Burkhart, B. D., Lamerson, P. M., and Gillham, N. W. (1987). Transmission of mitochondrial and chloroplast genomes in crosses of Chlamydomonas. *Proc. Natl. Acad. Sci. U. S. A.* 84, 2391–2395. doi: 10.1073/pnas.84.8.2391

Chakraborty, R., Tyagi, K., Kundu, S., Rahaman, I., Singha, D., Chandra, K., et al. (2018). The complete mitochondrial genome of Melon thrips, Thrips palmi (Thripinae): Comparative analysis. *PloS One* 13, e0199404. doi: 10.1371/journal.pone.0199404

Chat, J., Jáuregui, B., Petit, R. J., and Nadot, S. (2004). Reticulate evolution in kiwifruit (Actinidia, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *Am. J. Bot.* 91, 736–747. doi: 10.3732/ajb.91.5.736

Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43, 7762–7768. doi: 10.1093/nar/gkv784

Cho, S. E., Oh, J. Y., and Lee, D. H. (2023). The complete mitochondrial genome of Cladosporium anthropophilum (cladosporiaceae, dothideomycetes). *Mitochondrial. DNA B. Resour.* 8, 164–166. doi: 10.1080/23802359.2023.2167474

Cui, H., Ding, Z., Zhu, Q., Wu, Y., Qiu, B., and Gao, P. (2021). Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Sci. Rep.* 11, 1595. doi: 10.1038/s41598-020-80149-9

Del Valle-Echevarria, A. R., Kiełkowska, A., Bartoszewski, G., and Havey, M. J. (2015). The mosaic mutants of cucumber: A method to produce knock-downs of mitochondrial transcripts. *G3 (Bethesda).* 5, 1211–1221. doi: 10.1534/g3.115.017053

Del Valle-Echevarria, A. R., Sanseverino, W., Garcia-Mas, J., and Havey, M. J. (2016). Pentatricopeptide repeat 336 as the candidate gene for paternal sorting of mitochondria (Psm) in cucumber. *Theor. Appl. Genet.* 129, 1951–1959. doi: 10.1007/s00122-016-2751-x

Edris, R., Sultan, L. D., Best, C., Mizrahi, R., Weinstein, O., Chen, S., et al. (2024). Root primordium defective 1 encodes an essential PORR protein required for the splicing of mitochondria-encoded group II introns and for respiratory complex I biogenesis. *Plant Cell Physiol.* 65, 602–617. doi: 10.1093/pcp/pcad101

Faure, S., Noyer, J. L., Carreel, F., Horry, J. P., Bakry, F., and Lanaud, C. (1994). Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (Musa acuminata). *Curr. Genet.* 25, 265–269. doi: 10.1007/BF00357172

Garcia, L. E., Edera, A. A., Palmer, J. D., Sato, H., and Sanchez-Puerta, M. V. (2021). Horizontal gene transfers dominate the functional mitochondrial gene space of a holoparasitic plant. *New Phytol.* 229, 1701–1714. doi: 10.1111/nph.16926

Grienenberger, J. M. (2009). Plant mitochondrial RNA editing: the Strasbourg chapter. *IUBMB Life.* 61, 1110–1113. doi: 10.1002/iub.277

Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016

Hansen, A. K., Escobar, L. K., Gilbert, L. E., and Jansen, R. K. (2007). Paternal, maternal, and biparental inheritance of the chloroplast genome in Passiflora (Passifloraceae): implications for phylogenetic studies. *Am. J. Bot.* 94, 42–46. doi: 10.3732/ajb.94.1.42

He, P., Xiao, G., Liu, H., Zhang, L., Zhao, L., Tang, M., et al. (2018). Two pivotal RNA editing sites in the mitochondrial atp1mRNA are required for ATP synthase to produce sufficient ATP for cotton fiber cell elongation. *New Phytol.* 218, 167–182. doi: 10.1111/nph.14999

Huang, S., Li, L., Petereit, J., and Millar, A. H. (2020). Protein turnover rates in plant mitochondria. *Mitochondrion* 53, 57–65. doi: 10.1016/j.mito.2020.04.011

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, 241. doi: 10.1186/s13059-020-02154-5

Kang, B. C., Bae, S. J., Lee, S., Lee, J. S., Kim, A., Lee, H., et al. (2021). Chloroplast and mitochondrial DNA editing in plants. *Nat. Plants.* 7, 899–905. doi: 10.1038/s41477-021-00943-9

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

Kim, J. S., Jung, J. D., Lee, J. A., Park, H. W., Oh, K. H., Jeong, W. J., et al. (2006). Complete sequence and organization of the cucumber (Cucumis sativus L. cv. Baekmibaekdadagi) chloroplast genome. *Plant Cell Rep.* 25, 334–340. doi: 10.1007/s00299-005-0097-y

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Konan, J. A., Guyot, R., Koffi, K. K., Vroh-Bi, I., and Zoro, A. I. B. (2020). Molecular confirmation of varietal status in bottle gourd (Lagenaria siceraria) using genotyping-by-sequencing. *Genome* 63, 535–545. doi: 10.1139/gen-2020-0050

Kubo, N., and Arimura, S. (2010). Discovery of the rpl10 gene in diverse plant mitochondrial genomes and its probable replacement by the nuclear gene for chloroplast RPL10 in two lineages of angiosperms. *DNA Res.* 17, 1–9. doi: 10.1093/dnares/dsp024

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

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Lee, Y., and Park, J. K. (2022). Complete mitochondrial genome of Conuslischkeanus Weinkauff 1875 (Neogastropoda, Conidae) and phylogenetic implications of the evolutionary diversification of dietary types of Conus species. *Zookeys* 1088, 173–185. doi: 10.3897/zookeys.1088.78990

Lewis, S. E., Searle, S. M., Harris, N., Gibson, M., Lyer, V., Richter, J., et al. (2002). Apollo: a sequence annotation editor. *Genome Biol.* 3, RESEARCH0082. doi: 10.1186/gb-2002-3-12-research0082

Li, J., Tang, H., Luo, H., Tang, J., Zhong, N., and Xiao, L. (2023). Complete mitochondrial genome assembly and comparison of Camellia sinensis var. Assamica cv. Duntsa. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1117002

Li, J., Wang, D., Xue, P., Sun, H., Feng, Q., and Miao, N. (2021). The complete chloroplast genome of Scutellaria scordifolia (Labiatae). *Mitochondrial. DNA B. Resour.* 6, 84–85. doi: 10.1080/23802359.2020.1847621

Liang, L., Tang, W., Lian, H., Sun, B., Huang, Z., Sun, G., et al. (2022). Grafting promoted antioxidant capacity and carbon and nitrogen metabolism of bitter gourd seedlings under heat stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1074889

Liberatore, K. L., Dukowic-Schulze, S., Miller, M. E., Chen, C., and Kianian, S. F. (2016). The role of mitochondria in plant development and stress tolerance. *Free Radic. Biol. Med.* 100, 238–256. doi: 10.1016/j.freeradbiomed.2016.03.033

Liu, Y. T., Senkler, J., Herrfurth, C., Braun, H. P., and Feussner, I. (2023). Defining the lipidome of Arabidopsis leaf mitochondria: Specific lipid complement and biosynthesis capacity. *Plant Physiol.* 191, 2185–2203. doi: 10.1093/plphys/kiad035

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford nanopore minION sequencing and genome assembly. *Genomics Proteomics Bioinf.* 14, 265–279. doi: 10.1016/j.gpb.2016.05.004

Ma, L., Wang, Q., Zheng, Y., Guo, J., Yuan, S., Fu, A., et al. (2022). Cucurbitaceae genome evolution, gene function and molecular breeding. *Hortic. Res.* 9, uhab057. doi: 10.1093/hr/uhab057

Mandel, J. R., Ramsey, A. J., Holley, J. M., Scott, V. A., Mody, D., and Abbot, P. (2020). Disentangling complex inheritance patterns of plant organellar genomes: an example from carrot. *J. Hered.* 111, 531–538. doi: 10.1093/jhered/esaa037

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, pdb.prot5448. doi: 10.1101/pdb.prot5448

Møller, I. M., Rasmusson, A. G., and Van Aken, O. (2021). Plant mitochondria - past, present and future. *Plant J.* 108, 912–959. doi: 10.1111/tpj.15495

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Niu, R., Gao, X., Luo, J., Wang, L., Zhang, K., Li, D., et al. (2021). Mitochondrial genome of Aphis gossypii Glover cucumber biotype (Hemiptera: Aphididae). *Mitochondrial. DNA B. Resour.* 6, 922–924. doi: 10.1080/23802359.2021

Niu, Y., Zhang, T., Chen, M., Chen, G., Liu, Z., Yu, R., et al. (2023). Analysis of the complete mitochondrial genome of the bitter gourd (Momordica charantia). *Plants (Basel).* 12, 1686. doi: 10.3390/plants12081686

Ogihara, Y., Terachi, T., and Sasakuma, T. (1988). Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. U. S. A.* 85, 8573–8577. doi: 10.1073/pnas.85.22.8573

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Shao, C. S., Zhou, X. H., Miao, Y. H., Wang, P., Zhang, Q. Q., and Huang, Q. (2021). *In situ* observation of mitochondrial biogenesis as the early event of apoptosis. *iScience* 24, 103038. doi: 10.1016/j.isci.2021.103038

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345

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

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, e49. doi: 10.1093/nar/gkr1293

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PloS Comput. Biol.* 13, e1005595. doi: 10.1371/journal.pcbi.1005595

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

Wu, C. S., and Chaw, S. M. (2022). Evolution of mitochondrial RNA editing in extant gymnosperms. *Plant J.* 111, 1676–1687. doi: 10.1111/tpj.15916

Xu, M., Wang, Y., Zhang, M., Chen, M., Ni, Y., Xu, X., et al. (2023). Genome-wide identification of BES1 gene family in six cucurbitaceae species and its expression analysis in cucurbita moschata. *Int. J. Mol. Sci.* 24, 2287. doi: 10.3390/ijms24032287

Yang, H., Chen, H., Ni, Y., Li, J., Cai, Y., Wang, J., et al. (2023). Mitochondrial Genome Sequence of Salvia officinalis (Lamiales: Lamiaceae) Suggests Diverse Genome Structures in Cogeneric Species and Finds the Stop Gain of Genes through RNA Editing Events. *Int. J. Mol. Sci.* 24, 5372. doi: 10.3390/ijms24065372

Yang, J. X., Dierckxsens, N., Bai, M. Z., and Guo, Y. Y. (2023). Multichromosomal mitochondrial genome of paphiopedilum micranthum: compact and fragmented genome, and rampant intracellular gene transfer. *Int. J. Mol. Sci.* 24, 3976. doi: 10.3390/ijms24043976

Yang, Y., Lu, X., Yan, B., Li, B., Sun, J., Guo, S., et al. (2013). Bottle gourd rootstock-grafting affects nitrogen metabolism in NaCl-stressed watermelon leaves and enhances short-term salt tolerance. *J. Plant Physiol.* 170, 653–661. doi: 10.1016/j.jplph.2012.12.013

Yang, W., Zou, J., Yu, Y., Long, W., and Li, S. (2021). Repeats in mitochondrial and chloroplast genomes characterize the ecotypes of the Oryza. *Mol. Breed.* 41, 7. doi: 10.1007/s11032-020-01198-6

Yu, X., Duan, Z., Wang, Y., Zhang, Q., and Li, W. (2022). Sequence Analysis of the Complete Mitochondrial Genome of a Medicinal Plant, Vitex rotundifolia Linnaeus f. (Lamiales: Lamiaceae). *Genes (Basel).* 13, 839. doi: 10.3390/genes13050839

Yu, T., Tzeng, D. T. W., Li, R., Chen, J., Zhong, S., Fu, D., et al. (2019). Genome-wide identification of long non-coding RNA targets of the tomato MADS box transcription factor RIN and function analysis. *Ann. Bot.* 123, 469–482. doi: 10.1093/aob/mcy178

Zahoor, M., Ikram, M., Nazir, N., Naz, S., Batiha, G. E., Kamran, A. W., et al. (2021). A comprehensive review on the medicinal importance; biological and therapeutic efficacy of lagenaria siceraria (Mol.) (Bottle gourd) standley fruit. *Curr. Top. Med. Chem.* 21, 1788–1803. doi: 10.2174/1568026621666210701124628

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. *BMC Bioinf.* 14, 244. doi: 10.1186/1471-2105-14-244

Zhou, C., Wang, P., Zeng, Q., Zeng, R., Hu, W., Sun, L., et al. (2023). Comparative chloroplast genome analysis of seven extant Citrullus species insight into genetic variation, phylogenetic relationships, and selective pressure. *Sci. Rep.* 13, 6779. doi: 10.1038/s41598-023-34046-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 September 2024
DOI 10.3389/fpls.2024.1396635

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Micai Zhong,
Chinese Academy of Sciences, China
Yue Jing,
South China Agricultural University, China
Fuyan Liu,
Beijing Genomics Institute (BGI), China

*CORRESPONDENCE
Junqin Zhou
zhoujunqin@csuft.edu.cn
Xiaofeng Tan
T19781103@csuft.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 06 March 2024
ACCEPTED 31 July 2024
PUBLISHED 03 September 2024

CITATION
Gu Y, Yang L, Zhou J, Xiao Z, Lu M, Zeng Y and Tan X (2024) Mitochondrial genome study of *Camellia oleifera* revealed the tandem conserved gene cluster of *nad5–nads* in evolution.
Front. Plant Sci. 15:1396635.
doi: 10.3389/fpls.2024.1396635

Mitochondrial genome study of *Camellia oleifera* revealed the tandem conserved gene cluster of *nad5–nads* in evolution

Yiyang Gu[1,2†], Liying Yang[1,2,3†], Junqin Zhou[1,2,4*], Zhun Xiao[1,2,5], Mengqi Lu[1,2], Yanling Zeng[1,2] and Xiaofeng Tan[1,2*]

[1]Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha, China, [2]Academy of Camellia Oil Tree, Central South University of Forestry and Technology, Changsha, China, [3]Hunan Horticulture Research Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [4]College of Landscape Architecture, Central South University of Forestry and Technology, Changsha, China, [5]School of Foreign Languages, Changsha Social Work College, Changsha, China

Camellia oleifera is a kind of high-quality oil supply species. Its seeds contain rich unsaturated fatty acids and antioxidant active ingredients, which is a kind of high-quality edible oil. In this study, we used bioinformatics methods to decipher a hexaploid Camellia oil tree's mitochondrial (mt) genome based on second-generation sequencing data. A 709,596 bp circular map of *C. oleifera* mt genome was found for the first time. And 74 genes were annotated in the whole genome. Mt genomes of *C. oleifera* and three Theaceae species had regions with high similarity, including gene composition and gene sequence. At the same time, five conserved gene pairs were found in 20 species. In all of the mt genomes, most of *nad* genes existed in tandem pairs. In addition, the species classification result, which, according to the gene differences in tandem with nad5 genes, was consistent with the phylogenetic tree. These initial results provide a valuable basis for the further researches of *Camellia oleifera* and a reference for the systematic evolution of plant mt genomes.

KEYWORDS

***Camellia oleifera*, mitochondrial genome, Camellia, gene cluster, nad5-nads tandem genes**

Abbreviations: mt, mitochondrial; ORF, open reading frame; tRNA, transfer RNA; rRNA, ribosomal RNA; SSRs, Simple repeats.

1 Introduction

Camellia oleifera Abel., originating in China, is a non-wood forest that blooms and harvests in winter. *C. oleifera* is a kind of high-quality woody edible oil tree in Theaceae, which is famous for its rich unsaturated fatty acids in fruit (Tan, 2013). *Camellia* oil tree not only contains unsaturated fatty acids up to 90% but also is rich in antioxidant active ingredients such as squalene, sterol, vitamin E and polyphenols (He and He, 2002). *C. oleifera* with *Elaeis guineensis* Jacq., *Olea europaea* L. and *Cocos nucifera* L. are known as the world's four famous oil species. In recent years, in China, with the support of national policies, such as sustainable development, green development and targeted poverty alleviation, the development of *C. oleifera* has been widely supported due to its high economic value, wide use, comprehensive development, and utilization potential. *C. oleifera* is called "iron crop" and "green oil reservoir" by farmers (Wu et al., 2018). Studies on the development characteristics of *C. oleifera* fruits (Zhang et al., 2021), self-incompatibility of flowers (Zhou et al., 2020; Lu et al., 2021), and other economic potential exploitation have enabled researchers to have a further understanding of *C. oleifera*. With continuous progress of various technologies, plant genome research has become one of the most popular research topics. For researchers, the genome is a reference book for understanding a species or an organelle. At present, the chloroplast genome of *C. oleifera* has been reported (Wu et al., 2020), and the nuclear genome is also being carried out gradually. *C. oleifera* mitochondrial (mt) genome will also be the focus of the next step.

Mitochondria are important organelles in eukaryotic cells, which are closely related to energy generation, fatty acid synthesis, and active protein synthesis (Niu et al., 2022). Mitochondria are semi-autonomous organelles that encode their own functional genes and are regulated by nuclear genes. The mt genome of plants is more complex than that of animals and single-celled eukaryotes. The mt genomes of plants have different structures, gene contents, and DNA mutation rate strategies to meet their specific needs for physical, photosynthetic, or physiological and biochemical functions (Cui et al., 2009). Concurrently, cytoplasmic male sterility and species evolution in plants are closely related to the mitochondria (Wang et al., 2020). More than 200 plant mt genomes have been published, and most of them are crop species. However, there are only two articles on mt genome of Theaceae plants (Zhang et al., 2019; Rawal et al., 2020). The mt genome sequence is not as conserved as the chloroplast genome sequence, but some gene families are also conserved in the mt genome. There may even be concatenation of two genes or a group of genes. In Niu's study, two gene clusters (*rps12-nad3* and *rps3-rpl16*) were present in eight mitochondrial genomes (Niu et al., 2022). This is actually one of the dependencies on which all genomes can be assembled after sequencing.

With the development of sequencing technology, longer and more accurate reads can be obtained from the sequencing result, which enables many complex plant genomes to be gradually broken down. The sequencing results of one species include not only nuclear genome sequences but also mt genome sequences. Based on this, strategies and software extraction and assembly of mt genome sequences from whole-genome data are developing (King et al., 2014; Al-Nakeeb et al., 2017). MIA software (Green et al., 2008), for example, is an early mt genome assembly software. It takes mt genome sequences from the ancient human whole-genome sequence data and assembles them together. After then, software, such as MITObim (Hahn et al., 2013), ARC (Hunter et al., 2015), and NOVOPlasty (Dierckxsens et al., 2017), have been developed for every species other than people. Burns et al. (2015) also used MITObim to assemble the mt genome of *Cymbomonas tetramitiformis* when assembling its nuclear genome. Therefore, the genomes of complex species assembled from sequencing data can be used as a kind of data reference for research.

Therefore, the mt genome of *C. oleifera* was selected as the research object in this study. After obtaining all related reads from the sequencing data, the mt genome of *C. oleifera* was assembled by the MITObim software (v4.0.2) (Hahn et al., 2013). According to the sequence information of the genome, its structural characteristics and the situation of coding genes were annotated. The successful analysis of mt genome enables researchers to have a further understanding of *C. oleifera* and focus on the study of biological issues with more time and more energy.

2 Materials and methods

2.1 Materials and data background

The plant material *Camellia oleifera* cv. Huashuo was planted in the experimental field in Wangcheng district, Changsha, Hunan province (28°05′N, 113°2′E). Young leaves of adult *C. oleifera* trees were collected, preserved with liquid nitrogen, and sent to BGI Genomics for sequencing. The BGISEQ 500 Platform was used to build the second-generation sequencing database. The original data used in the experiment was extracted from the next-generation sequencing datasets of *Camellia oleifera*. All reads were trimmed and error corrected and then were provided for the assembly of mt genome of *C. oleifera* as data source.

2.2 Genome assembly

MITObim v4.0.2 (Hahn et al., 2013) was used for *C. oleifera* mt genome assembly. It can directly assemble the mt genomes of non-model species from DNA sequencing data embedded with MIRA and IMAGE modules. The mt genomes of *Camellia sinensis* var. *assamica* cv. Yunkang10 (YK10) (Zhang et al., 2019) and *Camellia sinensis* var. *assamica* cv. TV-1 (TV-1) (Rawal et al., 2020) were used as reference. All the paired-end sequencing reads of *C. oleifera* were mapped to the two mitochondrial reference sequences by BWA v0.7.12 (bwa index -p ref reference.fa) (Li and Durbin, 2009). According to the similarity of sequences, reads needed for mitochondrial assembly were captured preliminarily. Then, Samtools v1.9 was used to extract reads that were paired aligned to the reference sequence (bwa men *-t 16 ref.fa F.fq.gz R.fq.gz* | samtools view *-bF 12* | samtools sort *-@ 16 -m 1G -o output.bam*) (samtools view *-h output.bam*| tail *-n +4* | cut *-f 1*> map_reads.txt).

All reads were integrated to a new sequence file. This file was used as the input file of MITObim to assemble the mt genome of *C. oleifera*. Concurrently, *Camellia sinensis* mitochondrial sequences (Zhang et al., 2019; Rawal et al., 2020) were used as the main reference sequence in the assembly process.

2.3 Genome annotation and visualization

The GeSeq (Tillich et al., 2017) tool was used to annotate the mt genome of *C. oleifera*. Protein-coding genes, transfer RNA (tRNA), and ribosomal RNA (rRNA) genes were annotated by BLAST with the existing plant mt genome data in NCBI database (https://www.blast.ncbi.nlm.nih.gov), including *Camellia sinensis* var. *assamica* cv. Yunkang10 (MK574876.1) (Zhang et al., 2019), *Camellia sinensis* var. *assamica* cv. TV-1 (NC_043914.1) (Rawal et al., 2020), *Vitis vinifera* (NC_012119.1) (Goremykin et al., 2009), *Triticum aestivum* (NC_036024.1) (Cui et al., 2009), *Oryza sativa* subsp. *indica* (NC_007886.1) (Tian et al., 2006), *Zea mays* subsp. *mays* (NC_007982.1) (Clifton et al., 2004), *Glycine max* (NC_020455.1) (Chang et al., 2013), *Gossypium arboreum* (NC_035073.1) (Chen et al., 2017), *Ziziphus jujuba* (NC_029809.1), *Bupleurum falcatum* (NC_035962.1), *Boechera stricta* (NC_042143.1) (Li et al., 2018), and others. Concurrently, tRNA genes in the mt genome were annotated again with tRNA scan-SE tool (Lowe and Eddy, 1997). The annotated genes with coverage and match less than 60% were manually eliminated, and the repeated annotation results were compared. The final annotation results were drawn by Draw Organelle Genome Maps (OGDRAW v1.3.1) tool (Greiner et al., 2019) with a circular map and a linear map.

2.4 Repeat analysis and RNA-editing site prediction

The repeat sequence detection in the mt genome of *C. oleifera* was carried out by MISA v2.1 (Beier et al., 2017) and REPuter tools (Kurtz et al., 2001). Simple repeats (SSR) were verified by MISA, with the minimum number of nucleotide repeats setting as 8, 4, 4, 3, 3, and 3 for monomer, dimer, trimer, tetramer, pentamer, and hexamer, respectively (Zhang et al., 2019). At the same time, incomplete repeats of SSR interrupted by a few bases (spacing less than 100 or equal to 100) were screened, identified, and located. Forward and palindromic repeats are confirmed by REPuter with a minimum length of 50 nt and a minimum fault tolerance of 8 nt. The RNA-editing site was predicted by the PREP-Mt web tool (Mower, 2005). By reading the location information of genes in the annotation file, the base sequences of protein-coding genes were obtained from the genome. The gene sequence file was adjusted according to the format required by the tool, and the threshold was set at 0.2.

2.5 Phylogenetic analysis

The phylogenetic tree was constructed by IQ-Tree software (Nguyen et al., 2015), carried by TBtools (Chen et al., 2020). Published mt genome data of 19 species were selected from the NCBI database, including 16 dicotyledons as follows: *Arabidopsis thaliana* (Ath, NC_037304.1) (Unseld et al., 1997), *Brassica napus* (Bna, NC_008285.1) (Handa, 2003), *Bupleurum falcatum* (Bfa, NC_035962.1), *Camellia gigantocarpa* (Cgi, OP270590) (Lu et al., 2022), *Camellia sinensis* (Csi, MK574876.1; NC_043914.1) (Zhang et al., 2019; Rawal et al., 2020), *Capsicum annuum* (Can, NC_024624.1), *Carica papaya* (Cpa, NC_012116.1) (Magee et al., 2010), *Glycine max* (Gma, NC_020455.1) (Chang et al., 2013), *Gossypium arboretum* (Gar, NC_035073.1) (Chen et al., 2017), *Gossypium barbadense* (Gba, NC_028254.1) (Tang et al., 2015), *Malus domestica* (Mdo, NC_018554.1) (Goremykin et al., 2012), *Nicotiana tabacum* (Nta, NC_006581.1) (Sugiyama et al., 2005), *Rhazya stricta* (Rst, NC_024293.1) (Park et al., 2014), *Spinacia oleracea* (Sol, NC_035618.1) (Cai et al., 2017), and *Vitis vinifera* (Vvi, NC_012119.1) (Goremykin et al., 2009), two monocots: *Cocos nucifera* (Cnu, NC_031696.1) and *Triticum aestivum* (Tae, NC_036024.1) (Cui et al., 2009), and one gymnosperm: *Ginkgo biloba* (Gbi, NC_027976.1). These include two published species of the genus *Camellia*. The annotation information of all species was compared manually, including the *Camellia oleifera*. The sequences of 15 conserved genes were extracted, respectively, and aligned using Muscle v5 (Edgar, 2021) with default parameters. After integrating the comparison results, the portable IQ-tree software carried by TBtools (Chen et al., 2020) was used to build the phylogenetic tree, setting the model parameter at Auto and Boostrap value at 1,000 (iqtree *-s TBtools5888937064767651616.tmpIn -pre supergene.fa -bb 1000 -bnni -m MFP -nt AUTO*).

2.6 Collinearity analysis

MCScanxX (Wang et al., 2012) was used to analyze collinearity among the mt genomes of four Theaceae species. The sequences and GFF (general feature format) annotation files of two *Camellia sinensis* mt genomes were downloaded from the NCBI database (MK574876.1;NC_043914. 1) and the data number of *Camellia gigantocarpa* is OP270590. The sequence files and annotation files were modified to meet the input file format of One Step MCScanX software in TBtools (Chen et al., 2020). Running the software with the parameters (*-CPU for BlastP 2 -E-value 1e-10 -Num of BlastHits 5*) obtained the collinearity analysis results of *C. oleifera* and the other three, respectively. Then, the resulting data files were used as input files of Dual Systeny Plot software to draw the colinear map of mt genomes.

3 Results

3.1 Assembly and annotation

The next-generation sequencing yielded 525 G data files, of which 34,720,712 reads were compared to the reference genome. A 709,596-bp circle mitochondrial map of *C. oleifera* was obtained, and GC content reached 45.33% (Figure 1). A total of 42 protein-coding genes (including *orf*), 29 tRNA genes, and 2 rRNA genes (*rrnL* and *rrnS*)

were annotated in this mt genome. Among all protein-coding genes, Complex I (NADH dehydrogenase) family and Ribosomal protein (SSU) family exhibited the highest gene count. The total length of the exon (or CDS) region of the protein-encoding gene was 26,781 bp, including *sdh3*, *atp9*, *rps4*, *rps13*, *rps19*, and *orf102* with double copies, while the promoters of *nad2*, *rps4* (two copies), and *sdh3* were not "ATG." Among 29 tRNA genes, there are five copies of *trnM-CAU*, four copies of *trnnull-NNN*, three copies of *trnS-UGA*, and two copies of *trnD-GUC*, *trnN-GUU*, and *trnl-GAU* (Table 1).

3.2 Repeat analysis

A total of 530 SSRs were identified in the mt genome of *C. oleifera*. Among them, monomer, dimer, trimer, tetramer, pentamer, and hexamer accounted for 33.2%, 44.7%, 4.7%, 14.9%, 2.1%, and 0.4%, respectively (Table 2). Among the 176 monomer repeats, A/T accounted for the main proportion, reaching 86.9%, and C/G was only 13.1%. Among the dimer repeats, AT repeats up to eight times and TA repeats up to nine times, and there were only two hexamers (CTATCC and TTTCTA) (Supplementary Table 1). Furthermore, a total of 50 pairs of long repeat sequence (repeat unit >50 bp) were identified, including 20 pairs of forward repeats and 30 pairs of palindromic repeats (Table 3). The length of the shortest repeats was 139 bp, and the longest was 10,565 bp (Supplementary Table 2). Moreover, forward repeats mainly cluster in the earlier part of the genome sequence, while reverse repeats cluster in the later part.

3.3 RNA-editing sites analysis

With 35 protein-coding genes (including multiple copies of genes), 413 RNA-editing sites were predicted in the mt genome of *Camellia oleifera*. The *ccmB*, which belongs to Cytochrome C biogenesis, had the most editing sites (34), while rps14 had the

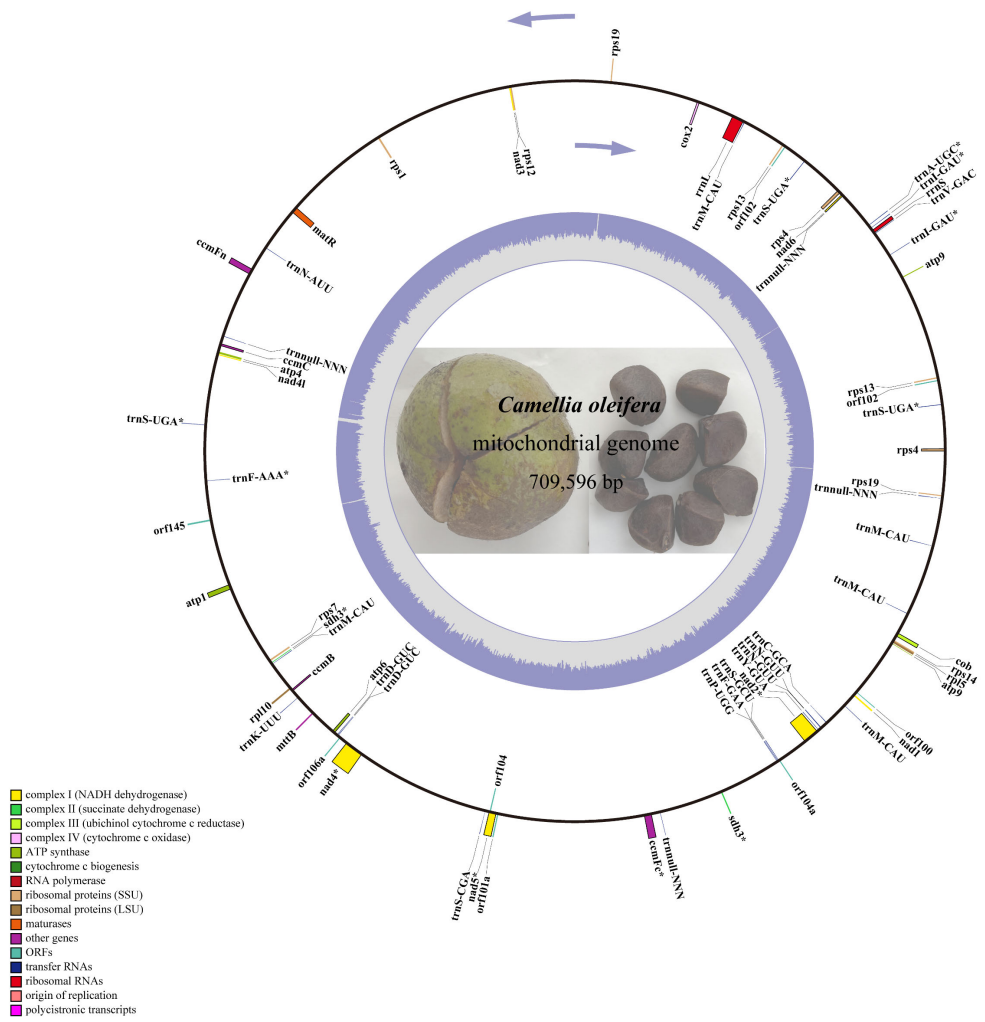


FIGURE 1
The circular map of *Camellia oleifera* cv. Huashuo mt genome. Gene map showing 74 annotated genes with different functional groups that are color-coded on the outer circle as transcribed clock-wise (outside) and transcribed counter clock-wise (inside). The inner circle indicates the GC content as gray purple plot.

10.3389/fpls.2024.1396635

TABLE 1 Gene content of mt genome with three *Camellia* species.

Category		*Camellia oleifera* cv. Huashuo		*Camellia sinensis* var. *assamica*		*Camellia gigantocarpa*
				TV-1	YK10	
Genome size (bp)		709,309		707,441	701,719 + 177,329	970,410
GC content (%)		45.33		45.75	45.63/45.81	45.00
Total genes number		74		66	73	66
Protein-coding genes number	Complex I (NADH dehydrogenase)	35	*nad1*, *nad2**, *nad3*, *nad4**, *nad4l*, *nad5**, *nad6*	31	44	44
	Complex II (succinate dehydrogenase)		*sdh3**(3x)			
	Complex III (ubiquinol cytochrome *c* reductase)		*cob*			
	Complex IV (cytochrome *c* oxidase)		*cox2*			
	Complex V (ATP synthase)		*atp1*, *atp4*, *atp6*, *atp9* (2x)			
	Cytochrome *c* biogenesis		*ccmB*, *ccmC*, *ccmFc**, *ccmFn*			
	Ribosomal proteins (SSU)		*rps1*, *rps4*(2x), *rps7*, *rps12*, *rps13*(2x), *rps14*, *rps19*(2x)			
	Ribosomal proteins (LSU)		*rpl5*, *rpl10*			
	Maturases		*matR*			
	Transport membrane proteins		*mttB*			
Transfer RNAs		29	*trnA-UGC**, *trnC-GCA*, *trnD-GUC* (2x), *trnF-AAA**, *trnF-GAA*, *trnK-UUU*, *trnM-CAU* (5x), *trnN-AUU*, *trnN-GUU* (2x), *trnP-UGG*, *trnS-CGA*, *trnS-GCU*, *trnS-UGA** (3x), *trnV-GAC*, *trnY-GUA*, *trnl-GAU** (2x), *trnnull-NNN* (4x)	29	24	19
Other genes		8	*orf100*, *orf101a*, *orf102 (2x)*, *orf104*, *orf104a*, *orf106a*, *orf145*	4	2	/
Ribosomal RNAs		2	*rrnL*, *rrnS*	2	3	3
Total SSRs		530		529	665	746
RNA-editing sites		413/35		429/31	478/44	483/44

*Gene with introns. Numbers of copies are shown in parenthesis for genes with multiple copies.

least (2) (Supplementary Table 3). By analyzing the relationship between the gene length and the number of RNA-editing sites, it was found that the longer the coding sequence, the more RNA-editing sites. However, there was no absolute linear relationship between them (Figure 2). All the RNA-editing were "C" to "U," and the number of editing sites at the second base of codon was the highest. There were predicted 267 sites (64.65%) and 20 sites (4.84%) at second base and first and second base of codon, respectively. Furthermore, no site at third base of codon was predicted (Supplementary Table 4). Among all editing sites, 105 sites (25.42%) enabled serine to convert into leucine, and 89 sites (21.55%) enabled proline to convert into leucine. They account for almost half of the total. However, the conversion of two glutamine and two arginine to terminators was predicted (Table 4).

10.3389/fpls.2024.1396635

TABLE 2 Statistics of SSR motifs in the *Camellia oleifera* cv. Huashuo mt genomes.

SSR-motif		Repeats/number	SSR number	SSR %
Monomer	A	8/33, 9/30, 10/4, 11/2, 12/2, 13/1 72)	176	33.2
	C	8/5, 9/3 (8)		
	G	8/14, 9/1 (15)		
	T	8/41, 9/26, 10/8,11/2, 12/2, 15/1, 16/1 (81)		
Dimer	AC	4/5 (5)	237	44.7
	AG	4/41, 5/9, 6/1, 7/2 (53)		
	AT	4/13, 5/4, 6/3, 7/3, 8/1, 9/1 (25)		
	CA	4/1 (1)		
	CG	4/3 (3)		
	CT	4/32, 5/11, 6/2 (45)		
	GA	4/23, 5/9, 6/1 (33)		
	GC	4/2 (2)		
	GT	4/5, 5/1 (6)		
	TA	4/17, 5/6, 6/2 (26)		
	TC	4/26, 5/3, 6/2 (31)		
	TG	4/7 (7)		
Trimer	AAC, AAG, AGC, GCT, TAT, TTG	4/1 (6)	25	4.7
	ACT, CTT, GAA, TTC	4/2 (8)		
	TAA, TTA	4/3 (6)		
	TCT	4/4 (4)		
	GAT	5/1 (1)		
Tetramer	AACA, AACC, AAGC, AAGG, AATA, AATG, ACTA, AGCT, AGTC, ATAA, ATGG, CAAT, CAGA, CCAG, CCGA, CCTT, CGGG, CTAG, CTCC, GAAG, GAGC, GCTT, GGCC, GTGA, TAGA, TAGT, TGAA, TTAC, TTAT, TTCC, TTTG	3/1 (31)	79	14.9
	AAAT, ATAG, ATTC, CAAA, CTTC, CTTG, GAAA, GAAT, GATT, GGAA	3/2 (20)		
	AAGA	3/3 (3)		
	AAAG, AGAA, TCTT	3/4 (12)		
	TTCT	3/5 (5)		
	CTTT	3/6 (6)		
	CCTT, TATC	4/1 (1)		
Pentamer	AATTA, ACTAG, ATAGG, CTAAT, GAAAT, GCCTT, TAAAG, TTAAT, TTATA	3/1 (9)	11	2.1
	TAGAG	3/2 (2)		
Hexamer	CTATCC, TTTCTA	3/1 (2)	2	0.4

3.4 Phylogenetic analysis of mt genomes

A total of 15 conserved genes were found from 20 mt genomes, including the *C.oleifera* mt genomes (*atp1*, *atp6*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFN*, *cob*, *cox2*, *nad3*, *nad4*, *nad4L*, *rps12*, *rpl5*, and *matR*). The protein sequences of 15 genes were compared and then tandem connected to obtain the super-sequence gene file to construct the phylogenetic tree. The ModelFinder program of IQ-Tree software

10.3389/fpls.2024.1396635

TABLE 3 Statistics of long repeat sequences in the *Camellia oleifera* cv. Huashuo mt genomes.

No.	Size (bp)	Palindromic Copy (P)		Size (bp)	Forward Copy (F)	
		Start 1	Start 2		Start 1	Start 2
1	174	17,713	364,406	722	0	87,394
2	162	54,865	684,866	582	727	88,125
3	191	105,254	648,172	433	1,332	88,732
4	281	152,801	709,308	447	1,764	89,165
5	236	153,082	709,073	139	2,244	89,668
6	196	153,452	708,740	487	2,387	89,807
7	322	153,664	708,406	221	2,872	90,293
8	279	154,100	708,013	3,371	3,092	90,514
9	205	154,375	707,811	10,565	6,621	94,050
10	222	154,698	707,473	639	17,179	104,607
11	268	154,920	707,207	4,060	17,884	105,321
12	532	155,184	706,680	406	21,940	109,367
13	202	155,716	706,479	3,366	22,336	109,762
14	414	156,046	705,936	159	25,700	113,113
15	1,108	156,454	704,843	3,883	25,857	113,261
16	585	157,561	704,260	162	54,865	177,382
17	253	158,141	704,013	139	54,957	647,939
18	637	158,392	703,379	171	78,225	409,425
19	2,444	159,024	700,935	302	172,510	681,489
20	2,045	161,607	698,753	151	309,545	309,848
21	2,622	163,649	696,135			
22	310	166,344	695,751			
23	420	166,714	695,271	**Long repeat sequences**		10,565 * 2 bp
24	6,180	167,250	688,964			
25	2,080	173,438	686,884	**Forward match (F)**		20 pairs
26	1,471	175,651	685,280			
27	2,834	177,197	682,379	**Palindromic match (P)**		30 pairs
28	397	330,626	501,796			
29	347	374,491	423,250			
30	302	681,489	689,582			

had tested 546 protein models, and HIVw+F+R3 was selected as the most suitable model. All species were clearly divided into three groups in the phylogenetic tree, including gymnosperm, monocotyledon, and dicotyledon, of which dicotyledon was the main group (Figure 3). *C. oleifera* with two cultivars of *C. sinensis* var. *assamica* (YK10 and TV-1) and *C. gigantocarpa* were grouped together in a branch with a bootstrap value of 100%. Species belonging to the same family were successfully grouped together.

3.5 Comparative analysis with *Camellia* species

At present, only three mt genomes (*C. sinensis* var. *assamica* cv. TV-1, *C. sinensis* cv. Yunkang10, and *C. gigantocarpa*) had been reported in the *Camellia* genus and even the Theaceae family. The mt genome of TV-1 and *C. gigantocarpa* had one single circular map of 707,441 and 970,410 bp, and YK10 had two complete circular maps (701,719 and 177,329 bp). *C. gigantocarpa* has the largest mt genome, and *C. gigantocarpa* and YK10 have the largest number of protein-coding genes (44), followed by *C. oleifera*. It was found that the succinate dehydrogenase group of *C. oleifera* and YK10 had the same gene composition. The NADH dehydrogenase group and ribosomal protein group of *C. oleifera* and TV-1 have the same gene composition. By comparing linear maps of four mt genomes, it was found that the arrangements of some genes in all genomes were similar (Figures 4, 5). The sites selected by the same color boxes in Figure 4 had almost identical gene member, and the sequences of these genes in the four genomes were almost the same. The difference was that the genes were arranged on the sense strand of *C. oleifera*, *C. gigantocarpa*, and TV-1, but on the anti-sense strand of the YK10. For example, the composition of gene clusters *atp1*, *rps7*, *rpl10*, *ccmB*, *mttB*, *atp6* and *nad4* was relatively conserved, and their arrangement was the same on mt genomes of *C. oleifera* and TV-1, but reversed on YK10's. The same was true of the other two gene clusters (*rps13*, *rrnL*, and *cox2*, and *rpl5*, *rps14*, and *cob*).

3.6 Conservation of tandem gene pairs on genomes

To study the conserved gene clusters on the mt genomes of all species, including *Camellia*, we mapped the linear mt genome structure map of 20 species (Figure 6). We found 17 conserved genes on 20 mt genomes, including 15 protein-coding genes and two ribosomal RNAs genes (*rrnL* and *rrnS*). By analyzing the location information and upstream and downstream genes of these 17 genes on the 20 mt genomes, we did not find their arrangement rules on the genome. But back to the location information of all the genes on 20 mt genomes, we found five conserved tandem gene pairs from 20 species, namely, *rps3-rpl6*, *rrn18-rrn5*, *rpl5-cob*, *nad3-rps12*, and *nad1-matR*. Sometimes, *rpl5-rps14-cob* tandem gene pairs existed on the genome instead of *rpl5-cob*.

In all the mt genomes, tandem gene pairs may be located on the positive-sense strand or the antisense strand. The arrangement of some tandem gene pairs was reversed on some mt genomes. For example, on the mt genomes of Bna, Ath, Cpa, and Gba, *nad3-rps2* gene pairs located on the positive-sense strand, but on the anti-sense strand of Vvi, Mdo, and Gma mt genomes. The sequence of *nad1–matR* tandem gene pairs was inversely complementary on the Ath and TV-1 mt genomes. However, some genes in these five tandem gene pairs were missing on the genomes of individual species. For example, *rpl6* is missing on Mdo and Sol mt genomes, and *rrn5* is missing on TV-1 and Col mt genomes. These

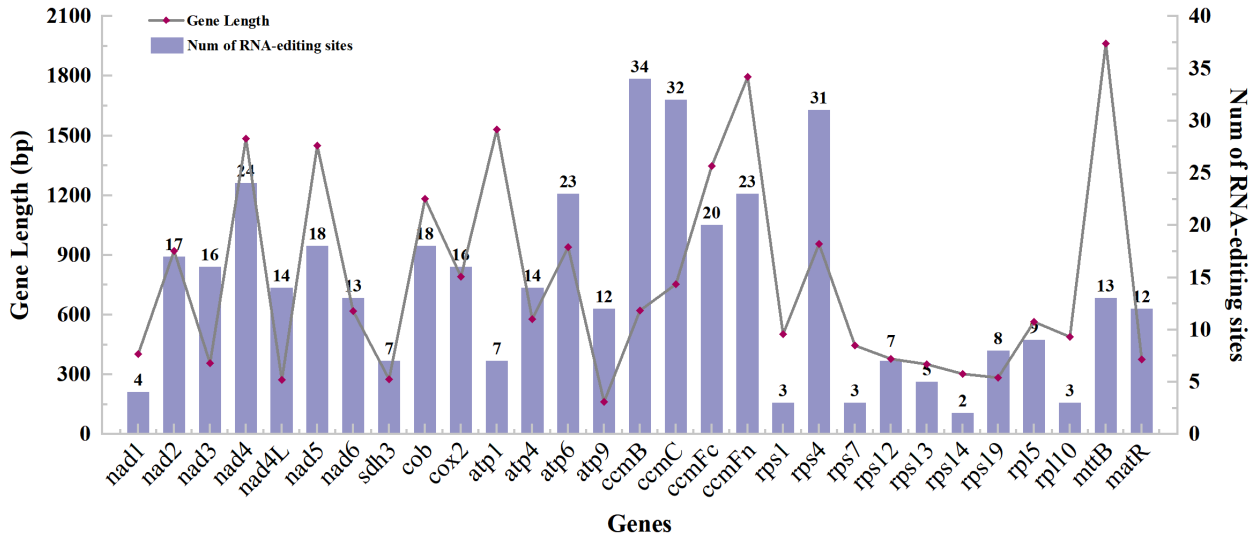


FIGURE 2
The gene length and predicted RNA-editing sites of protein-coding genes of the *Camellia oleifera* cv. Huashuo mt genome.

differences, such as the number and the type of genes annotated, may be related to the different annotation methods. Therefore, according to the arrangement of these five conserved gene pairs mentioned above, 20 species were divided into six groups. Bna, Ath, Cpa, Gba, and Gar formed group One. Vvi, Mdo, and Gma formed group Two. YK10, TV-1, Cgi, and Col were group Three. Rst, Bfa, Nta, and Can were group Four. Group Five only had Sol. Finally, Tae, Cnu, and Gbi formed group Six. The members of these six groups were consistent with the species branches in the evolutionary tree results. Therefore, these five gene pairs had a certain conservation in the evolutionary process.

3.7 Tandem rule of *nads* on mt genome

When analyzing the number and location information of all gene families on 20 genomes, several *nad–nad* gene pairs were found on each mt genome (Figure 7). For example, on Col, YK10, and TV-1, *nad4* and *nad5* were next to each other and also *nad2* and *nad1*. All *nads* with tandem arrangement in each species were identified, and the rules of *nad–nad* gene tandem were found in the six groups obtained from the previous classification. It was found that *nad1*, *nad5*, *nad6*, and *nad9* genes have the *nad–nad* gene tandem rule in group One, *nad1*, *nad2*, *nad4*, *nad5*, and *nad6* in group Two. In group Four, *nad1*, *nad2*, and *nad5* had the tandem rule and also *nad4L* and *nad5* in group Six. By integrating the results, it was found that all the mt genomes had the *nad5* gene, and there was also another *nad* gene next to it. So *nad5* always had the tandem rule during the evolution of the *nad* gene family. All the sequences of *nad* family genes were obtained from mt genomes, and the genes with a length less than 200 bp were removed. Finally, 233 *nad* genes were obtained for constructing the phylogenetic tree (Figure 8). Comparing the phylogenetic tree results of all nad subfamilies, the genetic distance of *nad5* was different with that of other subfamilies, which was divided into a separate group. At the same time, a species might have multiple *nad5* genes, and the *nad5* genes of all species were located in two branches of the evolutionary tree.

4 Discussion

The common *Camellia* oil tree known to us is a hexaploid plant with huge genomic data and complex structure (Gong et al., 2022). The chloroplast genome and mt genome could help to provide a more comprehensive auxiliary analysis of the genome evolution of *C. oleifera* (Li et al., 2021). With the advancement of sequencing methods, we have obtained more accurate sequences for the genome assembly. At the same time, original and simple methods are used to dig out the information from the genome sequence. So, this paper presents a complete mt genome of *C. oleifera* cv. Huashuo with a length of 709,596 bp, which is larger than most

TABLE 4 The amino acid transformation of RNA-editing sites in the *Camellia oleifera* cv. Huashuo mt genomes.

Amino acid	A-V	H-Y	L-F	P-F	P-L	P-S	Q-X	R-C	R-X	R-W	S-F	S-L	T-I	T-M
Count	12	23	13	20	89	33	2	34	2	19	50	105	6	5
Proportion(%)	2.91	5.57	3.15	4.84	21.55	7.99	0.48	8.23	0.48	4.60	12.11	25.42	1.45	1.21
Base position	2	1	1	1/2	2	1	1	1	1	1	2	2	2	2

of the known plant mt genomes. Our lab has reported the chloroplast genome of *C. oleifera* cv. Huashuo (Wu et al., 2020), and the goal of this study is to analyze the mt genome of *C. oleifera* based on the sequencing data.

4.1 Differences in gene quantity and conservation of gene arrangement among *Camellia* species

Compared with mt genomes of *C. sinensis* var. *assamica* (YK10, Zhang et al., 2019 and TV-1 Rawal et al., 2020) and *C. gigantocarpa*, the mt genome size of *C. oleifera* was 709,596 bp, which was similar to the mt genome of TV-1 (707,441 bp) (Table 1). The mt genome of YK10 and *C. gigantocarpa* was longer than theirs. According to all the previous studies, it was found that the evolution of plant mt genome was more complicated than that of chloroplast genome. The genome size, composition, and gene sequence varied greatly among different species, among different individuals within the same species, and even among different cells of the same individual (Li et al., 2011). Moreover, the two cultivated tea plants are from China and India. But even then, like the Bryophyte (Liu et al., 2014), the mitochondria genome remained structurally stable through evolution. Thus, all plants cannot be generalized; they need to have a case-by-case analysis.

A total of 35 protein-coding genes, 29 tRNAs, and 2 rRNAs were annotated in the mt genome of *C. oleifera*. The protein-encoding genes of TV-1 were four fewer than that of *C. oleifera*, and the total number of genes was four fewer than that of *C. oleifera*. The sequenced mt genomes of plants have been annotated to obtain approximately 35 protein-coding genes, such as 35 in *Vitis vinifera* (Goremykin et al., 2009), 36 in *Leucaena Trichandra* (Kovar et al., 2018), and 33 in *Dalbergia odorifera* (Hong et al., 2021). YK10 and Cgi has 9 protein-coding genes more than *C. oleifera*, but 5 and 10 tRNA genes fewer than the other two. Unseld et al. (1997) found that the number of genes was different among different species, which might be related to the number of ribosomal subunit genes and tRNAs (Li et al., 2011). Indeed, rpl genes in the *C. oleifera* mt genome (2 genes) and TV-1 (2 genes) are less than those in YK10 and Cgi (4 genes). In addition, the NADH dehydrogenase gene of *C.oleifera* was also lost compared with YK10 and Cgi. *C. oleifera*, Cgi, YK10, and TV-1 all belong to the Theaceae, but there is a great quantity variance in the number of annotated genes. It is possible that the methods used in genome annotation made this situation in closely related species. It is worth paying attention to the occurrence of such cases with comparative genomic analysis in the future. The mt of *C. oleifera* also contained eight ORF genes, which have been found in previous studies to be usually related to the "infertility" problem in plants (Chaumont et al., 1995). For example, the *T2urf13* in maize, *orf79* gene in rice, and *orf224*/*orf138* in rape

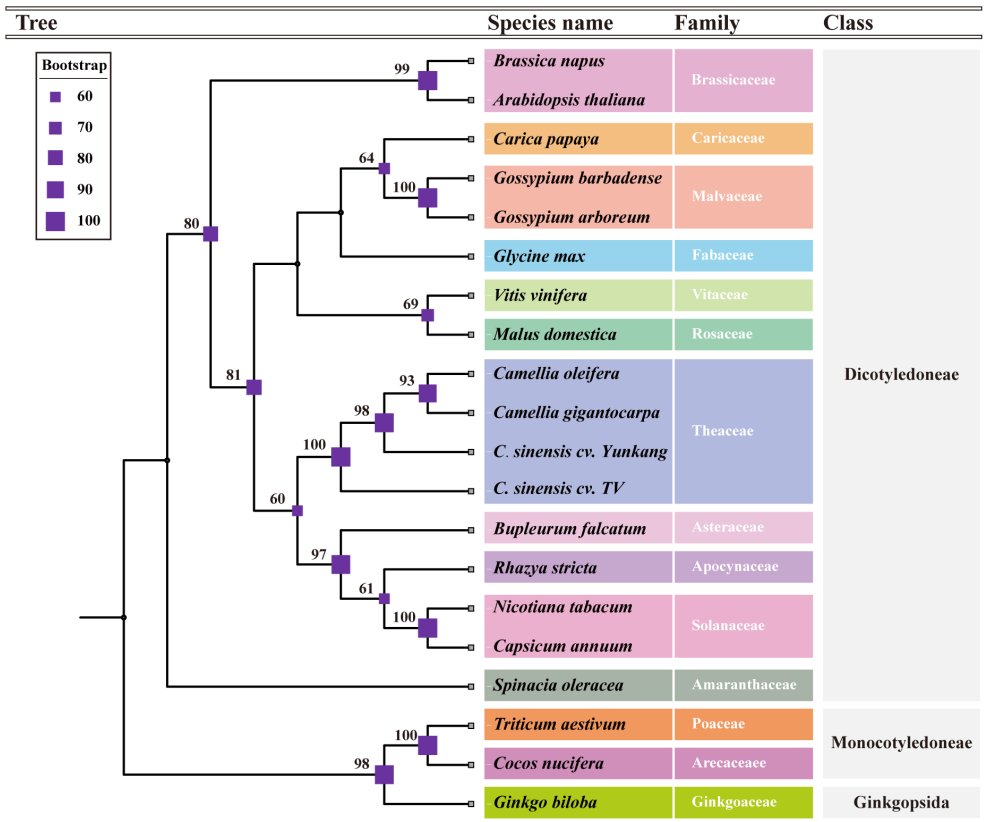


FIGURE 3
The phylogenetic tree based on the alignment of 18 other plants and *Camellia oleifera* cv. Huashuo mt genomes with bootstrap support values on each node. Different families can be distinguished by color.

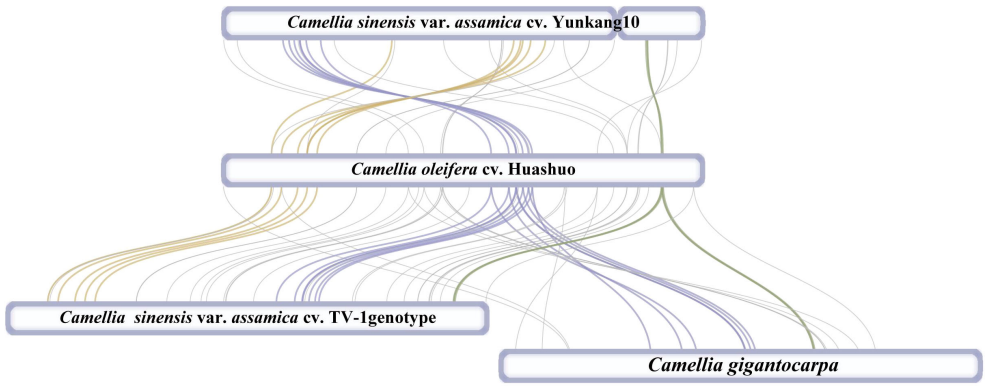


FIGURE 4
Collinearity analysis with mt genomes of four *Camellia* species. The genes indicated by the colors correspond to those in Figure 4.

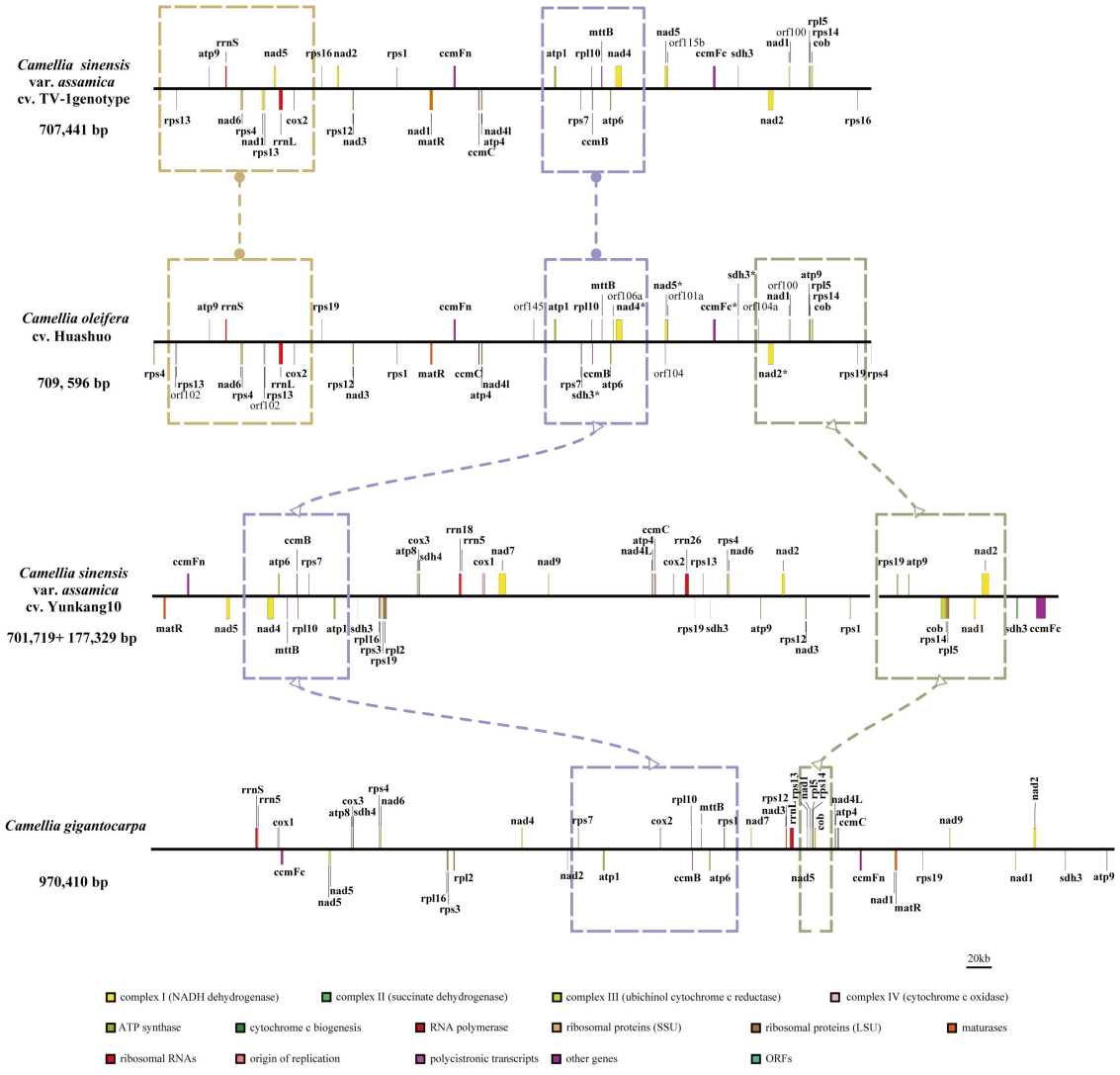


FIGURE 5
Comparison of mt genomes of four *Camellia* species (without transfer RNAs). Parts selected by the same color box have similar structure.

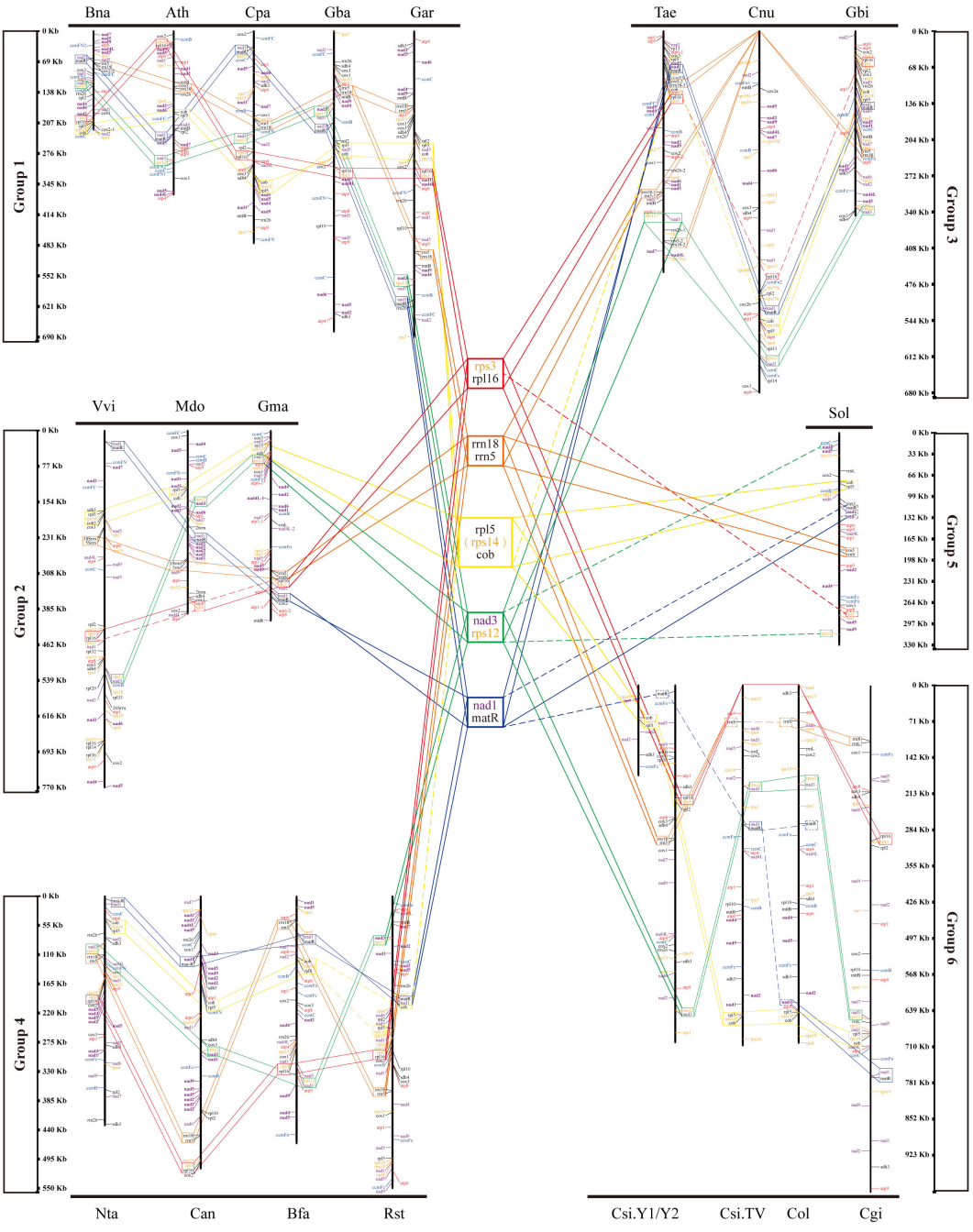


FIGURE 6
Gene localization in mt genomes of all species in the phylogenetic tree. Five gene clusters are in the middle of the figure.

are all associated with cytoplasmic male sterility (Duroc et al., 2005). The function of rich *orf* genes in *C. oleifera* mt can be explored later in combination with plant characteristics and nuclear genes.

RNA-editing sites can affect gene function by changing the protein folding pattern. The genome size and number of protein-coding genes of *C. oleifera* were larger than that of TV-1, but there were fewer RNA-editing sites. The number of protein-coding genes of *C. oleifera* and YKg10 differed by 9, but 65 differences in RNA-editing sites. These results confirmed that an increase in the number of bases in the mt genome did not lead to an increase in RNA-editing sites. At the same time, the number of editing sites of the same gene in different species is also different. In addition, when studying early terrestrial plants, Zhang found that the number of RNA-editing sites ranged from 0 to 2,152, and different types and numbers of RNA-editing sites contributed to the characteristics of plants (Zhang et al., 2020).

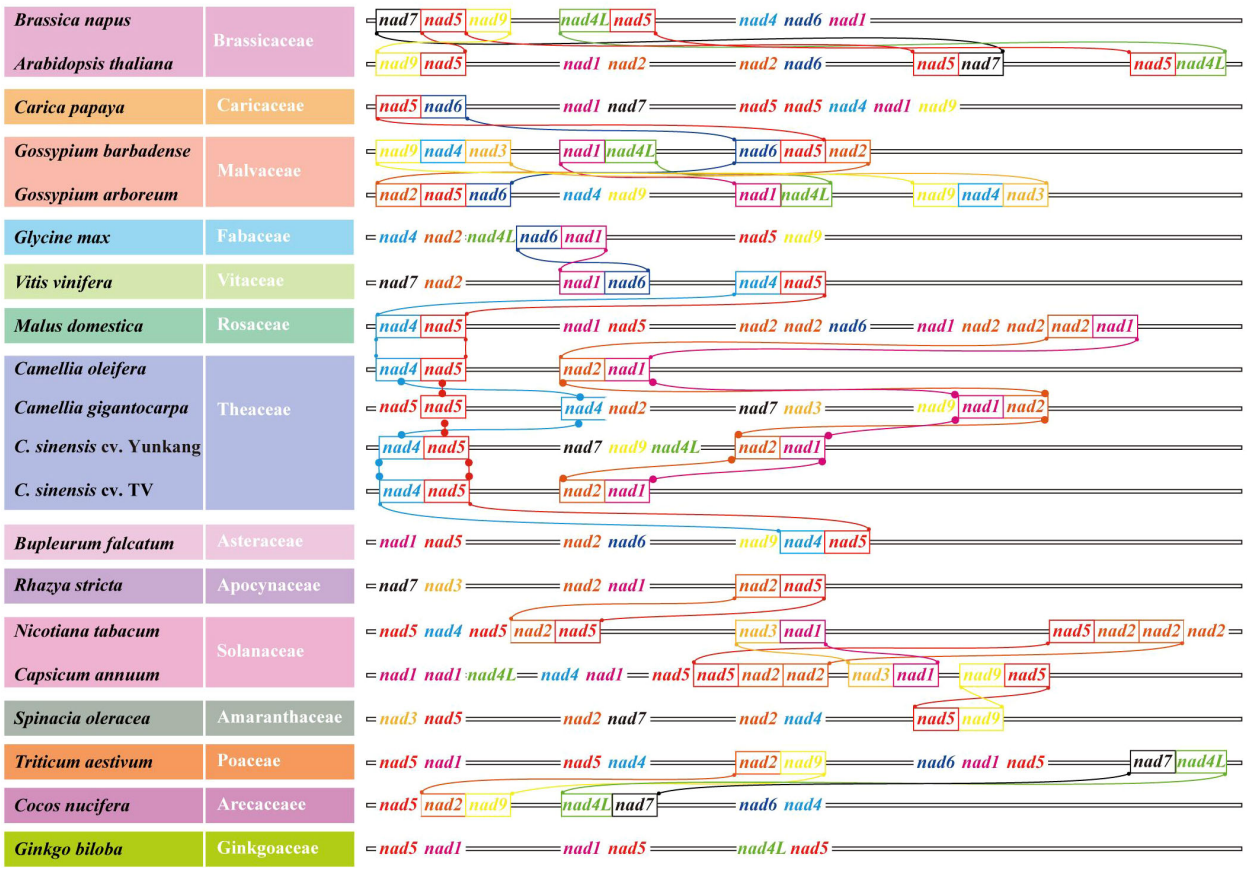


FIGURE 7
Nad–nad gene clusters in mt genomes of all species in the phylogenetic tree. Different *nad* genes were color coded. Identical tandem gene pairs between adjacent species were connected by curves.

4.2 Conserved gene clusters exist during the evolution of species, especially *nad5*

From the mt genome of *C. sinensis* and other previous studies, it is found that the species can be correctly classified by constructing an evolutionary tree through conserved genes. In this study, we screened 15 conserved genes from 20 mt genomes, and finally *C. oleifera*, *C. sinensis*, and *C. gigantocarpa* clustered together in one branch, and other species belonging to the same family also clustered together. We found that the number of conserved genes screened for evolutionary tree construction accounted for about one-third of the total number of genes (excluding tRNA genes). Concurrently, when comparing four mt genomes of *Camellia*, we also found that there were structurally conserved regions in the sequences of *C. oleifera* and other three.

Rearranging the position of genes in the mt genome can occur during evolution as the result of sequence break and sequence recombination. However, some highly conserved genes or gene clusters maintained their original evolutionary patterns throughout the process of evolution (Niu et al., 2022). In this study, five conserved gene clusters were identified across the 20 species: *rps3-rpl6*, *rrn18-rrn5*, *rpl5-cob*, *nad3-rps12*, and *nad1-matR*. But it was observed that some gene of these five gene pairs also were absent in some species. For instance, this study found that *S. oleracea* lacks the *rpl16* gene, which is also missing in *Melastoma dodecandrum* (Cai et al., 2017; Zhou et al., 2023). Additionally, *C. nucifera* lacks the *rps3*, which was relatively conserved in all species studied. So, this study grouped all species according to the conservativeness of all genes and found that the results of the grouping matched the position on the phylogenetic tree. Therefore, the absence of genes can be used to distinguish different evolutionary groups of species. The conservativeness of gene pairs can also be used to analyze the distance of genome evolution. Niu et al. found two conserved gene clusters (*rps12-nad3* and *rps3-rpl16*) in eight mt genomes (Niu et al., 2022). This result overlaps with the gene pairs found in this study, which shows that the conservativeness of gene clusters is universal in different species. It can be further explored during the study of mt genome.

When analyzing conservative genes, it was found that the *nad* gene has its own genetic rule. The *nad* gene is a part of the large enzyme complex of complex I, and it is active in the mitochondria (Wu et al., 2022). Complex I is one of the enzyme complexes necessary for oxidation phosphorylation. The mitochondrion is the main organelle that produces energy in the cell, which produces adenosine triphosphate (ATP), the main energy source of the cell, through the process of oxidative phosphorylation (Gualberto et al., 2013; Zhao et al., 2023). Therefore, for the whole life of an organism, the *nad* gene is a key role. When analyzing the genetic inheritance rules between different species, it showed that *nad* gene tandem

10.3389/fpls.2024.1396635

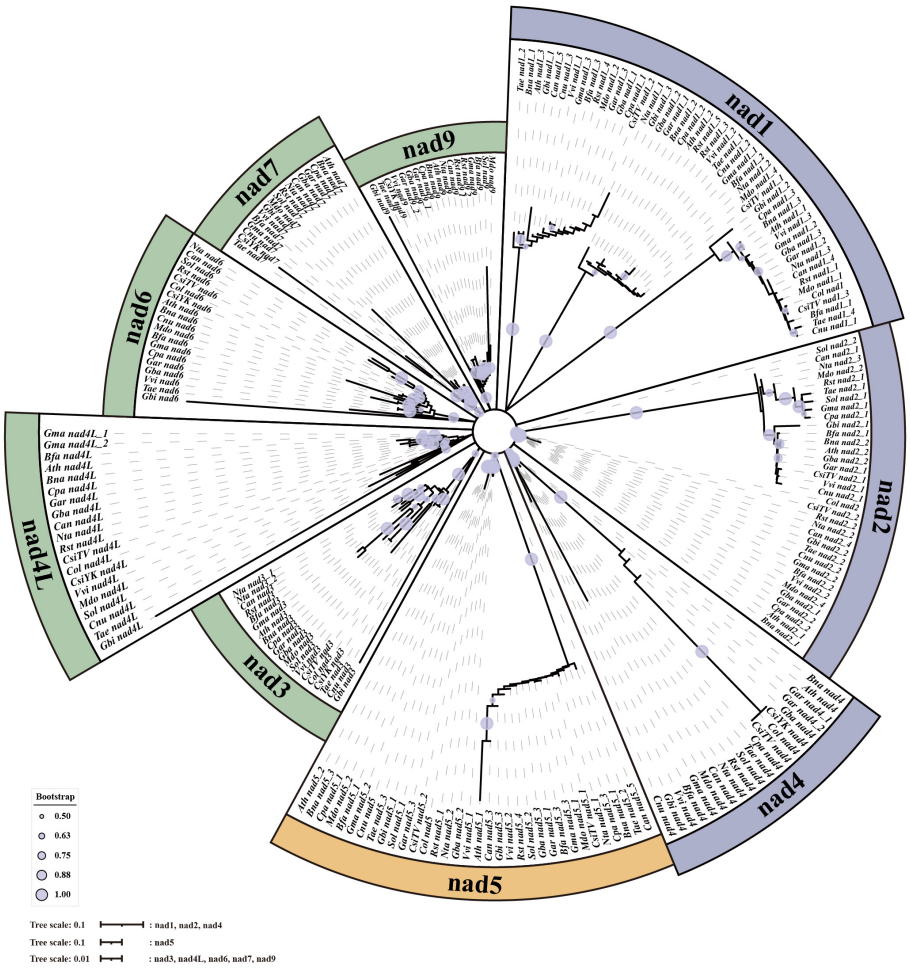


FIGURE 8
The phylogenetic tree of *nad* genes in mt genomes.

arranging was common in all the mt genomes. For example, *nad5–nad4* in *C. oleifera* and *C. sinensis*, *nad5–nad9*, *nad1–nad2*, and *nad5–nad7* in *A. thaliana* were located in adjacent locations in the mt genome (Unseld et al., 1997). With the phylogenetic tree, there were common *nad* tandem pairs in the species with close parental relationship. For example, *nad1/2/3/4/5/9/4L* in *G. barbadense* and *G. arboretum* has genomic tandem (Tang et al., 2015). Also, *nad1/2/3/4/5* in the mt genomes of *N. tabacum* and *C. annuum* existed in tandem (Sugiyama et al., 2005; Magee et al., 2010). In the mt genomes of gymnosperms, such as *G. biloba*, all *nad–nad* tandem gene pairs contain *nad5* (*nad5–nad1* and *nad5–nad4L*). It was found that, except for *Oryza minuta*, the mt genomes of most species contain *nad5* (Sajjad et al., 2016; Niu et al., 2022), and the *nad5* genes were arranged in tandem. Similarly, in the mt genome of *Dalbergia odorifera*, the *nad5* gene was found to be linked in tandem with *nad1*, as was the case with the *nad5* gene in *Leucaena trichandra* (Kovar et al., 2018; Zhou et al., 2021). Therefore, other *nad* genes may be required to induce or enhance the function of *nad5* in mitochondria life activity. *nad5* may have an ability to enhance its tandem genes' function.

4.3 Software parameters on gene annotation in mt genomes

When we annotated the *Camellia oleifera* mitochondrial genome, we found that the gene number of *nads* and other gene families was correlated with the software parameters that we set. At first, when the annotated genes were screened, the "coverage" and "match" were set to more than 40%. Twelve *nad* genes and all 48 protein-coding genes were obtained (Supplementary Figure 1). But when the two parameters were increased to 60% or even 90%, only seven *nad* genes were annotated (Supplementary Figure 1). In our manuscript, parameters with 60% were used. It was also found when we reannotated the mitochondrial genome of TV-1. In addition, the evolution-related regularities we found in our manuscript were mainly related to these seven *nad* genes. Combined with the subsequent analysis of the paper, we found that adjusting the values of coverage and matching degree during screening could help in obtaining more accurate results in comparative genomic analysis.

5 Conclusion

In summary, by sequencing and analyzing the mt genome of *Camellia oleifera*, we found the general tandem array of the *nad* gene family in the mt genome. Through the analysis of *nad* gene families in different species, we found that the *nad* gene with tandem relationship was species specific. In addition, it was noteworthy that the *nad5* has tandem gene pairs of *nad5–nads* in all species. The discovery of conserved genes and conserved gene pairs will provide a new direction for the study of the mitochondria.

Data availability statement

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0005996.

Author contributions

YG: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. LY: Data curation, Writing – original draft, Methodology. JZ: Writing – review & editing, Supervision, Conceptualization. ZX: Writing – original draft, Validation, Software. ML: Writing – original draft, Visualization, Investigation. YZ: Writing – review & editing, Project administration. XT: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

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 Hunan Province Innovative Province construction project (Project Number: 2021JC0007); and the Hunan Provincial Innovation Foundation for Postgraduate (Project Number: CX20220702).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1396635/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Gene annotation results with different parameters.

SUPPLEMENTARY TABLE 1
SSR motifs annotation of *C. oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 2
Long repeat sequences annotation of *C. oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 3
Statistics of RNA-editing sites in the *Camellia oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 4
RNA-editing sites annotation of *C. oleifera* cv. Huashuo mt genomes.

References

Al-Nakeeb, K., Petersen, T. N., and Sicheritz-Pontén, T. (2017). Norgal: extraction and *de novo* assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinf.* 18, 1–7. doi: 10.1186/s12859-017-1927-y

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

Burns, J. A., Paasch, A., Narechania, A., and Kim, E. (2015). Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. *Genome Biol. Evol.* 7, 3047–3061. doi: 10.1093/gbe/evv144

Cai, X. F., Jiao, C., Sun, H. H., Wang, X. L., Xu, C. X., Fei, Z. J., et al. (2017). The complete mitochondrial genome sequence of spinach, *Spinacia oleracea* L. *Mitochondrial. DNA B. Resour.* 2, 339–340. doi: 10.1080/23802359.2017.1334518

Chang, S. X., Wang, Y. K., Lu, J. J., Gai, J. Y., Li, J. J., Chu, P., et al. (2013). The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PloS One* 8, E56502. doi: 10.1371/journal.pone.0056502

Chaumont, F., Bernier, B., Buxant, R., Williams, M. E., Levings, 3. C.S., and Boutry, M. (1995). Targeting the maize T-urf13 product into tobacco mitochondria confers methomyl sensitivity to mitochondrial respiration. *P. Natl. A. Sci.* 92, 1167–1171. doi: 10.1073/pnas.92.4.1167

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., 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, Z., Nie, H., Grover, C. E., Wang, Y., Li, P., Wang, M., et al. (2017). Entire nucleotide sequences of *Gossypium raimondii* and *G.arboreum* mitochondrial genomes revealed A-genome species as cytoplasmic donor of the allotetraploid species. *Plant Biol.* 19, 484–493. doi: 10.1111/plb.12536

Clifton, S. W., Minx, P., Fauron, C. M.-R., Gibson, M., Allen, J. O., Sun, H., et al. (2004). Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* 136, 3486–3503. doi: 10.1104/pp.104.044602

Cui, P., Liu, H. T., Lin, Q., Ding, F., Zhuo, G. Y., Hu, S. N., et al. (2009). A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. *J. Genet.* 88, 299–307. doi: 10.1007/s12041-009-0043-9

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18–e18. doi: 10.1093/nar/gkw955

Duroc, Y., Gaillard, C., Hiard, S., DeFrance, M., Pelletier, G., and Budar, F. (2005). Biochemical and functional characterization of *ORF138*, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassiceae. *Biochimie* 87, 1089–1100. doi: 10.1016/j.biochi.2005.05.009

Edgar, R. C. (2021). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13, 6968. doi: 10.1038/s41467-022-34630-w

Gong, W. F., Xiao, S. X., Wang, L. K., Liao, Z. Y., Chang, Y. H., Mo, W. J., et al. (2022). Chromosome-level genome of *Camellia lanceoleosa* provides a valuable resource for understanding genome evolution and self-incompatibility. *Plant J.* 110, 881–898. doi: 10.1111/tpj.15739

Goremykin, V. V., Lockhart, P. J., Viola, R., and Velasco, R. (2012). The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant J.* 71, 615–626. doi: 10.1111/j.1365-313X.2012.05014.x

Goremykin, V. V., Salamini, F., Velasco, R., and Viola, R. (2009). Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* 26, 99–110. doi: 10.1093/molbev/msn226

Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L., Uhler, C., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134, 416–426. doi: 10.1016/j.cell.2008.06.021

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

Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2013). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016

Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41, e129–e129. doi: 10.1093/nar/gkt371

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* 31, 5907–5916. doi: 10.1093/nar/gkg795

He, F., and He, B. (2002). Cultural distribution and site classification for *Camellia oleifera*. *Sci. Silv. Sin.* 38, 64–72. doi: 10.1007/s11769-002-0026-8

Hong, Z., Liao, X. Z., Ye, Y. J., Zhang, N. N., Yang, Z. J., Zhu, W. D., et al. (2021). A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics* 22, 1–13. doi: 10.1186/s12864-021-07967-7

Hunter, S. S., Lyon, R. T., Sarver, B. A., Hardwick, K., Forney, L. J., Settles, M. L., et al. (2015). Assembly by reduced complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *Biorxiv* 014662. doi: 10.1101/014662

King, J. L., LaRue, B. L., Novroski, N. M., Stoljarova, M., BumSeo, S., Zeng, X. P., et al. (2014). High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic. Sci. Int.: Gen.* 12, 128–135. doi: 10.1016/j.fsigen.2014.06.001

Kovar, L., Nageswara-Rao, M., Ortega-Rodriguez, S., Dugas, D. V., Straub, S., Cronn, R., et al. (2018). PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. *Genome Biol. Evol.* 10, 2501–2517. doi: 10.1093/gbe/evy179

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., et al. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, J., Bi, C. W., Tu, J., and Lu, Z. H. (2018). The complete mitochondrial genome sequence of *Boechera stricta*. *Mitochondrial. DNA B. Resour.* 3, 896–897. doi: 10.1080/23802359.2018.1501323

Li, L., Hu, Y. F., He, M., Zhang, B., Wu, W., Cai, P. M., et al. (2021). Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of Camellia. *BMC Genomics* 22, 1–22. doi: 10.1186/s12864-021-07427-2

Li, S. S., Xue, L. F., Su, A. G., Lei, B. B., Wang, Y. M., and Hua, J. P. (2011). Progress on sequencing and alignment analysis of higher plant mitochondrial genomes. *J. China Agric. Univ.* 16, 22–27. doi: 10.3724/SP.J.1011.2011.00353

Liu, Y., Medina, R., and Goffinet, B. (2014). 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. *Mol. Biol. Evol.* 31, 2586–2591. doi: 10.1093/molbev/msu199

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Lu, C., Gao, L. Z., and Zhang, Q. J. (2022). A high-quality genome assembly of the mitochondrial genome of the oil-tea tree *camellia gigantocarpa* (Theaceae). *Diversity* 14, 850. doi: 10.3390/d14100850

Lu, M. Q., Zhou, J. Q., Liu, Y. Y., Yang, J., and Tan, X. F. (2021). *CoNPR1* and *CoNPR3. 1* are involved in SA-and MeSA-mediated growth of the pollen tube in *Camellia oleifera*. *Physiol. Plant.* 172, 2181–2190. doi: 10.1111/ppl.13410

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Sémon, M., Perry, A. S., et al. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20, 1700–1710. doi: 10.1101/gr.111955.110

Mower, J. P. (2005). PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinf.* 6, 1–15. doi: 10.1186/1471-2105-6-96

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Niu, L., Zhang, Y. Y., Yang, C. M., Yang, J., Ren, W., Zhong, X. F., et al. (2022). Complete mitochondrial genome sequence and comparative analysis of the cultivated yellow nutsedge. *Plant Genome* 15, e20239. doi: 10.1002/tpg2.20239

Park, S., Ruhlman, T. A., Sabir, J. S., Mutwakil, M. H. Z., Baeshen, M. N., Sabir, M. J., et al. (2014). Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics* 15, 405. doi: 10.1186/1471-2164-15-405

Rawal, H. C., Kumar, P. M., Bera, B., Singh, N. K., and Mondal, T. K. (2020). Decoding and analysis of organelle genomes of Indian tea (*Camellia assamica*) for phylogenetic confirmation. *Genomics* 112, 659–668. doi: 10.1016/j.ygeno.2019.04.018

Sajjad, A., Latif, K. A., Rahim, K. A., Muhammad, W., Kang, S. M., Aaqil, K. M., et al. (2016). Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. *PLoS One* 11, e0152937. doi: 10.1371/journal.pone.0152937

Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A., et al. (2005). The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genomics* 272, 603–615. doi: 10.1007/s00438-004-1075-8

Tan, X. F. (2013). Advances in the molecilar breeding of *Camellia oleifera*. J. *Cent. South Univ. Forest. Tech.* 43, 1–24. doi: 10.14067/j.cnki.1673-923x.2023.01.001

Tang, M. Y., Chen, Z. W., Grover, C. E., Wang, Y. M., Li, S. S., Liu, G. Z., et al. (2015). Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. *BMC Genomics* 16, 770. doi: 10.1186/s12864-015-1988-0

Tian, X. J., Zheng, J., Hu, S. N., and Yu, J. (2006). The rice mitochondrial genomes and their variations. *Plant Physiol.* 140, 401–410. doi: 10.1104/pp.105.070060

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

Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* 15, 57–61. doi: 10.1038/ng0197-57

Wang, R., Fang, Y. N., Wu, X. M., Qing, M., Li, C. C., Xie, K. D., et al. (2020). The miR399-*CsUBC24* module regulates reproductive development and male fertility in citrus. *Plant Physiol.* 183, 1681–1695. doi: 10.1104/pp.20.00129

Wang, Y. P., Tang, H. B., Debarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wu, F. Y., LI, Z., Liu, F. D., Wang, H., Shi, K., Sun, Y., et al. (2018). Comparison of photosynthetic characteristics of *Camellia oleifera* cultivars. *Nonwood. For. Res.* 36, 7–13. doi: 10.14067/j.cnki.1003-8981.2018.02.002

Wu, L. L., Li, J. A., Gu, Y. Y., Zhang, F. H., Li, Z., and Tan, X. F. (2020). Complete chloroplast genome sequences and phylogenetic analysis of three *Camellia oleifera* cultivars. *Int. J. Agric. Biol.* 24, 743–750. doi: 10.17957/IJAB/15.1495

Wu, Z. Q., Liao, X. Z., Zhang, X. N., Tembrock, L. R., and Broz, A. (2022). Genomic architectural variation of plant mitochondria— A review of multichromosomal structuring. *J. Syst. Evol.* 60, 160–168. doi: 10.1111/jse.12655

Zhang, F., Li, W., Gao, C. W., Zhang, D., and Gao, L. Z. (2019). Deciphering tea tree chloroplast and mitochondrial genomes of Camellia sinensis var. *assamica*. *Sci. Data* 6, 1–10. doi: 10.1038/s41597-019-0201-8

Zhang, F. F., Li, Z., Zhou, J. Q., Gu, Y. Y., and Tan, X. F. (2021). Comparative study on fruit development and oil synthesis in two cultivars of *Camellia oleifera*. *BMC Plant Biol.* 21, 1–16. doi: 10.1186/s12870-021-03114-2

Zhang, J., Fu, X. X., Li, R. Q., Zhao, X., Liu, Y., Li, M. H., et al. (2020). The hornwort genome and early land plant evolution. *Nat. Plants* 6, 107–118. doi: 10.1038/s41477-019-0588-4

Zhao, J. M., Chen, J., Xiong, Y., He, W., Xiong, Y. L., and Xu, Y. D. (2023). Organelle genomes of *Indigofera amblyantha* and *Indigofera pseudotinctoria*: comparative genome analysis, and intracellular gene transfer. *Ind. Crop Prod.* 198, 0926–6690. doi: 10.1016/j.indcrop.2023.116674

Zhou, H., Liao, X. Z., Ye, Y. J., Zhang, N. N., Yang, Z. J., Zhu, W. D., et al. (2021). A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics* 22, 672. doi: 10.1186/s12864-021-07967-7

Zhou, J. Q., Lu, M. Q., Yu, S. S., Liu, Y. Y., Yang, J., and Tan, X. F. (2020). In-depth understanding of *Camellia oleifera* self-incompatibility by comparative transcriptome, proteome and metabolome. *Int. J. Mol. Sci.* 21, 1600. doi: 10.3390/ijms21051600

Zhou, Y. Z., Zheng, R. Y., Peng, Y. K., Chen, J. M., Zhu, X. Y., Xie, K., et al. (2023). The first mitochondrial genome of Melastoma dodecandrum resolved structure evolution in Melastomataceae and micro inversions from inner horizontal gene transfer. *Ind. Crop Prod.* 205, 0926–6690. doi: 10.1016/j.indcrop.2023.117390

frontiers | Frontiers in Plant Science

TYPE Opinion
PUBLISHED 15 October 2024
DOI 10.3389/fpls.2024.1477240

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Ji Shi,
Peking University, China

*CORRESPONDENCE
Jun Tang
tangjun@caas.cn
Xuemin Wang
wangxuemin@caas.cn

RECEIVED 07 August 2024
ACCEPTED 30 September 2024
PUBLISHED 15 October 2024

CITATION
Tang J and Wang X (2024) Strategies for RNA m^{6}A modification application in crop improvement.
Front. Plant Sci. 15:1477240.
doi: 10.3389/fpls.2024.1477240

Strategies for RNA m^{6}A modification application in crop improvement

Jun Tang* and Xuemin Wang*

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

KEYWORDS

RNA modification, gene expression, m^{6}A modification, m^{6}A application, crop improvement, m^{6}A editing

Introduction

The improvement of crop yield and quality is an eternal theme to face climate change and population growth. The key to improving crop varieties lies in precisely manipulating gene expression. Recent advancements in CRISPR/Cas9 technology have made gene knockout increasingly straightforward, yet for genes related to important agronomic traits, it is crucial to regulate their expression levels appropriately. Complete knockout often results in defects in other aspects. In addition, many agronomic traits require upregulation of target gene expression for their improvement. Therefore, the development of novel methods for precise upregulation or downregulation of gene expression, without altering gene protein sequences or introducing new genome fragments, will significantly bolster the technical foundation for crop genetic improvement.

N^6-methyladenosine (m^{6}A) is the most abundant and reversible internal chemical modification in eukaryotic mRNA, which is installed, removed, and recognized by methyltransferases (writers), demethylases (erasers), and m^{6}A-binding proteins (readers), respectively (Tang et al., 2023). Currently, two types of m^{6}A methyltransferases have been identified in plants: multiprotein complexes and a single protein. The multiprotein complex includes MTA, MTB, FIP37, VIRILIZER (VIR), HAKAI, and HIZ2 (HAKAI interacting zinc finger protein 2), which catalyze the majority of m^{6}A modifications in mRNA (Parker et al., 2021; Ruzicka et al., 2017; Shen et al., 2016; Zhang et al., 2022; Zhong et al., 2008). The single protein FIONA1 also exhibits methyltransferase activity in Arabidopsis (Wang et al., 2022; Xu et al., 2022), catalyzing approximately 10% of m^{6}A modifications in mRNA. Several m^{6}A demethylases, which belong to the Fe (II)/α - kg dependent dioxygenase superfamily, have been identified in plants, including Arabidopsis AtALKBH10B and AtALKBH9B (Martinez-Perez et al., 2017), rice OsALKBH9 (Tang et al., 2024), and tomato SlALKBH2 (Zhou et al., 2019). m^{6}A is recognized by m^{6}A-binding proteins, such as ECTs in Arabidopsis, which contains the YTH domain. In plants, the ratio of m^{6}A/A in poly A$^+$ RNA varied among different tissues, with a range of 0.36–0.75% in Arabidopsis and 0.52–0.67% in rice, suggesting its high abundance (Wang et al., 2024). At the transcriptome level, m^{6}A sites are primarily enriched within the 3′-untranslated region (3′ UTR), followed by the coding DNA sequence (CDS) and 5′-untranslated region (5′ UTR). Recent studies have

demonstrated the crucial roles of m^{6}A in regulating gene expression in plants, primarily by influencing mRNA stability, translation, and 3′ UTR processing (Tang et al., 2023). Among them, mRNA stability regulation is one of the primary functions of m^{6}A, which involves two aspects: acceleration of RNA decay or preservation of RNA stability, depending on the specific m^{6}A-binding proteins. Recent studies combining proteomics and m^{6}A analysis have shown that m^{6}A in the untranslated regions is negatively correlated with protein abundance, suggesting that m^{6}A in UTR is likely to inhibit protein abundance in plants (Li et al., 2024). Therefore, manipulating m^{6}A could lead to an increase in protein abundance. m^{6}A modifications have also been found to play crucial roles in plant biology, such as embryo development, floral transition, stem cell fate determination, pollen development, fruit ripening, photomorphogenesis, circadian clock, nitrate signaling, and responses to biotic and abiotic stress (Tang et al., 2023). Given the high abundance and crucial roles of m^{6}A modifications in gene expression regulating, altering m^{6}A modifications in genes holds promise as a strategy for enhancing crop agronomic traits.

Manipulating m^{6}A modification to improve crop agronomic traits

In mammals, the m^{6}A demethylase FTO, known as an obesity gene, plays a crucial role in regulating body weight. Researchers have genetically engineered rice and potato to express FTO. In field experiments, the yield and biomass of genetically engineered rice and potato increased by approximately 50% (Yu et al., 2021). Further research indicates that the expression of FTO in rice promotes root growth, tiller bud formation, photosynthetic efficiency, and drought resistance, and these phenotypes are dependent on the m^{6}A demethylase activity of FTO (Yu et al., 2021). In strawberries, inhibiting the expression of m^{6}A methyltransferase genes *FveMTA* or *FveMTB* can delay fruit ripening, while upregulating the expression of either gene can accelerate fruit ripening, suggesting that manipulating m^{6}A modification can regulate the ripening time of strawberries (Zhou et al., 2021). The above research suggests that manipulating the m^{6}A modification level overall can enhance the agronomic traits of crops.

In addition to altering the overall m^{6}A modification level, researchers have also endeavored to edit m^{6}A modification levels on individual genes. In plants, a precise editing system for plant mRNA m^{6}A has been successfully developed by fusing dCas13a with the plant m^{6}A methyltransferase core complex MTA-MTB or the mammalian demethylase ALKBH5 (Shi et al., 2024). By specifically editing the m^{6}A modification on the *SHR* transcript, which is a key gene for root development, it was found that an increase in the m^{6}A modification level of the *SHR* transcript can promote the enlargement of the aboveground and root parts of plants, increase leaf area, plant height, biomass, and grain yield, thereby promoting plant growth (Shi et al., 2024). In cotton, similar m^{6}A editing tools have also been developed, combining CRISPR/dCas13(Rx) with the methyltransferase GhMTA (Targeted RNA Methylation Editor, TME) or the demethyltransferase GhALKBH10 (Targeted RNA Demethylation Editor, TDE) (Yu et al., 2024). Using TME editor, the m^{6}A level of *GhDi19* transcript increased, and the plants with increased m^{6}A levels results in a significant increase in root length and enhanced drought resistance. Both works indicate that manipulating the m^{6}A modification level of key genes can regulate plant phenotype and improve agronomic traits.

Strategies for improving crop traits through m^{6}A modification

According to current research, three strategies have been proposed for enhancing crop agronomic traits through m^{6}A modification. (1) Altering RNA m^{6}A modification regulatory proteins (RMRPs) (methyltransferases, demethylases, and recognition proteins) and their interacting proteins. Given that m^{6}A regulatory proteins typically regulate a multitude of substrate genes, this strategy exhibits a high degree of randomness and uncertainty in improving agronomic traits, and may frequently result in other phenotypic abnormalities; (2) Conducting m^{6}A editing by fusing Cas13 (dCas13) with RNA-modified methyltransferase or demethylase in specific mRNA. However, the RNA-based m^{6}A editing strategy poses limitations in crop breeding applications: the m^{6}A editing vector must be maintained in the offspring, otherwise the editing effect of m^{6}A cannot be preserved. Currently, they are not suitable for practical breeding applications; (3) Editing m^{6}A motif of specific gene to remove m^{6}A modifications. For example, using a base editor to accurately replace m^{6}A-modified adenine on DNA. Regarding the m^{6}A modification site on the UTR, the CRISPR/Cas9 system can be employed to disrupt the m^{6}A modification motif on the UTR, thereby eliminating the m^{6}A modification. This strategy involves directly altering m^{6}A motifs on DNA, enabling the production of m^{6}A editing materials without transgenic vectors in offspring, which can be utilized in breeding applications.

Steps for m^{6}A editing

Steps 1: drawing single-base resolution m^{6}A modification map

For the purpose of m^{6}A editing, accurate m^{6}A modification maps at single-base resolution must be generated initially. With the advancements in m^{6}A detection technology, various single base resolution m^{6}A sequencing methods have been reported, among them, m^{6}A-selective allyl chemical labeling and sequencing (m^{6}A-SAC-seq) (Hu et al., 2022), Nanopore direct RNA sequencing (DRS) (Zheng et al., 2020) and Glyoxal and nitrite-mediated deamination of unmethylated adenosine and sequencing (GLORI-seq) (Liu et al., 2022) are three methods that have broad application prospects (Figure 1). Recent research has reported the transcriptome-wide m^{6}A maps at single-base resolution in various tissues of rice (plumule dark, plumule light, seedling at 8 days, seedling at 2 weeks, panicle, flag leaf at 10 days after anthesis, endosperm at 10 days after anthesis, and embryo at 10 days after anthesis) and Arabidopsis (seedling, shoot, root, rosette leaf,

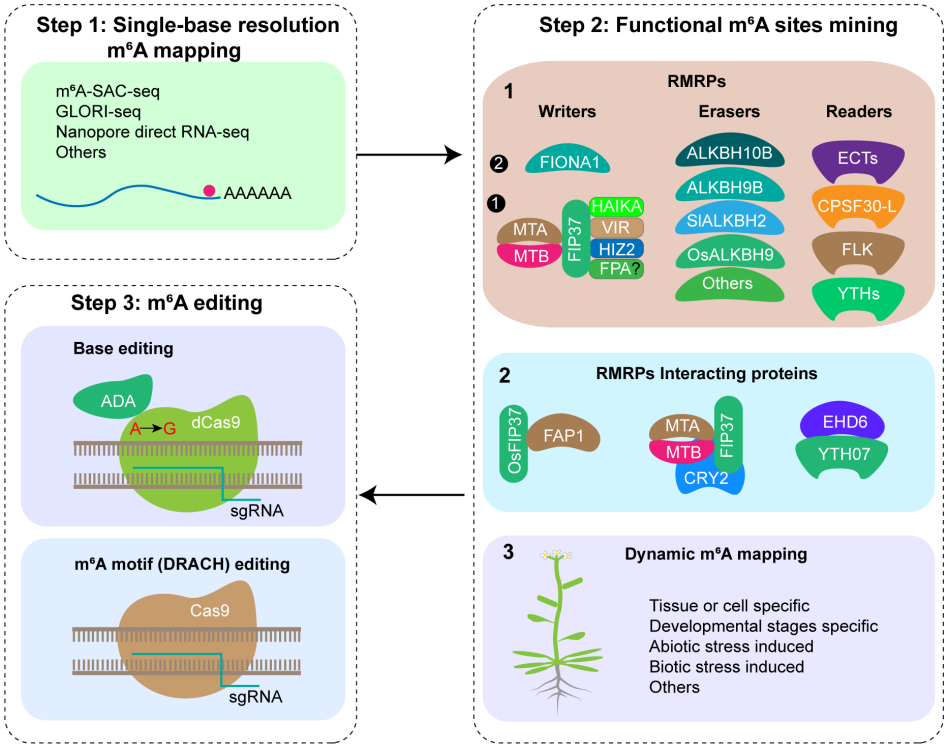


FIGURE 1
Putative steps of m^{6}A editing in crop traits improvement. Step 1: Mapping mRNA m^{6}A modification at single-base resolution; Step 2: Functional m^{6}A sites mining by functional study of RMRPs, RMRPs interacting proteins, and dynamic m^{6}A modification mapping; Step 3: m^{6}A editing using CRISPR/Cas9 system or base editor.

cauline leaf, stem, flower, silique, and seed) using m^{6}A-SAC-seq, identifying a total of 205,691 m^{6}A sites distributed across 22,574 genes in rice, and 188,282 m^{6}A sites across 19,984 genes in Arabidopsis, offering comprehensive resources for investigating single-base resolution m^{6}A sites and their functions in plants (Wang et al., 2024). Given the advancements in SAC-seq, DRS, and GLORI-seq, there are no technical limitations to m^{6}A single base resolution sequencing, and single-base resolution maps for other crops can be constructed rapidly.

Step 2: identification of functional m^{6}A sites

Since there are more than ten thousands m^{6}A sites present in the plant transcriptome, which ones have regulatory functions remains a key question to be clarified. Currently, three strategies are employed for identifying functional m^{6}A sites (Figure 1). (1) Functional study of RMRPs. By utilizing mutants, RNA interference, or overexpression techniques for RMRPs, analyzing the phenotypes of relevant plant materials at different growth and development stages and in response to various biotic and abiotic stresses. Next, m^{6}A sequencing and corresponding molecular biology methods are employed to ultimately confirm the m^{6}A modification sites associated with the observed phenotypes. Utilizing RMRPs, several m^{6}A modification sites with crucial functions have been identified, such as m^{6}A in *STM* and *WUS* mRNA regulating shoot stem cell fate in Arabidopsis (Shen et al., 2016), *FT*, *SPL3*, and *SPL9* mRNA regulating floral transition in Arabidopsis (Duan et al., 2017), and *OsYUCCA3*, *TDR*, and *GAMYB* mRNA regulating rice pollen development (Cheng et al., 2022; Tang et al., 2024), etc. In the future, we should enhance the application of genetic materials, comprehensively study their functions under various growth and development stages, environmental stress and stimulation, and different pest and disease stress in the field, and explore important functional m^{6}A sites. (2) Functional study of proteins interacting with RMRPs. Interaction proteins of RMRPs might be involved in the processes of installing, removing, or recognizing m^{6}A modifications at specific sites on specific mRNAs. Several studies have demonstrated the selection of specific m^{6}A sites by interacting proteins of RMRPs. In Arabidopsis, CRY2 recruits the m^{6}A writer complex (MTA/MTB/FIP37) to the photobody under blue light, thereby promoting m^{6}A modification of circadian clock associated 1 (*CCA1*) and enhancing its mRNA stability (Wang et al., 2021). In rice, OsFIP37 is recruited by the RNA-binding protein OsFAP1 to deposit the m^{6}A modification on *OsYUCCA3* mRNA, thereby regulating male germ cell miosis (Cheng et al., 2022). EARLY HEADING DATE6 (EHD6) recruits the m^{6}A reader YTH07 and sequesters *OsCOL4* mRNA into phase-separated ribonucleoprotein condensates, thereby promoting rice flowering (Cui et al., 2024). (3) Exploring key m^{6}A modification sites through dynamic m^{6}A modification mapping. By mapping the changes in m^{6}A modification maps under various tissues, developmental stages of the same tissue,

and stress treatments of crops, m^{6}A modification sites specific to corresponding processes are expected to be identified, which may have relevant regulatory functions.

Step 3: m^{6}A editing using CRISPR/Cas9 system or base editor

After generating a single-base resolution m^{6}A modification map and identifying functional m^{6}A modification sites, m^{6}A motif of specific gene can be edited using the CRISPR/Cas9 system or base editor (Figure 1). After gene editing, screening transgenic positive seedlings, identifying mutations, and confirming m^{6}A modification changes, m^{6}A-edited plants were obtained. Subsequently, phenotype analysis will be conducted and transgenic free plants will be screened for subsequent breeding applications.

Discussion

Fine-tuning gene expression is of great significance for crop improvement. m^{6}A, the most abundant modification in mRNAs, plays crucial roles in regulating gene expression, positioning it as a potential tool for manipulating gene expression. With the advent of new m^{6}A sequencing methods, mapping m^{6}A at single base resolution is no longer an obstacle. The primary limitation of m^{6}A application is the insufficient mining of functional m^{6}A sites. The most effective approach to explore functional m^{6}A sites in crops involves studying the biological functions of m^{6}A methyltransferases, demethylases, or m^{6}A binding proteins in crop development, biotic and abiotic stress response, and other processes. However, fewer RMRPs have been identified in crops so far, and future research should strengthen the identification and functional research of crop RMRPs. In summary, because of the vital roles of m^{6}A in regulating gene expression post-transcriptionally, the m^{6}A editing strategy holds significant potential for crop improvement.

Author contributions

JT: Writing – original draft, Writing – review & editing. XW: Writing – original draft, 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 earmarked fund for China Agriculture Research System (CARS-34) and the National Crop Germplasm Resources Center (NCGRC-63).

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

References

Cheng, P., Bao, S., Li, C., Tong, J., Shen, L., and Yu, H. (2022). RNA N^6-methyladenosine modification promotes auxin biosynthesis required for male meiosis in rice. *Dev. Cell.* 57, 246–259.e4. doi: 10.1016/j.devcel.2021.12.014

Cui, S., Song, P., Wang, C., Chen, S., Hao, B., Xu, Z., et al. (2024). The RNA binding protein EHD6 recruits the m^{6}A reader YTH07 and sequesters OsCOL4 mRNA into phase-separated ribonucleoprotein condensates to promote rice flowering. *Mol. Plant* 17, 935–954. doi: 10.1016/j.molp.2024.05.002

Duan, H. C., Wei, L. H., Zhang, C., Wang, Y., Chen, L., Lu, Z., et al. (2017). ALKBH10B is an RNA N^6-methyladenosine demethylase affecting arabidopsis floral transition. *Plant Cell* 29, 2995–3011. doi: 10.1105/tpc.16.00912

Hu, L., Liu, S., Peng, Y., Ge, R., Su, R., Senevirathne, C., et al. (2022). m^{6}A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol.* 40, 1210–1219. doi: 10.1038/s41587-022-01243-z

Li, S.-T., Ke, Y., Zhu, Y., Zhu, T.-Y., Huang, H., Li, L., et al. (2024). Mass spectrometry-based proteomic landscape of rice reveals a post-transcriptional regulatory role of N^6-methyladenosine. *Nat. Plants.* 10, 1201–1214. doi: 10.1038/s41477-024-01745-5

Liu, C., Sun, H., Yi, Y., Shen, W., Li, K., Xiao, Y., et al. (2022). Absolute quantification of single-base m^{6}A methylation in the mammalian transcriptome using GLORI. *Nat. Biotechnol.* 41, 355–366. doi: 10.1038/s41587-022-01487-9

Martinez-Perez, M., Aparicio, F., Lopez-Gresa, M. P., Belles, J. M., Sanchez-Navarro, J. A., and Pallas, V. (2017). Arabidopsis m^{6}A demethylase activity modulates viral infection of a plant virus and the m^{6}A abundance in its genomic RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10755–10760. doi: 10.1073/pnas.1703139114

Parker, M. T., Knop, K., Zacharaki, V., Sherwood, A. V., Tome, D., Yu, X., et al. (2021). Widespread premature transcription termination of arabidopsis thaliana NLR genes by the spen protein FPA. *Elife* 10, e65537. doi: 10.7554/eLife.65537

Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., et al. (2017). Identification of factors required for m^{6}A mRNA methylation in arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* 215, 157–172. doi: 10.1111/nph.14586

Shen, L., Liang, Z., Gu, X., Chen, Y., Teo, Z. W., Hou, X., et al. (2016). N^6-methyladenosine RNA modification regulates shoot stem cell fate in arabidopsis. *Dev. Cell* 38, 186–200. doi: 10.1016/j.devcel.2016.06.008

Shi, C., Zou, W., Liu, X., Zhang, H., Li, X., Fu, G., et al. (2024). Programmable RNA N^6-methyladenosine editing with CRISPR/dCas13a in plants. *Plant Biotechnol. J.* 22, 1867–1880. doi: 10.1111/pbi.14307

Tang, J., Chen, S., and Jia, G. (2023). Detection, regulation, and functions of RNA N^6-methyladenosine modification in plants. *Plant Commun.* 4, 100546. doi: 10.1016/j.xplc.2023.100546

Tang, J., Lei, D., Yang, J., Chen, S., Wang, X., Huang, X., et al. (2024). OsALKBH9-mediated m^{6}A demethylation regulates tapetal PCD and pollen exine accumulation in rice. *Plant Biotechnol. J.* 22, 2410–2423. doi: 10.1111/pbi.14354

Wang, C., Yang, J., Song, P., Zhang, W., Lu, Q., Yu, Q., et al. (2022). FIONA1 is an RNA N^6-methyladenosine methyltransferase affecting Arabidopsis photomorphogenesis and flowering. *Genome Biol.* 23, 40. doi: 10.1186/s13059-022-02612-2

Wang, G., Li, H., Ye, C., He, K., Liu, S., Jiang, B., et al. (2024). Quantitative profiling of m^6A at single base resolution across the life cycle of rice and Arabidopsis. *Nat. Commun.* 15, 4881. doi: 10.1038/s41467-024-48941-7

Wang, X., Jiang, B., Gu, L., Chen, Y., Mora, M., Zhu, M., et al. (2021). A photoregulatory mechanism of the circadian clock in Arabidopsis. *Nat. Plants* 7, 1397–1408. doi: 10.1038/s41477-021-01002-z

Xu, T., Wu, X., Wong, C. E., Fan, S., Zhang, Y., Zhang, S., et al. (2022). FIONA1-mediated m^6A modification regulates the floral transition in arabidopsis. *Adv. Sci. (Weinh)*, e2103628. doi: 10.1002/advs.202103628

Yu, L., Alariqi, M., Li, B., Hussain, A., Zhou, H., Wang, Q., et al. (2024). CRISPR/dCas13(Rx) derived RNA N^6-methyladenosine (m^6A) dynamic modification in plant. *Advanced Sci.* e2401118. doi: 10.1002/advs.202401118

Yu, Q., Liu, S., Yu, L., Xiao, Y., Zhang, S., Wang, X., et al. (2021). RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* 39, 1581–1588. doi: 10.1038/s41587-021-00982-9

Zhang, M., Bodi, Z., Mackinnon, K., Zhong, S., Archer, N., Mongan, N. P., et al. (2022). Two zinc finger proteins with functions in m^6A writing interact with HAKAI. *Nat. Commun.* 13, 1127. doi: 10.1038/s41467-022-28753-3

Zheng, H. X., Zhang, X. S., and Sui, N. (2020). Advances in the profiling of N^6-methyladenosine (m^6A) modifications. *Biotechnol. Adv.* 45, 107656. doi: 10.1016/j.biotechadv.2020.107656

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., et al. (2008). MTA is an arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20, 1278–1288. doi: 10.1105/tpc.108.058883

Zhou, L., Tang, R., Li, X., Tian, S., Li, B., and Qin, G. (2021). N^6-methyladenosine RNA modification regulates strawberry fruit ripening in an ABA-dependent manner. *Genome Biol.* 22, 168. doi: 10.1186/s13059-021-02385-0

Zhou, L., Tian, S., and Qin, G. (2019). RNA methylomes reveal the m^6A-mediated regulation of DNA demethylase gene SlDML2 in tomato fruit ripening. *Genome Biol.* 20, 156. doi: 10.1186/s13059-019-1771-7

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 13 November 2024
DOI 10.3389/fpls.2024.1480588

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Xiaolu Liu,
University of Science and Technology Beijing, China
Zhongbao Zhang,
Beijing Academy of Agricultural and Forestry Sciences, China
Chongrong Wang,
Guangdong Academy of Agricultural Sciences (GDAAS), China

*CORRESPONDENCE
Jiaming Zhang
zhangjiaming@itbb.org.cn

RECEIVED 14 August 2024
ACCEPTED 24 October 2024
PUBLISHED 13 November 2024

CITATION
Tan D, Fu L, Yu Y, Sun X and Zhang J (2024) Screening method and metabolic analysis of plant anti-aging microorganisms via ammonia-induced senescence in the duckweed *Wolffia microscopica*. *Front. Plant Sci.* 15:1480588. doi: 10.3389/fpls.2024.1480588

Screening method and metabolic analysis of plant anti-aging microorganisms via ammonia-induced senescence in the duckweed *Wolffia microscopica*

Deguan Tan[1,2], Lili Fu[1,2], Ying Yu[1,2], Xuepiao Sun[1,2] and Jiaming Zhang[1,2]*

[1]National Key Laboratory for Tropical Crop Breeding, Institute of Tropical Bioscience and Biotechnology, Sanya Research Institute of Chinese Academy of Tropical Agricultural Sciences, Haikou, China, [2]Hainan Key Laboratory of Microbiological Resources, Hainan Institute for Tropical Agricultural Resources, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

Ammonium is the preferred N nutrition over nitrate for some plant species, but it is toxic to many other plant species and induces senescence at high concentrations. The duckweed *Wolffia microscopica* (Griff.) Kurz is the smallest and fast-growing angiosperm. It is highly sensitive to ammonium and has a short lifespan on media containing 0.5 mM or higher ammonia. This feature makes it a potential model plant to screen for anti-aging microorganisms. By co-culturing *W. microscopica* with endophytic microorgainisms isolated from rubber tree, we screened out an *Aspergillus sclerotiorum* strain ITBB2-31 that significantly increased the lifespan and the biomass of *W. microscopica*. Interestingly, both filter-sterilized and autoclaved exudates of ITBB2-31 increased the lifespan of *W. microscopica* cultures from 1 month to at least 7 months. Meanwhile, the exudates also showed strong anti-aging effects on cassava and the rubber tree leaves and increased chlorophyll contents by 50% - 350%. However, high contents of filter-sterilized exudates inhibited the growth of *W. microscopica* while extending its lifespan, indicating that there were heat-sensitive growth-inhibiting agents in the exudates as well. Comparative metabolome analysis of the filter-sterilized and autoclaved exudates revealed multiple heat-stable anti-aging and heat-sensitive growth-inhibiting compounds. Our results suggest that *W. microscopica* can be served as a rapid and efficient model plant to screen for plant anti-aging microorganisms.

KEYWORDS

duckweeds, anti-aging microorganism, stay-green, screening model, endophytes, comparative metabolome

1 Introduction

Leaf aging is one of the important factors that influence crop yields. The stay-green characters of leaves extend the duration of active photosynthesis and are positively correlated with yields and stress tolerance in many crops, including maize, wheat, rice, soybean, sorghum, oats, and cassava (Rawson et al., 1983; Sasaki and Ishii, 1992; Bekavac et al., 2002; Joshi et al., 2007; Kusaba et al., 2013; Thomas and Ougham, 2014). The stay-green genotypes of maize, for example, do not lose the green color of their leaves until physiological maturity, while the non-stay-green genotypes start losing the green color of their leaves approximately 30 days after flowering (Bekavac et al., 2002). Stay-green alleles enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns (Borrell et al., 2014). Exogenous application of cytokinins on leaves of winter wheat improved stay-green characteristics and thus increased grain yield in heat stress (Yang et al., 2016). Therefore, stay-green traits have been used in breeding programs of many grain crops. Through intensive selection, varieties with both higher yields and longer duration of greenness in a range of grain crops have been commercialized (Thomas and Ougham, 2014). Stay-green is also important to root crops. For example, cassava plant ideally has a leaf life of 15–20 weeks and an optimal leaf area index (LAI) of 4 to 6 according to a simulation model for substantially increased cassava yields with longer leaf life (Cock et al., 1979). In most cultivars, however, leaf life is much shorter, and replacement of leaves requires energy input and creates a competitive sink for photosynthetic assimilates, thus reducing root yields. Introduction of a senescence-inducible isopentyl transferase gene into cassava extended leaf greenness, and thus increased the yield (Zhang et al., 2010). The importance of stay-green in agriculture and the relevant molecular mechanisms that regulate stay-green traits have been well reviewed (Kusaba et al., 2013; Thomas and Ougham, 2014; Abdelrahman et al., 2017).

The stay-green characters are usually obtained by natural mutation and/or artificial engineering. An alternative way may be the application of beneficial microorganisms to plants. Beneficial microorganisms have been isolated from many plants as endophytes. The interactions between hosts and endophytes involve flows of various compounds, of which some are highly toxic to the hosts (Powell and Petroski, 1992), however, in many cases, the endophytes are beneficial to hosts by increasing abiotic and biotic stress resistances, and thus improve the adaptation of the plants (Schardl et al., 2004). Some grass species from coastal and geothermal habitats require symbiotic fungal endophytes for salt and heat tolerance (Rodriguez and Redman, 2008). The ability of symbiotic microorganisms to confer stress tolerance to plants may provide a novel strategy to cope with the problems in modern agriculture. In this paper, we report a rapid and efficient method for screening of anti-aging microorganisms, based on an early aging duckweed species *Wolffia microscopica* using *in vitro* culture system.

Abbreviations: MH, Modified Hoagland medium; FM, fermented MH broth; MS, Mass spectrometry; LC-MS/MS, Liquid chromatography tandem mass spectrometry; Rf, Relative abundance in the filter-sterilized medium; Ra, Relative abundance in the autoclaved medium.

Duckweeds belong to a globally distributed family Lemnaceae. They are free floating monocots and are currently classified into 5 genera and 37 species (Landolt, 1986; Appenroth et al., 2013). Duckweed clones are collected from all over the world and are maintained at Duckweed Stock Collection Centers situated in Zurich, Switzerland; Jena, Germany; Rutgers, New Brunswick, USA; and Chengdu and Haikou, China. The most accepted maintenance media is the modified Hoagland medium (Hoagland and Arnon, 1950), N, E, and SH media supplemented with 1% sucrose and solidified with 0.6% agar (Appenroth, 2015; Sree and Appenroth, 2016). While most clones can be stored for three months before subculture, some clones should be subcultured more often, such as *W. microscopica*.

W. microscopica is one of the fast growing duckweed species with a biomass doubling time of only one day (Ziegler et al., 2015; Appenroth et al., 2017). Its plant body is simplified into a tiny frond with a size of only around 1 mm (Figure 1). However, its axenic stock cultures senesce more rapidly, resulting in a shorter maintenance cycle compared to the other species in the duckweed family, thus requiring to be subcultured more often. This species was once lost completely from all duckweed collections in the year 2009; however, it was rediscovered in 2013 (Sree and Appenroth, 2014).

The rubber tree (*Hevea brasiliensis*) is widely cultivated in tropical areas (Heng and Joo, 2017) where there exist rich environmental microorganisms, and its trunk bark is regularly cut for harvesting latex, thus making it more susceptible to external microbial invasion. In addition, the rubber tree mainly reproduces through bud grafting (Heng and Joo, 2017), which keeps the spread of endophytes among its offsprings. These make the rubber tree abundant in endophytic microorganisms. Many endophytic fungal strains from the rubber tree have been discovered, among which *Ascomycota* species accounts for 97%, and the sapwood has a greater diversity of endophytes than the leaves (Gazis and Chaverri, 2010). Our research team has also isolated many endophytic fungi from the rubber tree, and found that some strains can remarkably inhibit the growth of the pathogenic fungus *Colletotrichum gloeosporioides* Penz. Sace and *Fusarium oxysporum* Cubense (Zheng et al., 2009a), and another strain called ITBB2-1 exhibited high salt tolerance (Zheng et al., 2009c). These findings indicate that endophytic fungi of the rubber tree have some unique biological functions.

In this study, we found that *W. microscopica* is sensitive to ammonia-induced senescence. The early-aging phenomena can be served as a reporter to screen for anti-aging microorganisms by co-culturing it with candidate microorganisms or by supplementation of the fermented medium of microorganisms into the culture medium of the plant.

2 Materials and methods

2.1 Plant materials and culture media

The axenic *Wolffia microscopica* (Griff.) Kurz clones 2001 and 2008 were provided by Prof. Klaus-J. Appenroth at University of Jena, Germany and Dr. K. Sowjanya Sree at Central University of Kerala,

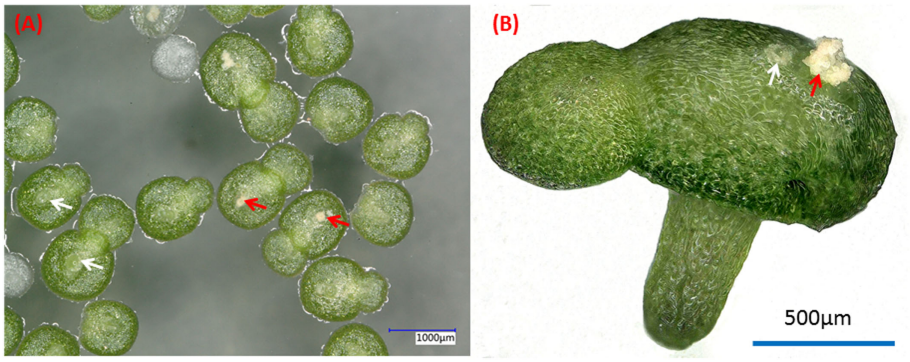


FIGURE 1
Dorsal **(A)** and side **(B)** views of typical two-fronds plant of *W. microscopica* 2001. White arrows indicate the pistils and red arrows indicate the stamens.

India. The clones are maintained at the Institute of Tropical Bioscience and Biotechnology (ITBB), Chinese Academy of Tropical Agricultural Sciences (CATAS). The basic subculture medium was a modified Hoagland medium (MH) originally designated by Hoagland and Arnon (1950) supplemented with 29 mM sucrose and 6 g L^{-1} agar (Biotechnology grade, Beijing Solarbiao Science & Technology co., Ltd, Beijing, China) for solidification. The components of the MH medium was previously described (Fu et al., 2017). Other media that were tested in this research include N (Appenroth et al., 1996), E (Cleland and Briggs, 1967), and SH (Schenk and Hildebrandt, 1972) media (Supplementary Table S1). The media were autoclaved at 121°C for 20 min.

2.2 Stay-green verification of the stock cultures of *W. microscopia*

Ten plants of *W. microscopia* clones 2001 and 2008, respectively, were inoculated in the centre of each 6-cm plate containing MH medium solidified with 6 g L^{-1} agar at 5 d intervals, 6 plates were inoculated each time for each clone. The plates were incubated in a growth chamber (BIC-250, Boxun Instrument, Shanghai, China) with 16 h photoperiod at 100 μmol·m^{-2}·s^{-1} photosynthetic active radiation (PAR) and 25 ± 1°C. The status of the plants was examined every other day until two months, and then fronds were harvested and weighed. Three parallel experiments were performed.

2.3 Screening of anti-aging fungal strains by coculturing *W. microscopica* with candidate fungi

The endophytic fungal strains were previously isolated from the rubber tree (Zheng et al., 2009a). The fungi were maintained on Potato Dextrose and Agar (PDA) medium at room temperature. Ten fronds of *W. microscopia* clone 2001 were inoculated in the centre of each 6-cm plate containing MH medium. Fungal spores and/or mycelia were co-cultured on the edge of the same plate. The plates that were only inoculated with *W. microscopia* were used as control. All plates were incubated in the growth chamber (BIC-250, Boxun, Shanghai) with 16 h photoperiod at 100 μmol·m^{-2}·s^{-1} photosynthetic active radiation (PAR) and 25 ± 1°C for 30 d. The status of the plants was examined every other day.

2.4 Temperature-dependence of growth-promoting effect of fungal strain ITBB2-31

The growth-promoting experiments of *Aspergillus sclerotiorum* strain ITBB2-31 was carried out as described above. The plates were incubated in growth chambers (BIC-250, Boxun, Shanghai) with 16 h photoperiod at 100 μmol·m^{-2}·s^{-1} photosynthetic active radiation (PAR) for 30 d at 20°C, 25°C, and 30°C, respectively. The fronds in each plate was wrapped in tinfoil and dried at 105°C overnight and weighed. The significance of weight differences was analyzed using IBM SPSS Statistics Version 24 (IBM Corporation, New York, USA).

2.5 Anti-aging effects of ITBB2-31 fermented broth on *W. microscopica*

The fungal strain ITBB2-31 was inoculated in MH liquid medium supplemented with 29 mM sucrose and incubated at 28°C in the dark for 10 d. The fermented broth was filtered with 50 μm nylon membrane to remove the mycelium. The filtrate was then passed through filter paper, and filter-sterilized with 0.22 μm filter units (Millipore, Bedford, USA), and then added into the sterilized MH medium with concentrations of 10%, 20%, 30%, and 50% (v/v), respectively. To test whether the functional element in the exudate was temperature resistant, the filtrate was added into MH medium with above concentrations and autoclaved at 121°C for 20 min. The autoclaved media were poured into 6-cm plates. *W. microscopia* clones 2001 and 2008 were inoculated in the centre of the plates and incubated in the growth-chamber with 16 h photoperiod at 100 μmol·m^{-2}·s^{-1} photosynthetic active radiation (PAR) and 25°C for one month. The plants in each plate were

separately wrapped in tinfoil and dried at 105°C overnight and weighed. The significance of weight differences between treatments was analyzed using IBM SPSS Statistics Version 24.

2.6 Anti-aging effect of ITBB2-31 on the rubber tree and cassava leaves

The anti-aging effect of ITBB2-31 was tested using the leaves of test-tube plants of rubber tree and cassava following the method described by Zhang et al (Zhang et al., 2010). Fully expanded, green, and healthy leaves collected from rubber tree plants grown in the greenhouse were also used. The leaves were put in 15-cm petri-dishes on filter paper wetted with MH medium supplemented by 0%, 10%, 20%, and 30% fermented MH broth of ITBB2-31. Dark-induced senescence was performed by incubating the leaves in the dark at 25°C for two weeks. The leaf aging status was recorded by photography every day. To quantify the anti-senescence effect, leaf samples were homogenized in liquid nitrogen and dissolved in 100% (v/v) acetone as described in a method (Lichtenthaler, 2010). The samples were incubated at room temperature for 30 min and centrifuged to remove cell debris. Chlorophyll *a* and *b* content of the supernatants was determined by spectrophotometer and calculated according to the equation as previously described (Lichtenthaler, 1987). The significance of the differences between treatments was tested by one-way ANOVA and LSD test using IBM SPSS Statistics Version 24.0 (IBM Corporation, New York, USA).

2.7 DNA extraction and polymerase chain reaction and phylogenetic analysis

Genomic DNA of fungal strain ITBB2-31 was isolated using a fungal DNA isolation kit (Tiangen Biotech, Beijing, China). The internal transcribed spacer (ITS) sequence was amplified as previously described (Zheng et al., 2009c) with two primers HNP76 (5′-TCCGTAGGTGAACCTGCGG-3′) and HNP77 (5′-TCCTCCGCTTATTGATATGC-3′). The amplified fragment was sequenced on both strands at BGI (Beijing Genomic Institute), Shenzhen, China. For phylogenetic analysis, reference sequences were downloaded from GenBank. The sequences were aligned with MacVector 15.0.2, the unaligned 5′ and 3′ sequences were removed. Phylogenetic tree was generated using Mega7 (Kumar et al., 2016). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The tree with the highest log likelihood is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated, and there was a total of 453 positions in the final dataset.

2.8 LC-ESI-MS/MS analysis of the fermented media

Both filter-sterilized and autoclaved fermented media of ITBB2-31 were respectively analyzed using liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS). For liquid chromatograph (LC), an Agilent 1290 (Agilent Technologies, CA) with a Waters column ACQUITY UPLCBEH Amide (1.8 μm, 2.1 mm × 100 mm) was used. The column temperature was set to 35°C, and 2 μL of sample was injected. The elution buffers for positive ion model were 0.1% formic acid in water (buffer A) and 0.1% formic acid in acetonitrile (buffer B); 95% of buffer A was supplied for the first minute of run, then the profile of buffer B was increased gradually until reaching 95% of the elution buffer at 16 minutes, and reduced to 5% at 18 minutes. For negative ion model, 2 mM ammonia acetate (buffer A) and acetonitrile (buffer B) were used, following the same elution schedule as above. The flow rate was 0.4 mL/min. For mass spectrometry, an Agilent 6545QTOF with control software LC/MS Data Acquisition, Version B.08.00 was used. The ion source temperature was 320°C. The nitrogen gas flow and sheath gas flow were 8 L/min and 12 L/min, respectively. Sheath gas temperature was 350°C. The auxiliary pressure was set to 3500 V and 4000 V for negative and positive models, respectively. Mass spectra of the parent ions and the subsequent fragmented ions were scanned over the range of 50–1100 m/z. The data was analyzed using software MSDIAL (version 2.54) and the identities of the metabolites were determined by searching against the in-built MS/MS reference libraries including metlin, MassBank, MoNA, and HMDB. Relative abundance of a compound was calculated using the equation: the area of the compound/the total area of all compounds × the number of compounds.

3 Results

3.1 Ammonium-induced senescence observed by *in vitro* culture of *W. microscopica*

The MH medium is one of the commonly used media to retain the axenic stock cultures of duckweed species. However, one species, *W. microscopica* did not perform well on MH medium. This species grew at the beginning, and reached a colony size of approximately 0.8-1.3 cm in diameter in two weeks, and then approached senescence (Figure 2A). Most colonies died within one month with a few exceptions in some colonies, and all plants died without exceptions in two months (Figure 2A). Therefore, the lifespan of *in vitro* cultures of *W. microscopica* on MH medium was only one month. In contrast, the other species in the duckweed family can be subcultured at three-month intervals on MH medium. This is probably the reason why *W. microscopica* was easy lost from the duckweed stock collections (Sree and Appenroth, 2014).

To investigate the reason of early senescence in *W. microscopica* stock cultures, four commonly used media including MH, SH, E, and N media were used to culture *W. microscopica* clone 2001. The plants grew the fastest on E medium, followed by N medium, and slowest on SH and MH media. In the meantime, the plants on MH and SH media began senescence rapidly and mostly turned pale white within three weeks and could not resume growth on fresh medium, while the plants on E and N media were still prosperous (Figure 2B). By comparing the components in the four media, we

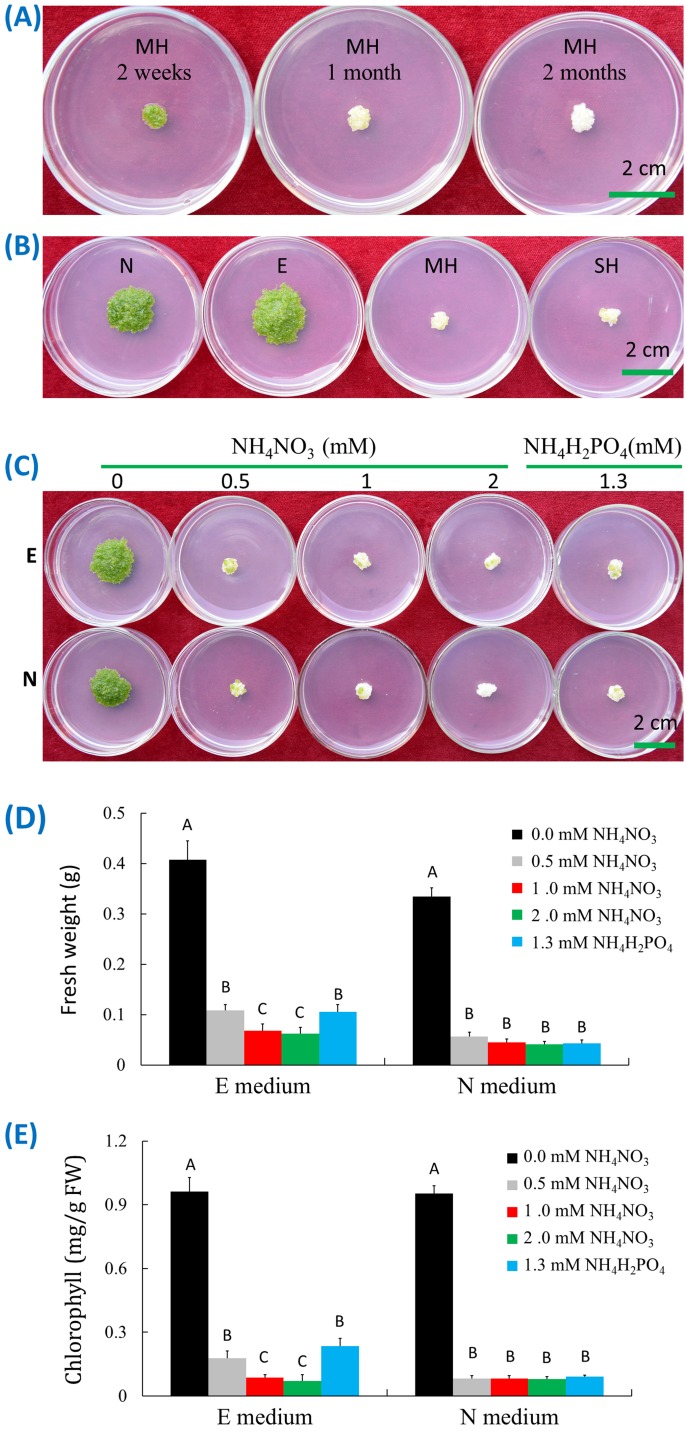


FIGURE 2
Ammonia-induced senescence of *W. microscopica* 2001. **(A)** Growth status on MH medium at different time points as indicated on the plates; **(B)** Growth status on different culture media as indicated on the plates; **(C)** Growth status on E and N media supplemented with different types and concentrations of ammonia salts; **(D)** Fresh weight analysis of plant colonies in **(C)**; **(E)** Chlorophyll content analysis of plant colonies in **(C)**. The culture temperature was 25°C and the culture time was four weeks if not indicated. The significance of data was analyzed with one-way ANOVA and LSD test. Different letters on columns indicate significance at 1% level.

found that both MH and SH media contained ammonia nitrogen (Supplementary Table S1), which were 1 mM NH_4NO_3 and 1.3 mM $(NH_4)H_2PO_4$ in MH and SH media, respectively, while E and N media contained only nitrate.

Ammonia-induced senescence has been reported in both mammals and plants (Lin et al., 2002; Gorg et al., 2015). To investigate whether the early senescence of *W. microscopica* on MH and SH media was caused by ammonia-induced senescence,

W. microscopica clones 2001 and 2008 were cultured on E and N media supplemented with different concentrations of ammonium ion. Results indicated that supplementation of 0.5 mM NH_4^+ in both E and N media was enough to drive *W. microscopica* to die within four weeks (Figure 2C), the higher the NH_4^+ concentration, the quicker for *W. microscopica* to die. The fresh weights were reduced by 73% - 85% in E medium and 83% - 88% in N medium depending on the concentration of ammonium salts (Figure 2D). The chlorophyll contents were reduced by 81% - 92% in the E and N media after addition of ammonium salts (Figure 2E). These results indicated that the early senescence of *W. microscopica* was induced by ammonia ion.

3.2 Screening of anti-aging microorganisms

Twenty-five endophytic microbial strains isolated from rubber tree (Zheng et al., 2009a) were tested for their anti-aging effect on *W. microscopica*. One strain designated as ITBB2-31 showed strong anti-aging activity. When *W. microscopia* clones was co-cultured with ITBB2-31, its growth was significantly promoted, and the co-cultured plant colonies grew to a larger diameter and kept growing after one month of inoculation at all tested temperatures (Figure 3A). In contrast, the control colonies that were not co-cultured had stopped growth in approximately two weeks and died in one month (Figure 3A). The dry weight of co-cultured colonies was 3.5-4.5 times the weights of the controls depending on the culture temperatures, and the highest dry weight (45.6 mg) was obtained at 25°C (Figure 3B).

3.3 Molecular identification of ITBB2-31

The internal transcribed spacer sequence (ITS) of strain ITBB2-31 was amplified by PCR method. The amplified product was 609 bp. BLAST searches against the NCBI database resulted in best hit to *Aspergillus sclerotiorum* Huber strain ATCC16892, which is the type strain of *A. sclerotiorum* and was isolated from fruit of apple (*Malus sylvestris*) by G.A. Huber in Oregon, USA (Huber, 1933). Phylogenetic analysis indicated that ITBB2-31 belonged to the yellow aspergilli, the *Aspergillus* section *Circumdati* (Visagie et al., 2014), and formed a clade with *A. sclerotiorum* strains and *A. subramanianii*, *A. bridgeri*, *A. persii*, *A. salwaensis* and *A. roseoglobulosus* strains with 98% bootstrap support (Supplementary Figure S1). Pairwise alignments indicated that the ITS sequence of strain ITBB2-31 had 100% identities with strains ATCC16892, NW3, NRRL35202, NRRL415, NRRL35024, DTO129-F5, and CCF3434; and 99.6% - 99.8% identities compared to *A. scelrotiorum* strain ANDEF08, *A. subramanianii* strains DTO129G4, NRRL6161 and DTO245E4, *A. bridgeri* strain NRRL1300, and *A. persii* strain CBS112795. The next closely related strains in the same clade were *A. salwaensis* DTO297B3T with 99.2% identities and *A. roseoglobulosus* strain CBS112800 with 97.1% identities. The sequence identities compared to the strains in other clades were less than 92.3%. Taken together, ITBB2-31 is an *A. scelrotiorum* strain.

3.4 The fermented broth of ITBB2-31 presents anti-aging activity for *W. microscopia*

It is obvious that the anti-aging and growth-promoting activity of ITBB2-31 did not require physical contact between the plants and fungal colonies (Figure 3A), because the fungus and plants were inoculated far away, and in most cases the fungus and plants remained separated in one month. Therefore, there must be some components secreted into the medium or some volatiles released to the air by the fungus. We tested the medium in this research, either autoclaved or filter-sterilized fermented MH medium (FM hereafter) was added to the fresh MH medium in different proportions before inoculation of the plants. Results indicated that both autoclaved and filter-sterilized FM prevented senescence and promoted the growth of *W. microscopia*. The plants of both clones 2001 and 2008 kept alive and continued to grow at all tested FM concentrations (Figure 4A). In contrast, the plants of both clones that grew on the control medium died in one month (Figure 4A). Therefore, the anti-aging factor was thermo-stable and was secreted into the culture medium.

Besides the anti-aging and growth-promoting effect, supplementation with 30% or higher concentrations of the filter-sterilized FM resulted in lower growth-rate of *W. microscopia* as compared to 10% filter-sterilized FM, although the plants stayed green (Figure 4A). This inhibition effect was significant when the proportion of the filter-sterilized FM was increased to 50% (Figures 4B, C). However, the autoclaved FM present very weak inhibition effect and the decrease of dry weight at high concentrations was not statistically significant (Figures 4B, C), and the dry weights of plant colonies grown on 50% of autoclaved FM were significantly higher than those grown on 50% filter-sterilized FM for both clones 2001 and 2008 (Figures 4B, C). These results indicated that the components that inhibited plant growth were sensitive to high temperatures (e.g. 121°C).

Moreover, the exudates of ITBB2-31 presented anti-aging activity to *W. microscopica* for many months. *W. microscopica* clones 2001 and 2008 were successfully retained on MH medium supplemented with 10% autoclaved FM for more than 7 months at 25°C without subculture (Figure 5).

3.5 Anti-aging activity of ITBB2-31 against dark-induced senescence of the rubber tree and cassava leaves

The FM of fungal strain ITBB2-31 presented strong anti-aging effects to detached leaves of *in vitro* plantlets of the rubber tree (Figure 6A) and cassava (Figure 6B), and the rubber tree leaves collected from the field (Figure 6C). Obvious senescence was shown on the control leaves of *in vitro* rubber tree and cassava after 15 days of dark treatment, and the chlorophyll contents of the leaves were only 0.79 mg/g and 0.70 mg/g FW for the rubber tree and cassava, respectively. Adding 10% FM significantly reduced senescence and increased chlorophyll contents by 53.6% and 106% for *in vitro* rubber tree and cassava leaves, respectively. The higher the concentration of FM, the higher the anti-aging effect (Figure 6D).

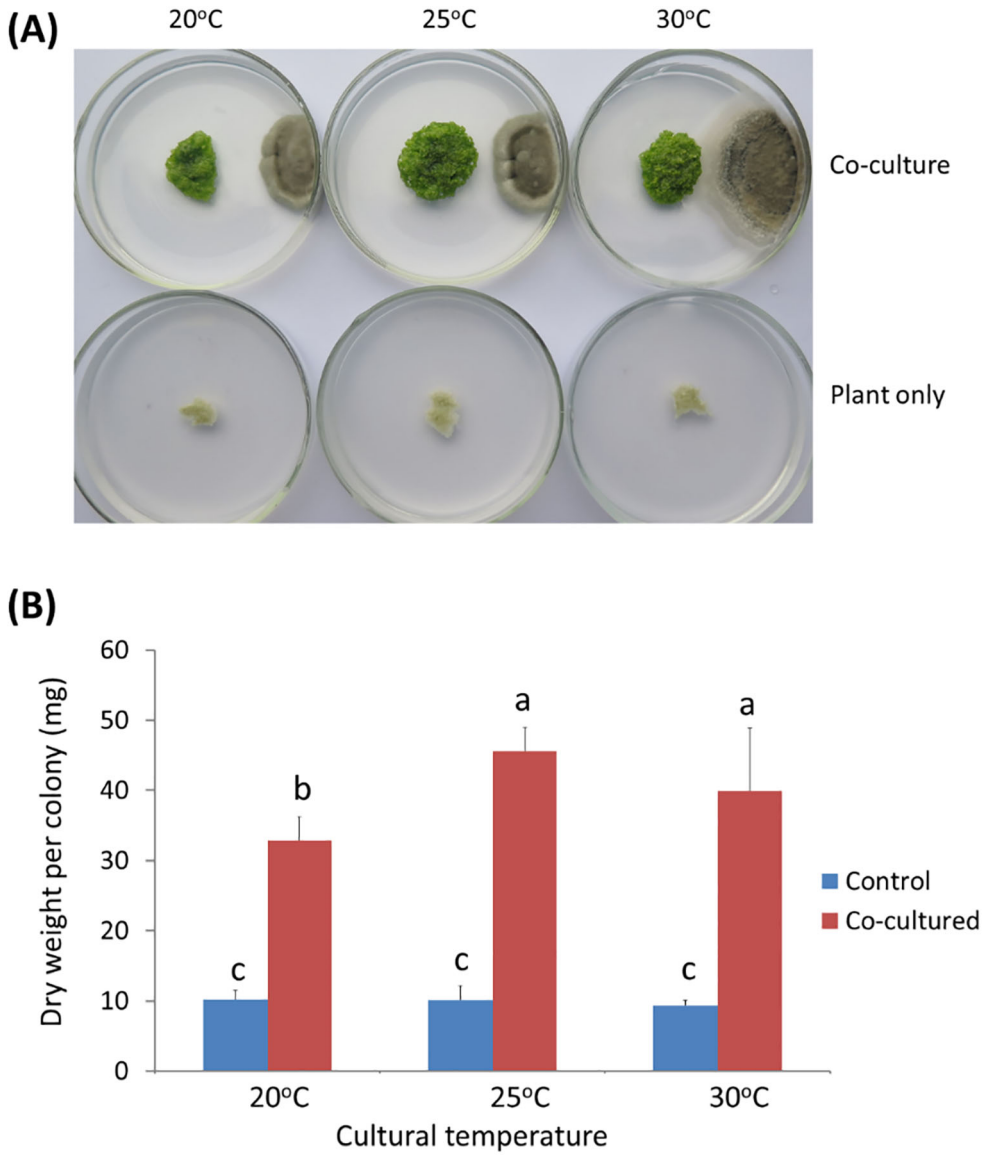


FIGURE 3
Anti-aging effects of endophytic fungal strain ITBB2-31 by co-culture with *W. micoscopica* 2001. **(A)** Representative plates of co-culture and control after incubation for one month under different temperatures; **(B)** Statistic analysis of the dry weights of *W. micorscopica* colonies. The significance of differences was tested by one-way ANOVA and LSD test. Different letters above columns indicate significant differences at 5% level.

The anti-aging effect of ITBB2-31 on the rubber tree leaves collected from the field was also significant. The control leaves turned brown with chlorophyll content of 1.40 mg/g FW after 15 days of dark treatment. The leaves treated with 10% FM of ITBB2-31 had 50% green area, and the total chlorophyll content was 1.94 mg/g FW, significantly higher than that of control leaves (Figure 6D). When the FM concentration was increased to 20%, the anti-aging effect increased significantly, and the chlorophyll content increased to 2.51 mg/g FW. Further increasing of FM concentration did not significantly increase the chlorophyll content further. Taken together, the anti-aging strain ITBB2-31 screened by *W. microscopica*-ammonia system is effective to other plants.

3.6 Comparative metabolome analysis reveals both anti-aging and growth-inhibiting agents in the fermented medium of ITBB2-31

LC-ESI-MS/MS identified 470 compounds in the filter-sterilized FM (Supplementary Table S2), among which 368 compounds disappeared after autoclaving (Figure 7A), while 1141 new compounds that were supposed to be driven from heat-degradation were identified in the autoclaved FM (Figure 7A; Supplementary Table S3). The anti-aging agents are heat-stable (Figure 4) and are supposed to present in both filter-sterilized and autoclaved FMs with similar abundances. In total, 102 compounds are shared by both

10.3389/fpls.2024.1480588

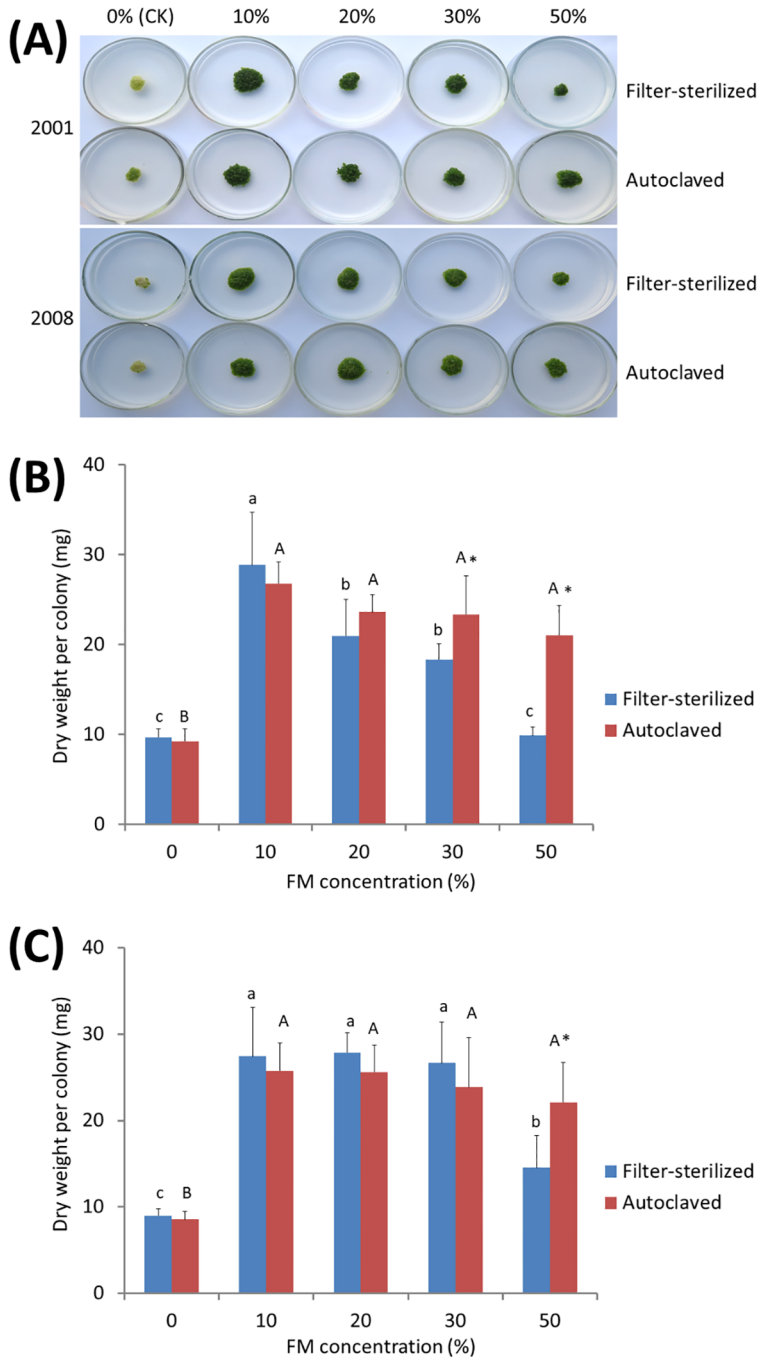


FIGURE 4
Anti-aging effects of the exudates of fungal strain ITBB2-31 on *W. microscopica* clones 2001 and 2008. Both autoclaved and filter-sterilized fermented MH broth (FM) were supplemented to the MH medium with different proportions as indicated in the figure. **(A)** growth status of the two strains after one month of culture at 25°C; **(B, C)** statistical analysis of dry weights of clones 2001 **(B)** and 2008 **(C)**. The significance of differences among the proportions of filter-sterilized (lower case letters) and the autoclaved (Upper case letters) FM was tested by one-way ANOVA and LSD methods. Different letters above columns indicate significant differences at 5% level of significance. The differences between the filter-sterilized and the autoclaved FM at the same proportion was tested with Independent-Samples T-test, and (*) above the columns indicates significance at 5% level.

treatments (Figure 7A; Supplementary Table S4). They are classified into ten subgroups, including lipids and lipid-like molecules (Fu et al., 2017), organic acids and derivatives (Appenroth, 2015), Organoheterocyclic compounds (Sree and Appenroth, 2016), Benzenoids (Zhang et al., 2010), and others (Zheng et al., 2009c) (Figure 7B). Seventeen compounds have similar relative abundances in both treatments (0.75 < Rf/Ra < 1.5, Rf (Relative abundance of Filter-sterilized FM), Ra (Relative abundance of Autoclaved FM), Supplementary Table S4) and are potentially the anti-aging candidates, out of which, some have been reported to have anti-

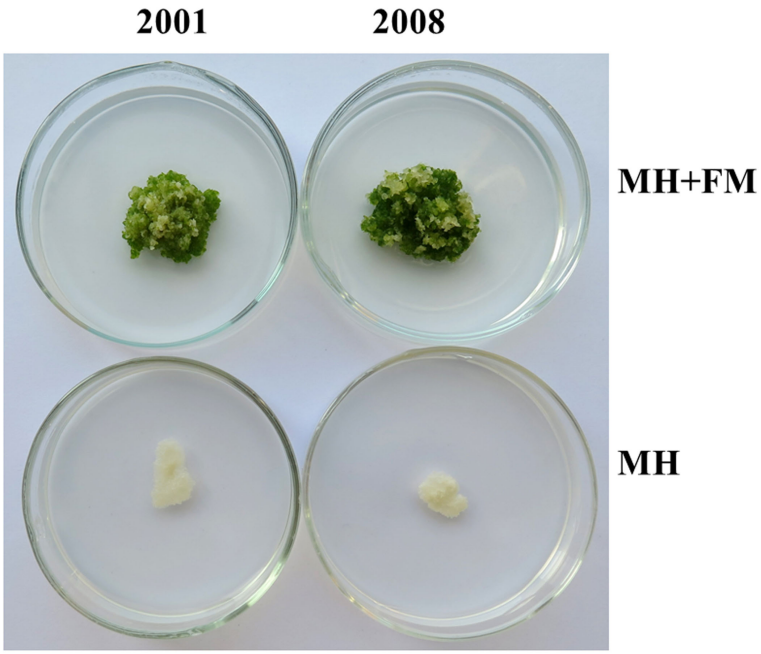


FIGURE 5
Long lasting anti-aging effects of the exudates of fungal strain ITBB2-31 on *W. microscopica* clones 2001 and 2008. Representative plates of *M. microscopica* clones 2001 and 2008 incubated on MH medium supplemented with 10% FM at 25°C for seven months. The medium without addition of FM was used as control.

aging effects in animals and/or plants, including 3-(2,4-dihydroxypentyl)-8-hydroxy-6-methoxyisochromen-1-one (an isocoumarin derivative) (Powers et al., 1989), indole-3-acetaldehyde (McClerklin et al., 2018), leupeptin (Carlin et al., 1994), muramic acid (Lehtonen et al., 1997), alpha-lapachone (Cho et al., 2006), and fatty acyls (Yu et al., 2020). Therefore, the anti-aging effects may come from a combination of compounds.

Compounds that inhibited the growth of *W. microscopica* are sensitive to heat-treatment and presented weak inhibition effect in the autoclaved FM (Figure 4), therefore, the growth-inhibiting agents are supposed to have significantly reduced abundance in the autoclaved FM. We set a cutoff value of 90% reduction of relative abundance in the autoclaved FM, which is equivalent to Rf/Ra > 9. Among the compounds that are shared in both treatments, 13 compounds mainly organic acids, alkaloid derivatives, and nucleoside analogues, have a reduction of > 90% in the autoclaved FM (Supplementary Table S4), some of which have been reported to have toxic and/or growth-inhibiting effects, including S-adenosyl-L-homocysteine (Penyalver et al., 2009), 4-hydroxyquinoline (Chung et al., 1989; Inderjit, Bajpai and Rajeswari, 2010), ginkgolide A (Weakley et al., 2011), phenylalanine (Nguyen et al., 2014; Cao et al., 2019), sparteine-15-carboxylic acid (Ashcroft et al., 1991), 3-hydroxypropanoic acid (Yung et al., 2016), and 2-hydroxyglutaric acid (Bjugstad et al., 2001; Keyser et al., 2008). Therefore, the growth-inhibiting effect of strain ITBB2-31 may come from a combination of compounds, similar to the anti-aging effect.

4 Discussion

4.1 *W. microscopica* is the duckweed species most sensitive to ammonium-induced senescence

W. microscopia is distributed in the Indian subcontinent. It was first named *Grantia microscopica* (Griff. Ex Voigt) (Voigt, 1845; Griffith, 1851), and was later identified as a member of the genus Wolffia Horkel (Kurz, 1866). Many investigations have been carried out on this plant since its identification (Hegelmaier, 1885; Maheshwari SC and Chauhan OS, 1963; Maheshwari SC and Venkataraman, 1966; Roy and Dutt, 1967; Bakshi et al., 1979; Khurana et al., 1986; Landolt, 1986). However, this species could have been causing trouble since its discovery (Hegelmaier, 1885), and could not be found for decades at places where it was discovered (Sree et al., 2015). The axenic clones of *W. microscopica* were once lost completely from all duckweed collections in the year 2009 (Sree and Appenroth, 2014), however, rediscovered from the lakes at Patan, Ambapur, Sughad and Vadasma in Gujarat, India and from that at Jessore, Bangladesh in 2013 (Sree and Appenroth, 2014). We obtained two of the newly collected clones 2001 and 2008 from K-J Appenroth in 2014, and immediately realized that they were more difficult to maintain than the other duckweed species.

In this research, we tested four commonly used duckweed culture media and found that MH and SH were not good media to keep *W. microscopia* alive for as long time as the other duckweed species. *W. microscopica* clones 2001 and 2008 performed similarly on these media

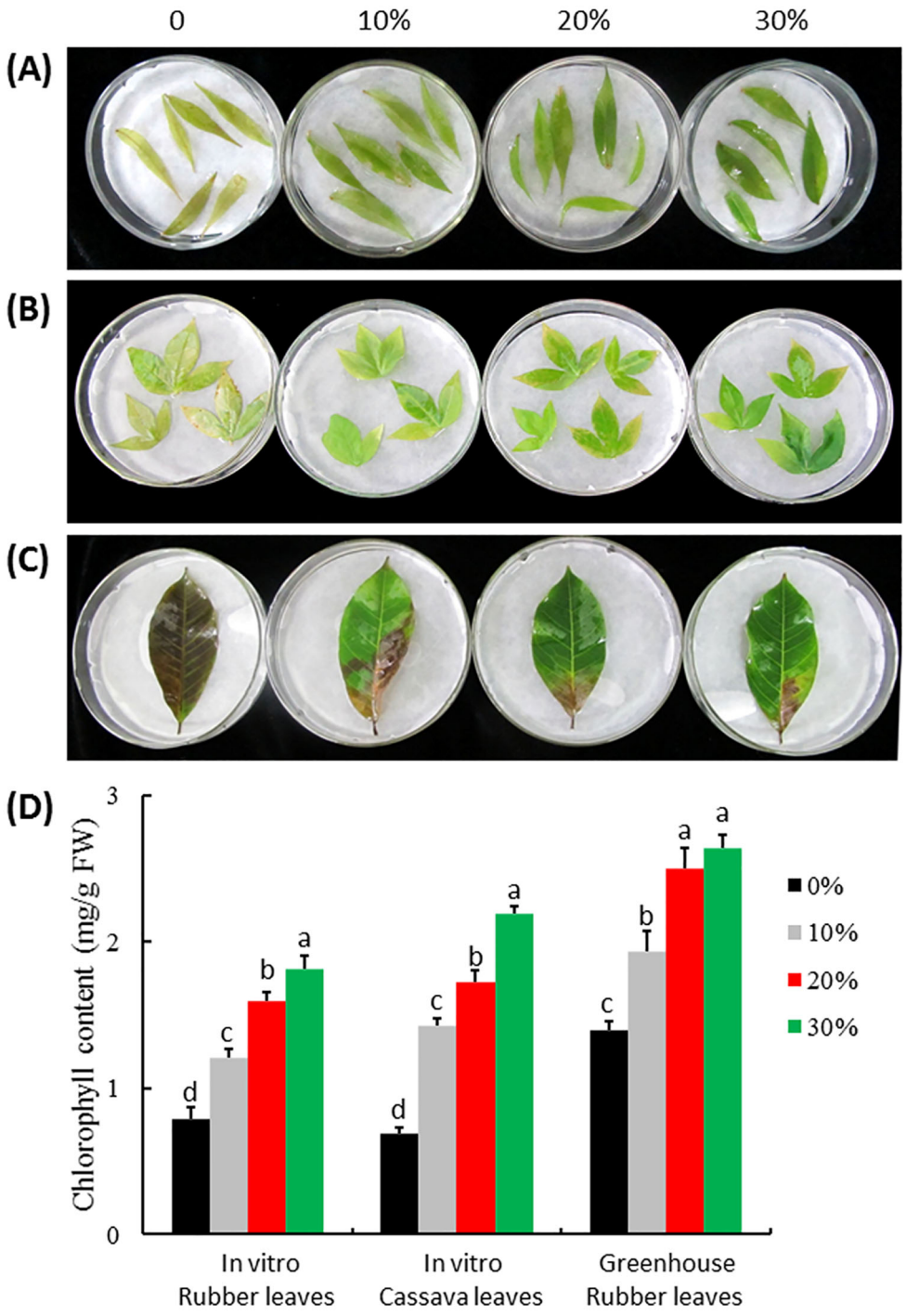


FIGURE 6
Anti-aging effect of ITBB2-31 on detached leaves of rubber tree and cassava. Leaves were detached from *in vitro* plantlets of rubber tree variety Reyan7-33-97 **(A)** and cassava variety SC5 **(B)** and from field plants of rubber tree Reyan7-33-97 **(C)**; The leaves were incubated on different concentrations of autoclaved FM at 25°C for 15 days in the dark; **(D)** Statistical analysis of chlorophyll contents in the leaves. Different letters above columns indicate significant difference at 5% significance level.

and grew rapidly in the beginning, and then stopped growth and died within one month (Figure 2). In contrast, the other species can be subcultured at three-month intervals without problem. By comparing the components in the four media, we found that both MH and SH media contained ammonium ion, while E and N media that did not facilitate early senescence did not contain ammonium ion. Therefore, ammonia was supposed to be the inducer for the early-senescence in *W. microscopica*. This hypothesis was further confirmed to be right by supplementing ammonia in the E and N media that do not contain ammonium ion (Figure 2C). A concentration of 0.5 mM of ammonia was enough to induce early senescence in both strains of *W. microscopica* (Figures 2D, E).

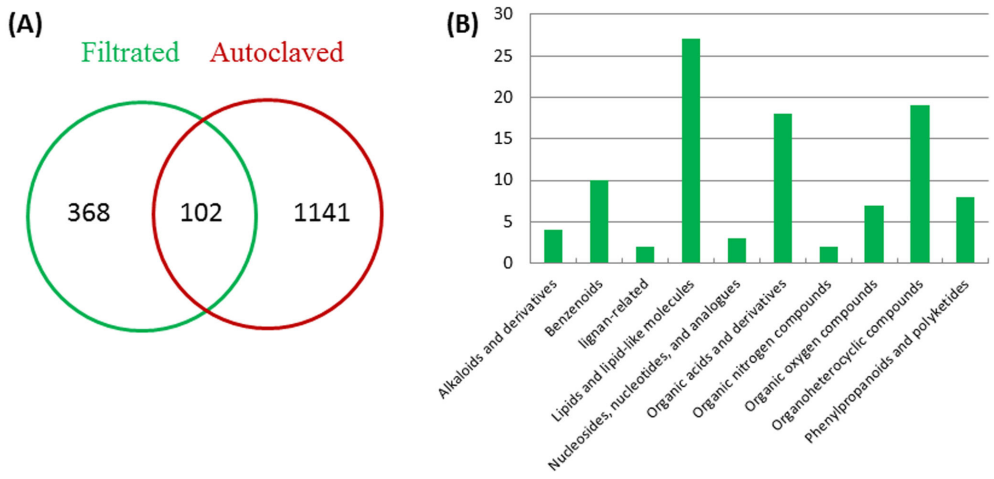


FIGURE 7
Classification of metabolites in the filter-sterilized and autoclaved FM. **(A)** numbers of shared and distinct compounds in the filter-sterilized and autoclaved FM; **(B)** classification of compounds shared by two media.

4.2 The high sensitivity of *W. microscopica* to ammonia-induced senescence makes it a useful tool to screen for plant anti-aging microorganisms

There have been many reports on plant growth-promoting effects of endophytic microorganisms, which has been well reviewed (Aly et al., 2011; Ma et al., 2011; Santoyo et al., 2016). These endophytes are highly diverged and include both fungi (Harzallah et al., 2012; Larriba et al., 2015; Khan et al., 2017) and bacteria (Shi et al., 2010; Paz et al., 2012; Quecine et al., 2012; You et al., 2012). Some endophytic strains previously isolated from the rubber tree have been shown to have antifungal activities (Zheng et al., 2009a), and have a potential in biological control of plant pathogenic diseases (Tan et al., 2015).

In this research, we used *W. microscopica* to screen the endophytic fungal strains isolated from rubber tree (Zheng et al., 2009b; Tan et al., 2015) for plant anti-aging microorganisms, and found one of the strains, ITBB2-31 that strongly postponed aging and promoted the growth of *W. microscopica* (Figures 2-5). The stay-green period of the axenic stock cultures was extended from less than one month to at least seven months (Figure 5). This strain also extended the lifespan of the other duckweed strains in the genus Wolffia. Interestingly, besides the anti-aging and growth promoting effects, the fermented broth of ITBB2-31 inhibited the growth of *W. microscopica* at high concentrations, which was detoxified by autoclaving (Figure 4), suggesting that ITBB2-31 has both promotion and inhibition effects on plant growth.

4.3 Anti-aging activity of microorganisms sreened by the *W. microscopica*-Ammonia system is effective to inhibit senescence of the other plant species

The anti-aging microorganism ITBB2-31 was screened using the *W. microscopica*-ammonia system. To test whether it is effective to inhibit the senescence of other plant species, we used a dark-detached system to induce senescence of the leaves of rubber tree and cassava. Darkness is an extreme light condition and is often used to induce rapid and synchronous senescence in detached leaves (Weaver and Amasino, 2001), and the dark-detached system has been widely used as a senescence model to study the age-triggered senescence (Song et al., 2014; Wei et al., 2017). Our results indicated that the exudates of ITBB2-31 was effective to inhibit dark-induced senescence of detached leaves of rubber tree and cassava (Figure 6), suggesting that ammonia-induced senescence and dark-induced senescence may have shared signal transduction pathway.

4.4 Potential anti-aging and growth-inhibiting agents in ITBB2-31

Multiple compounds with either anti-aging or growth-inhibiting activities were identified through comparative metabolome analysis of filter-sterilized and autoclaved media of strain ITBB2-31. Both activities are suggested to come from a combination of compounds, since multiple compounds in the metabolites have been reported to have anti-aging or growth-inhibiting activities. In this research, a total of 1188 compounds were identified by LC-ESI-MS/MS analysis, and 102 compounds were shared by both filter-sterilized and autoclaved FMs (Figure 7A; Supplementary Table S4). To narrow down our search for the anti-aging compounds, we supposed the anti-aging agents to have similar relative abundance in both treatments, and set a cutoff value of $0.75 < Rf/Ra < 1.5$, since the anti-aging agents are heat-stable (Figure 4). Seventeen compounds lay in the search scope, and biological functions of them were mined in literatures. Results indicated that some of them have been reported to have anti-aging effects in animals and/or plants, including 3-(2,4-dihydroxypentyl)-8-hydroxy-6-methoxyisochromen-1-one (an isocoumarin derivative) (Powers et al., 1989), indole-3-acetaldehyde (McClerklin et al., 2018), leupeptin (Carlin et al., 1994), muramic acid (Lehtonen et al., 1997), alpha-lapachone (Cho et al., 2006), and fatty acyls (Yu

et al., 2020). Therefore, the anti-aging effects may come from a combination of compounds. Isocoumarin and derivatives have been shown to inhibit the activity of a variety of serine proteases (Powers et al., 1989), and potentially inhibit senescence. Indole-3-acetaldehyde is a precursor of indole-3-acetic acid (IAA) in Arabidopsis and is converted to IAA through a novel pathway (McClerklin et al., 2018). Leupeptin specifically inhibit the proteolytic activity in senescent cell lysates and potentially inhibit senescence of plants (Carlin et al., 1994). Muramic acid interacts with lectin in plants (Ayouba et al., 1991), and were detected in more percentage in young people than in the olds (Lehtonen et al., 1997). A plant alpha-lapachone has been proven to have antivascular activity in animal cells (Garkavtsev et al., 2011) and antifungal activity in plants (Cho et al., 2006). Benzoic acid derivatives from *Bjerkandera adusta* has been shown to modulate the proteostasis network, and likely to be anti-aging agents (Georgousaki et al., 2020). Fatty acyls are potential biomarkers related to the anti-aging effect of ginsenoside Rb1 (GRb1), an active ingredient of traditional Chinese medicine *Panax ginseng* C. A. Meyer (Yu et al., 2020).

To narrow down our search for potential growth-inhibiting agents, we set a cutoff value of 90% reduction of relative abundance in the autoclaved FM compared to that in the filter-sterilized FM, which is equivalent to Rf/Ra > 9, since compounds that inhibited the growth of *W. microscopica* are sensitive to heat-treatment and presented only weak inhibition effect in the autoclaved FM (Figure 4). Among the compounds shared in both treatments, 13 compounds lay in our search scope, some of which have been reported to have toxic and/or growth-inhibiting effects, including S-adenosyl-L-homocysteine (Penyalver et al., 2009), 4-Hydroxyquinoline (Chung et al., 1989; Inderjit, Bajpai and Rajeswari, 2010), Ginkgolide A (Weakley et al., 2011), Phenylalanine (Nguyen et al., 2014; Cao et al., 2019), sparteine-15-carboxylic acid (Ashcroft et al., 1991), 3-hydroxypropanoic acid (Yung et al., 2016), and 2-hydroxyglutaric acid (Bjugstad et al., 2001; Keyser et al., 2008). S-adenosyl-L-homocysteine is a nucleotide analogue, its hydrolase gene ahcY is required for optimal growth of *Agrobacterium radiobacter* K84 (Penyalver et al., 2009), suggesting that S-adenosyl-L-homocysteine is a growth inhibitor. In our research, S-Adenosyl-L-homocysteine has a high content before autoclaving, but only 4% left after autoclaving, suggesting that it may take part in the growth inhibition of *W. microscopica*. 4-Hydroxyquinoline is an inhibitor of NADH-ubiquinone reductase in the respiratory chain of mitochondria (Chung et al., 1989); a similar compound, 8-hydroxyquinoline (HQ) is exuded from the roots of *Centaurea diffusa*, and reduce growth of the other plants (Inderjit, Bajpai and Rajeswari, 2010). Ginkgolide A inhibits vascular smooth muscle proliferation and reduces neointimal hyperplasia in a mouse model (Weakley et al., 2011). Phenylalanine inactivates kinases (Nguyen et al., 2014) and intestinal digestive enzymes (Cao et al., 2019) activities, and thus inhibits development. Sparteine derivatives are well known toxic alkaloids (Pothier et al., 1998), and have been shown to inhibit growth of monocots (Parmaki et al., 2018). 3-hydroxypropanoic acid has been shown to inhibit the growth of *Escherichia coli* (Yung et al., 2016).

Besides the compounds mentioned above, some other metabolites in the fermented media may have also taken part in either the anti-aging or the growth-inhibiting activities, since the cutoff values we used in the comparative analysis were relatively arbitrary to narrow down the search, however, they contribute only to a larger combination of the compounds.

4.5 Possible mechanism of ammonia-induced senescence in plants

Ammonium is often the preferred form of N nutrition for some higher plants such as eucalyptus (Pfautsch et al., 2009) and rice (Song et al., 2013), and to some extent delays leaf aging (Diaz et al., 2008; Aguera et al., 2010). However, ammonium inhibits the growth of many other plant species, including *Arabidopsis thaliana* (Podgorska et al., 2017). The toxicity of ammonium is associated with changes in the cellular redox state (Podgorska et al., 2015), and uncoupling of electron transport and respiration inhibition (Hachiya and Noguchi, 2011). The duckweed species *W. microscopica* is extremely sensitive to ammonium toxicity. In this study, a concentration of 0.5 mM of ammonium was enough to inhibit the growth of *W. microscopica* and induce its senescence. Moreover, the ammonium-induced senescence seemed to be associated with dark-induced senescence. The exudates of fungal strain ITBB2-31 that inhibited ammonia-induced senescence, inhibited dark-induced senescence of the rubber tree and cassava leaves (Figure 6). Therefore, ammonia- and dark-induced senescence must have shared their signal transduction pathway. Ammonia-induced senescence has been reported in both mammals and plants (Chen and Kao, 1996; Gorg et al., 2015). Dark-induced senescence of detached maize leaves and water-stress induced senescence of detached rice leaves were associated with accumulation of ammonium ion (Chen and Kao, 1996; Lin and Kao, 1998). A decrease of glutamine synthase GS activity attributed to ammonium accumulation in maize and rice leaves (Chen and Kao, 1996; Lin and Kao, 1998; Chien and Kao, 2000), which in turn, increased tissue sensitivity to ethylene and accelerated leaf senescence (Lin and Kao, 1998). However, a later report showed that ammonia accumulation in detached rice leaves did not change tissue sensitivity to ethylene, and ruled out the possible involvement of ethylene in ammonia-induced senescence (Lin et al., 2002). Calcium effectively reduced ammonium-promoted senescence of detached rice leaves, suggesting that ammonium-induced senescence may be mediated through blocking the entrance of calcium ions into the cytosol (Hung and Kao, 1998).

5 Conclusions

W. microscopica was showed to be highly sensitive to ammonia-induced senescence, and can be served as a model plant to screen for plant anti-aging microorganisms. By co-culturing *W. microscopica* with endophytic microorgainisms isolated from rubber tree, an *Aspergillus sclerotiorum* strain ITBB2-31 that dramatically increased

the lifespan and the biomass of *W. microscopica* was selected. Interestingly, both filter-sterilized and autoclaved exudates of this fungal strain prolonged the lifespan of *W. microscopica* cultures from 1 month to at least 7 months. Moreover, the anti-aging effect of this fungal exudate on the rubber tree and cassava leaves was significant, with an increase of chlorophyll contents by 50% - 350%. Additionally, high contents of filter-sterilized exudates remarkably restrained the growth of *W. microscopica* while extending its lifespan, but high contents of autoclaved exudates presented a weak inhibition effect of its growth. These results indicated that there were both heat-sensitive growth-inhibiting and heat-stable anti-aging agents in the exudates. Comparative metabolome analysis of the filter-sterilized and autoclaved exudates revealed multiple heat-stable anti-aging and heat-sensitive growth-inhibiting compounds. Our findings will be useful in large scale screening of microbes and compounds with anti-aging function, and will be beneficial to agriculture.

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

DT: Investigation, Methodology, Validation, Writing – original draft. LF: Investigation, Validation, Writing – original draft. YY: Investigation, Validation, Writing – original draft. XS: Methodology, Validation, Writing – original draft. JZ: Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported in part by the Natural Science Foundation of China (32271932, 32271913), the Project of National Key Laboratory for Tropical Crop Breeding (NKLTCB202322), the Natural Science Foundation of Hainan Province (322RC761), and the Central Public-Interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630052020020, 1630052020022).

Acknowledgments

We are thankful to Prof. Klaus-J. Appenroth at University of Jena, Germany and Dr. K. Sowjanya Sree at Central University of Kerala, India for providing the duckweed strains 2001 and 2008. Thanks also to Mr. Zhiwei Xu at the Hainan Entry-Exit Inspection and Quarantine Bureau and Dr. Fandong Kong at the Institute of Tropical Bioscience and Biotechnology, CATAS for their technical help in Mass spectrometry. Mr. Long Xu, Mr. Zhenyu Zhang, and Miss Yuepan Ren provided technical help in some of the research.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1480588/full#supplementary-material

References

Abdelrahman, M., El-Sayed, M., Jogaiah, S., Burritt, D. J., and Tran, L. P. (2017). The "STAY-GREEN" trait and phytohormone signaling networks in plants under heat stress. *Plant Cell Rep.* 36, 1009–1025. doi: 10.1007/s00299-017-2119-y

Aguera, E., Cabello, P., and de la Haba, P. (2010). Induction of leaf senescence by low nitrogen nutrition in sunflower (*Helianthus annuus*) plants. *Physiol. Plant* 138, 256–267. doi: 10.1111/j.1399-3054.2009.01336.x

Aly, A. H., Debbab, A., and Proksch, P. (2011). Fungal endophytes: unique plant inhabitants with great promises. *Appl. Microbiol. Biotechnol.* 90, 1829–1845. doi: 10.1007/s00253-011-3270-y

Appenroth, K.-J. (2015). Useful methods 3: Media for *in vitro*-cultivation of duckweed. *Newslett. Community Duckweed Res. Appl.* 3, 169–203.

Appenroth, K.-J., Borisjuk, N., and Lam, E. (2013). Telling duckweed apart: Genotyping technologies for the Lemnaceae. *Chin. J. Appl. Environ. Biol.* 19, 1–10. doi: 10.3724/SP.J.1145.2013.00001

Appenroth, K.-J., Sree, K. S., Böhm, V., Hammannd, S., Vetter, W., Leiterer, M., et al. (2017). Nutritional value of duckweeds (*Lemnaceae*) as human food. *Food Chem.* , 217, 266–273. doi: 10.1016/j.foodchem.2016.08.116

Appenroth, K.-J., Teller, S., and Horn, M. (1996). Photophysiology of turion formation and germination in Spirodela polyrhiza. *Biol. Plantarum* 38, 95–106. doi: 10.1007/BF02879642

Ashcroft, F. M., Kerr, A. J., Gibson, J. S., and Williams, B. A. (1991). Amantadine and sparteine inhibit ATP-regulated K-currents in the insulin-secreting beta-cell line, HIT-T15. *Br. J. Pharmacol.* 104, 579–584. doi: 10.1111/j.1476-5381.1991.tb12472.x

Ayouba, A., Chatelain, C., and Rouge, P. (1991). Legume lectins interact with muramic acid and N-acetylmuramic acid. *FEBS Lett.* 289, 102–104. doi: 10.1016/0014-5793(91)80918-S

Bakshi, I. S., Abad Farooqi, A. H., and Maheshwari, S. C. (1979). Control of circadian rhythm in nitrate reductase activity in *Wolffia microscopica* Griff. *Plant Cell Physiol.* 20, 957–963. doi: 10.1093/oxfordjournals.pcp.a075890

Bekavac, G., Stojaković, M., Ivanović, M., Jocković, Đ, Vasić, N., Purar, B., et al. (2002). Relationships of stay green trait in maize. *Genetika* 34, 33–40. doi: 10.2298/GENSR0201033B

Bjugstad, K. B., Zawada, W. M., Goodman, S., and Freed, C. R. (2001). IGF-1 and bFGF reduce glutaric acid and 3-hydroxyglutaric acid toxicity in striatal cultures. *J. Inherit Metab. Dis.* 24, 631–647. doi: 10.1023/A:1012706908779

Borrell, A. K., van Oosterom, E. J., Mullet, J. E., George-Jaeggli, B., Jordan, D. R., Klein, P. E., et al. (2014). Stay-green alleles individually enhance grain yield in sorghum under drought by modifying canopy development and water uptake patterns. *New Phytol.* 203, 817–830. doi: 10.1111/nph.2014.203.issue-3

Cao, Y., Liu, S., Yang, X., Guo, L., Cai, C., and Yao, J. (2019). Effects of dietary leucine and phenylalanine on gastrointestinal development and small intestinal enzyme activities in milk-fed holstein dairy calves. *Biosci. Rep.*, 39, BSR20181733. doi: 10.1042/BSR20181733

Carlin, C., Phillips, P. D., Brooks-Frederich, K., Knowles, B. B., and Cristofalo, V. J. (1994). Cleavage of the epidermal growth factor receptor by a membrane-bound leupeptin-sensitive protease active in nonionic detergent lysates of senescent but not young human diploid fibroblasts. *J. Cell Physiol.* 160, 427–434. doi: 10.1002/jcp.1041600305

Chen, S. J., and Kao, C. H. (1996). Ammonium accumulation in relation to senescence of detached maize leaves. *Bot. Bull. Acad. Sin.* 37, 255–259. doi: 10.1016/S0304-3770(96)01067-4

Chien, H., and Kao, C. H. (2000). Accumulation of ammonium in rice leaves in response to excess cadmium. *Plant Sci.* 156, 111–115. doi: 10.1016/S0168-9452(00)00234-X

Cho, J. Y., Kim, H. Y., Choi, G. J., Jang, K. S., Lim, H. K., Lim, C. H., et al. (2006). Dehydro-alpha-lapachone isolated from Catalpa ovata stems: activity against plant pathogenic fungi. *Pest Manag Sci.* 62, 414–418. doi: 10.1002/ps.v62:5

Chung, K. H., Cho, K. Y., Asami, Y., Takahashi, N., and Yoshida, S. (1989). New 4-hydroxypyridine and 4-hydroxyquinoline derivatives as inhibitors of NADH-ubiquinone reductase in the respiratory chain. *Z Naturforsch. C J. Biosci.* 44, 609–616. doi: 10.1515/znc-1989-7-811

Cleland, C. F., and Briggs, W. R. (1967). Flowering responses of the long-day plant *Lemna gibba* G3. *Plant Physiol.* 42, 1553–1561. doi: 10.1104/pp.42.11.1553

Cock, J. H., Franklin, D., Sandoval, G., and Juri, P. (1979). The ideal cassava plant for maximum yield. *Crop Sci.* 19, 271–279. doi: 10.2135/cropsci1979.0011183X001900020025x

Diaz, C., Lemaitre, T., Christ, A., Azzopardi, M., Kato, Y., Sato, F., et al. (2008). Nitrogen recycling and remobilization are differentially controlled by leaf senescence and development stage in Arabidopsis under low nitrogen nutrition. *Plant Physiol.* 147, 1437–1449. doi: 10.1104/pp.108.119040

Fu, L., Huang, M., Han, B., Sun, X., Sree, K. S., Appenroth, K. J., et al. (2017). Flower induction, microscope-aided cross-pollination, and seed production in the duckweed Lemna gibba with discovery of a male-sterile clone. *Sci. Rep.* 7, 3047. doi: 10.1038/s41598-017-03240-8

Garkavtsev, I., Chauhan, V. P., Wong, H. K., Mukhopadhyay, A., Glicksman, M. A., Peterson, R. T., et al. (2011). Dehydro-alpha-lapachone, a plant product with antivascular activity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11596–11601. doi: 10.1073/pnas.1104225108

Gazis, R., and Chaverri, P. (2010). Diversity of fungal endophytes in leaves and stems of wild rubber trees (*Hevea brasiliensis*) in Peru. *Fungal Ecol.* 3, 240–254. doi: 10.1016/j.funeco.2009.12.001

Georgousaki, K., Tsafantakis, N., Gumeni, S., Lambrinidis, G., Gonzalez-Menendez, V., Tormo, J. R., et al. (2020). Biological evaluation and in silico study of benzoic acid derivatives from Bjerkandera adusta targeting proteostasis network modules. *Molecules* 25, 666. doi: 10.3390/molecules25030666

Gorg, B., Karababa, A., Shafigullina, A., Bidmon, H. J., and Haussinger, D. (2015). Ammonia-induced senescence in cultured rat astrocytes and in human cerebral cortex in hepatic encephalopathy. *Glia* 63, 37–50. doi: 10.1002/glia.22731

Griffith, W. (1851). *Notulae ad plantas asiaticas, vol. 3* (Calcutta, 492 India: Bishop College Press).

Hachiya, T., and Noguchi, K. (2011). Integrative response of plant mitochondrial electron transport chain to nitrogen source. *Plant Cell Rep.* 30, 195–204. doi: 10.1007/s00299-010-0955-0

Harzallah, D., Sadrati, N., Zerroug, A., Dahamna, S., and Bouharati, S. (2012). Endophytic fungi isolated from wheat (*Triticum durum Desf.*): evaluation of their antimicrobial activity, antioxidant activity and host growth promotion. *Commun. Agric. Appl. Biol. Sci.* 77, 245–248. doi: mdl-23878980

Hegelmaier, F. (1885). *Wolffia microscopica. . Botanische Zeitung* 43, 241–249.

Heng, T., and Joo, G. (2017);Rubber in *Encyclopedia of applied plant sciences*, *2nd ed.*Eds. B. Thomas, B. Murray and D. Murphy (Academic Press: Oxford), 402–409.

Hoagland, D. R., and Arnon, D. I. (1950). The water-culture method for growing plants without soil. *California Agric. Experiment Station Circular* 347, 1–32.

Huber, G. A. (1933). *Aspergillus sclerotiorum, n.* sp., and its relation to decay of apples. *Phytopathology* 23, 3.

Hung, K. T., and Kao, C. H. (1998). Ammonium, calcium, and leaf senescence in rice. *Plant Growth Regul.* 26, 63–66. doi: 10.1023/A:1006012827917

Inderjit, Bajpai, D., and Rajeswari, M. S. (2010). Interaction of 8-hydroxyquinoline with soil environment mediates its ecological function. *PloS One* 5, e12852. doi: 10.1371/journal.pone.0012852

Joshi, A. K., Kumari, M., Singh, V. P., Reddy, C. M., Kumar, S., Rane, J., et al. (2007). Stay green trait: variation, inheritance and its association with spot blotch resistance in spring wheat (Triticum aestivum L.). *Euphytica* 153, 59–71. doi: 10.1007/s10681-006-9235-z

Keyser, B., Glatzel, M., Stellmer, F., Kortmann, B., Lukacs, Z., Kolker, S., et al. (2008). Transport and distribution of 3-hydroxyglutaric acid before and during induced encephalopathic crises in a mouse model of glutaric aciduria type 1. *Biochim. Biophys. Acta* 1782, 385–390. doi: 10.1016/j.bbadis.2008.02.008

Khan, A. R., Ullah, I., Waqas, M., Park, G. S., Khan, A. L., Hong, S. J., et al. (2017). Host plant growth promotion and cadmium detoxification in Solanum nigrum, mediated by endophytic fungi. *Ecotoxicol Environ. Saf.* 136, 180–188. doi: 10.1016/j.ecoenv.2016.03.014

Khurana, J. P., Tamot, B. K., and Maheshwari, S. C. (1986). Induction of flowering in a duckweed, Wolffia microscopica, under non-inductive long days, by 8-hydroxyquinoline. *Plant Cell Physiol.* 27, 373–376. doi: 10.1016/0303-7207(86)90135-8

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

Kurz, S. (1866). Enumeration of Indian *lemnaceae. J. Linn Soc. Bot.* 9, 264–268. doi: 10.1111/j.1095-8339.1866.tb01286.x

Kusaba, M., Tanaka, A., and Tanaka, R. (2013). Stay-green plants: what do they tell us about the molecular mechanism of leaf senescence. *Photosynth Res.* 117, 221–234. doi: 10.1007/s11120-013-9862-x

Landolt, E. (1986). "Biosystematic investigations in the family of duckweeds (*Lemnaceae*) Vol. 2," in *The family of* Lemnaceae*: a monographic study. – Morphology, karyology, ecology, geographic distribution, nomenclature, descriptions* (Eidgenössische Technische Hochschule Zürich;, Zürich).

Larriba, E., Jaime, M. D., Nislow, C., Martin-Nieto, J., and Lopez-Llorca, L. V. (2015). Endophytic colonization of barley (*Hordeum vulgare*) roots by the nematophagous fungus Pochonia chlamydosporia reveals plant growth promotion and a general defense and stress transcriptomic response. *J. Plant Res.* 128, 665–678. doi: 10.1007/s10265-015-0731-x

Lehtonen, L., Eerola, E., and Toivanen, P. (1997). Muramic acid in human peripheral blood leucocytes in different age groups. *Eur. J. Clin. Invest.* 27, 791–792. doi: 10.1046/j.1365-2362.1997.1950732.x

Lichtenthaler, H. K. (1987). "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," in *Methods in enzymology*, vol. 148. (New York: Academic Press), 350–382.

Lichtenthaler, H. (2010). *Chlorophyll and carotenoid determination (after Lichtenthaler 1987)*, a practical instruction.

Lin, J.-N., and Kao, C. H. (1998). Water stress, ammonium, and leaf senescence in detached rice leaves. *Plant Growth Regul.* 25, 165–169. doi: 10.1023/A:1006014618721

Lin, C.-C., Ting Hsu, Y., and Kao, C.-H. (2002). Ammonium ion, ethylene, and NaCl-induced senescence of detached rice leaves. *Plant Growth Regul.* 37, 85–92. doi: 10.1023/A:1020359825056

Ma, Y., Prasad, M. N., Rajkumar, M., and Freitas, H. (2011). Plant growth promoting rhizobacteria and endophytes accelerate phytoremediation of metalliferous soils. *Biotechnol. Adv.* 29, 248–258. doi: 10.1016/j.biotechadv.2010.12.001

Maheshwari SC, S. C., and Chauhan OS, O. S. (1963). *In vitro* control of flowering in Wolffia microscopica. *Nature* 198, 99–100. doi: 10.1038/198099b0

Maheshwari SC, S. C., and Venkataraman, R. (1966). Induction of flowering in a duckweed Wolffia microscopica by a new kinin, zeatin. *Planta* 70, 304–306. doi: 10.1007/BF00396494

McClerklin, S. A., Lee, S. G., Harper, C. P., Nwumeh, R., Jez, J. M., and Kunkel, B. N. (2018). Indole-3-acetaldehyde dehydrogenase-dependent auxin synthesis contributes to virulence of Pseudomonas syringae strain DC3000. *PloS Pathog.* 14, e1006811. doi: 10.1371/journal.ppat.1006811

Nguyen, T., Coover, R. A., Verghese, J., Moran, R. G., and Ellis, K. C. (2014). Phenylalanine-based inactivator of akt kinase: design, synthesis, and biological evaluation. *ACS Med. Chem. Lett.* 5, 462–467. doi: 10.1021/ml500088x

Parmaki, S., Vyrides, I., Vasquez, M. I., Hartman, V., Zacharia, I., Hadjiadamou, I., et al. (2018). Bioconversion of alkaloids to high-value chemicals: Comparative analysis of newly isolated lupanine degrading strains. *Chemosphere* 193, 50–59. doi: 10.1016/j.chemosphere.2017.10.165

Paz, I. C., Santin, R. C., Guimaraes, A. M., Rosa, O. P., Dias, A. C., Quecine, M. C., et al. (2012). Eucalyptus growth promotion by endophytic Bacillus spp. *Genet. Mol. Res.* 11, 3711–3720. doi: 10.4238/2012.August.17.9

Penyalver, R., Oger, P. M., Su, S., Alvarez, B., Salcedo, C. I., Lopez, M. M., et al. (2009). The S-adenosyl-L-homocysteine hydrolase gene ahcY of *Agrobacterium radiobacter* K84 is required for optimal growth, antibiotic production, and

biocontrol of crown gall disease. *Mol. Plant Microbe Interact.* 22, 713–724. doi: 10.1094/MPMI-22-6-0713

Pfautsch, S., Rennenberg, H., Bell, T. L., and Adams, M. A. (2009). Nitrogen uptake by Eucalyptus regnans and *Acacia* spp. - preferences, resource overlap and energetic costs. *Tree Physiol.* 29, 389–399. doi: 10.1093/treephys/tpn033

Podgorska, A., Burian, M., Rychter, A. M., Rasmusson, A. G., and Szal, B. (2017). Short-term ammonium supply induces cellular defence to prevent oxidative stress in Arabidopsis leaves. *Physiol. Plant* 160, 65–83. doi: 10.1111/ppl.12538

Podgorska, A., Ostaszewska, M., Gardestrom, P., Rasmusson, A. G., and Szal, B. (2015). In comparison with nitrate nutrition, ammonium nutrition increases growth of the frostbite1 Arabidopsis mutant. *Plant Cell Environ.* 38, 224–237. doi: 10.1111/pce.12404

Pothier, J., Cheav, S. L., Galand, N., Dormeau, C., and Viel, C. (1998). A comparative study of the effects of sparteine, lupanine and lupin extract on the central nervous system of the mouse. *J. Pharm. Pharmacol.* 50, 949–954. doi: 10.1111/j.2042-7158.1998.tb04013.x

Powell, R. G., and Petroski, R. J. (1992). Alkaloid toxins in endophyte-infected grasses. *Natural Toxins* 1, 163–170. doi: 10.1002/nt.2620010304

Powers, J. C., Kam, C. M., Narasimhan, L., Oleksyszyn, J., Hernandez, M. A., and Ueda, T. (1989). Mechanism-based isocoumarin inhibitors for serine proteases: use of active site structure and substrate specificity in inhibitor design. *J. Cell Biochem.* 39, 33–46. doi: 10.1002/jcb.240390105

Quecine, M. C., Araujo, W. L., Rossetto, P. B., Ferreira, A., Tsui, S., Lacava, P. T., et al. (2012). Sugarcane growth promotion by the endophytic bacterium *Pantoea agglomerans 33.1*. *Appl. Environ. Microbiol.* 78, 7511–7518. doi: 10.1128/AEM.00836-12

Rawson, H. M., Hindmarsh, J. H., Fisher, R. A., and Stockman, Y. M. (1983). Changes in leaf photosynthesis with plant ontogeny and relationships with yield per ear in wheat cultivars and 120 progeny. *Aust. J. Plant Physiol.* 10, 503–514. doi: 10.1071/PP9830503

Rodriguez, R., and Redman, R. (2008). More than 400 million years of evolution and some plants still can't make it on their own: plant stress tolerance via fungal symbiosis. *J. Exp. Bot.* 59, 1109–1114. doi: 10.1093/jxb/erm342

Roy, R. P., and Dutt, B. (1967). Cytology of *wolffia microscopica* kurz. *Cytologia* 32, 270–272. doi: 10.1508/cytologia.32.270

Santoyo, G., Moreno-Hagelsieb, G., Orozco-Mosqueda Mdel, C., and Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiol. Res.* 183, 92–99. doi: 10.1016/j.micres.2015.11.008

Sasaki, H., and Ishii, R. (1992). Cultivar differences in leaf photosynthesis of rice bred in Japan. *Photosynthesis Res.* 32, 139–146. doi: 10.1007/BF00035948

Schardl, C. L., Leuchtmann, A., and Spiering, M. J. (2004). Symbioses of grasses with seedborne fungal endophytes. *Annu. Rev. Plant Biol.* 55, 315–340. doi: 10.1146/annurev.arplant.55.031903.141735

Schenk, R. U., and Hildebrandt, A. C. (1972). Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. *Can. J. Bot.* 50, 199–204. doi: 10.1139/b72-026

Shi, Y., Lou, K., and Li, C. (2010). Growth and photosynthetic efficiency promotion of sugar beet (*Beta vulgaris L.*) by endophytic bacteria. *Photosynth Res.* 105, 5–13. doi: 10.1007/s11120-010-9547-7

Song, W., Sun, H., Li, J., Gong, X., Huang, S., Zhu, X., et al. (2013). Auxin distribution is differentially affected by nitrate in roots of two rice cultivars differing in responsiveness to nitrogen. *Ann. Bot.* 112, 1383–1393. doi: 10.1093/aob/mct212

Song, Y., Yang, C., Gao, S., Zhang, W., Li, L., and Kuai, B. (2014). Age-triggered and dark-induced leaf senescence require the bHLH transcription factors PIF3, 4, and 5. *Mol. Plant* 7, 1776–1787. doi: 10.1093/mp/ssu109

Sree, K. S., and Appenroth, K. J. (2014). Rediscovery of *wolffia microscopica* (Griff.) kurz. *Newslett. Int. Steering Committee Duckweed Res. Appl.* , 3, 2–4.

Sree, K. S., and Appenroth, K. J. (2016). Useful methods 4: Stock cultivation of duckweed. *Newslett. Community Duckweed Res. Appl.* 4, 204–237.

Sree, K. S., Maheshwari, S. C., Bokac, K., Khuranad, J. P., Keresztesc, A., and Appenroth, K. J. (2015). The duckweed *Wolffia microscopica*: A unique aquatic monocot. *Flora* 210, 31–39. doi: 10.1016/j.flora.2014.10.006

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi: 10.1093/oxfordjournals.molbev.a040023

Tan, D., Fu, L., Han, B., Sun, X., Zheng, P., and Zhang, J. (2015). Identification of an endophytic antifungal bacterial strain isolated from the rubber tree and its application in the biological control of banana Fusarium wilt. *PloS One* 10, e0131974. doi: 10.1371/journal.pone.0131974

Thomas, H., and Ougham, H. (2014). The stay-green trait. *J. Exp. Bot.* 65, 3889–3900. doi: 10.1093/jxb/eru037

Visagie, C. M., Varga, J., Houbraken, J., Meijer, M., Kocsube, S., Yilmaz, N., et al. (2014). Ochratoxin production and taxonomy of the yellow aspergilli(*Aspergillus section circumdati*). *Stud. Mycol* 78, 1–61. doi: 10.1016/j.simyco.2014.07.001

Voigt, J. O. (1845). *Hortus suburbanus calcuttensis* (Calcutta, India: Bishop College Press).

Weakley, S. M., Wang, X., Mu, H., Lu, J., Lin, P. H., Yao, Q., et al. (2011). Ginkgolide A-gold nanoparticles inhibit vascular smooth muscle proliferation and migration *in vitro* and reduce neointimal hyperplasia in a mouse model. *J. Surg. Res.* 171, 31–39. doi: 10.1016/j.jss.2011.03.018

Weaver, L. M., and Amasino, R. M. (2001). Senescence is induced in individually darkened Arabidopsis leaves, but inhibited in whole darkened plants. *Plant Physiol.* 127, 876–886. doi: 10.1104/pp.010312

Wei, B., Zhang, W., Chao, J., Zhang, T., Zhao, T., Noctor, G., et al. (2017). Functional analysis of the role of hydrogen sulfide in the regulation of dark-induced leaf senescence in *Arabidopsis*. *Sci. Rep.* 7, 2615. doi: 10.1038/s41598-017-02872-0

Yang, D., Li, Y., Shi, Y., Cui, Z., Luo, Y., Zheng, M., et al. (2016). Exogenous cytokinins increase grain yield of winter wheat cultivars by improving stay-green characteristics under heat stress. *PloS One* 11, e0155437. doi: 10.1371/journal.pone.0155437

You, Y. H., Yoon, H., Kang, S. M., Shin, J. H., Choo, Y. S., Lee, I. J., et al. (2012). Fungal diversity and plant growth promotion of endophytic fungi from six halophytes in Suncheon Bay. *J. Microbiol. Biotechnol.* 22, 1549–1556. doi: 10.4014/jmb.1205.05010

Yu, S., Xia, H., Guo, Y., Qian, X., Zou, X., Yang, H., et al. (2020). Ginsenoside Rb1 retards aging process by regulating cell cycle, apoptotic pathway and metabolism of aging mice. *J. Ethnopharmacol* 255, 112746. doi: 10.1016/j.jep.2020.112746

Yung, T. W., Jonnalagadda, S., Balagurunathan, B., and Zhao, H. (2016). Transcriptomic analysis of 3-hydroxypropanoic acid stress in Escherichia coli. *Appl. Biochem. Biotechnol.*, 178, 527–543. doi: 10.1007/s12010-015-1892-8

Zhang, P., Wang, W. Q., Zhang, G. L., Kaminek, M., Dobrev, P., Xu, J., et al. (2010). Senescence-inducible expression of isopentenyl transferase extends leaf life, increases drought stress resistance and alters cytokinin metabolism in cassava. *J. Integr. Plant Biol.* 52, 653–669. doi: 10.1111/j.1744-7909.2010.00956.x

Zheng, P., He, J., Chang, K., Zhang, S., Tan, D., Sun, X., et al. (2009a). Isolation and identification of endophytic fungi from rubber tree and their antagonism to plant pathogens. *Chin. J. Trop. Crops* 30, 832–837. doi: 10.3969/j.issn.1000-2561.2009.06.020

Zheng, P., Tan, D., Sun, X., and Zhang, J. (2009c). Morphology and phylogenetic position of an endophytic fungus ITBB2-1 from rubber tree. *Chin. J. Trop. Crops* 30, 314–319. doi: 10.3969/j.issn.1000-2561.2009.03.013

Ziegler, P., Adelmann, K., Zimmer, S., Schmidt, C., and Appenroth, K. J. (2015). Relative *in vitro* growth rates of duckweeds (*Lemnaceae*) - the most rapidly growing higher plants. *Plant Biol. (2015)* 17(Suppl1), 33–41. doi: 10.1111/plb.2014.17.issue-s1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 November 2024
DOI 10.3389/fpls.2024.1499024

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Ake Liu,
Changzhi University, China
Pengbo Xu,
Shanghai Jiao Tong University, China
Pingchuan Deng,
Northwest A&F University, China

*CORRESPONDENCE
Yang Chen
chenyangrz@126.com

RECEIVED 20 September 2024
ACCEPTED 21 October 2024
PUBLISHED 13 November 2024

CITATION
Zhang C, Li H, Yin J, Han Z, Liu X and Chen Y (2024) Pan-genome wide identification and analysis of the *SAMS* gene family in sunflowers (*Helianthus annuus* L.) revealed their intraspecies diversity and potential roles in abiotic stress tolerance. *Front. Plant Sci.* 15:1499024. doi: 10.3389/fpls.2024.1499024

Pan-genome wide identification and analysis of the *SAMS* gene family in sunflowers (*Helianthus annuus* L.) revealed their intraspecies diversity and potential roles in abiotic stress tolerance

Chun Zhang[1], Haoyu Li[2], Jiamin Yin[1], Zhibin Han[1], Xinqi Liu[1] and Yang Chen[1]*

[1]Department of Agronomy, Hetao College, Bayannur, China, [2]Bayannur Modern Agriculture and Animal Husbandry Development Center, Bayannur, China

Introduction: S-adenosylmethionine (SAM), a key molecule in plant biology, plays an essential role in stress response and growth regulation. Despite its importance, the SAM synthetase *(SAMS)* gene family in sunflowers *(Helianthus annuus L.)* remains poorly understood.

Methods: In this study, the *SAMS* genes were identified from the sunflower genome. Subsequently, the protein properties, gene structure, chromosomal location, cis-acting elements, collinearity, and phylogeny of the SAMS gene family were analyzed by bioinformatic methods. Finally, the expression patterns of *SAMS* genes in different tissues, under different hormonal treatment and abiotic stress were analyzed based on transcriptome data and qRT-PCR.

Results: This study identified 58 SAMS genes across nine cultivated sunflower species, which were phylogenetically classified into seven distinct subgroups. Physicochemical properties and gene structure analysis showed that the *SAMS* genes are tightly conserved between cultivars. Collinearity analysis revealed segmental duplications as the primary driver of gene family expansion. The codon usage bias analysis suggested that natural selection substantially shapes the codon usage patterns of sunflower *SAMS* genes, with a bias for G/C-ending high-frequency codons, particularly encoding glycine, leucine, and arginine. Analysis of the cis-regulatory elements in promoter regions, implied their potential roles in stress responsiveness. Differential expression patterns for HanSAMS genes were observed in different tissues as well as under hormone treatment or abiotic stress conditions by analyzing RNA-seq data from previous studies and qRT-PCR data in our current study. The majority of genes demonstrated a robust response to BRA and IAA treatments in leaf tissues, with no significant expression change observed in roots, suggesting the response of *HanSAMS* genes to hormones is tissue-specific. Expression analyses under abiotic stresses demonstrated diverse expression profiles of *HanSAMS* genes,

with *HanSAMS5* showing significant upregulation in response to both drought and salt stresses.

Discussion: This comprehensive genomic and expression analysis provides valuable insights into the *SAMS* gene family in sunflowers, laying a robust foundation for future functional studies and applications in crop improvement for stress resilience.

KEYWORDS

S-adenosylmethionine synthase, pan-genome, sunflowers, abiotic stresses, qRT-PCR

1 Introduction

Plants have developed sophisticated and adaptable mechanisms to adjust to challenging environments, involving a spectrum of morphological, physiological, and molecular changes (Ahuja et al., 2010). They frequently employ strategies such as strengthening and preserving the integrity of biological membranes, along with boosting the production of antioxidant enzymes, to endure stresses from cold, drought, and high salinity (Yang et al., 2013; Mehari et al., 2021). The cultivated sunflower (*Helianthus annuus* L.) is a prominent oil crop with global significance, renowned for its resilience in adverse environmental conditions,which is originally domesticated by Native Americans in North America, and later introduced to Europe and subsequently became a vital crop worldwide (Zukovsky, 1950; Mantenese et al., 2006). Nevertheless, the cultivation of sunflowers faces various challenges, with drought and salinity being prominent abiotic stressors (Rele and Mohile, 2003; Keeley et al., 2021). A multitude of gene families, including S-adenosyl-L-methionine synthase (SAMS), are integral to the complex regulatory networks that dictate plant stress responses, impacting their growth and bolstering their resilience to harsh conditions (He et al., 2019).

SAMS genes are distinguished by the presence of a methionine-binding site in their N-terminal domain and an ATP-binding motif in their C-terminal domain. These enzymes catalyze the formation of SAM (S-Adenosyl-L-methionine) through the condensation of methionine with ATP, playing a crucial role in essential biological pathways within eukaryotic cells (Heidari et al., 2020). Numerous *SAMS* genes have been identified by researchers (Ahuja et al., 2010; Yang et al., 2013; Heidari et al., 2020). In *Arabidopsis thaliana*, there exist four *SAMS* genes, with *AtSAMS3* demonstrating predominant expression within pollen tissues (Yang et al., 2013). The suppression of *OsSAMS1*, 2, and 3 in rice (*Oryza sativa*) led to alterations in histone modifications and DNA methylation patterns, which in turn triggered a delay in flowering time (Li et al., 2011). Espartero et al. observed that the expression of *SAMS* in tomatoes (*Solanum lycopersicum*) was altered in response to salt stress (Heidari et al., 2020). Similarly, in cucumbers (*Cucumis sativus*), salt stress induced the expression of *SAMS*, implicating its role in the modulation of associated stress-response mechanisms (Roje, 2006; Bürstenbinder et al., 2007). In soybean (*Glycine max*), the expression profiles of the *SAMS* gene family exhibited significant variation in the face of drought and waterlogging stress, yet displayed relative stability under treatments involving sodium chloride (NaCl) and low temperatures (Jang et al., 2012; Ma et al., 2017). The gene *GhSAMS2* has emerged as a promising candidate for the genetic enhancement of upland cotton's resistance to multiple abiotic stresses (Gupta et al., 2013). The overexpression of *CsSAMS1* and its interaction with *CsCDPK6* resulted in the stimulation of ethylene and polyamines biosynthesis, ultimately improving salt stress tolerance in transgenic tobacco (*Nicotiana tabacum*) plants (Zhu et al., 2021). Overexpressing *Medicago sativa subsp. Falcata SAMS1* in transgenic tobacco plants increased their tolerance to cold stress by enhancing oxidation and polyamine synthesis (Guo et al., 2014).

Pan-genomic analysis, now a prevalent approach, is utilized to assess genetic variability within species, explore gene flow between species, and examine the processes of domestication and crop improvement (Hübner et al., 2019; Li et al., 2010; Gao et al., 2019; Tao et al., 2021; Tettelin et al., 2005). A single reference genome might not capture the full spectrum of genetic diversity that evolves over time within a species, possibly leading to the exclusion of many important genes. While the *SAMS* gene family has been extensively researched in *A. thaliana*, rice, cotton, and tomato, there is a pronounced shortfall in studies on the *SAMS* genes in sunflowers, particularly in relation to their pan-genome diversity and how their expression patterns react to abiotic stresses such as cold, drought, and salinity, as well as to external hormonal signals (He et al., 2019; Heidari et al., 2020; Sun et al., 2022).

Abbreviations: *SAMS*, S-adenosylmethionine synthase; *HanSAMS*, S-adenosylmethionine synthase in *Helianthus annuus* L; MW, Molecular weight; pI, Isoelectric points; GRAVY, Grand average of hydropathicity; CDS, Coding sequence; CUB, codon usage bias; GC1, GC2, and GC3, GC content in the base composition of codon 1st, 2nd, and 3rd positions; RSCU, Relative synonymous codon usage; Enc, Effective number of codons; GC12, The average of GC1 and GC2; GC3s, GC content of 3rd synonymous codons; qRT-PCR, quantitative RT-PCR.

In this research, we conducted a comprehensive, genome-wide identification of *SAMS* genes utilizing the sunflower pan-genome. A total of 58 *SAMS* genes were discovered across nine cultivated sunflower varieties. We investigated their phylogenetic relationships, gene structures, motifs, *cis*-elements, and the secondary and tertiary structures of the corresponding proteins. Additionally, we analyzed the codon usage bias in these 58 *SAMS* genes, employing neutrality plot, ENc-plot, PR2-plot, and the Relative Synonymous Codon Usage (RSCU) method. Building on this, we extracted gene expression data for the *SAMS* gene from a variety of conditions, including exposure to abiotic stresses and treatments with external hormones. Furthermore, we performed a systematic analysis of the *SAMS* gene expressions, with a particular focus on their expression patterns under drought and salt stress conditions, using quantitative real-time PCR (qRT-PCR). These results provided comprehensive genomic information of sunflower *SAMS* gene family, enhancing our understanding of their roles in stress response and potentially contributing to the development of sunflower varieties with improved stress tolerance.

2 Materials and methods

2.1 Identification of *SAMS* genes

Protein sequences of SAMSs from *A. thaliana* were obtained from the TAIR (https://www.Arabidopsis.org/). Genome and annotation files of *Helianthus annuus* XRQ) was downloaded from Ensembl plants (https://plants.ensembl.org/index.html). *Helianthus annuus* (HA89), *Helianthus annuus* (HA300), *Helianthus annuus* (IR), *Helianthus annuus* (LR1), *Helianthus annuus* (OQP8), *Helianthus annuus* (PI659440), *Helianthus annuus* (PSC8) and *Helianthus annuus* (RHA438) were downloaded from NCBI. The Hidden Markov Model (HMM) (PF02772, PF02773, PF00438) of S-adenosylmethionine synthase was downloaded from the Pfam database (https://pfam.xfam.org/), and were used to search protein databases by HMMER in TBtools-II (Chen et al., 2023) with an E-value<1e−5. Subsequently, all putative *SAMS* genes shared the three HMM domains were validated by batch-CD search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), Pfam, and HMMER (https://www.ebi.ac.uk/Tools/hmmer/) databases (Potter et al., 2018; Mistry et al., 2021; Wang et al., 2023). The *SAMS* genes in the XRQ cultivar are named using Latin abbreviations coupled with their chromosomal positions in the XRQ reference genome. For instance, the designation *XRQ-HanSAMS1* indicates that XRQ represents the cultivar name, Han refers to Helianthus annuus, and the numeral in SAMS corresponds to the gene's ordered position on the chromosome, listed from the smallest to the largest. Other cultivars' genes keep their names but get a SAMS number based on where they group with *XRQ-HanSAMS* genes in the evolution tree. Furthermore, the biochemical parameters of HanSAMS were determined using the ProtParam tool (https://web.expasy.org/protparam/) (Gasteiger et al., 2005). Finally, the subcellular localizations of *HanSAMS* were predicted using the WoLF PSORT (https://wolfpsort.hgc.jp/). The NPS@: SOPMA secondary structure (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) was used to predict the secondary structures of HanSAMS proteins. SWISS-MODEL (https://swissmodel.expasy.org/) was employed to 3D protein structure prediction and PyMOL software was used to draw 3D structures of SAMS proteins (PyMOL molecular graphics system; http://www.pymol.org) (DeLano, 2002).

2.2 Phylogenetic, gene structure, *cis*-element, motif and collinear analysis

Multiple sequences alignments of the full-length SAMS protein sequences was performed using the ClustalX (Larkin et al., 2007). The Neighbor-joining (NJ) tree was constructed by MEGA7.0 with the amino acid substitution Poisson model and 1000 bootstrap replicates test model (Kumar et al., 2016). The exon-intron structure of the *SAMS* genes was analysed using GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). Conserved domains of SAMS proteins were analysed by MEME (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi). The upstream 2000 bp sequences relative to the start codon of each *SAMS* gene were obtained to analyze the promoter regions, and the *cis*-elements within these regions were predicted using the PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). We employed BLASTP to identify homologous genes, with key parameters set to an e-value threshold of 1e-3 and a maximum of 10 target sequences. To identify collinear genes using MCScanX (Wang et al., 2012), we applied the default parameters, include an E_VALUE of 1e-05 and a MAX_GAPS count of 25. The nonsynonymous substitution rate/synonymous substitution rate (Ka/Ks) values were calculated via the DnaSP 6.0 application released by Universitat de Barcelona.

2.3 Estimation of codon bias

A Python-compiled custom program was used to calculate the genomic composition of the *SAMS* gene family across nine cultivated sunflower varieties, determining the total GC content (GCall) as well as the GC content at the first (GC1), second (GC2), the average GC content at the first and second positions (GC12) and third (GC3) codon positions within the coding DNA sequences (CDS). Additionally, we utilized the software CodonW v1.4.4 (http://codonw.sourceforge.net) to assess the relative synonymous codon usage (RSCU), count the number of effective codons (ENc), and calculate the codon adaptation index (CAI), also determining the length of the amino acid sequences. Furthermore, we conducted a series of analyses to visualize the codon usage bias and neutrality: the Neutrality plot, the PR2 plot, the ENc-plot, and the RSCU plot were all generated using R software.

2.4 Analysis of RNA-seq data of *HanSAMS*

Hormonal response expression data (NCBI accession number SRP092742) were sourced from the SunExpress V1 database, which provides a comprehensive resource for exploring the expression patterns of genes under various conditions in sunflowers. The FPKM values for all *XRQ-HanSAMS* genes were extracted and subsequently processed using TBtools-II to create heatmaps.

2.5 Plant cultivation, treatments, RNA isolation, and qRT-PCR

The sunflower salt-tolerant inbred line 19S05 was used to explore the influence of salt and drought stress on sunflower seedlings. We sowed high-quality sunflower seeds in a perforated plastic container filled with nutrient-rich soil, ensuring they received regular watering every three days to support their healthy development. The plants were grown under a controlled photoperiod of 16 hours of light followed by 8 hours of darkness, all within a stable room temperature range of 21 to 25 degrees Celsius (Song et al., 2024b). Once the sunflower seedlings reached the four-true-leaf stage, seedings were treated with 150 mM NaCl solution and 15% PEG6000 solution, respectively. The leaves were then collected at 0 h, 1 h, 3 h, 6 h, 12 h, and 24 h, immediately frozen in liquid nitrogen, and stored at −80°C. The total RNA isolation and purification of samples were performed using an RNAprep Pure Plant Plus Kit (rich in polysaccharides and polyphenolics) (Tiangen, Beijing, China). The RNA isolation for gene expression was done in biological replicates for each sample analyzed. RNA integrity was visualized by 1% agarose gel electrophoresis. The concentration and purity of RNAs (OD260/OD280>1.95) were determined with a NanoDrop Onemicrovolume UVvis spectrophotometer (NanoDrop Technologies, DE, USA). Further, 1 ug of total RNA was reverse transcribed in a 20 ul reaction volume using a PrimeScript RT reagent kit with a gDNA eraser (Code No.6210A, Takara, Beijing, China) following the manufacturer's instructions to remove traces of contaminant DNA and prepare cDNA. 1 μg of purified total RNA was reverse transcribed into the first strand cDNA that was used to qRT-PCR. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis was used to analyze the expression level of the identified *HanSAMSs*. The standard qRT-PCR with SYBR Premix Ex Taq II (TaKaRa, Beijing, China) was repeated at least three times on a CFX96 real-time System (BioRad, Beijing). Subsequently, Cycling parameters were 95°C for 30 s, 40 cycles of 95°C for 5 s, and 60°C for 30 s. For melting curve analysis, a program including 95°C for 15 s, followed by a constant increase from 60°C to 95°C, was included following the PCR cycles. Primer Premier 6.0 software were used to designed the specific primers of *HanSAMS* genes according to their gene sequences, listed in Supplementary Table S1. Results were analyzed by the $2^{-\triangle\triangle Ct}$ method using the *HanActin* as the endogenous reference gene (He et al., 2019).

3 Results

3.1 Pangenome-wide identification of *SAMS* gene family in sunflowers

A total of 58 SAMS genes are identified based on the nine sunflowers genomes, including 7 *XRQ-HanSAMS*, 6 *HA89-HanSAMS*, 7 *HA300-HanSAMS*, 6 *IR-HanSAMS*, 7 *LR1-HanSAMS*, 6 *PI659440-HanSAMS*, 6 *PSC8-HanSAMS*, 6 *OQP8-HanSAMS*, and 7 *RHA438-HanSAMS* (Table 1). The physicochemical properties of the *SAMS* genes were presented in Table 1. Their protein sequence length ranged from 390 to 391 aa, with a molecular weight (MW) varying from 42583.22 to 43012.92 Da. The isoelectric points (pI) of the protein ranged from 5.47 to 5.97. The grand average of hydropathicity (GRAVY) of the proteins ranged from -0.291 to -0.357, all were the hydrophobic proteins. Secondary structure prediction analysis revealed that the proteins encoded by all the genes were predominantly composed of α-helices, β-turns, random coils, and extended chains (Supplementary Table S2). Among these, random coils were the most abundant structural element, accounting for 40.26% to 45.9% of the secondary structure. α-helices were the next most common, representing a proportion of 30.51% to 37.69%. Extended chains followed with a composition of 13.85% to 16.92%. The least prevalent structure was β-turns, which constituted only 6.92% to 8.97% of the total secondary structure content. Tertiary structure prediction showed that seven SAMS proteins from the reference genome XRQ were matching prediction s-adenosylmethionine synthase 2, which including 2 diphosphomethylphosphonic acid adenosyl ester and 2 potassium ion (Supplementary Figure S1).

3.2 Phylogenetic and evolution analysis of *HanSAMS* gene

The phylogenetic analysis of nine cultivated sunflower SAMS proteins were performed to examine their relationships. Based on the constructed phylogenetic tree, the *SAMS* genes could be classified into two major clades with seven groups (SAMS1-SAMS7) (Figure 1). Clade I, which is specific to sunflower *SAMS* genes, was found to branch into three main divisions, with each division containing a pair of distinct *SAMS* genes: SAMS1 with SAMS5, SAMS6 with SAMS2, and SAMS4 with SAMS7 (Figure 1). It is proposed that the WGT-1 event around 38-50 million years ago (Badouin et al., 2017) was possibly responsible for generating three homologs within the sunflower *SAMS* gene family, establishing the three principal branches. Subsequently, the WGD-2 event, occurring approximately 29 million years ago (Badouin et al., 2017), is believed to have caused the duplication of each branch, resulting in two copies per branch and shaping the present structure of the clade I gene family. Notably, the clade II only consists of the SAMS3 group and is uniquely distributed on a separate branch and forms a striking cluster with three homologs from *A. thaliana*. This

10.3389/fpls.2024.1499024

TABLE 1 The information of the identified HanSAMS gene family in nine sunflowers.

Gene Name	Gene ID	Chr	Start	End	Number of amino acids (aa)	Molecular weight (Da)	Theoretical pI	Grand average of hydropathicity (GRAVY)
XRQ-HanSAMS1	HanXRQr2_Chr01g0040721	Chr01	141178594	141179766	390	42968.88	5.97	-0.346
XRQ-HanSAMS2	HanXRQr2_Chr02g0076781	Chr02	151379397	151381775	390	42667.34	5.67	-0.337
XRQ-HanSAMS3	HanXRQr2_Chr05g0218761	Chr05	123974636	123977188	391	42737.51	5.65	-0.291
XRQ-HanSAMS4	HanXRQr2_Chr07g0301771	Chr07	118073833	118076445	390	42637.31	5.58	-0.319
XRQ-HanSAMS5	HanXRQr2_Chr11g0515381	Chr11	179525398	179526763	390	42768.54	5.86	-0.341
XRQ-HanSAMS6	HanXRQr2_Chr13g0586041	Chr13	83908346	83910875	390	42583.22	5.73	-0.316
XRQ-HanSAMS7	HanXRQr2_Chr14g0659171	Chr14	154860282	154862804	390	42640.32	5.58	-0.303
HA89-HanSAMS1	HanHA89Chr01g0035721	Chr01	141952252	141953424	390	42968.88	5.97	-0.346
HA89-HanSAMS2	HanHA89Chr02g0072371	Chr02	151050725	151053081	390	42667.34	5.67	-0.337
HA89-HanSAMS3	HanHA89Chr05g0193641	Chr05	123509230	123511775	391	42737.51	5.65	-0.291
HA89-HanSAMS4	HanHA89Chr07g0265261	Chr07	118227621	118230059	390	42637.31	5.58	-0.319
HA89-HanSAMS6	HanHA89Chr13g0512461	Chr13	83922587	83925120	390	42583.22	5.73	-0.316
HA89-HanSAMS7	HanHA89Chr14g0584821	Chr14	156049129	156051559	390	42640.32	5.58	-0.303
HA300-HanSAMS1	HanHA300Chr01g0033191	Chr01	138801146	138802318	390	43012.92	5.97	-0.344
HA300-HanSAMS2	HanHA300Chr02g0063931	Chr02	148706034	148708390	390	42667.34	5.67	-0.337
HA300-HanSAMS3	HanHA300Chr05g0178901	Chr05	117579400	117581945	391	42737.51	5.65	-0.291
HA300-HanSAMS4	HanHA300Chr07g0248451	Chr07	115405490	115407929	390	42637.31	5.58	-0.319
HA300-HanSAMS5	HanHA300Chr11g0422801	Chr11	174480957	174482129	390	42768.54	5.86	-0.341
HA300-HanSAMS6	HanHA300Chr13g0480381	Chr13	82039590	82042123	390	42583.22	5.73	-0.316
HA300-HanSAMS7	HanHA300Chr14g0536951	Chr14	144837421	144839851	390	42640.32	5.58	-0.303
IR-HanSAMS2	HanIRChr02g0089631	Chr02	151201584	151206312	390	42667.34	5.67	-0.337
IR-HanSAMS3	HanIRChr05g0235181	Chr05	126120008	126122677	391	42737.51	5.65	-0.291
IR-HanSAMS4	HanIRChr07g0325441	Chr07	118275055	118277458	390	42637.31	5.58	-0.319
IR-HanSAMS5	HanIRChr11g0553961	Chr11	179753853	179758008	390	42768.54	5.86	-0.341
IR-HanSAMS6	HanIRChr13g0638071	Chr13	82142037	82144570	390	42583.22	5.73	-0.316
IR-HanSAMS7	HanIRChr14g0715241	Chr14	158625950	158628504	390	42640.32	5.58	-0.303
LR1-HanSAMS1.1	HanLR1Chr00c0365g0744971	–	50299	51471	390	42968.88	5.97	-0.346
LR1-HanSAMS1.2	HanLR1Chr00c0566g0760211	–	47071	48243	390	43012.92	5.97	-0.344
LR1-HanSAMS2	HanLR1Chr02g0066821	Chr02	151075783	151076955	390	42667.34	5.67	-0.337
LR1-HanSAMS4	HanLR1Chr07g0247581	Chr07	117507373	117508545	390	42637.31	5.58	-0.319
LR1-HanSAMS5	HanLR1Chr11g0424191	Chr11	179222692	179223864	390	42768.54	5.86	-0.341
LR1-HanSAMS6	HanLR1Chr13g0482441	Chr13	81444379	81445551	390	42594.16	5.47	-0.311
LR1-HanSAMS7	HanLR1Chr14g0547181	Chr14	157346616	157349011	390	42638.3	5.58	-0.318
OQP8-HanSAMS1	HanOQP8Chr01g0034171	Chr01	167373744	167374916	390	42952.82	5.97	-0.353
OQP8-HanSAMS2	HanOQP8Chr02g0077651	Chr02	166481855	166484211	390	42667.34	5.67	-0.337
OQP8-HanSAMS4	HanOQP8Chr07g0255101	Chr07	117648767	117651187	390	42637.31	5.58	-0.319
OQP8-HanSAMS5	HanOQP8Chr11g0424921	Chr11	177950410	177951582	390	42768.54	5.86	-0.341
OQP8-HanSAMS6	HanOQP8Chr13g0481321	Chr13	83037810	83040343	390	42583.22	5.73	-0.316

(Continued)

10.3389/fpls.2024.1499024

TABLE 1 Continued

Gene Name	Gene ID	Chr	Start	End	Number of amino acids (aa)	Molecular weight (Da)	Theoretical pI	Grand average of hydropathicity (GRAVY)
OQP8-HanSAMS7	HanOQP8Chr14g0544331	Chr14	155795294	155797687	390	42640.32	5.58	-0.303
PI659440-HanSAMS1	HanPI659440Chr00c05g0713751	–	1573106	1574278	390	42972.81	5.97	-0.357
PI659440-HanSAMS2	HanPI659440Chr02g0085791	Chr02	157900681	157906226	390	42667.34	5.67	-0.337
PI659440-HanSAMS3	HanPI659440Chr05g0204041	Chr05	125469449	125472052	391	42737.51	5.65	-0.291
PI659440-HanSAMS5	HanPI659440Chr11g0438411	Chr11	185595937	185599063	390	42768.54	5.86	-0.341
PI659440-HanSAMS6	HanPI659440Chr13g0489401	Chr13	28919635	28922154	390	42608.23	5.55	-0.312
PI659440-HanSAMS7	HanPI659440Chr14g0565851	Chr14	148426212	148428354	390	42640.32	5.58	-0.303
PSC8-HanSAMS1	HanPSC8Chr01g0039531	Chr01	147977048	147980576	390	42952.82	5.97	-0.353
PSC8-HanSAMS2	HanPSC8Chr02g0074491	Chr02	156100025	156102667	390	42667.34	5.67	-0.337
PSC8-HanSAMS3	HanPSC8Chr05g0211181	Chr05	124940843	124943410	391	42737.51	5.65	-0.291
PSC8-HanSAMS4	HanPSC8Chr07g0292091	Chr07	117662755	117665348	390	42637.31	5.58	-0.319
PSC8-HanSAMS5	HanPSC8Chr11g0496671	Chr11	179223032	179225152	390	42784.58	5.86	-0.327
PSC8-HanSAMS7	HanPSC8Chr14g0632221	Chr14	161860094	161862616	390	42640.32	5.58	-0.303
RHA438-HanSAMS1	HanRHA438Chr01g0041581	Chr01	144767362	144769020	390	42968.88	5.97	-0.346
RHA438-HanSAMS2	HanRHA438Chr02g0088081	Chr02	153746776	153749342	390	42667.34	5.67	-0.337
RHA438-HanSAMS3	HanRHA438Chr05g0227851	Chr05	123778849	123781427	391	42737.51	5.65	-0.291
RHA438-HanSAMS4	HanRHA438Chr07g0311691	Chr07	118251355	118254279	390	42637.31	5.58	-0.319
RHA438-HanSAMS5	HanRHA438Chr11g0527231	Chr11	178138655	178140891	390	42784.58	5.86	-0.327
RHA438-HanSAMS6	HanRHA438Chr13g0596681	Chr13	82615728	82618267	390	42583.22	5.73	-0.316
RHA438-HanSAMS7	HanRHA438Chr14g0670101	Chr14	155470650	155474584	390	42640.32	5.58	-0.303

finding suggests that the SAMS3 group may share a common ancient ancestor with *A. thaliana* and appears to have not undergone the most recent whole-genome duplication event, due to lacking the partnered SAMS group that are found clustered together in other SAMS groups. To explore the expansion mechanism of the *HanSAMS* gene family, we analyzed gene duplication events in sunflowers using the reference genome XRQ. We found that the *HanSAMS* genes are distributed across seven chromosomes, with one gene per chromosome (see Figure 2), and no tandem clusters were identified. Subsequently, we investigated the gene collinearity within sunflowers and identified 12 pairs of duplicated genes (Figure 2), suggesting that whole genome duplication (WGD) is the primary driver behind the expansion of the *HanSAMS* gene family. The Ka/Ks values were all lower than 1 for the duplicated genes (Table 2), indicated that the *SAMS* gene family in sunflower has predominantly experienced purifying selection. The interspecies collinearity analysis of the *HanSAMS* gene families among XRQ and other eight sunflowers was further performed, and it was found that there were 134 pairs of collinearity, including 19 pairs of HA89, 19 pairs of OQP8, 18 pairs of HA300, 18 pairs of IR, 18 pairs of RHA438, 16 pairs of PSC8, 14 pairs of LR1 and 12 pairs of PI659440 (Figure 3). The collinear relationship among HA89, OQP8 and XRQ genes is the strongest, followed by HA300, IR, RHA438, and the least in PI659440, which may reflect the divergency among nine different cultivated sunflowers.

3.3 Genes structure and subcellular localization analysis

Examination of gene structures revealed that all *SAMS* gene in nine sunflowers have only one exon and are devoid of introns (Figure 4A). The *HA300-HanSAMS1*, *HA300-HanSAMS5*, *HA89-HanSAMS1*, *LR1-HanSAMS1.1*, *LR1-HanSAMS1.2*, *LR1-HanSAMS2*, *LR1-HanSAMS4*, *LR1-HanSAMS5*, *LR1-HanSAMS6*, *OQP8-HanSAMS1*, *OQP8-HanSAMS5*, *PI659440-HanSAMS1*, and *XRQ-HanSAMS1* were all lacked 5' and 3' untranslated regions (UTRs). The subcellular localization prediction suggested that the majority of SAMS proteins are predominantly found in the cytoskeleton, with the exception of SAMS3, which is localized in the cytoplasm, as shown in Figure 4B. The motifs of SAMS protein sequences were predicted using the MEME server, and all members of the *SAMS* contain motif1-motif10 (Supplementary Figure S2A), indicating highly conserved between different SAMS and different cultivars. The motif2 and motif5 were s-adenosylmethionine synthase domain (central domain), motif1 was s-adenosylmethionine synthase domain (N-terminal domain), motif3 and motif4 were s-

10.3389/fpls.2024.1499024

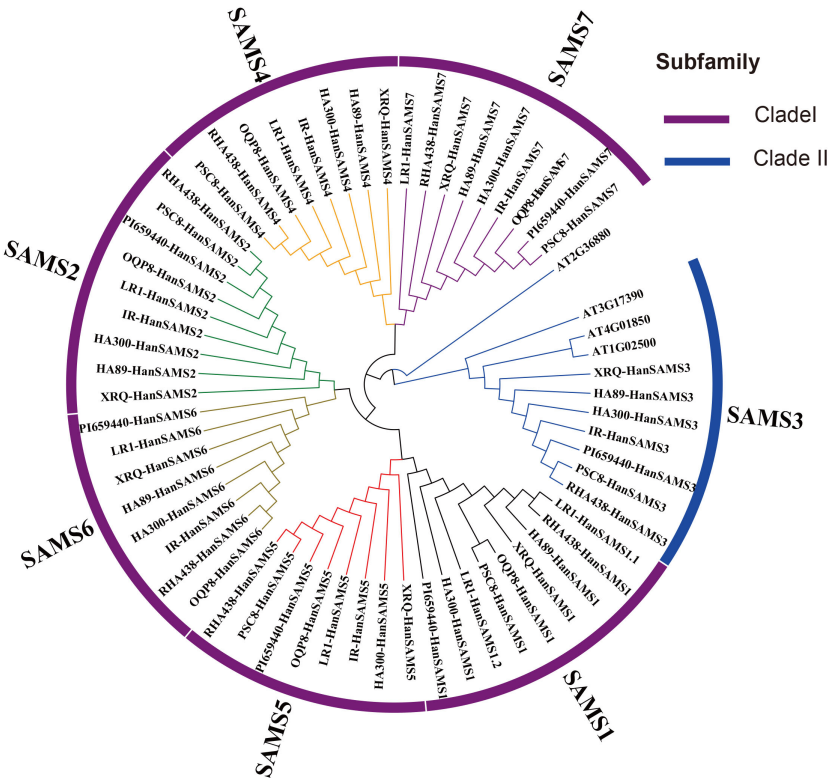


FIGURE 1
Identification of *HanSAMS* genes in nine cultivated sunflowers. Phylogenetic tree of nine cultivated sunflowers and Arabidopsis *SAMS* genes.

adenosylmethionine synthase domain (C-terminal domain), about 50 amino acid residues long and is considered a key element (Supplementary Figure S2B).

3.4 Analysis of the codon usage bias of *HanSAMS* genes

The codon usage bias (CUB) of *SAMS* gene family in nine sunflowers species was investigated by analyzing the GC, GC1, GC2, and GC3 content (Supplementary Table S3). The GC content of the *HanSAMS* genes among the nine sunflower genomes ranged from 48.95% to 52.59%, with all group of *SAMS* and group SAMS3 having the lowest value 50% (Table 3). The GC1 content of all *HanSAMS* genes and the GC3 content of 91.38% of *HanSAMS* genes across nine sunflower species exceeded 50%, while the GC2 content remained below 50%. This suggests a notable variation in base composition at different positions and a pronounced bias towards G/C-rich start and stop codons. Although CUB across all *HanSAMS* genes was generally weak, as indicated by ENc values ranging from 41.76 to 53.62, there were variations among different SAMS groups. Specifically, HanSAMS2, HanSAMS4, and HanSAMS7 exhibited lower ENc values (<50%), suggesting a stronger preference for certain codons compared to HanSAMS groups 1, 3, 5, and 6 (see Supplementary Table S3 for details). The ENc-plots of *HanSAMS* genes exhibit deviations from the

TABLE 2 Ka Ks analysis of *HanSAMS* duplicated genes in XRQ genome.

Seq 1	Seq 2	Ka	Ks	Ka/Ks
XRQ-HanSAMS1	*XRQ-HanSAMS2*	0.044	1.530	0.029
XRQ-HanSAMS1	*XRQ-HanSAMS6*	0.052	1.196	0.043
XRQ-HanSAMS2	*XRQ-HanSAMS6*	0.013	0.625	0.021
XRQ-HanSAMS2	*XRQ-HanSAMS3*	0.085	2.911	0.029
XRQ-HanSAMS4	*XRQ-HanSAMS7*	0.009	0.504	0.018
XRQ-HanSAMS4	*XRQ-HanSAMS5*	0.034	1.088	0.031
XRQ-HanSAMS4	*XRQ-HanSAMS1*	0.051	2.056	0.025
XRQ-HanSAMS4	*XRQ-HanSAMS2*	0.033	1.448	0.023
XRQ-HanSAMS4	*XRQ-HanSAMS6*	0.038	1.724	0.022
XRQ-HanSAMS4	*XRQ-HanSAMS3*	0.069	4.345	0.016
XRQ-HanSAMS5	*XRQ-HanSAMS1*	0.031	0.817	0.038
XRQ-HanSAMS5	*XRQ-HanSAMS2*	0.027	1.233	0.022
XRQ-HanSAMS5	*XRQ-HanSAMS6*	0.030	1.166	0.025
XRQ-HanSAMS6	*XRQ-HanSAMS3*	0.095	2.458	0.039
XRQ-HanSAMS7	*XRQ-HanSAMS5*	0.038	1.347	0.028
XRQ-HanSAMS7	*XRQ-HanSAMS1*	0.052	2.098	0.025

(Continued)

TABLE 2 Continued

Seq 1	Seq 2	Ka	Ks	Ka/Ks
XRQ-HanSAMS7	*XRQ-HanSAMS2*	0.033	1.303	0.025
XRQ-HanSAMS7	*XRQ-HanSAMS6*	0.038	1.486	0.025
XRQ-HanSAMS7	*XRQ-HanSAMS3*	0.073	1.725	0.042
Average		0.044	1.635	0.028

expected curve, suggesting that natural selection predominantly influences CUB (Supplementary Figure S3).

The neutrality curve analysis of GC12 and GC3 values of nine sunflowers *HanSAMS* gene family revealed positive correlation between GC12 and GC3, with R values ranging from 0.11(IR-HanSAMS) to 0.72(HA89-HanSAMS) and the regression coefficients varying from 0.0576(IR-HanSAMS) to 0.18(HA89-HanSAMS), indicated that CUB of *HansSAMS* genes was mainly affected by natural selection (Supplementary Figure S4). The PR2-plot analysis reveals the distribution of the third base at the codon. The results show an uneven distribution of scatters across the four regions (Figure 5). Scatters in the top and bottom are predominantly in the lower half, indicating a preference for T at the third position. Those on the left and right are mostly in the left half, indicating a preference for C at the third position. Comparison among the quadrants shows the highest number of scatters in the quadrant three, suggesting a preference for C/T at the third position, implying that natural selection is the primary factor leading to CUB.

RSCU (relative synonymous codon usage) is a pivotal metric that quantifies CUB by comparing the observed frequency of each synonymous codon to its expected frequency under equal usage. The RSCU values of the *HanSAMS* genes were calculated and the results showed there were 22 codons shared by all nine sunflower cultivars with RSCU values greater than 1, of which 15 codons end with C/G (Supplementary Table S4; Figure 6). Conversely, low-frequency codons, which end in A/U, were also prevalent (22 of 35), indicating a bias for these codons in the gene family. The top 3 codons with the largest average RSCU value were encode Arg (AGG with RSCU 2.36), Gly (GGU with RSCU 2.29) and Leu (CUU with RSCU 2.00). The RSCU value varied among cultivars but were generally similar, suggesting a consistent pattern of codon usage across the *SAMS* gene family.

3.5 *Cis*-element analysis

Promoter *cis*-acting elements are crucial for regulating gene expression. We utilized PlantCARE to identify *cis*-acting elements in the promoter regions of 58 *HanSAMS* genes (Supplementary

FIGURE 2
Chromosomal distribution and collinear relationships of the *HanSAMS* family. The collinearity of all genes within the sunflower is connected by gray background lines, and the collinearity where the SAMS gene is located is marked with yellow lines.

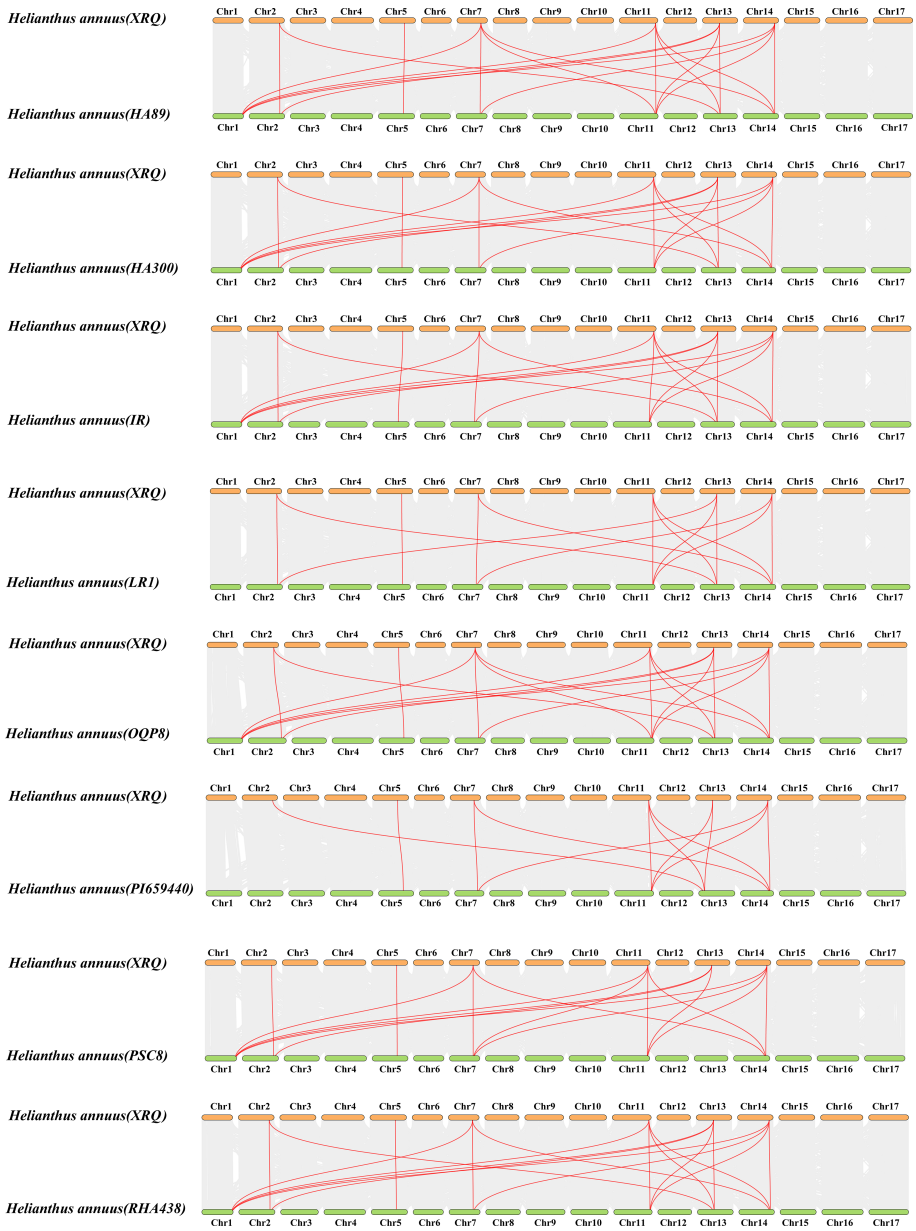


FIGURE 3
Collinearity analysis of *HanSAMS* gene families between *XRQ* and other eight cultivated sunflowers. gray lines indicate all synteny blocks in the sunflower genome, and the red lines indicate duplicated *SAMS* gene pairs, the chromosome number is indicated at the top or bottom of each chromosome.

Table S5, Supplementary Figure S5). Our statistical analysis showed these elements involved in various plant processes, including growth, development, hormone response, light response, and stress response (Figure 7A). Notably, stress response elements predominated in the *HanSAMS* promoters (Figure 7A). In detail analysis, 302 MYB binding site, 191 anaerobic induction (ARE element),186 MYC binding site and 149 stress response element (STRE element) were predicted with high frequency in the promoter region of *HanSAMS* genes (Figure 7B; Supplementary Table S5). Additionally, 624 light response-related elements were identified, such as MRE (n=105), GT1-motif (n=103), G-box (n=94), Box 4 (n=87) (Supplementary Table S5). Hormone response-related *cis*-regulatory elements were also observed, such as 108 salicylic acid responsiveness (TCA, as-1), 92 abscisic acid responsiveness (ABRE), 78 gibberellin-responsiveness, 74 MeJA-responsiveness (CGTCA-motif, TGACG-motif), and 41 ethylene response elements (ERE) (Figure 7B; Supplementary Table S5). We also noted that while most *SAMS* genes within the same group shared similar element distributions across different cultivars, some cultivars exhibited distinct differences (see marked with black boxes in Figure 7B). For instance, the *SAMS3* gene in the IR cultivar contained 12 salicylic acid response elements, which is significantly higher than other cultivars by at least three folds. These differences may be linked to the cultivars' adaptability to environmental

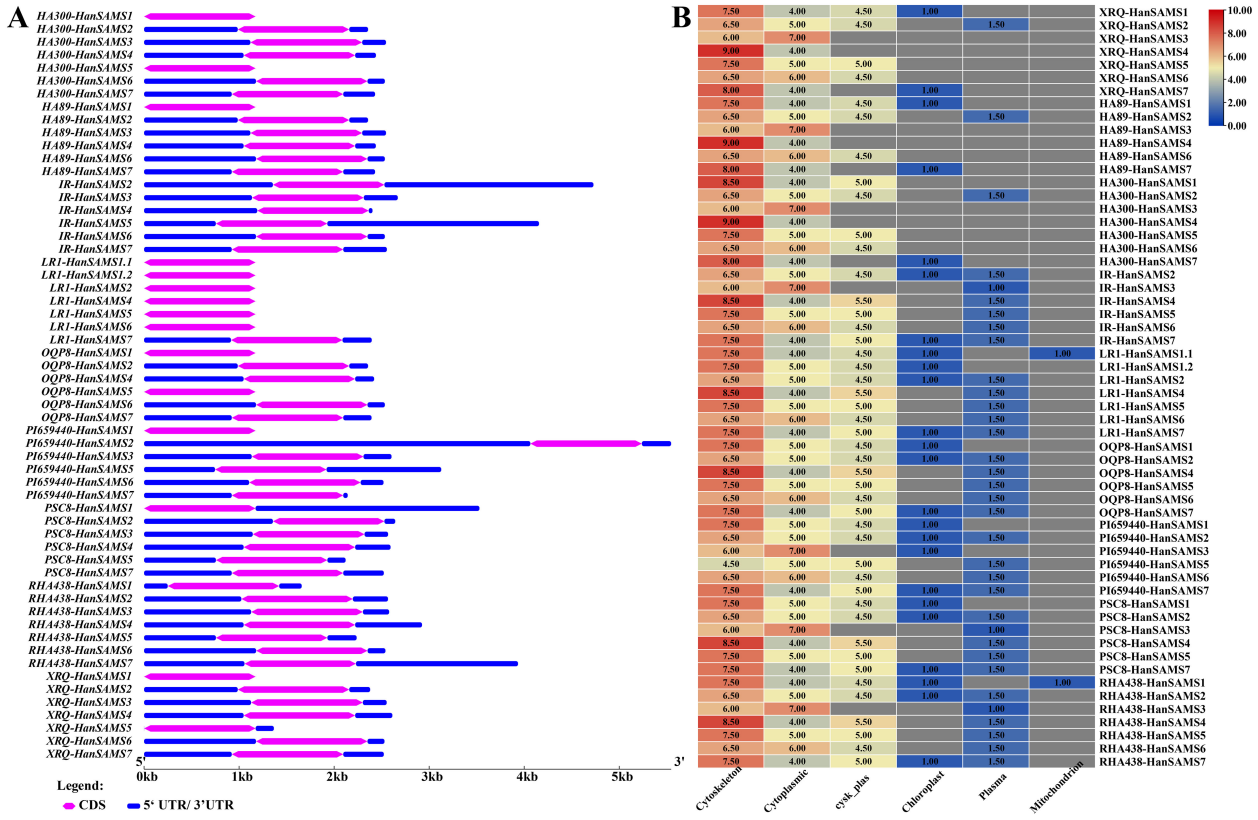


FIGURE 4
Gene structure and subcellular localization of *HanSAMS* genes in nine cultivated sunflowers. **(A)** Comparison of the gene structures of *SAMS* genes in nine cultivated sunflowers. **(B)** Comparison of subcellular localization of *HanSAMS* genes in nine cultivated sunflowers. The values in heatmap represents sorting signals for each candidate locations. Cysk_Plas, cytoskeleton and plasma membrane.

stresses and functional selection during evolution. Collectively, our findings suggest that *SAMS* genes are likely broadly involved in the regulation of hormones and stress responses.

3.6 Expression patterns of *HanSAMS* genes in different tissues

HanSAMS genes may have different functions in the growth and development of sunflowers. To determine the spatial expression pattern of *HanSAMS* genes in sunflowers, we measured the expression levels of seven *XQR-HanSAMS* genes from three tissues (roots, stems, leaves) using qRT-PCR. As shown in Figure 8, seven *HanSAMS* genes were expressed in all the tissues. Among them, *HanSAMS1*, *HanSAMS3* and *HanSAMS5* had similar expression patterns and were expressed highest in leaves. Meanwhile, the expression of *HanSAMS4* and *HanSAMS7* was higher in stems than in the other three tissues. *HanSAMS2* and *HanSAMS6*, exhibited relatively high expression levels in the root. The results suggested that these genes showed a tissue-specific expression pattern and may play different roles in the growth and development of sunflowers.

TABLE 3 Average GC content and ENC values of *HanSAMS* genes in nine sunflowers.

Group	GC	GC1s	GC2s	GC3s	GC12	ENC
HanSAMS1	49.60	56.98	40.21	51.59	48.60	53.09
HanSAMS2	51.65	58.70	41.33	54.93	50.01	48.88
HanSAMS3	49.05	57.18	40.00	49.96	48.59	51.74
HanSAMS4	51.19	57.91	40.54	55.12	49.23	45.19
HanSAMS5	52.48	57.65	40.74	59.06	49.19	51.30
HanSAMS6	51.13	57.52	41.33	54.56	49.43	52.53
HanSAMS7	51.05	58.21	40.04	54.93	49.12	42.23

3.7 Expression analysis of *HanSAMS* genes under different hormonal treatment

Promoter analysis revealed there are many hormonal response elements (Figure 7A) the RNA-seq data available for cultivated sunflower (XRQ) were examined and to elucidate the expression patterns of *SAMS* genes in response to hormones. A responsive pattern was observed across all *HanSAMS* genes upon IAA treatment in both leaves and roots (Figure 9), suggesting a pronounced sensitivity to auxin signaling (Song et al., 2024a). In leaf tissues, the *HanSAMS* gene family—excluding *HanSAMS1*—demonstrated a significant upregulation in expression following

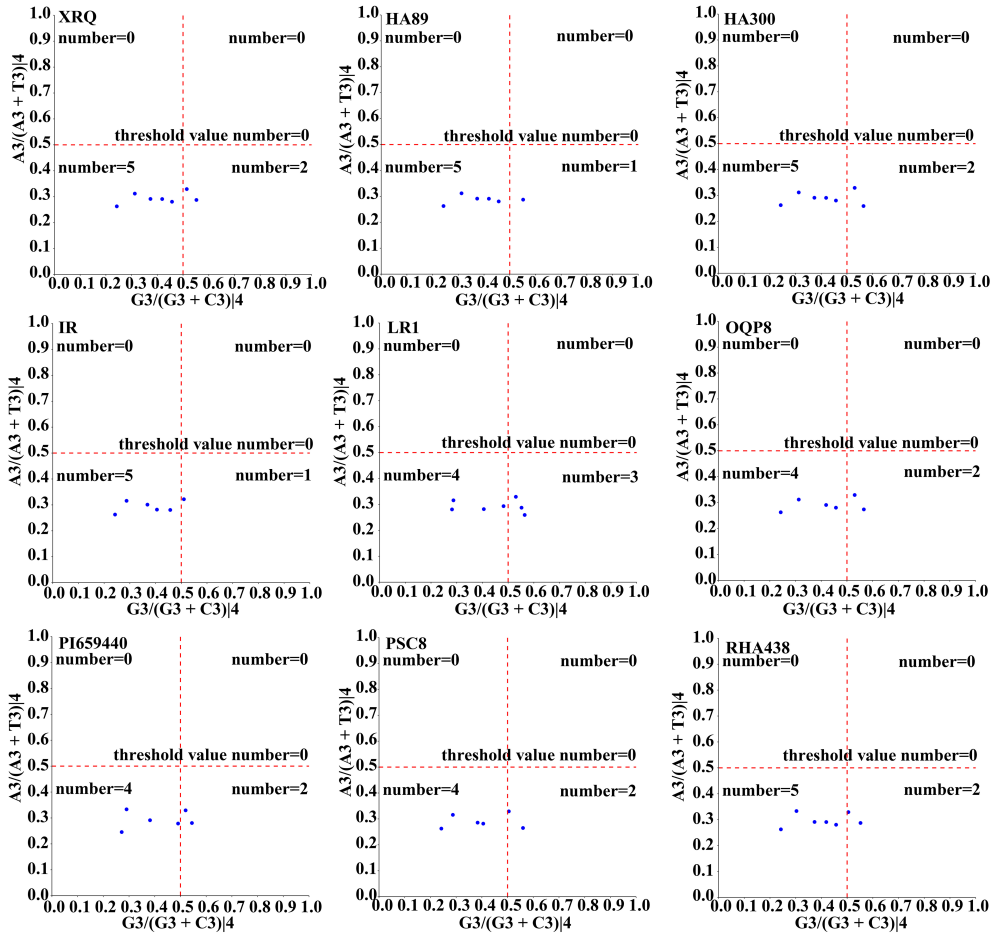


FIGURE 5
PR2-plot analysis of the *HanSAMS* gene family in nine sunflowers. A3/(A3+T3)|4 and G3/(G3+C3)|4 represents the four-codon degenerate amino acids.

BRA (brassinosteroids) treatment, as depicted in Figure 9A. This underscores their crucial role in the regulatory pathways activated by BRA. However, in root tissues, all *HanSAMS* genes under BRA treatment conditions did not exhibit a significant increase compared to the control samples (Figure 9B). For other different hormone treatments, we also observed the different expression patterns in different tissues. For instance, *HanSAMS3* demonstrates its highest expression levels in leaves following MeJA treatment (Figure 9A), whereas in roots, the ABA treatment elicits its peak expression (Figure 9B). These results indicate that the *HanSAMS* gene expression is modulated in a tissue-specific manner in response to hormonal signals.

3.8 Expression patterns of *HanSAMS* genes under drought and salt stresse

Considering that the *cis*-elements responding to various stress existed in the promoter sequences of *HanSAMS* genes (Figure 7A), we conducted a quantitative analysis of the *HanSAMS* gene using qRT-PCR to examine their expression profiles under drought and salt stresses (Figure 10). Our findings revealed that the *HanSAMS* genes exhibited distinct expression patterns at various time intervals (0 h, 1 h, 3 h, 6 h, 12 h, 24 h) following exposure to drought and salt stresses. All of the *HanSAMS* genes showed increased expression levels at different times under stresses, and some differences were extremely significant when compared with the untreated group (CK, 0h). In the case of drought treatment, four of the seven *HanSAM* genes, including *HanSAMS3*, *HanSAMS4*, *HanSAMS5* and *HanSAMS6*, showed the highest upregulation at 12th hour, while *HanSAMS1* showed the highest upregulation at the 6th hour. Notably, we observed that the expression level of the *HanSAMS5* gene under drought stress is the highest among all *SAMS* genes (exceeding 55-fold at the 12th hour). In the case of salt treatment, five of the seven *HanSAM* genes, including *HanSAMS2*, *HanSAMS4*, *HanSAMS5*, *HanSAMS6* and *HanSAMS7*, showed the highest upregulation at the 3th hours and gradually downregulated thereafter, while *HanSAMS3* was up-regulated to highest point at 6 th of treatment. Notably, we also observed that the expression level of the *HanSAMS5* gene under salt stress is the highest among all *SAMS* genes (exceeding 10-fold at the 3th hour). In summary, most of *HanSAMS* genes exhibit responsiveness to

10.3389/fpls.2024.1499024

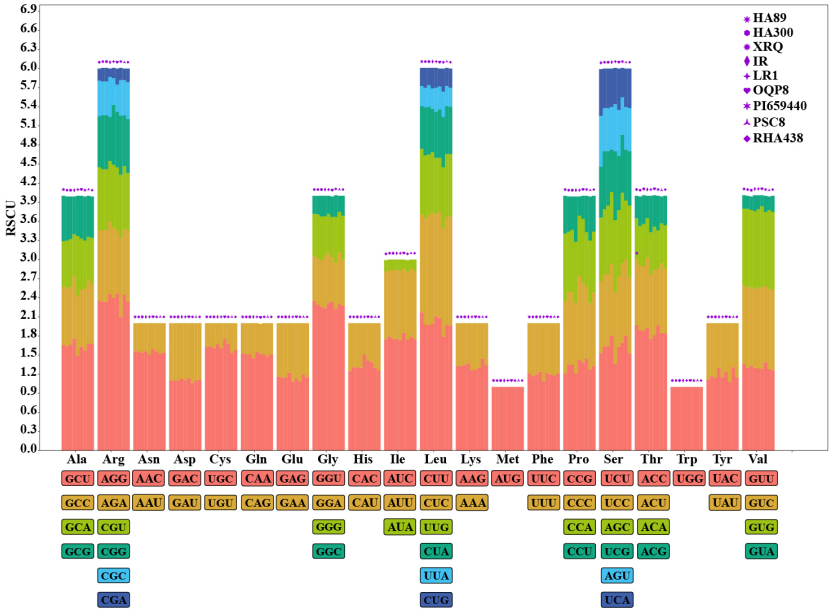


FIGURE 6
RSCU analysis of the *HanSAMS* gene family in nine cultivated sunflower species.

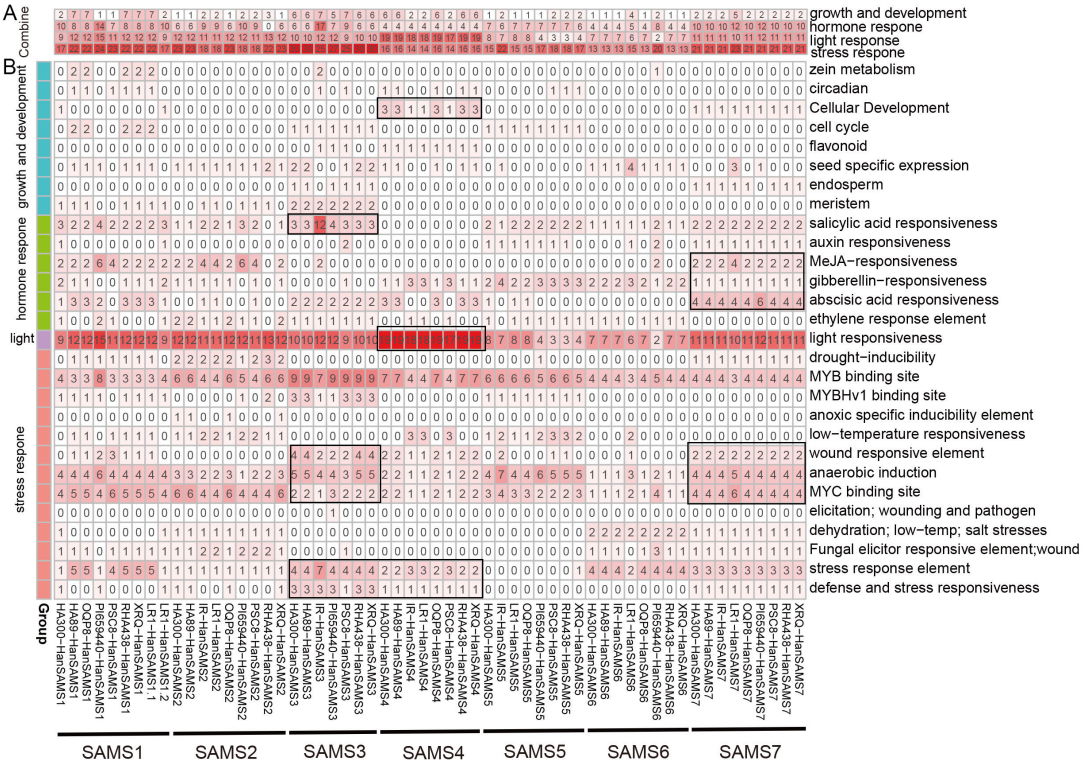


FIGURE 7
Analysis of the cis-element of *HanSAMS* genes. **(A)** Classification of cis-elements of *HanSAMS* promoters into four main groups: growth and development, hormone response, light responses and stress response. **(B)** Detail analysis of cis-elements in four groups for each *HanSAMS* gene promoter. The color intensity and number in each square indicate the number of each type of cis-element in the promoter region of the indicated gene. The distribution patterns of genes within the SAMS group that we are mentioned are marked with black boxes.

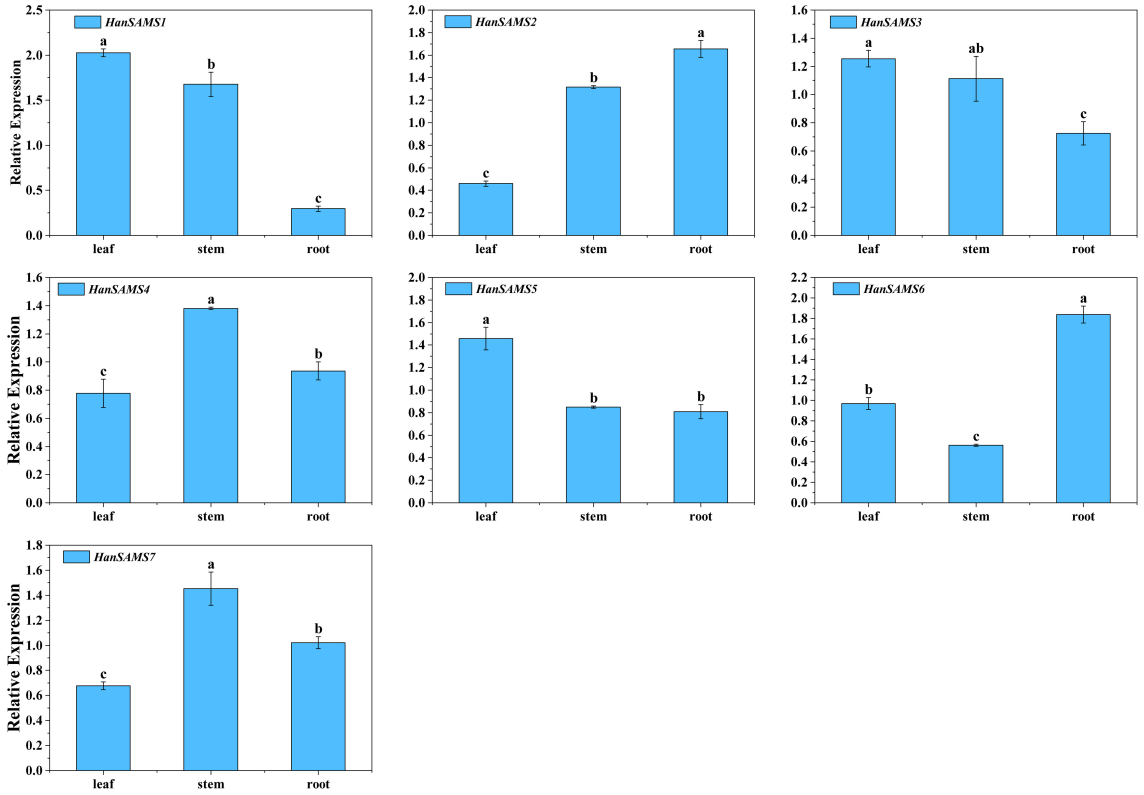


FIGURE 8
Expression profiles of *HanSAMS* genes in leaf, stem and root. a, b, c, bar indicates a significant difference among the different tissues (Significant differences were determined using the Duncan's method of univariate ANOVA with a significance level of $P < 0.05$).

both salt and drought stress treatments, with a more rapid response observed for salt stress (peak at 3th hour) compared to drought stress (peak at 12th hour).

4 Discussion

S-adenosylmethionine (SAM) is produced through the catalysis of methionine and adenosine triphosphate (ATP) by the enzyme S-adenosylmethionine synthetase (*SAMS*) (Fontecave et al., 2004). *SAM* genes play a crucial role in various cellular pathways, including those associated with ethylene and polyamine biosynthesis, methionine metabolism, as well as transmethylation and transsulfuration processes (Chen et al., 2016; Sauter et al., 2013). In the current study, *SAMS* genes have been analyzed by an extensive use of bioinformatics, such as *Arabidopsis* (4), *rice* (3), *tomato* (4), *Eggplant* (4), *Triticum urartu* (3), *Barley* (4), *Sorghum* (3), *Medicago truncatula* (5), *Soybean* (9) (Heidari et al., 2020). In this study, 7 *XRQ-HanSAMS*, 6 *HA89-HanSAMS*, 7 *HA300-HanSAMS*, 6 *IR-HanSAMS*, 7 *LR1-HanSAMS*, 6 *PI659440-HanSAMS*, 6 *PSC8-HanSAMS*, 6 *OQP8-HanSAMS*, and 7 *RHA438-HanSAMS* genes were identified in nine sunflower genomes, respectively. The phylogenetic analysis of the *HanSAMS* genes were performed to examine their relationships, the results indicated that they could be divided into seven groups (SAMS1-SAMS7) (Figure 1). It is not the case that every group encompasses all species. Only SAMS1 and SAMS2 are present in all nine varieties, while SAMS3 is found in only seven, suggesting the genetic diversity among different cultivars. The gene structure analysis revealed that all the 58 *HanSAMS* genes were intron-less and contain only one exon (Figure 4A), which is consistent with the results of previous studies in other species (Sun et al., 2022; Kilwake et al., 2023). Furthermore, the *cis*-elements analysis in the promoter region of the *HanSAMS* genes indicated that they might be primarily involved in the plant hormonal signals, light, and abiotic stresses responsiveness (Figure 7), this is similar to the findings of *cis*-elements in plants such as *Arabidopsis* and *Triticum aestivum* (Cheng et al., 2012; Shen et al., 2003). Our findings also suggest that one or two cultivars show a number notably different with other cultivars while most of *HanSAMS* genes within the same group share similar type and number of regulatory elements across cultivars, The results of the Ka/Ks analysis indicate that the *SAMS* gene family in nine sunflowers has predominantly experienced purifying selection throughout its evolutionary history.

Codon bias plays a complex role in the formation of gene mutation and the results of selection, but it is also important for the structure, function and expression of genes encoding proteins that are closely linked, and affects evolution (Chen et al., 2004; Hartl et al., 1994; Hershberg and Petrov, 2008). The codon usage bias of sunflower *SAMS* gene families in nine cultivated species, ENc-plot, PR2-plot and neutrality curve analysis indicated that codon usage bias formation of sunflower *SAMS* gene families may be the result of

10.3389/fpls.2024.1499024

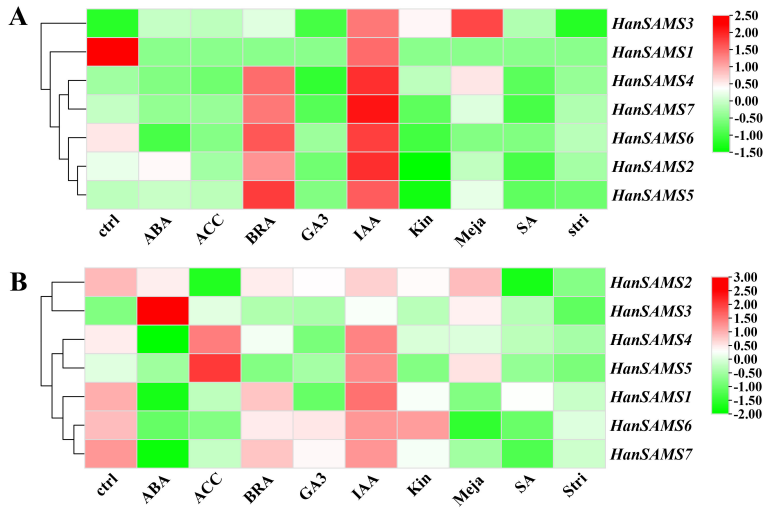


FIGURE 9
The expression of *HanSAMS* in different tissues and under different abiotic stresses. **(A)** *HanSAMS* gene expression in leaves under exogenous hormone treatment (SRP092742); **(B)** HanSAMS gene expression in roots under exogenous hormone treatment (SRP092742). ctrl, control; ABA, abscisic acid; ACC, Ethylene; BRA, Brassinosteroids; GA3, Gibberellic Acid 3; IAA, Indole Acetic Acid; Kin, Kinetin; Meja, Methyl-Jasmonate; SA, Salicylic acid; Stri, Strigolactone.

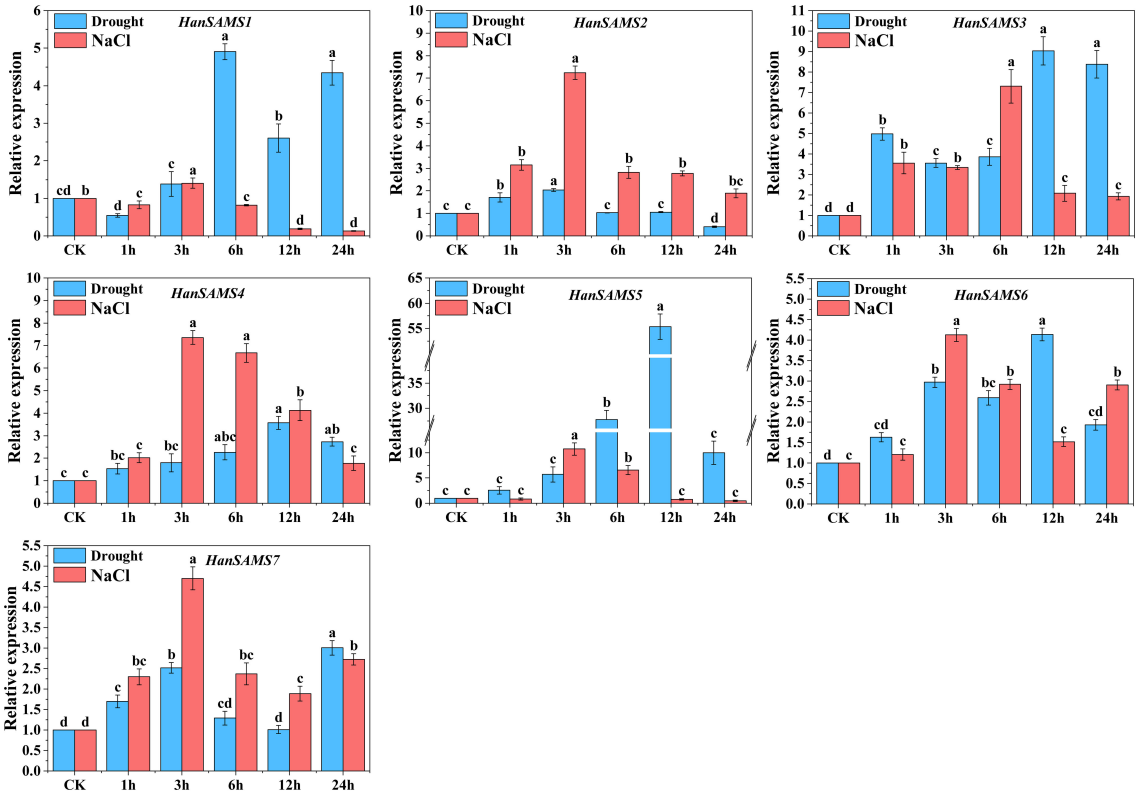


FIGURE 10
Expression patterns of *HanSAMS* genes under drought and NaCl stress treatments. a, b, c, d bar indicates a significant difference between the experimental treatments and control (CK) treatment (Significant differences were determined using the Duncan's method of univariate ANOVA with a significance level of $P < 0.05$).

base mutations, natural selection and other factors. Through RSCU analysis, it was found that high frequency codons in sunflower *SAMS* gene families of nine cultivated species preferred G/C ending, and the codon with the largest RSCU value encodes Leucine (Leu, CUU), Glycine (Gly, GGU), and Arginine (Arg, AGG). The codon bias of the plant genome can be analyzed and studied by a correlation index, and the frequency of codon usage between species at the order and family level is different; thus, the genetic relationship between species can be analyzed by a correlation index (Puigbo et al., 2008; Chen et al., 2014; Wang et al., 2011).

Previous research has indicated that *SAMS* genes are often activated by a variety of hormonal treatments and abiotic stresses. For instance, *AtSAMS3* and *AtSAMS4* are upregulated under biotic stress and brassinosteroid (BR) treatment, but downregulated in response to abiotic stresses such as salt, heat, and temperature stress, as well as ABA application (Heidari et al., 2020). In this study, we discovered that *HanSAMS* genes exhibit high expression levels in sunflower leaves when subjected to brassinolide (BRA) and indole-3-acetic acid (IAA) treatments, as revealed by previously published RNA-seq data (Figure 9). BRs are emerging as a plant hormone of significant importance due to their role in stress responses, including extreme temperatures and drought (Brewer et al., 2013; Ha et al., 2014; Nolan et al., 2020). Recent studies have shown that the overexpression of BRL3, a vascular BR receptor, enhances drought responses without hindering growth in *Arabidopsis* (Fàbregas et al., 2018). ABA and IAA are known as a hormone responsive to abiotic stresses such as drought, heat, low temperature, radiation and salt stress (Vishwakarma et al., 2017). However, in root tissues, all *HanSAMS* genes exhibiting reduced expression under BRA and IAA treatment conditions (Figure 9), indicating that the *HanSAMS* gene expression is modulated in a tissue-specific manner in response to hormonal signals.

In our promoter analysis, a multitude of MYB-related elements were identified in *HanSAMS* genes (Figure 7), Several studies have highlighted MYB as a crucial transcription factor associated with plant drought resistance and a key player in the transcriptional regulatory network governing plant responses to drought and salt stress (Espartero et al., 1994; Baldoni and Genga, 2015; Leng and Zhao, 2020; Sun et al., 2022). The overexpression of the *SAMS* gene from *Lycoris radiata* in *E. coli* has been shown to enhance plant tolerance to salt stress (Li et al., 2013). Given the presence of drought and salt responsive *cis*-elements in the promoter regions of *HanSAMS* genes, we performed a qRT-PCR analysis to assess all 7 *XQR-HanSAMS* expression dynamics under drought and salt stress conditions (Figure 10). Our results indicated that *HanSAMS* genes displayed unique expression profiles at different time points following stress exposure. Notably, *HanSAMS5* showed the most significant upregulation under both stress types, with over 55-fold increase at the 12th hour for drought and over 10-fold at the 3th hour for salt stress. This suggests that *HanSAMS5* may play a crucial role in the plant's response to adverse environmental conditions.

5 Conclusion

The study identified 58 *HanSAMS* genes in nine sunflowers through whole-genome bioinformatics analysis. The identified *HanSAMS* genes are distributed across seven chromosomes, exhibiting a conserved exon-intron structure devoid of introns. Phylogenetic analysis has uncovered that the sunflower SAMS genes have expanded due to recent WGT-1 and WGD-2 events, resulting in three homologous branches, each comprising two discrete SAMS groups. The analysis of codon usage bias revealed a pronounced preference for high-frequency codons ending in G or C, notably those encoding glycine, leucine, and arginine, highlighting the significant role of natural selection in shaping the evolution of the *HanSAMS* genes. A detailed promoter analysis revealed a wealth of stress-responsive *cis*-elements, suggesting their regulatory roles in stress tolerance. Moreover, expression profiling under hormonal stimuli and abiotic stresses, especially the marked upregulation of *HanSAMS5*, points to its pivotal role in managing multiple abiotic stresses. Collectively, these findings provide valuable insights into the functional diversity and evolutionary dynamics of the *SAMS* genes in sunflowers, laying a robust foundation for future research aimed at enhancing sunflower stress resilience through genetic improvement strategies.

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

CZ: Visualization, Data curation, Writing – review & editing, Writing – original draft. HL: Validation, Investigation, Writing – review & editing. JY: Visualization, Data curation, Writing – review & editing. ZH: Writing – original draft, Methodology, Formal analysis. XL: Software, Investigation, Writing – original draft. YC: Writing – review & editing, Validation, Supervision, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Natural Science Foundation of Inner Mongolia Autonomous Region (2023LHMS03033), Hetao College Talent Fund (HYRC202319), the Science and Technology Research Project of Hetao College (HYHB202302), and the Science and Technology Innovation Research Team Project of Hetao College (HTKCT-A202405).

Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. We also thanks Beijing OmicsGang Technologies Co., LTD. for bioinformatics training and help for this 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1499024/full#supplementary-material

References

Ahuja, I., de Vos, R. C. H., Bones, A. M., and Hall, R. D. (2010). Plant molecular stress responses face climate change. *Trends Plant Science.* 15, 664–674. doi: 10.1016/j.tplants.2010.08.002

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380

Baldoni, E., and Genga, A. (2015). Plant MYB transcription factors: Their role in drought response mechanisms. *Int. J. Mol. Sci.* 16, 15811–15851. doi: 10.3390/ijms160715811

Brewer, P. B., Koltai, H., and Beveridge, C. A. (2013). Diverse roles of strigolactones in plant development. *Mol. Plant* 6, 18–28. doi: 10.1093/mp/sss130

Bürstenbinder, K., Rzewuski, G., Wirtz, M., Hell, R., and Sauter, M. (2007). The role of methionine recycling for ethylene synthesis in *Arabidopsis*. *Plant J.* 49, 238–249. doi: 10.1111/j.1365-313x.2006.02942.x

Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., and McAdams, H. H. (2004). Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl. Acad. Sci. United States America.* 101, 3480–3485. doi: 10.1073/pnas.0307827100

Chen, H., Sun, S., Norenburg, J. L., and Sundberg, P. (2014). Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). *PloS One* 9, e85631. doi: 10.1371/journal.pone.0085631

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Chen, Y., Zou, T., and McCormick, S. (2016). S-adenosylmethionine synthetase 3 is important for pollen tube growth. *Plant Physiol.* 172, 244–253. doi: 10.1104/pp.16.00774

Cheng, M. C., Hsieh, E. J., Chen, J. H., Chen, H. Y., and Lin, T. P. (2012). *Arabidopsis* RGLG2, functioning as a RING E3 ligase, interacts with AtERF53 and negatively regulates the plant drought stress response. *Plant Physiol.* 158, 363–375. doi: 10.1104/pp.111.189738

DeLano, W. L. (2002). The PyMOL molecular graphics system. Proteins. *Structure Function Bioinf.* 30, 442–454.

Espartero, J., Pintor-Toro, J. A., and Pardo, J. M. (1994). Differential accumulation of S-adenosylmethionine synthetase transcripts in response to salt stress. *Plant Mol. Biol.* 25, 217–227. doi: 10.1007/BF00023239

Fàbregas, N., Lozano-Elena, F., Blasco-Escámez, D., Tohge, T., Martínez-Andújar, C., Albacete, A., et al. (2018). Overexpression of the vascular brassinosteroid receptor BRL3 confers drought resistance without penalizing plant growth. *Nat. Communication.* 9, 4680. doi: 10.1038/s41467-018-06861-3

Fontecave, M., Atta, M., and Mulliez, E. (2004). S-adenosylmethionine: Nothing goes to waste. *Trends Biochem. Sci.* 29, 243–249. doi: 10.1016/j.tibs.2004.03.007

Gao, L., Gonda, I., Sun, H. H., Ma, Q. Y., 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

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the expasy server," in *The Proteomics Protocols Handbook*. Ed. J. M. Walker (Totowa, NJ, USA: Humana Press Inc.), 571–607. doi: 10.1385/1-59259-890-0:571

Guo, Z., Tan, J., Zhuo, C., Wang, C., Xiang, B., and Wang, Z. (2014). Abscisic acid, H_2O_2 and nitric oxide interactions mediated cold-induced S-adenosylmethionine synthetase in Medicago sativa subsp. Falcata that confers cold tolerance through up-regulating polyamine oxidation. *Plant Biotechnol. J.* 12, 601–612. doi: 10.1111/pbi.12166

Gupta, K., Dey, A., and Gupta, B. (2013). Plant polyamines in abiotic stress responses. *Acta Physiologiae Plantarum.* 35, 2015–2036. doi: 10.1007/s11738-013-1239-4

Ha, C. V., Leyva-Gonzalez, M. A., Osakabe, Y., Tran, U. T., Nishiyama, R., Watanabe, Y., et al. (2014). Positive regulatory role of strigolactone in plant responses to drought and salt stress. *Proc. Natl. Acad. Sci. United States America.* 111, 851–856. doi: 10.1073/pnas.1322135111

Hartl, D. L., Moriyama, E. N., and Sawyer, S. A. (1994). Selection intensity for codon bias. *Genetics.* 138, 227–234. doi: 10.1093/genetics/138.1.227

He, M. W., Wang, Y., Wu, J. Q., Shu, S., Sun, J., and Guo, S. R. (2019). Isolation and characterization of S-Adenosylmethionine synthase gene from cucumber and responsive to abiotic stress. *Plant Physiol. Biochem.* 141, 431–445. doi: 10.1016/j.plaphy.2019.06.006

Heidari, P., Mazloomi, F., Nussbaumer, T., and Barcaccia, G. (2020). Insights into the SAM Synthetase Gene Family and Its Roles in Tomato Seedlings under Abiotic Stresses and Hormone Treatments. *Plants.* 9, 586. doi: 10.3390/plants9050586

Hershberg, R., and Petrov, D. A. (2008). Selection on codon bias. *Annu. Rev. Genet.* 42, 287–299. doi: 10.1146/annurev.genet.42.110807.091442

Hu, B., Jin, J. P., Guo, A. Y., Hang, H., Luo, J. C., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

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

Jang, S. J., Wi, S. J., Choi, Y. J., An, G., and Park, K. Y. (2012). Increased polyamine biosynthesis enhances stress tolerance by preventing the accumulation of reactive oxygen species: T-DNA mutational analysis of Oryza sativa lysine decarboxylase-like protein 1. *Molecules Cells* 34, 251–262. doi: 10.1007/s10059-012-0067-5

Keeley, S. C., Cantley, J. T., and Gallaher, T. J. (2021). The "evil tribe" spreads across the land: a dated molecular phylogeny provides insight into dispersal, expansion, and biogeographic relationships within one of the largest tribes of the sunflower family (Vernonieae: Compositae). *Am. J. Botany.* 108, 505–519. doi: 10.1002/ajb2.1614

Kilwake, J. W., Umer, M. J., Wei, Y. Y., Mehari, T. G., Magwanga, R. O., Xu, Y. C., et al. (2023). Genome-wide characterization of the SAMS gene family in cotton unveils the putative role of *ghSAMS2* in enhancing abiotic stress tolerance. *Agronomy* 13, 612. doi: 10.3390/agronomy13020612

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

Larkin, M. A., lackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics.* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Leng, P., and Zhao, J. (2020). Transcription factors as molecular switches to regulate drought adaptation in maize. *Theor. Appl. Genet.* 133, 1455–1465. doi: 10.1007/s00122-019-03494-y

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, 325–327. doi: 10.1093/nar/30.1.325

Li, R., Li, Y., Zheng, H., Luo, R., Zhu, H., Li, Q., et al. (2010). Building the sequence map of the human pan-genome. *Nat. Biotechnol.* 28, 57–63. doi: 10.1038/nbt.1596

Li, W., Han, Y., Tao, F., and Chong, K. (2011). Knockdown of SAMS genes encoding S-adenosyl-l-methionine synthetases causes methylation alterations of DNAs and histones and leads to late flowering in rice. *J. Plant Physiol.* 168, 1837–1843. doi: 10.1016/j.jplph.2011.05.020

Li, X. D., Xia, B., Wang, R., Xu, S., Jiang, Y. M., Yu, F. B., et al. (2013). Molecular cloning and characterization of S-adenosylmethionine synthetase gene from *Lycoris radiata*. *Mol. Biol. Rep.* 40, 1255–1263. doi: 10.1007/s11033-012-2168-9

Ma, C., Wang, Y., Gu, D., Nan, J., Chen, S., and Li, H. (2017). Overexpression of S-adenosyl-L-methionine synthetase 2 from sugar beet M14 increased arabidopsis tolerance to salt and oxidative stress. *Int. J. Mol. Sci.* 18, 847. doi: 10.3390/ijms18040847

Mantenese, A., Medan, D., and Hall, A. (2006). Achene structure, development and lipid accumulation in sunflower cultivars differing in oil content at maturity. *Ann. Botany.* 97, 999–1010. doi: 10.1093/aob/mcl046

Mehari, T. G., Xu, Y., Umer, M. J., Shiraku, M. L., Hou, Y., Wang, Y., et al. (2021). Multi-omics-based Identification and functional characterization of gh_A06G1257 proves its potential role in drought stress tolerance in gossypium hirsutum. *Front. Plant Science.* 12. doi: 10.3389/fpls.2021.746771

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

Nolan, T., Vukasinovi, N., Liu, D., Russinova, E., and Yina, Y. (2020). Brassinosteroids: multidimensional regulators of plant growth, development, and stress responses. *Plant Cell.* 32, 295–318. doi: 10.1105/tpc.19.00335

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, W200–W204. doi: 10.1093/nar/gky448

Puigbo, P., Bravo, I. G., and Garcia-Vallve, S. (2008). E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinf.* 9, 65. doi: 10.1186/1471-2105-9-65

Rele, A. S., and Mohile, R. B. (2003). Effect of mineral oil, sunflower oil, and coconut oil on prevention of hair damage. *J. Cosmetic Science.* 54, 175–192.

Roje, S. (2006). S-Adenosyl-l-methionine: Beyond the universal methyl group donor. *Phytochemistry.* 67, 1686–1698. doi: 10.1016/j.phytochem.2006.04.019

Sauter, M., Moffatt, B., Saechao, M. C., Hell, R., and Wirtz, M. (2013). Methionine salvage and S-adenosylmethionine: Essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* 451, 145–154. doi: 10.1042/BJ20121744

Shen, Y. G., Zhang, W. K., He, S. J., Zhang, J. S., Liu, Q., and Chen, S. Y. (2003). An EREBP/AP2-type protein in Triticum aestivum was a DRE-binding transcription factor induced by cold, dehydration and ABA stress. *Theor. Appl. Genet.* 106, 923–930. doi: 10.1007/s00122-002-1131-x

Song, H. F., Ji, X. C., Wang, M. Y., Li, J., Wang, X., Meng, L. Y., et al. (2024a). Genome-wide identification and expression analysis of the Dof gene family reveals their involvement in hormone response and abiotic stresses in sunflower (*Helianthus annuus* L.). *Gene.* 910, 148336. doi: 10.1016/j.gene.2024.148336

Song, H., Wang, M., Shen, J., Wang, X., Qin, C., Wei, P., et al. (2024b). Physiological and transcriptomic profiles reveal key regulatory pathways involved in cold resistance in sunflower seedlings. *Genomics* 116, 110926. doi: 10.1016/j.ygeno.2024.110926

Sun, F., Ma, J., Wang, P., and Yang, Y. (2022). Genome-wide Identification of the SAMS gene family in upland cotton (*Gossypium hirsutum* L.) and expression analysis in drought stress treatments. *Genes.* 13, 860. doi: 10.3390/genes13050860

Tao, Y. F., Luo, H., Xu, J. B., Cruickshank, A., Zhao, X. R., Teng, F., et al. (2021). Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat. Plants.* 7, 766–773. doi: 10.1038/s41477-021-00925-x

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome. *Proc. Natl. Acad. Sci. United States America.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., et al. (2017). Abscisic acid signaling and abiotic stress tolerance in plants: A review on current knowledge and future prospects. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.00161

Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., et al. (2023). The conserved domain database in 2023. *Nucleic Acids Res.* 51, D384–D388. doi: 10.1093/nar/gkac1096

Wang, M., Liu, Y. S., Zhou, J. H., Chen, H. T., Ma, L. N., Ding, Y. Z., et al. (2011). Analysis of codon usage in Newcastle disease virus. *Virus Genes.* 42, 245–253. doi: 10.1007/s11262-011-0574-z

Wang, Y., Tang, H., eBarry, 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, e49. doi: 10.1093/nar/gkr1293

Yang, L., Zhang, Y., Zhu, N., Koh, J., Ma, C., Pan, Y., et al. (2013). Proteomic analysis of salt tolerance in sugar beet monosomic addition line M14. *J. Proteome Res.* 12, 4931–4950. doi: 10.1021/pr400177m

Zhu, H., He, M., Jahan, M. S., Wu, J., Gu, Q., Shu, S., et al. (2021). CsCDPK6, a CsSAMS1-interacting protein, affects polyamine/ethylene biosynthesis in cucumber and enhances salt tolerance by overexpression in tobacco. *International journal of molecular sciences* 22 (20), 11133. doi: 10.3390/ijms222011133

Zukovsky, P. M. (1950). *Cultivated Plants and Their Wild Relatives* Vol. 596 (London, UK: Commonwealth Agriculture Bureau).

frontiers | Frontiers in Plant Science

TYPE Data Report
PUBLISHED 12 December 2024
DOI 10.3389/fpls.2024.1513364

OPEN ACCESS

EDITED BY
Changmian Ji,
Chinese Academy of Tropical Agricultural Sciences, China

REVIEWED BY
Chun-Lei Xiang,
Chinese Academy of Sciences (CAS), China
Roohaida Othman,
Universiti Kebangsaan Malaysia, Malaysia

*CORRESPONDENCE
Reni Lestari
reni.lestari@brin.go.id

RECEIVED 18 October 2024
ACCEPTED 18 November 2024
PUBLISHED 12 December 2024

CITATION
Lestari R, Magandhi M, Hariri MR, Noviady I, Nugroho A and Indriani F (2024) Characterization of the complete chloroplast genome of the endangered and endemic bornean fruit *Artocarpus tamaran* Becc. *Front. Plant Sci.* 15:1513364. doi: 10.3389/fpls.2024.1513364

Characterization of the complete chloroplast genome of the endangered and endemic bornean fruit *Artocarpus tamaran* Becc

Reni Lestari[1]*, Mahat Magandhi[1], Muhammad Rifqi Hariri[2], Ikhsan Noviady[1], Aditya Nugroho[1] and Fitri Indriani[1]

[1]Research Center for Applied Botany, National Research and Innovation Agency, Bogor, Indonesia, [2]Research Center for Biosystematics and Evolution, National Research and Innovation Agency, Bogor, Indonesia

KEYWORDS

conservation, illumina, Moraceae, plastid genome, underutilized fruit

1 Introduction

Artocarpus tamaran Becc. is a member of the *Artocarpus* genus of the Moraceae family, comprising 74 plant species (POWO, 2024). The species tree may attain a height of 45 m and a stem diameter of 1 m, with buttresses up to 3 m in height (Kochummen, 2000). The species is endemic to Borneo, occurring in Sarawak, Sabah, Kalimantan, and Brunei Darussalam, specifically in low land to the hilly mixed Dipterocarpaceae forest, beside the river, on sandstone, clay, and alluvial substrate (POWO, 2024; Jarrett, 1959). It has also been recorded in the primary or old secondary forests and logged forests at 20 m to 1800 m above sea level (Jarrett, 1959). According to the Red List category of the International United Conservation Nations (IUCN), *Artocarpus tamaran* is classified as Vulnerable A2c according to the Red List category of the International Union for Conservation of Nature (IUCN, 2024). The species is endangered due to habitat loss, which has been converted into plantations, logged, burnt down, and climate affected such as in Sabah, Sarawak, and Kalimantan (IUCN, 2024; POWO, 2024). The species is utilized for fiber material sourced from the bark, which is used to produce cloth and hats (Kulip, 2003; Fern 2014), fresh fruit, and edible seed after being boiled or roasted (Lim, 2012). The stem, referred to as "terap" in local terminology, has potential applications in construction (Kochummen, 2000). The log and timber prices of the species were 22.90 USD m^{-3} and 50.88 USD m^{-3}, respectively (Karmini et al., 2020).

The chloroplast genome displays a quadripartite structure and is circular. The structure comprises a large single-copy region (LSC) and a small single-copy region (SSC), separated by a pair of inverted repeats (IRs), with some exceptions noted where the loss of an IR or the SSC has occurred. The size of the chloroplast genome in terrestrial plants ranges from 19 to 217 kb, with the IRs generally measuring 20-26 kb in length (http://www.ncbi.nlm.nih.gov/genome/organelle). The chloroplasts proteome consists of around 3000 proteins that play roles in

photosynthesis, and the biosynthesis of fatty acids, amino acids, hormones, vitamins, nucleotides, and secondary metabolites (Dobrogojski et al., 2020). The advancement and utilization of chloroplast genome engineering technology may inform the investigation of chloroplast gene functions, gene editing, gene expression regulation, and genome analysis (An et al., 2022). Regulation of chloroplast gene expression in chloroplast genome engineering is employed to achieve high-value industrial targets, improve photosynthetic capacity, and biofortify food crops (Boynton et al., 1988). This study presents the results of the chloroplast genome sequencing of the *A. tamaran* species.

2 Method

2.1 Plant material, DNA extraction and sequencing

A sample of *A. tamaran* was obtained from the living collection of Bogor Botanical Gardens in West Java, designated with collector number IN577. The plant sample originated from Central Kalimantan. Genomic DNA was extracted from fresh leaves utilizing the CTAB (cetyltrimethylammonium bromide) method as described by Doyle and Doyle (1987). The initial quantification and purity of DNA were evaluated using a Nanodrop 2000 (Thermo Scientific) and visualized through agarose gel electrophoresis with 1% TBE agarose. The Qubit dsDNA HS Assay Kit (Thermo Scientific) was utilized for enhanced DNA quantification accuracy. The integrity of DNA was assessed utilizing the 4150 TapeStation (Agilent).

Genomic DNA was utilized as the input for library preparation. The genomic DNA was enzymatically fragmented to obtain the required insert size. The fragmented DNA was ligated with MGI-compatible adapters, each containing a unique barcode for each sample. PCR was performed to amplify the library. The quality and quantity of library samples were assessed using Tape Station and Qubit Fluorometer, respectively. The amplified library samples underwent circularization, and the resulting circular DNA served as input for the DNB formation process. The DNBs were loaded onto the flow cell, and sequencing was conducted for 612 cycles (PE300) utilizing the MGI DNBSEQ-G400.

2.2 Chloroplast genome assembly and annotation

Quality control was conducted to evaluate the quality of reads utilizing FASTQC software version 0.11.8 (Andrews, 2010).

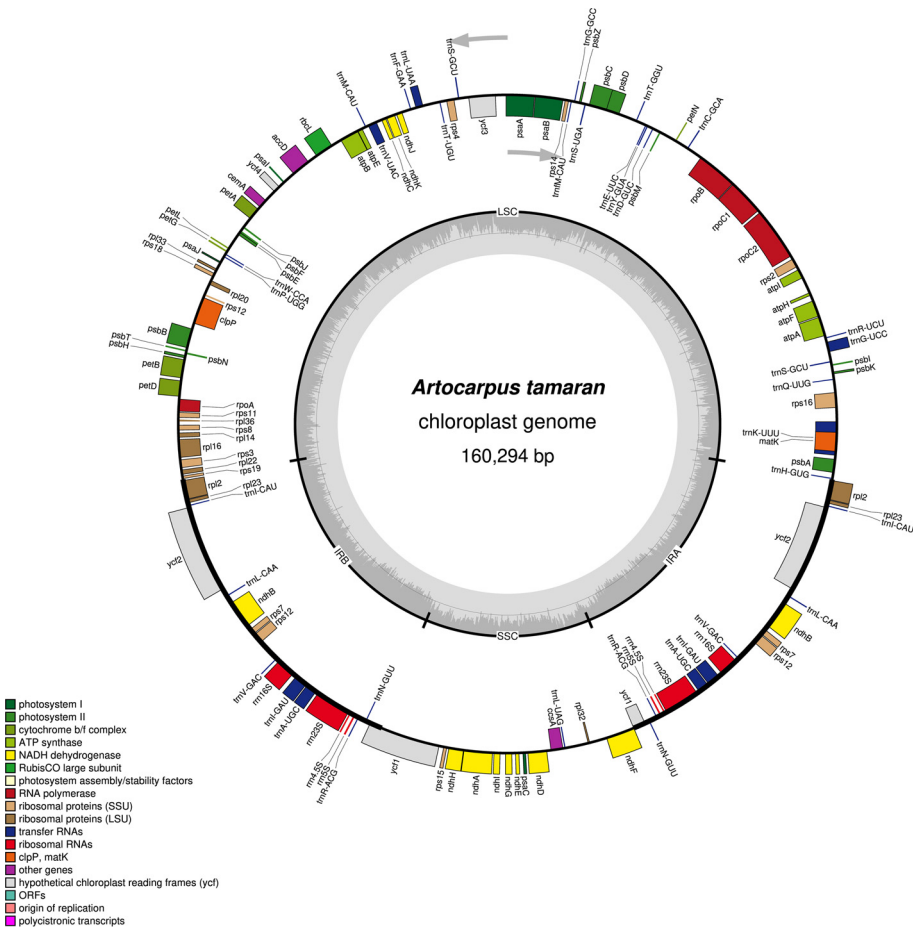


FIGURE 1
Gene map of the *A. tamaran* chloroplast genome.

Low-quality bases (less than 30), adapters, nucleotide position biases at the 3' and 5' ends, and sequence contamination were eliminated through trimming and filtering with Trimmomatic version 0.39. The parameters used were TruSeq3-PE.fa:2:30:10, SLIDING WINDOW:4:28, LEADING:28, TRAILING:28, and MINLEN:20 (Bolger et al., 2014). The clean reads were then assembled using GetOrganelle version 1.7.7.1 (Jin et al., 2020). Annotation was conducted with CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate) (Shi et al., 2019), utilizing the cp genome of *Artocarpus gomezianus* Wall. ex Trécul (accession number: NC_080592) as a reference. This was followed by manual verification in Unipro Ugene v. 45.1 (Okonechnikov et al., 2012) and NCBI Genomic Workbench v. 3.8.2 (Kuznetsov and Bollin, 2021). The circular genome was visualized with Organellar GenomeDRAW (OGDRAW) via the MPI-MP Chlorobox (Greiner et al., 2019).

3 Results

The complete chloroplast genome of *A. tamaran* has been successfully assembled, measuring 160,294 bp and exhibiting a quadripartite structure comprising four regions: the large single-copy (LSC) region, the small single-copy (SSC) region, and two inverted repeats (IR) regions (Figure 1). The LSC region has a length of 88,789 bp, the SSC region measures 20,015 bp, and each IR region is 25,745 bp. The genome exhibits a total GC content of 36%, with the highest concentration observed in the IR regions at 46.2%, followed by the LSC region at 34.2% and the SSC region at 28.9%. A total of 129 genes, comprising 110 unique genes, were annotated in the *A. tamaran* chloroplast genome. The identified genes comprised 84 protein-coding genes (77 unique), 37 tRNAs (29 unique), and 8 rRNAs (4 unique). Of the 129 genes analyzed, 14 exhibited a single intron, while three genes (*rps*12, *ycf*3 and *clp*P) contained two introns (see Table 1).

TABLE 1 List of genes in the *A. tamaran* chloroplast genome.

Functional category	Group of Gene	Name of Gene
Self-replication	rRNA	*rrn*16[d], *rrn*23[d], *rrn*4.5[d], *rrn*5[d]
	tRNA	*trn*K-UUU*, *trn*Q-UUG, *trn*S-GCU[d], *trn*R-UCU, *trn*C-GCA, *trn*D-GUC, *trn*Y-GUA, *trn*E-UUC, *trn*T-GGU, *trn*S-UGA, *trn*G-GCC, *trn*M-CAU, *trn*T-UGU, *trn*L-UAA*, *trn*F-GAA, *trn*fM-CAU, *trn*W-CCA, *trn*G-UCC, *trn*P-UGG, *trn*H-GUG, *trn*L-CAA[d], *trn*V-GAC[d], *trn*V-UAC*, *trn*I-GAU[d]*, *trn*A-UGC[d]*, *trn*R-ACG[d], *trn*N-GUU[d], *trn*L-UAG, *trn*I-CAU[d]
	Large subunit ribosomal proteins (LSU)	*rpl*14, *rpl*16*, *rpl*2[d]*, *rpl*20, *rpl*22, *rpl*23[d], *rpl*32, *rpl*33, *rpl*36
	Small subunit ribosomal proteins (SSU)	*rps*11, *rps*12[d]**, *rps*14, *rps*15, *rps*16*, *rps*18, *rps*19, *rps*2, *rps*3, *rps*4, *rps*7[d], *rps*8
	DNA-dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1*, *rpo*C2
	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
	Subunits of NADH-dehydrogenase	*ndh*A*, *ndh*B[d]*, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
Photosynthesis	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*M, *psb*N, *psb*T, *psb*Z, *ycf*3**
	Subunits of cytochrome b/f complex	*pet*A, *pet*B*, *pet*D*, *pet*G, *pet*L, *pet*N
	Subunit rubisco	*rbc*L
	Subunit of acetyl-CoA-carboxylase	*acc*D
	C-type cytochrome synthesis gene	*ccs*A
Other function	Protease	*clp*P**
	Maturase	*mat*K
	Envelope membrane protein	*cem*A
Unknown function	Conserved open reading frames	*ycf*1[d], *ycf*2[d], *ycf*4

[d], gene duplication; *, single intron; **, double intron.

Data availability statement

This study analyzes datasets available in the NCBI Short Read Archive (SRA) under accession number SRR31020103 (https://www.ncbi.nlm.nih.gov/sra/SRR31020103). The BioProject and Bio-Sample numbers are PRJNA1173771 and SAMN44319506, respectively. The chloroplast genome sequence of *A. tamaran* has been deposited in the NCBI under accession number PQ493654.

Author contributions

RL: Writing – original draft, Funding acquisition, Conceptualization. MM: Writing – original draft, Methodology, Formal analysis, Data curation. MRH: Writing – review & editing, Project administration, Funding acquisition. IN: Writing – review & editing. AN: Writing – review & editing, Formal analysis, Data curation. FI: Writing – review & editing, Data curation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study received funding from the Expedition and Exploration Funding Batch 2 for the fiscal year 2022, facilitated by the Deputy for Research and Innovation Facilitation, the National Research and Innovation Agency (BRIN), under contract number 2860/II.7/HK.01.00/8/2022.

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

References

An, Y., Wang, Y., Wang, X., and Xiao, J. (2022). Development of chloroplast transformation and gene expression regulation technology in land plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1037038

Andrews, S. (2010). FastQC: a Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (accessed September 27, 2024).

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., and Shark, K. B. (1988). Chloroplast transformation in *chlamydomonas* with high velocity microprojectiles. *Science* 240, 1534–1538. doi: 10.1126/science.2897716

Dobrogojski, J., Adamiec, M., and Lucinski, R. (2020). The chloroplast genome: a review. *Acta Physiol. Plant* 42, 98. doi: 10.1007/s11738-020-03089-x

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Fern, K. (2014). Useful tropical plants database: Artocarpus tamaran Becc. Available online at: https://tropical.theferns.info/viewtropical.php?id=Artocarpus+tamaran (accessed October 1, 2024).

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

IUCN (2024). The IUCN Red List of Threatened Species Version 2024-1. Available online at: https://www.iucnredlist.org (Accessed 4 October 2024).

Jarrett, F. M. (1959). Studies in *Artocarpus* and allied genera III. A Revision of *Artocarpus* subgenus *Artocarpus*. *J. Arnold. Arbor.* 40, 113–155. doi: 10.5962/p.186026

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

Karmini,, Karyati,, and Widiati, K. Y. (2020). Short Communication: The ecological and economic values of secondary forest on abandoned land in Samarinda, East Kalimantan Province, Indonesia. *Biodiversitas* 21, 5550–5558. doi: 10.13057/biodiv/d211164

Kochummen, K. M. (2000). "*Artocarpus* J. R. & G. Forster. nom. conserve," in *Tree Flora of Sabah and Sarawak*. Eds. E. Soepadmo and L. G. Saw (Sabah Forestry Department, Forest Research Institute Malaysia, and Sarawak Forestry Department, Kuala Lumpur), 187–212.

Kulip, J. (2003). An ethnobotanical survey of medicinal and other useful plants of Muruts in Sabah, Malaysia. *Telopea* 10, 81–98. doi: 10.7751/telopea20035608

Kuznetsov, A., and Bollin, C. J. (2021). "NCBI genome workbench: desktop software for comparative genomics, visualization, and GenBank data submission," in *Multiple Sequence Alignment: Methods and Protocols*. Ed. K. Katoh (Humana Press, New York), 261–295.

Lim, T. K. (2012). "Artocarpus tamaran," in *Edible Medicinal and Non Medicinal Plants: Volume 3, Fruits* (Springer, New York), 353–355.

Okonechnikov, K., Golosova, O., Fursov, M., Ugene Team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091

POWO (2024). Plants of the World Online (Kew: Royal Botanic Gardens). Available online at: http://www.plantsoftheworldonline.org (Accessed 15 October 2024).

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 December 2024
DOI 10.3389/fpls.2024.1511597

OPEN ACCESS

EDITED BY
Xuming Li,
Department of Scientific Affairs, Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Huolin Luo,
Nanchang University, China
Licao Cui,
Jiangxi Agricultural University, China
Shuangzhan Huang,
Jilin University, China

*CORRESPONDENCE
Haoyuan Chen
393229436@qq.com
Changbing Huang
cbhuang@szai.edu.cn

RECEIVED 15 October 2024
ACCEPTED 26 November 2024
PUBLISHED 13 December 2024

CITATION
Chen H, Li Q, Cheng P, Yan T, Dong C, Hou Z, Zhu P and Huang C (2024) Identification and analysis of major latex protein (*MLP*) family genes in *Rosa chinensis* responsive to *Botrytis cinerea* infection by RNA-seq approaches.
Front. Plant Sci. 15:1511597.
doi: 10.3389/fpls.2024.1511597

Identification and analysis of major latex protein (*MLP*) family genes in *Rosa chinensis* responsive to *Botrytis cinerea* infection by RNA-seq approaches

Haoyuan Chen[1]*, Qingkui Li[1], Peilei Cheng[2], Taotao Yan[2], Chunlan Dong[1], Zhe Hou[2], Peihuang Zhu[2] and Changbing Huang[2]*

[1]College of Horticultural Science and Technology, Suzhou Polytechnic Institute of Agriculture, Suzhou, China, [2]College of Landscape Engineering, Suzhou Polytechnic Institute of Agriculture, Suzhou, China

Roses (*Rosa chinensis*) are among the most cherished ornamental plants globally, yet they are highly susceptible to infections by *Botrytis cinerea*, the causative agent of gray mold disease. Here we inoculated the resistant rose variety 'Yellow Leisure Liness' with *B. cinerea* to investigate its resistance mechanisms against gray mold disease. Through transcriptome sequencing, we identified 578 differentially expressed genes (DEGs) that were significantly upregulated at 24, 48, and 72 hours post-inoculation, with these genes significantly enriched for three defense response-related GO terms. Further domain analysis of the genes in these GO terms reveal that 21 DEGs contain the Bet v 1 family domain, belonging to the major latex protein (*MLP*) gene family, suggesting their potential key role in rose disease resistance. Furthermore, we systematically identified 46 *RcMLP* genes in roses and phylogenetically categorized them into two distinct subfamilies: group I and II. Genomic duplication analysis indicates that tandem duplication is the main driver for the expansion of the *RcMLP* family, and these genes have undergone by purifying selection. Additionally, detailed analyses of gene structure, motif composition, and promoter regions reveal that *RcMLP* genes contain numerous stress-responsive elements, with 32 *RcMLP* genes harboring fungal elicitor/wound-responsive elements. The constructed potential transcription factor regulatory network showed significant enrichment of the ERF transcription factor family in the regulation of *RcMLP* genes. Gene expression analysis reveal that DEGs are mainly distributed in subfamily II, where four highly expressed genes (*RcMLP13*, *RcMLP28*, *RcMLP14*, and *RcMLP27*) are identified in a small branch, with their fold change exceeding ten folds and verified by qRT-PCR. In summary, our research

results underscore the potential importance of the *RcMLP* gene family in response to *B. cinerea* infection and provide comprehensive basis for further function exploration of the *MLP* gene family in rose resistance to fungal infections.

KEYWORDS

major latex protein, rose, RNA-Seq, tandem duplication, *Botrytis cinerea*

1 Introduction

Rosa chinensis, commonly known as the Chinese rose, is highly valued as both a garden ornamental and a source for cut flowers. It holds a pivotal role in horticulture and the economy, especially being indispensable in perfumery and cosmetic industries (Bendahmane et al., 2013; Mileva et al., 2021; Shang et al., 2024). However, the splendor and economic worth of roses are frequently threatened by various pathogens, with *Botrytis cinerea* being a particularly catastrophic fungus, notorious for causing gray mold disease across over 1000 plant species (Dean et al., 2012). This necrotrophic fungus can rapidly lead to significant crop losses, affecting not only the aesthetic appeal of roses but also their marketability (Ullah et al., 2024).

Given the enduring significance of roses in the floriculture industry, there is a pressing need to unravel the molecular mechanisms underlying their defense against such pathogens. Recent research has gradually uncovered how roses combat diseases through complex molecular mechanisms. For instance, the WRKY gene family has demonstrated potential in enhancing the disease resistance of roses (Liu et al., 2019, Liu et al., 2023). Major latex-like proteins (MLPs) homologs can be divided into three groups: MLPs, Bet v 1s, and pathogenesis-related proteins class 10 (PR-10s), one of the 17 members of the PR family (Fujita and Inui, 2021). These regulatory proteins are responsive to biotic and abiotic stress and involved in various physiological and biochemical processes, including responses to drought, salt, plant hormones, and pathogen infections (Fujita and Inui, 2021). They have been identified in many plants, such as *Malus domestica* (Yuan et al., 2020), *Cucumis sativus* (Kang et al., 2023), peanut (Li et al., 2023), Populus (Sun et al., 2024). A recent study identified the *MLP* gene family in *Pyrus bretschneideri*, with *PbrMLP* genes deemed as vital candidates for resistance to *Colletotrichum fructicola* in pears (Su et al., 2024a). Furthermore, a handful of studies have established the PR-10 proteins' role as a signaling module in defense against *Botrytis cinerea*, outlining regulatory mechanisms such as the 'Ethylene-MPK8-ERF.C1-PR' module and the RhERF005/RhCCCH12–RhPR10.1 module, which mediate resistance and cytokinin-induced defense responses, respectively (Deng et al., 2024; Liu et al., 2024). Proteomic analysis has uncovered the role of four PR10 proteins and a plasma membrane aquaporin in the rose defense against *Botrytis cinerea* infection (Li et al., 2024). These studies highlight the potential of *MLP* gene family as significant regulators in response to fungal infections. However, a comprehensive identification of *MLP* genes in roses and an in-depth exploration of their expression patterns in response to *Botrytis cinerea* infection remain underexplored.

In this study, leveraging publicly available genomic data, we analyzed the transcriptional changes in rose petals during *B. cinerea* infection, identifying 578 significantly upregulated DEGs at 24, 48, and 72 hours post-infection. Gene Ontology (GO) analysis of these DEGs reveals a significant enrichment of defense response-related terms among *RcMLP* genes. Furthermore, we systematically identified and analyzed 46 *RcMLP* genes, including their gene structure, motif composition, evolutionary relationships, chromosomal localization, collinearity analysis, and transcription factor regulatory networks, highlighting the potential significance of the *MLP* gene family in rose resistance to fungal infection. Our research not only provides new insights into the genomic evolutionary characteristics of the *MLP* gene family but also lays an essential foundation for exploring their expression patterns and functional mechanisms under pathogen invasion.

2 Materials and methods

2.1 Plant materials and fungal growth and plant infection

Rose cultivars were grown in greenhouses in Suzhou Polytechnic Institute of Agriculture (Jiangsu, China), under the controlled conditions of 75% relative humidity and 25°C under a 16 h light and 8 h dark photoperiod. An *B. cinerea* strain was isolated from typical diseased petals and then stored in 15% (v/v) glycerol at −80°C. *B. cinerea* was cultured on potato glucose agar (PDA) media in 25°C growth chamber for about 2 weeks. The inoculum was prepared by harvesting *B. cinerea* spores with ddH_2O and then suspending in 1/2 PDB to a final concentration of 1×10^5 conidia/mL. For inoculation, *B. cinerea* inoculum were sprayed evenly on the flowers, with 1 mL inoculum per flower, then the inoculated flowers were covered with plastic bags to ensuren100% humidity. A comprehensive screening for resistance to *B. cinerea* was conducted

across all cultivars with stable phenotypic traits in the germplasm repository at Suzhou Agricultural Vocational College under uniform irrigation and fertilization practices. This systematic evaluation led to the identification of two distinct rose varieties with markedly different responses to the disease: "Yellow Leisure Liness" which demonstrated high resistance, and "Miyaho," which exhibited sensitivity to gray mold. Control and infected petals for two cultivars were individually sampled in a randomized manner at 0 h, 24 h, 48 h and 72 h, with three biological repeats at each time point. Petals were immediately frozen in liquid nitrogen at the time of harvesting and stored at −80°C.

2.2 RNA-seq library construction

Following the OMEGA RNA kit protocol, total RNA was extracted from rose petals with three replicates. Subsequently, the purity, concentration and integrity of RNA samples were examined by NanoDrop, Qubit 2.0, Agilent 2100, *etc.*, respectively. Qualified RNA was processed for library construction. Subsequently, the first-strand of cDNA was synthesized with fragmented mRNA as template and random hexamers as primers, followed by second-strand synthesis with addition of PCR buffer, dNTPs, RNase H and DNA polymerase I. Purification of cDNA was processed with AMPure XP beads. Then, Double-strand cDNA was subjected to end repair. Adenosine was added to the end and ligated to adapters. AMPure XP beads were applied here to select fragments within size range of 300–400 bp. Finally, cDNA library was obtained by certain rounds of PCR on cDNA fragments generated from on step. In addition, to ensure the quality of library, Qubit 2.0 and Agilent 2100 were used to examine the concentration of cDNA and insert size. The libraries were then sequenced by on Illumina platform with PE150 mode.

2.3 Reads processing and differentially expressed genes analysis

It is crucial to ensure the quality of the reads before moving onto following analysis. Low quality sequences, primers were removed by BMKCloud (www.biocloud.net) and clean data were collected. Then, clean reads were aligned to the reference genome *R. chinensis* 'Old Blush' (v2.0) by HISAT2 v2.2.1 software (Kim et al., 2015). The *R. chinensis* genome was downloaded from Genome Database for Rosaceae (https://www.rosaceae.org/). The read count for each gene was determined by the StringTie v2.2.0 (Pertea et al., 2015), and the FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) was used to quantification the expression level of each gene. Furthermore, DEGs were identified based on the criteria of an absolute log2Ratio > 1 and the false discovery rate (FDR) < 0.01, utilized the R package DESeq2 (Love et al., 2014). Additionally, GO enrichment and KEGG pathway analyses were performed to explore the biological functions of the DEGs, utilizing clusterProfiler (Yu et al., 2012) and KOBAS software (Xie et al., 2011), respectively.

2.4 Genome-wide identification and characterization of *RcMLP* genes

To identify candidate members of the *MLP* genes within the *R. chinensis* 'Old Blush' (v2.0) genome, the hidden Markov model (HMM) profile for the Bet_v_1 (PF000407) domain was retrieved from the Pfam database (http://pfam.xfam.org/) database. All rose genes were then screened using HMMER 3.0 (http://hmmer.org/), with an E-value threshold of < 1e−10, and those containing the Bet_v_1 domain were designated as candidate *RcMLP* genes. Further validation of the conserved domain in these candidate *RcMLP* genes was conducted using the SMART database (https://smart.embl.de/) and the NCBI-CDD platform (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to ensure accuracy. Characteristics of the *RcMLP* genes, including amino acid length, molecular weight (MW), theoretical isoelectric point (PI), and grand average of hydropathicity (GRAVY), were analyzed using the ExPASy website (http://web.expasy.org/protparam/) (Artimo et al., 2012). Additionally, the predicted subcellular locations of these genes were determined using the Plant-mPLoc tool (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/).

2.5 Multiple sequences alignment, phylogenetic analysis, and tertiary structure prediction of *RcMLP* genes

To explore the evolutionary relationships of *MLPs*, sequences from *R. roxburghii*, *Oryza sativa*, and *Arabidopsis thaliana* were retrieved from the CNCB (https://ngdc.cncb.ac.cn/gwh/Assembly/84056/show), Phytozome (https://phytozome-next.jgi.doe.gov/), and TAIR (https://www.arabidopsis.org/) databases, respectively. Subsequently, the complete MLP protein sequences from these four species were aligned using MAFFT (Katoh and Toh, 2008) v7.4.1. An un-rooted phylogenetic tree was constructed using the maximum likelihood (ML) method in MEGA 11 (Kumar et al., 2018) with 1000 replicates boot-strap test. The tree was further refined and visualized on the Evolview v3 (Subramanian et al., 2019) platform (https://www.evolgenius.info/evolview/). Additionally, the conserved domains Bet_v_1 of all *RcMLPs* were annotated and visualized using the ggMSA software (Zhou et al., 2022) with default parameters. The tertiary structure of typical Bet_v_1 domain was retrieved from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/).

2.6 Chromosome location and gene syntenic analysis of *RcMLP* genes

The physical locations of the *RcMLP* genes across various chromosomes were determined from the gff annotation of the rose genome and visualized using TB-tools with advanced Circos options (Chen et al., 2020). Gene duplication events among the *RcMLP* genes were identified using MCScanX (Wang et al., 2012). Furthermore, the rates of nonsynonymous substitution (Ka), synonymous substitution

(Ks), and the Ka/Ks ratio for the duplicated *RcMLP* genes were calculated with KaKs_Calculator 3.0 (Wang et al., 2010). Genes with a Ka/Ks ratio greater than 1 are considered to be under positive selection; those with a Ka/Ks ratio equal to 1 are considered neutral; and those with a Ka/Ks ratio less than 1 are considered to be under negative or purifying selection.

2.7 Gene structure, conserved motif and *cis*-element analysis of *RcMLP* genes

The exon and intron sequences of the *RcMLP* genes were extracted from the gff annotation file of the rose genome. Following this, the web-based tool MEME v5.5.0 (http://meme-suite.org/tools/meme) was utilized to identify conserved motifs, employing all default parameters. Subsequently, PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to scan the 2-kilobase pair upstream regions of each *RcMLP* gene for the presence of potential *cis*-acting regulatory elements.

2.8 Transcription factor regulatory network analysis of *RcMLP* genes

Potential regulatory interactions involving transcription factors (TFs) within the 2-kilobase pair upstream regions of candidate *RcMLP* genes were predicted using the Plant Transcriptional Regulatory Map (PTRM, http://plantregmap.gao-lab.org/), with *Arabidopsis thaliana* as the reference species and a screening threshold of $P \leq 1e^{-5}$. Visual representations of the predicted TF networks using Cytoscape software (Shannon et al., 2003). The wordcloud is generated by the ggplot2 package. Furthermore, we also conducted protein-protein interaction (PPI) networks analysis by RcMLP protein sequences by STRING (https://cn.string-db.org/).

2.9 Quantitative real-time PCR analysis

To validate the results from the RNA-Seq assay, 6 DEGs with great alterations were chosen and validated by qRT-PCR. The flowers of 'Yellow Leisure Liness' were evenly sprayed with *B. cinerea*, and the petals were collected at 0 h, 24 h, 48 h and 72 h to detected the expression. The primers for the candidate DEGs and *GAPDH* gene were designed by Primer 5.0 software and are shown in Supplementary Table S1. Following the standard protocol of the ABI7500 system, the amplification programs for candidate genes in triplicate were validated by qRT-PCR, and the relative quantitative method ($2^{-\triangle\triangle CT}$) was used to calculate the fold changes to define the expression levels of target genes (Schefe et al., 2006).

3 Results

3.1 Transcriptome profiling of rose in response to *B. cinerea* infection

Through inoculation with *B. cinerea*, we have identified two different resistant rose cultivars, 'Yellow Leisure Liness' and 'Miyako' (Figure 1A). The 'Yellow Leisure Liness' cultivar exhibited exceptional resistance, with no noticeable disease spots on the petals even 48 hours post-inoculation (hpi). In contrast, the 'Miyako' cultivar had begun to show expanding lesions after 24 hpi. The 'Miyako' cultivar's spots were significantly more severe than those on 'Yellow Leisure Liness' at all teat time points, suggesting that 'Yellow Leisure Liness' has a decreased susceptibility to the pathogen, marking it as a resistant cultivar.

To elucidate the resistance mechanisms of the 'Yellow Leisure Liness' rose cultivar, we selected uninfected and treated at 24, 48, and 72 hpi petals for comprehensive transcriptomic sequencing analysis. The sequencing yielded a total of 52,906,936 reads, with at least 81.14% aligning to the reference genome (Supplementary Table S2). This dataset allowed us to identify DEGs in infected petals compared to the controls, with criteria set at a log2 fold change (FC) > 1 and FDR < 0.01.

At 24 hpi, we identified 1,797 DEGs, with 1,211 upregulated and 586 downregulated (Figures 1B, E; Supplementary Table S3). At 48 hpi, the number of significantly altered DEGs was 1,200, including 850 upregulated and 350 downregulated (Figures 1C, F; Supplementary Table S4). By 72 hpi, the number of DEGs had risen to 2,449, with 1,381 upregulated and 1,068 downregulated (Figures 1D, G; Supplementary Table S5). These DEGs are considered part of the rose's response repertoire to *Botrytis cinerea* infection.

3.2 *MLP* genes are involved in rose resistance to *B. cinerea*

By comparing the upregulated DEGs at 24 hpi, 48 hpi, and 72 hpi, 578 significantly upregulated DEGs were identified across all three time points (Figure 2A). GO enrichment analysis of these 578 DEGs revealed significant enrichment in three infection response-related pathways: the abscisic acid-activated signaling pathway, defense response, and response to biotic stimulus (Figure 2B). Protein domain analysis of the 31 DEGs annotated in three pathways (part of genes shared) revealed that the 23 genes with the pathogenesis-related protein Bet v 1 family (PF000407) domain, belonging to the *MLP* gene family (Figure 2C), suggesting their potential role in plant disease resistance. It is noteworthy that *RcMLP9*, *RcMLP7*, *RcMLP6*, *RcMLP32*, *RcMLP8*, and *RcMLP14* exhibited significant differential expression at 24 hpi with a log2 fold change >5 and FDR value < 0.01. Similarly, at 48 hpi, *RcMLP9*, *RcMLP32*, *RcMLP7*, *RcMLP6*, *RcMLP8*, *RcMLP16*, *RcMLP14*, *Chr4g0410081*, *RcMLP28*, *RcMLP13*, *RcMLP27*, and *RcMLP31* demonstrated significant differential expression and *RcMLP9*, *RcMLP32*, and *RcMLP6* were identified as significant differential expression genes at 72 hpi (Supplementary Table S6).

3.3 Systematic identification of *MLP* genes family in rose

In order to comprehensive study the roles of *MLPs* in resisting rose pathogens, we carried out the genome-wide identification of

10.3389/fpls.2024.1511597

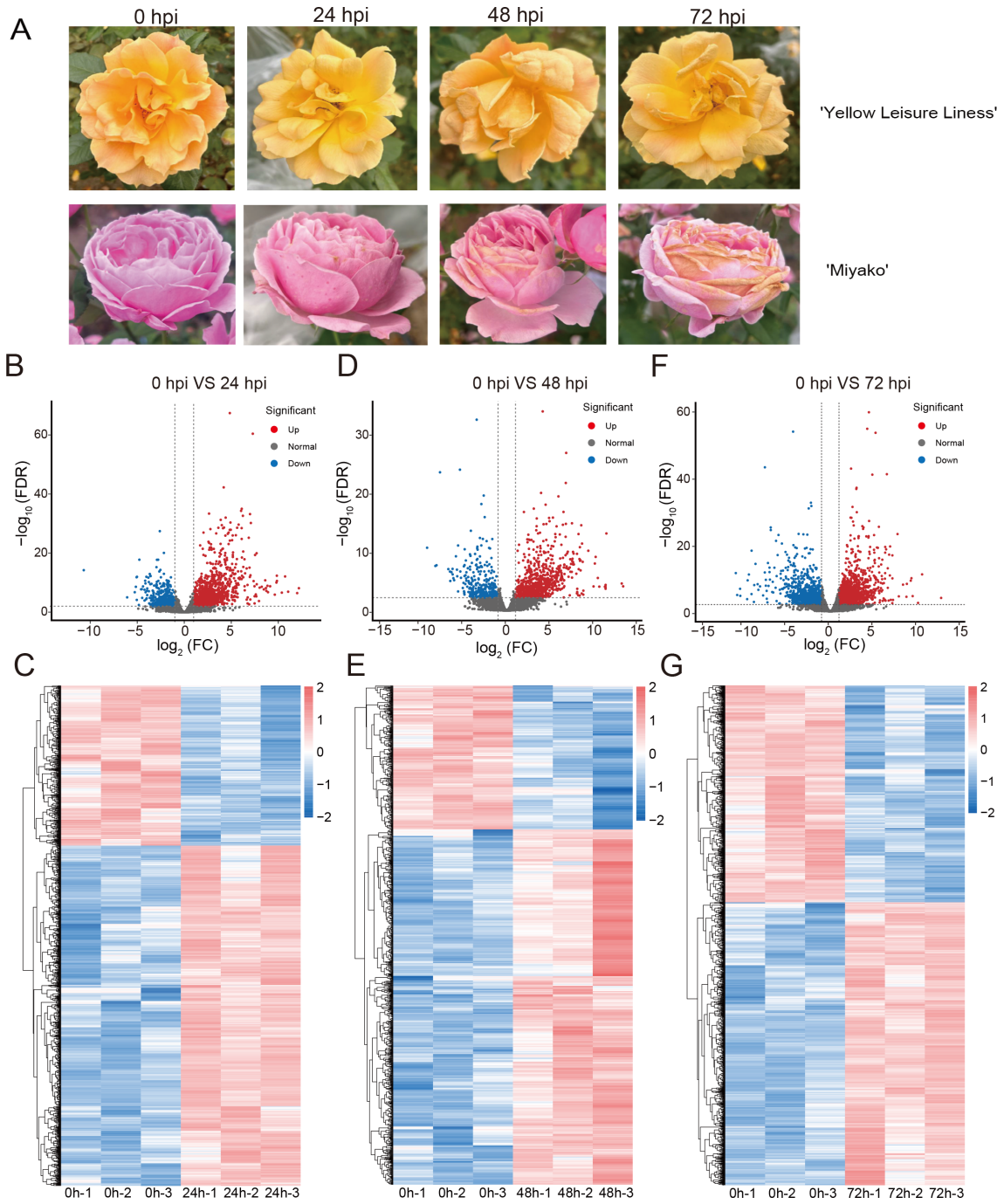


FIGURE 1
Development of *B. cinerea* on rose petal and analysis of DEGs in roses across control and time-based infections. **(A)** Variably severe disease lesions observed in two rose cultivars post-inoculation under different time of treatment. The volcano plots show DEGs of 'Yellow Leisure Liness' cultivar between uninfected rose petals (0 hour) and infected by *B. cinerea* at 24 hours **(B)**, 48 hours **(D)**, and 72 hours **(F)**, based on the criteria of log2 fold change (FC) > 1 and FDR value < 0.01. Heatmaps show the expression levels of DEGs in different samples at 24 hours **(C)**, 48 hours **(E)**, and 72 hours **(G)** post-infection with *B. cinerea*. Data were homogenized by Z-score.

MLP gene family in rose. Utilizing the pathogenesis-related protein Bet v 1 family domain (PF000407, abbreviated as Bet v 1) as a signature motif, we identified 46 *RcMLP* genes in the rose genome (Supplementary Table S7). These genes are named sequentially from *RcMLP1* to *RcMLP46* based on their chromosomal positions. The members of the *RcMLP* gene family exhibit significant differences in amino acid length and physicochemical properties. The number of amino acids ranges from 124 (*RcMLP41*) to 227 (*RcMLP19*), molecular weight of proteins varies between 14118.77 (*RcMLP41*) and 25532.37 Da (*RcMLP19*), the theoretical isoelectric point ranges from 4.7 (*RcMLP38*) to 9.15 (*RcMLP39*), and the Grand Average of Hydrophobicity Index (GRAVY) values range

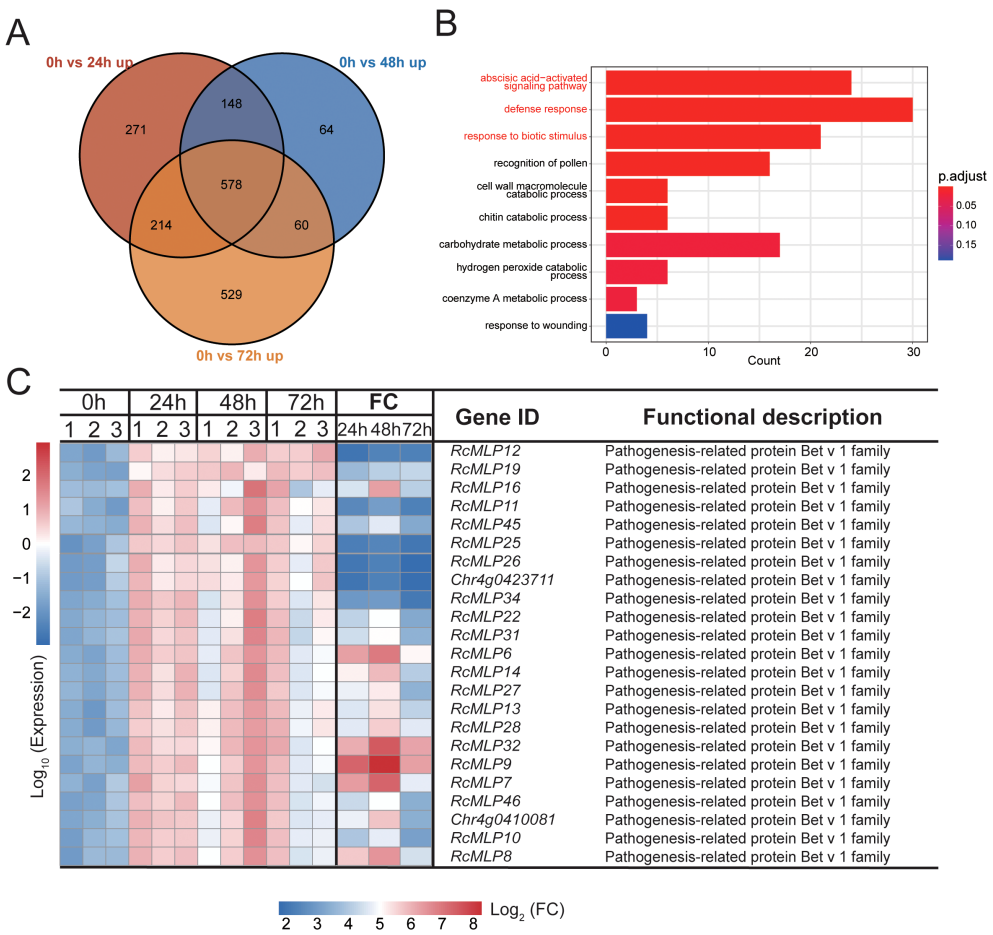


FIGURE 2
Identification of candidate *B. cinerea* infection-regulated rose genes. **(A)** Venn diagram depicting number and overlap among upregulated DEGs from each time point. **(B)** The bar plot illustrates the top 10 Biological Process (BP) GO terms enriched with commonly upregulated DEGs in **(A)**, highlighting the three most significant infection-related terms in red, and the genes enriched within these terms are considered potential candidates. **(C)** Heatmap depicting the expression patterns and protein domain annotations of the candidate genes identified in **(B)**. Data normalized using Z-score method. FC represents fold change between control and different infection time points.

from −0.611 (*RcMLP23*) to 0.113 (*RcMLP16*). Predictive analysis of subcellular localization indicates that all *RcMLPs* were located in the cytoplasm.

3.4 Phylogenetic analysis of the *RcMLP* gene family

The conserved domains and phylogenetic relationships of RcMLP proteins were explored through multiple alignment Bet v1 domain sequences. The results of the multiple sequence alignments indicate that *RcMLP* could be distinctly divided into two groups, namely group I and group II (Figure 3A). The 3-D structural analysis of this domain reveals three α-helices (short α1 and α2, and a long, flexible α3), seven β-strands (β1 to β7), and nine loops (L1 to L9), as illustrated in Figure 3B, based on the AlphaFold database. Group I has some amino acid deletions compared to group II, including the absence of Valine (V), Isoleucine (I) and Proline (P) residues at positions 13-15 in the L1 loop, the absence of Aspartic acid (D) at position 94 in L7 loop, the absence of V at position 127 in L9 loop, and the absence of Alanine (A) at 136 and Glycine (G) at 137 in α3. Both group I and group II have a conserved Gly-rich loop GDG[G/T][V/A]G[T/S][I/V]K located in the L4 Loop, which connects β2 and β3, which is crucial for the endonuclease activity of MLP (Fernandes et al., 2013).

To study the evolutionary relationship between the *MLP* gene family in different plant species, we combined the protein sequences of 45 *RcMLPs* from rose, 36 *RrMLPs* from *R. roxburghii*, 26 *AtMLPs* from *A. thaliana* and 9 *OsMLPs* from rice (Supplementary Table S12) to construct a ML phylogenetic tree. The results showed that a total of 117 *MLP* genes can be phylogenetically divided into three subfamilies: I, II, and III (Figure 3C). 14 *RcMLP* genes in rose were clustered within group I, aligning with the members of group I depicted in Figure 3A and present across all four species, suggesting a possible common ancestry among them. 32 *RcMLP* genes were clustered within group II and this group is uniquely comprised of MLPs indigenous to the genus Rosa, with no homolog genes detected in *A. thaliana* or *O. sativa*. Group III is only found in rice and no any member were found for rose, indicating the evolutionary variations between monocotyledons and dicotyledons.

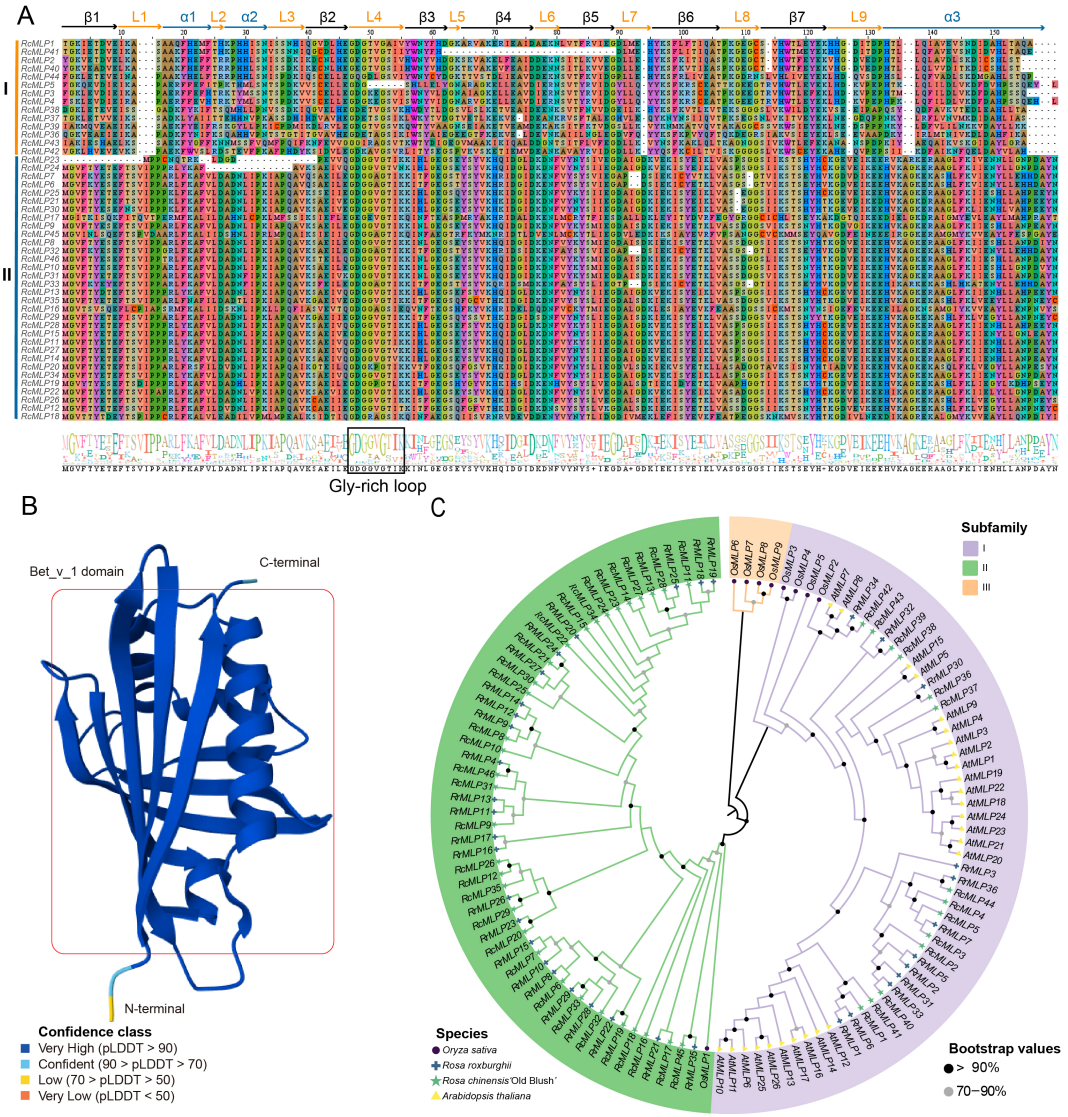


FIGURE 3
Evolution analysis of *RcMLPs* in rose. **(A)** The multiple sequence alignment of *RcMLPs* Bet v1 domain sequences. Sequence logo of Bet v1 domain, generated by WebLogo. We have marked the positions of the 3-D structure predicted in **(B)** atop the aligned 1D sequence diagrams. **(B)** The 3D structure model of Bet v1 domain predicted by AlphaFold3. **(C)** The ML phylogenetic tree of 117 *MLPs* from four species.

3.5 Gene duplication and collinearity analysis of *RcMLP* genes

Our genomic analysis of roses revealed that *RcMLP* genes are dispersed across five out of the seven chromosomes, with a significant clustering on chromosome 4 (Figure 4A). Using MCScanX, we discovered 33 tandem and 2 segmental duplicates in the rose *RcMLP* gene family (Supplementary Table S8), suggesting that tandem duplication events may have contributed to the expansion of the *RcMLP* gene family in rose genome. Of particular note in subfamily II (Figure 3C), tandem duplicated genes account for a substantial 84.38%, and an overwhelming 93.75% of these genes are clustered on chromosome 4, with the exception of *RcMLP45* and *RcMLP46*. This distribution strongly suggests that these RcMLPs have emerged from distinctive tandem duplication events which specific to genus Rosa. Furthermore, intraspecific collinearity analysis identified only one gene pair, *RcMLP16* and *RcMLP45*, suggesting a minor impact of segmental duplication on the expansion of *RcMLP* genes (Figure 4A). The Ka/Ks ratio, a measure of evolutionary pressure, was analyzed for *RcMLP* gene duplicates and results showed Ka/Ks ratios were below 1 (Supplementary Table S9), suggesting purifying selection in *RcMLP* evolution.

To further explore the possible evolutionary processes of the *RcMLP* genes among species, we analyzed the collinearity of the *MLP* family genes between rose with *R. roxburghii*, *O. sativa*, and *A. thaliana*, respectively (Figure 4B). The collinearity analysis between *R. chinensis* and *R. roxburghii* indicated the *RcMLPs* has the most synteny with the *RrMLPs*, exhibiting a predominant one-to-one homozygosity. Comparisons with *A. thaliana* uncovered extensive of chromosomal rearrangements, identifying only four syntenic gene pairs: *RcMLP37-AtMLP17* (*AT1G70890*), *RcMLP36-*

10.3389/fpls.2024.1511597

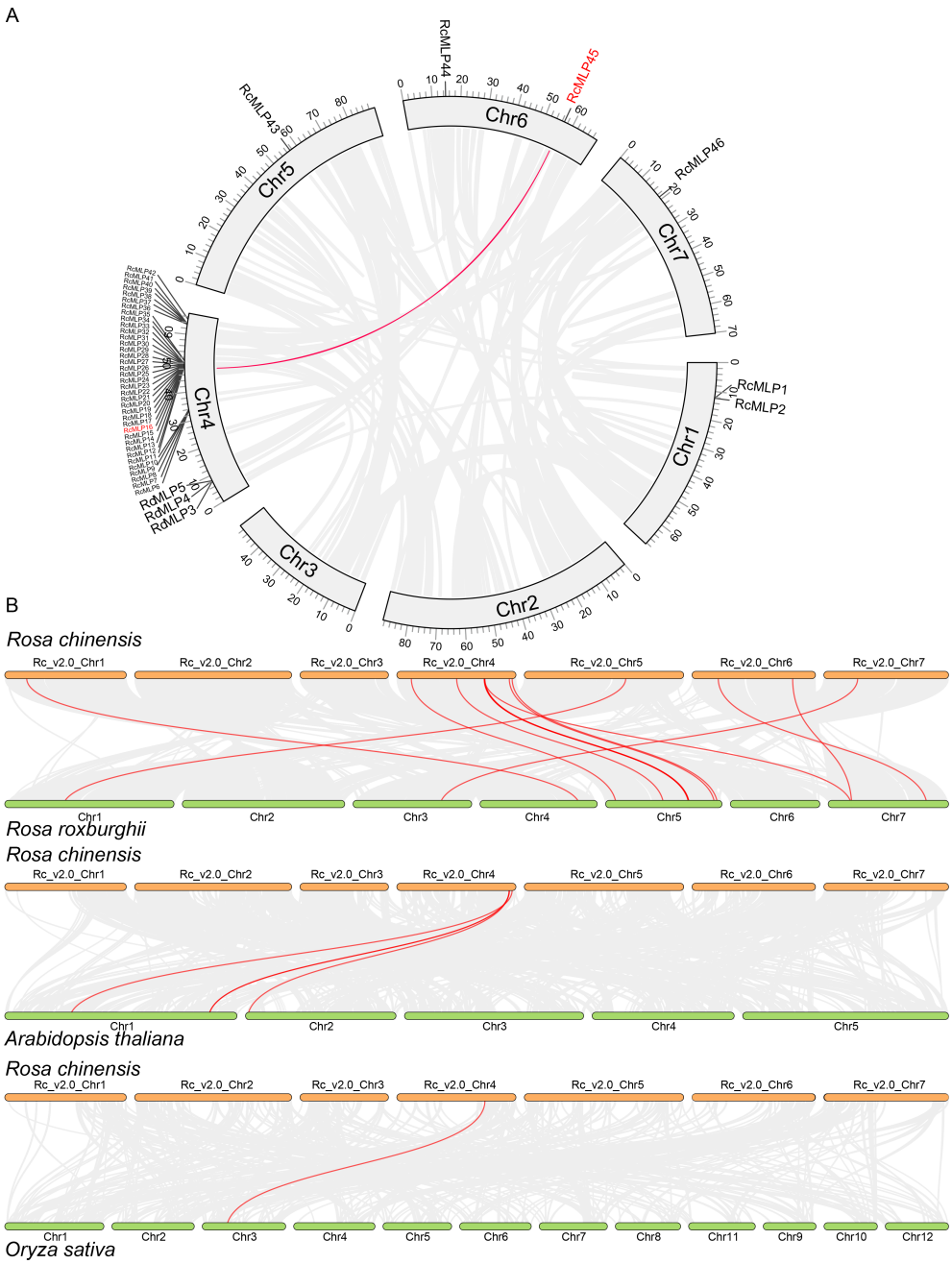


FIGURE 4
Collinearity analysis of *RcMLP* gene family. **(A)** Chromosomal locations and their synteny of *RcMLP* genes in rose. Gray lines indicate all synteny blocks in the rose genome and red line indicates segmental duplication of *RcMLP* genes. **(B)** Synteny analysis *RcMLP* genes between rose and *R. roxburghii*, *A. thaliana*, and *O. sativa*, respectively. The syntenic *MLP* gene pairs were highlighted in red lines.

AtMLP19 (*AT2G01530*), *RcMLP40-AtMLP16* (*AT1G70880*), and *RcMLP42-AtMLP8* (*AT1G24020*), and they were all in group I subfamily. Notably, there was only one collinear gene *RcMLP32*, classified under the group II subfamily, exhibited syntenic relationship with the rice gene *OsMLP1* (*LOC_Os03g18850*). These results demonstrated that the rapid evolution and contribution of species-specific genes tandem duplication in the expansion of the *MLP* gene family.

3.6 Gene structure, conserved motifs and promoter analysis *RcMLP* gene family

We used the protein sequences of 45 *RcMLPs* from rose to construct a ML phylogenetic tree. The results showed that *RcMLP* genes can be phylogenetically divided into two subfamilies: I and II (Figure 5A; Supplementary Figure S1). To elucidate the structural attributes and potential functions of the *MLP* gene family in rose, we

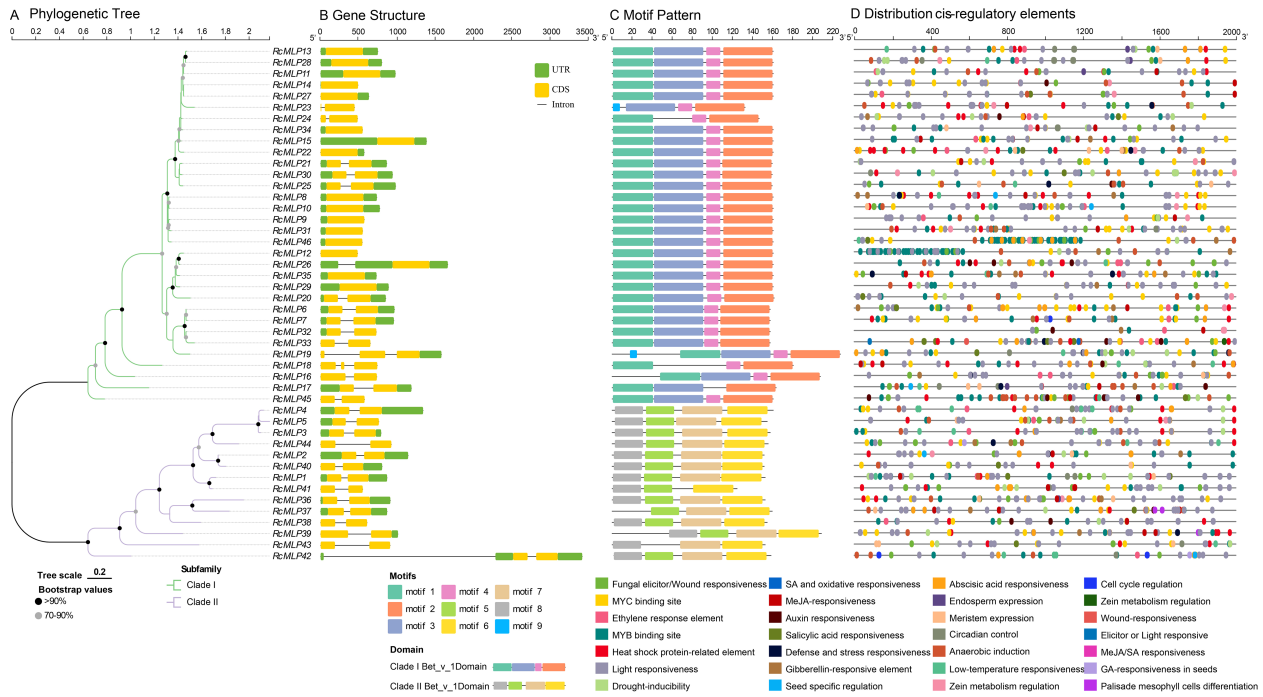


FIGURE 5
The phylogenetic tree, gene structures, protein motifs, and *cis*-regulatory elements analysis among *RcMLPs*. **(A)** The phylogenetic tree of *RcMLPs* was built with full-length protein sequences by the maximum likelihood (ML) method. **(B)** The gene structures of *RcMLP* gene family genes. Yellow boxes, green boxes and black horizontal lines displayed the CDS regions, UTR regions and introns, respectively. **(C)** The 9 conserved motifs of *RcMLP* gene family were represented by boxes with different colors. **(D)** *Cis*-regulatory elements (CREs) distribution of each *RcMLP* was shown.

undertook a comprehensive analysis of gene structures and motif compositions. Exon/intron structure analysis revealed that among the 46 *RcMLPs*, the coding sequences (CDSs) of 30 (~65.22%) were interrupted by introns (Figure 5B). Of these, 27 *RcMLPs* has two exons, and 3 *RcMLPs* has three exons. These variations may have arisen from changes during gene duplication events. Genes clustered within the same subgroup exhibited similar exon/intron structures. Using motif-based sequence analysis tool (MEME), motifs were identified and sequence logos for motifs 1 through 9 were created to predict the structural features of the *RcMLP* proteins and to identify conserved amino acid residues (Figure 5C; Supplementary Figure S2). Notably, members of the same subgroup shared similar motif composition. The *RcMLP* proteins functional domain Bet_v_1 were also shown in Supplementary Figure S3. The most of *RcMLPs* in subfamily I has motifs 1 to 4, with only two genes *RcMLP19* and *RcMLP23* has motif 5, whereas the majority of *RcMLPs* in the subfamily II has motifs 6 to 9. This result was consistent with previous studies (Feki et al., 2023).

Furthermore, we predicted *cis*-regulatory elements (CREs) within the 2-kilobase pair upstream regions of the *RcMLP* genes using PlantCARE. A total of 1,495 CREs were predicted in the promoter regions of *RcMLPs* (Supplementary Table S10) and the representative CREs were shown in Figure 5D. All CREs can be divided into four main categories. The first category is related to stress responsiveness (627), mainly including MYC binding site (132, 21.05%), MYB binding site (122, 19.46%), stress response element (102, 16.27%), anaerobic induction (92, 14.67%), fungal elicitor/wound responsiveness (52, 8.29%), and drought inducibility (41, 6.54%). Our focus is on the fungal elicitor/wound responsive elements (W Box), which are associated with 32 *RcMLP* genes. Notably, the genes *RcMLP45*, *RcMLP46*, and *RcMLP7* contain at least three of these elements each. The second category is related to hormone responsiveness (323), including abscisic acid (ABA) responsiveness (123, 38.08%), MeJA responsiveness (67, 20.74%), salicylic acid (SA) responsiveness (53, 16.41%), gibberellin (GA) responsiveness (34, 10.53%), and ethylene response element (28, 8.67%). Among them, the promoter regions of 41 and 25 *RcMLPs* abundantly displayed ABA responsiveness (ABRE) and MeJA responsiveness (CGTCA motif and TGACG motif), respectively. The *RcMLP12* gene contained the highest number of ABA regulatory elements with totaling of 11 elements, while the *RcMLP37* gene has the most MeJA regulatory elements, with 8 elements. Additionally, we found that the *RcMLP17* and *RcMLP24* genes contained the greatest variety of hormone responsiveness elements, amounting to six types, a broad spectrum of functions within roses. The third category is related to light responsiveness (459). CREs associated with light-responsive elements (such as G-box Box 4, GT1-motif, TCT-motif et al.) were present in the promoter region of all *RcMLPs*. The remaining category is related to growth and development responsiveness (86), mainly including zein metabolism regulation (25), meristem (22), and circadian (21), and the genes *RcMLP28*, *RcMLP13*, and *RcMLP17* contain more than five of these elements. In summary, these analyses collectively offer insights into the physiological functions of *RcMLPs* in rose under both normal and stress conditions.

3.7 Regulatory network analysis of *RcMLP* genes

To predict the transcription factors (TFs) potentially regulating the *RcMLP* genes, we conducted an analysis of the CREs within the promoter regions of these genes using the Plant Transcriptional Regulatory Map. Our analysis identified a total of 41 putative TFs with 4,360 binding sites, suggesting a complex regulatory landscape for the *RcMLP* gene family (Figure 6A; Supplementary Table S11). The TFs were abundant in C2H2 (42), MYB (37), Dof (36), ERF (36), MIKC_MADS (33), AP2 (31), and B3 (31) (Figure 6B), and the least abundant TF families distributed only a few members, such as EIL (2), FAR1 (2), RAV (2), VOZ (2), and LFY (1) (Supplementary Table S11). According to the prediction results, *RcMLP39* has the highest number of regulatory factors among all *RcMLP* genes, with a total of 24 TFs, followed by *RcMLP19* and *RcMLP37*, each having 23 TFs (Supplementary Table S11). In addition, the top ten enriched TF gene families predicted to be involved in regulating *RcMLPs* were identified, including C2H2, MYB, Dof, ERF, MIKC_MADS, AP2, B3, BBR-BPC, NAC, and bHLH (Figure 6C). Notably, the ERF family showed the broadest regulatory influence across various *RcMLP* members with the most enriched TFs (total 1,040 members) (Figure 6D).Our predictions aligned with the known RhERF005/RhCCCH12–RhPR10.1 module, which is implicated in cytokinin-induced defense responses to *B. cinerea* in roses (Liu et al., 2024).Overall, the predicted TFs regulatory network of *RcMLPs* suggests their potential roles in responses to biotic stresses and network interactions.

3.8 Expression patterns of *RcMLP* gene family response to *B. cinerea*se

MLPs enhance plant resistance to pathogens by inducing the expression of disease resistance-related genes, playing a crucial role in plant responses to diseases. At all treatment time, 21 *RcMLP* genes were significantly upregulated, and *RcMLP9* has the highest fold change at 48 hpi (Figure 2C). Phylogenetic analysis of the 46 *RcMLPs* revealed that in subfamily II, 19 genes were significantly upregulated at all three-time points treatment (Figure 7A), accounting for 90.5% of the significantly upregulated genes, suggesting that this clade may play a key role in responding to *B. cinerea* infection. Among them, *RcMLP13*, *RcMLP28*, *RcMLP11*, *RcMLP14*, and *RcMLP27* exhibited relatively high expression levels

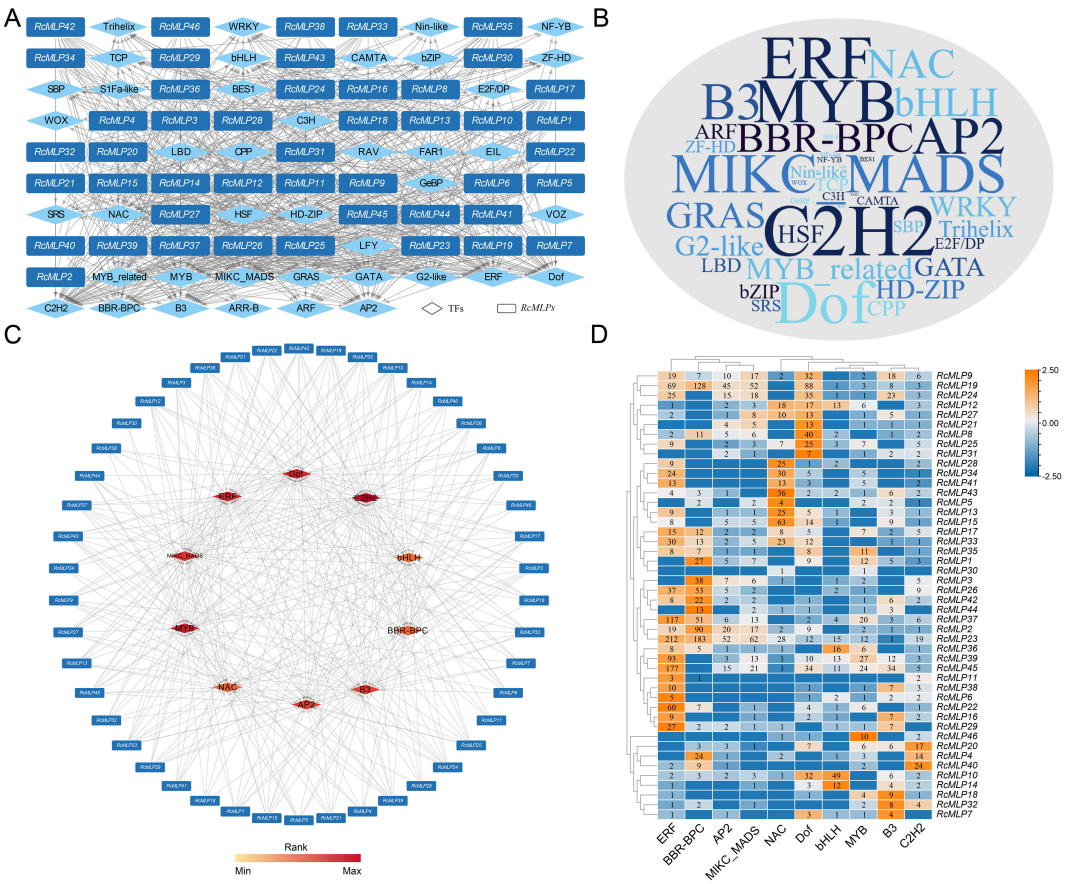


FIGURE 6
The putative TFs regulatory network analysis of *RcMLPs*. **(A)** All putative transcription factors (TFs) were represented by diamond nodes, *RcMLPs* were represented by rectangle nodes. The putative TFs and related mediated *RcMLPs* were linked by grey lines. **(B)** Wordcloud for TFs. Font size is positively correlated with the gene number of corresponding TFs. **(C)** Top 10 highly enriched and targeted *RcMLPs* were shown and the darker the color shows highly enriched. **(D)** The number of top 10 enriched TFs distributions across the targeted *RcMLP* genes.

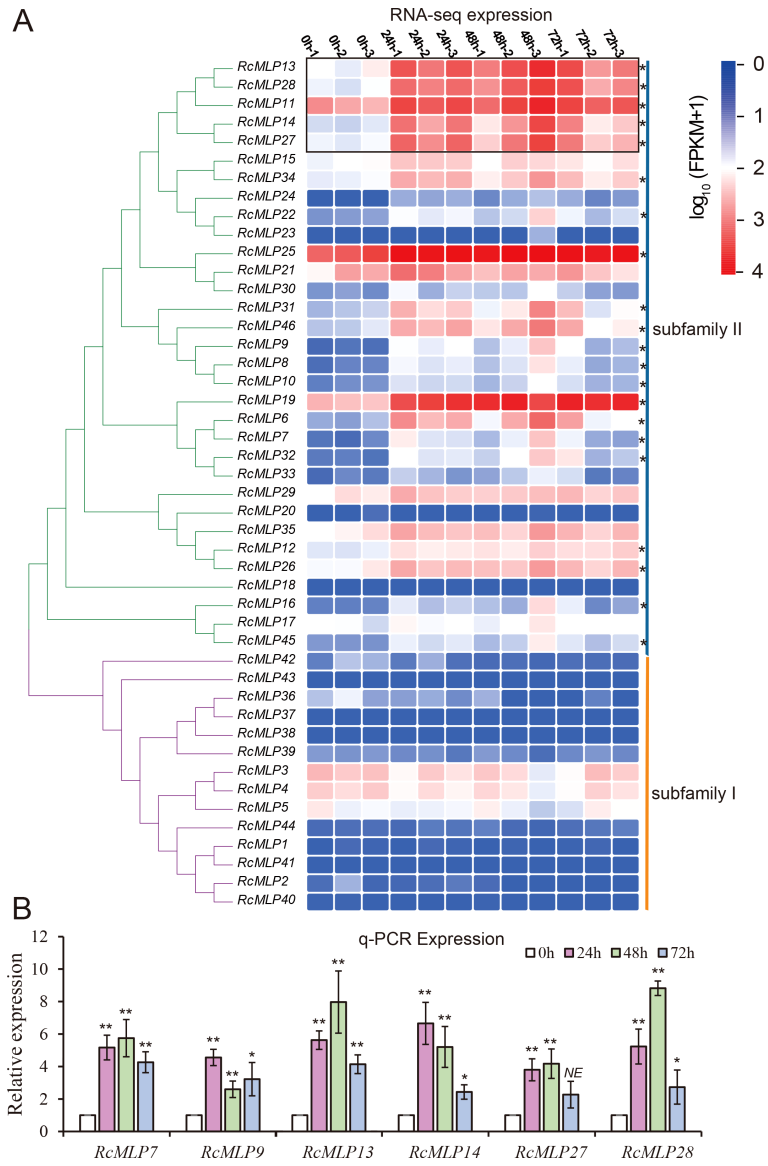


FIGURE 7
Expression patterns of *RcMLP* family at different time points during *B. cinerea* infection of rose. **(A)** The heatmap shows the expression levels of 46 genes encoding MLP gene family in rose petals at 0 hours (uninfected control) and at 24, 48, and 72 hours post-*B. cinerea* infection. **(B)** qRT-PCR results of eight selected *RcMLPs* genes under *B. cinerea* infection stress. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, **** represents $p < 0.0001$.

at all treatment time and were clustered together. Notably, *RcMLP13*, *RcMLP14*, *RcMLP27*, and *RcMLP28* at 48 hpi and *RcMLP14* at 24 hpi had a fold change more than 10. To verify the expression data obtained by RNA sequencing, qRT-PCR was used to examine the expression patterns of six *RcMLP* genes selected under *B. cinerea* infection stress (Figure 7B). The expression levels of the *RcMLP* genes under *B. cinerea* infection stress were largely consistent with the transcriptome data.

4 Discussion

Despite the significant economic implications of gray mold disease on roses, the intricacies of rose defense mechanisms during *B. cinerea* infection are not well understood. In this study, we conducted transcriptome sequencing on petals of the resistant rose cultivar at 24, 48, and 72 hpi with *B. cinerea* to elucidate the temporal defense responses (Figure 1). The analysis revealed that 578 DEGs were significantly upregulated at all three time points (Figure 2A). Notably, these DEGs were enriched in the *MLP* gene family containing the Bet_v_1 domain (Figure 2B).

The *MLP* gene family is crucial in mediating both biotic and abiotic stress responses in plants (Agarwal and Agarwal, 2014; Liu and Ekramoddoullah, 2006). The Bet_v_1 domain, a distinct feature of MLP proteins, is integral to their function (Kofler et al., 2014; Song et al., 2020). This domain facilitates small-molecule binding, thereby regulating the immune system and enabling *MLP* to participate in defense mechanisms against various biotic stresses (Jain et al., 2016;

Radauer et al., 2008). *MLP* were initially discovered in the *Papaver somniferum* and have since been identified in multiple plant species (Song et al., 2020). For instance, 10 *PR-10* genes have been identified in *P. bretschneideri* (Su et al., 2024a), and 13 members have been identified in *Phaseolus vulgaris* (Feki et al., 2023). These gene families play critical roles in plant growth, development, and responses to both biotic and abiotic stresses. The *MLP* genes likely enhance plant disease resistance by regulating specific defense pathways, such as allergen signaling and PR protein responses (Hendrich et al., 2024; Melnikova et al., 2024; Metwally et al., 2024). Furthermore, some *MLP* genes may work synergistically with other defense-related genes, forming a complex immune network. These genes not only function in *B. cinerea* resistance but may also confer broad-spectrum resistance to other pathogens (Castro et al., 2016; Varveri et al., 2024). Similarly, research has shown that *MLP* genes play pivotal roles in various stresses by modulating or participating in plant defense pathways to enhance resistance (Longsaward and Viboonjun, 2024; Sun et al., 2024, Sun et al., 2023). In *Nicotiana benthamiana*, *NbMLP28* is upregulated through the jasmonic acid (JA) signaling pathway, contributing to plant defense responses and significantly enhancing resistance to Potato virus Y, improving disease resistance in tobacco (Song et al., 2020). In peach, *PpMLP10* is significantly upregulated during cold storage in response to cold stress, and its overexpression enhances membrane stability and reduces damage, increasing cold tolerance in peach (Ma et al., 2024). In *Brassica napus*, *BnMLP6* is a key defense gene, encoding a protein that interacts with NPF5.12 at the plasma membrane and endoplasmic reticulum, fortifying the plant's root barrier by increasing suberin deposition to limit *Verticillium longisporum* infection and spread (Dolfors et al., 2024). These findings indicate that the *MLP* gene family broadly participates in plant responses to diverse biotic and abiotic stresses.

However, research on *MLP* in roses remains limited. To further investigate the role of *MLP* genes in rose's response to *B. cinerea* infection, we conducted a systematic identification of the *MLP* gene family in roses and identified 46 *MLP* genes (Figure 3). Multiple sequence alignment revealed that *MLP* genes across different evolutionary branches retain the core Bet_v_1 domain, indicating that *MLP* genes have conserved biological functions across evolution, participating in the regulation of plant immune responses to pathogens (Schenk et al., 2009). Phylogenetic analysis of the Rosa genus has revealed a unique subset of *MLP* genes, Subfamily II, which are indigenous to Rosa and not found in *A. thaliana* or *O. sativa* (Figure 3). This suggests a distinct evolutionary path for MLP genes in Rosa. The high proportion of tandem duplicated genes within Subfamily II, particularly their clustering on chromosome 4 (Figure 4A), indicates that these genes likely emerged from specific tandem duplication events unique to Rosa. This clustering and duplication could be linked to the genus's adaptation, possibly contributing to antifungal capabilities. Additionally, functional characterization through gene expression, indicating the group II subfamily genes tends to respond to *B. cinerea* infections, suggesting their roles in the adaptation and survival of Rosa species in their natural habitats. The absence of these genes in *A. thaliana* and *O. sativa* implies that their functions may be specific to Rosa or have evolved for unique roles within this genus. Further research, including a broader species comparison and functional studies, is needed to substantiate this hypothesis and understand the role of these genes in the biology of Rosa species. Comparative genomics across more Rosaceae family members and other plant families could provide insights into the conservation and divergence of these genes.

Additionally, *cis*-regulatory element analysis revealed that these genes contain functional elements related to hormone response, environmental response, and growth and development (Figure 5). Particularly, the presence of ABA response elements (ABRE) suggests that these genes could be targets of ABA signaling. Previous studies have demonstrated that *MLP* genes are induced under abiotic stress conditions and enhance plant resistance by promoting ABA signaling (Wang et al., 2016). *PsnMLP5* was activated after ABA treatment, suggesting its regulatory role under abiotic stress conditions via the ABA pathway (Sun et al., 2024). In tobacco, *NtMLP423* regulates drought resistance via the ABA pathway, and its overexpression significantly improves drought tolerance (Liu et al., 2020). To further analysis how the *MLP* gene family responds to *B. cinerea* infection, we predicted the transcription factor binding sites in *MLP* genes and found that NAC, bHLH, BBR-BPC, and ERF gene families might be involved in regulating *MLP* genes (Figure 6). These transcription factors are likely to work in concert with the *MLP* gene family to regulate plant immune responses and play key roles in specific signaling pathways, providing new insights into the regulatory mechanisms of rose's response to *B. cinerea* infection (Liu et al., 2020; Song et al., 2020). The ERF family of transcription factors (TFs) exhibited the broadest regulatory influence across various *RcMLP* members, which aligns with previous studies that have implicated ERF TFs in the regulation of plant defense responses (Deng et al., 2024; Liu et al., 2024).

In the expression analysis results, the upregulation of *RcMLP* genes in response to *B. cinerea* highlights their critical role in rose defense mechanisms. The significant expression of subfamily II genes, particularly *RcMLP13*, *RcMLP14*, *RcMLP27*, and *RcMLP28* (Figure 7), indicates their potential as key regulators in the resistance against gray mold disease. Future studies should focus on functional validation of these genes and their roles in other environmental stresses, offering new strategies and directions for crop improvement and breeding projects.

5 Conclusions

This study investigates the transcriptomic profiling of the resistant rose cultivars, 'Yellow Leisure Liness', in response to *Botrytis cinerea* infection, revealing the pivotal role of the *RcMLP* gene family in fungal resistance. We systematically identified and classified 46 *RcMLP* genes into two subfamilies, and gene duplication analysis indicates that tandem duplication contributes to the expansion of the *MLP* gene family, especially for group II subfamily. All *RcMLP* genes were showed evidence of purifying selection. Detailed structural and promoter analyses, along with the regulatory network construction, reveals a complex interplay of transcription factors, particularly the ERF family, in modulating *MLP* gene. Expression analysis confirms that the upregulation of 21 *RcMLP* genes in response to *B. cinerea* are all belonging to the group II subfamily, suggesting their potential roles in disease resistance.

This comprehensive analysis of the rose transcriptome and *MLP* gene family enhances our understanding of the molecular mechanisms underlying rose resistance to gray mold disease and provides a foundation for future breeding efforts.

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

HC: Writing – original draft, Conceptualization, Data curation, Investigation, Formal analysis. QL: Funding acquisition, Project administration, Resources, Writing – review & editing. PC: Validation, Visualization, Writing – review & editing. TY: Writing – review & editing, Formal analysis, Supervision. CD: Writing – review & editing, Methodology, Software. ZH: Data curation, Software, Supervision, Writing – review & editing. PZ: Conceptualization, Investigation, Methodology, Writing – review & editing. CH: Funding acquisition, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this program is provided by the Jiangsu Forestry Science and Technology Innovation and Promotion Project (LYKJ (2021) 02), the "JBGS" Project of Seed Industry Revitalization in Jiangsu Province (JBGS (2021) 091), Suzhou Science and Technology Project (SNG2022062).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1511597/full#supplementary-material

SUPPLEMENTARY FIGURE 1
An maximum-likelihood tree showing the MLP phylogenetic topology.

SUPPLEMENTARY FIGURE 2
Motif sequence LOGO found in RcMLP proteins.

SUPPLEMENTARY FIGURE 3
RcMLP proteins domains.

References

Agarwal, P., and Agarwal, P. K. (2014). Pathogenesis related-10 proteins are small, structurally similar but with diverse role in stress signaling. *Mol. Biol. Rep.* 41, 599–611. doi: 10.1007/s11033-013-2897-4

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal [; Research Support, Non-U.S. Gov't. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Bendahmane, M., Dubois, A., Raymond, O., and Bris, M. L. (2013). Genetics and genomics of flower initiation and development in roses. *J. Exp. Bot.* 64, 847–857. doi: 10.1093/jxb/ers387

Castro, A., Vidal, S., and Ponce de Leon, I. (2016). Moss Pathogenesis-Related-10 Protein Enhances Resistance to Pythium irregulare in Physcomitrella patens and Arabidopsis thaliana. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00580

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

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi: 10.1111/j.1364-3703.2011.00783.x

Deng, H., Pei, Y., Xu, X., Du, X., Xue, Q., Gao, Z., et al. (2024). Ethylene-MPK8-ERF.C1-PR module confers resistance against *Botrytis cinerea* in tomato fruit without compromising ripening. *New Phytol.* 242, 592–609. doi: 10.1111/nph.19632

Dolfors, F., Ilback, J., Bejai, S., Fogelqvist, J., and Dixelius, C. (2024). Nitrate transporter protein NPF5.12 and major latex-like protein MLP6 are important defense factors against Verticillium longisporum. *J. Exp. Bot.* 75, 4148–4164. doi: 10.1093/jxb/erae185

Feki, K., Tounsi, S., Jemli, S., Boubakri, H., Saidi, M. N., Mrabet, M., et al. (2023). Genome-wide identification of PR10 family members and expression profile analysis of PvPR10 in common bean (Phaseolus vulgaris L.) in response to hormones and several abiotic stress conditions. *Plant Growth Regul.* 102, 279–295. doi: 10.1007/s10725-023-00997-z

Fernandes, H., Michalska, K., Sikorski, M., and Jaskolski, M. (2013). Structural and functional aspects of PR-10 proteins [Review. *FEBS J.* 280, 1169–1199. doi: 10.1111/febs.12114

Fujita, K., and Inui, H. (2021). Review: Biological functions of major latex-like proteins in plants. *Plant Sci.* 306, 110856. doi: 10.1016/j.plantsci.2021.110856

Hendrich, J. M., Reuter, A., Jacob, T. P., Kara, H., Amer, S., Rodel, K., et al. (2024). Allergenicity and structural properties of new Cor a 1 isoallergens from hazel identified in different plant tissues. *Sci. Rep.* 14, 5618. doi: 10.1038/s41598-024-55856-2

Jain, D., Khandal, H., Khurana, J. P., and Chattopadhyay, D. (2016). A pathogenesis related-10 protein CaARP functions as aldo/keto reductase to scavenge cytotoxic aldehydes. *Plant Mol. Biol.* 90, 171–187. doi: 10.1007/s11103-015-0405-z

Kang, Y., Tong, J., Liu, W., Jiang, Z., Pan, G., Ning, X., et al. (2023). Comprehensive analysis of major latex-like protein family genes in cucumber (Cucumis sativus L.) and their potential roles in phytophthora blight resistance. *Int. J. Mol. Sci.* 24 (1), 784. doi: 10.3390/ijms24010784

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9, 286–298. doi: 10.1093/bib/bbn013

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–U121. doi: 10.1038/nmeth.3317

Kofler, S., Ackaert, C., Samonig, M., Asam, C., Briza, P., Horejs-Hoeck, J., et al. (2014). Stabilization of the dimeric birch pollen allergen Bet v 1 impacts its immunological properties. *J. Biol. Chem.* 289, 540–551. doi: 10.1074/jbc.M113.518795

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, 1547–1549. doi: 10.1093/molbev/msy096

Li, R., Yao, J., Ming, Y., Guo, J., Deng, J., Liu, D., et al. (2024). Integrated proteomic analysis reveals interactions between phosphorylation and ubiquitination in rose response to Botrytis infection. *Hortic. Res.* 11, uhad238. doi: 10.1093/hr/uhad238

Li, J., Zeng, R., Huang, Z., Gao, H., Liu, S., Gao, Y., et al. (2023). Genome-wide characterization of major latex protein gene family in peanut and expression analyses under drought and waterlogging stress. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1152824

Liu, X., Cao, X., Chen, M., Li, D., and Zhang, Z. (2024). Two transcription factors, RhERF005 and RhCCCH12, regulate rose resistance to Botrytis cinerea by modulating cytokinin levels. *J. Exp. Bot.* 75, 2584–2597. doi: 10.1093/jxb/erae040

Liu, J.-J., and Ekramoddoullah, A. K. M. (2006). The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiol. Mol. Plant Pathol.* 68, 3–13. doi: 10.1016/j.pmpp.2006.06.004

Liu, X., Li, D., Zhang, S., Xu, Y., and Zhang, Z. (2019). Genome-wide characterization of the rose (Rosa chinensis) WRKY family and role of RcWRKY41 in gray mold resistance. *BMC Plant Biol.* 19, 522. doi: 10.1186/s12870-019-2139-6

Liu, H., Ma, X., Liu, S., Du, B., Cheng, N., Wang, Y., et al. (2020). The Nicotiana tabacum L. major latex protein-like protein 423 (NtMLP423) positively regulates drought tolerance by ABA-dependent pathway. *BMC Plant Biol.* 20, 475. doi: 10.1186/s12870-020-02690-z

Liu, X., Zhou, X., Li, D., Hong, B., Gao, J., and Zhang, Z. (2023). Rose WRKY13 promotes disease protection to Botrytis by enhancing cytokinin content and reducing abscisic acid signaling. *Plant Physiol.* 191, 679–693. doi: 10.1093/plphys/kiac495

Longsaward, R., and Viboonjun, U. (2024). Genome-wide identification of rubber tree pathogenesis-related 10 (PR-10) proteins with biological relevance to plant defense. *Sci. Rep.* 14, 1072. doi: 10.1038/s41598-024-51312-3

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, X., Gong, C., An, R., Li, Y., Cheng, N., Chen, S., et al. (2024). Characterisation of the MLP genes in peach postharvest cold storage and the regulatory role of PpMLP10 in the chilling stress response. *Int. J. Biol. Macromol* 266, 131293. doi: 10.1016/j.ijbiomac.2024.131293

Melnikova, D. N., Finkina, E. I., Potapov, A. E., Danilova, Y. D., Toropygin, I. Y., Matveevskaya, N. S., et al. (2024). Structural and immunological features of PR-10 allergens: focusing on the major alder pollen allergen aln g 1. *Int. J. Mol. Sci.* 25 (9), 4965. doi: 10.3390/ijms25094965

Metwally, R. A., Taha, M. A., El-Moaty, N. M. A., and Abdelhameed, R. E. (2024). Attenuation of Zucchini mosaic virus disease in cucumber plants by mycorrhizal symbiosis. *Plant Cell Rep.* 43, 54. doi: 10.1007/s00299-023-03138-y

Mileva, M., Ilieva, Y., Jovtchev, G., Gateva, S., Zaharieva, M. M., Georgieva, A., et al. (2021). Rose flowers-A delicate perfume or a natural healer. *Biomolecules* 11 (1), 127. doi: 10.3390/biom11010127

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, 290. doi: 10.1038/nbt.3122

Radauer, C., Lackner, P., and Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* 8, 286. doi: 10.1186/1471-2148-8-286

Schefe, J. H., Lehmann, K. E., Buschmann, I. R., Unger, T., and Funke-Kaiser, H. (2006). Quantitative real-time RT-PCR data analysis:: current concepts and the novel "gene expression's C_T difference" formula. *J. Mol. Medicine-Jmm* 84, 901–910. doi: 10.1007/s00109-006-0097-6

Schenk, M. F., Cordewener, J. H., America, A. H., Van't Westende, W. P., Smulders, M. J., and Gilissen, L. J. (2009). Characterization of PR-10 genes from eight Betula species and detection of Bet v 1 isoforms in birch pollen. *BMC Plant Biol.* 9, 24. doi: 10.1186/1471-2229-9-24

Shang, J., Feng, D., Liu, H., Niu, L., Li, R., Li, Y., et al. (2024). Evolution of the biosynthetic pathways of terpene scent compounds in roses. *Curr. Biol.* 34, 3550–3563 e3558. doi: 10.1016/j.cub.2024.06.075

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 [Article; Proceedings Paper. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Song, L., Wang, J., Jia, H., Kamran, A., Qin, Y., Liu, Y., et al. (2020). Identification and functional characterization of NbMLP28, a novel MLP-like protein 28 enhancing Potato virus Y resistance in Nicotiana benthamiana. *BMC Microbiol.* 20, 55. doi: 10.1186/s12866-020-01725-7

Su, Z., Han, C., Qiao, Q., Li, C., Dong, H., Wang, X., et al. (2024a). Genome-wide analysis of the family 10 plant pathogenesis-related proteins in Pyrus bretschneideri and functional analysis of PbrMLP for Colletotrichum fructicola resistance. *Horticulture Adv.* 2, 21. doi: 10.1007/s44281-024-00037-4

Subramanian, B., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275. doi: 10.1093/nar/gkz357

Sun, X., Li, Y., Sun, Y., Wu, Q., and Wang, L. (2024).). Genome-Wide Characterization and Expression Analyses of Major Latex Protein Gene Family in Populus simonii x P. nigra. *Int. J. Mol. Sci.* 25 (5), 2748. doi: 10.3390/ijms25052748

Sun, Z., Meng, L., Yao, Y., Zhang, Y., Cheng, B., and Liang, Y. (2023). Genome-wide evolutionary characterization and expression analysis of major latex protein (MLP) family genes in tomato. *Int. J. Mol. Sci.* 24 (19), 15005. doi: 10.3390/ijms241915005

Ullah, I., Yuan, W., Khalil, H. B., Khan, M. R., Lak, F., Uzair, M., et al. (2024). Understanding Botrytis cinerea infection and gray mold management: a review paper on deciphering the rose's thorn. *Phytopathol. Res.* 6 (1), 42. doi: 10.1186/s42483-024-00262-9

Varveri, M., Papageorgiou, A. G., and Tsitsigiannis, D. I. (2024). Evaluation of biological plant protection products for their ability to induce olive innate immune mechanisms and control colletotrichum acutatum, the causal agent of olive anthracnose. *Plants (Basel).* 13 (6), 878. doi: 10.3390/plants13060878

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, e49. doi: 10.1093/nar/gkr1293

Wang, Y., Yang, L., Chen, X., Ye, T., Zhong, B., Liu, R., et al. (2016). Major latex protein-like protein 43 (MLP43) functions as a positive regulator during abscisic acid responses and confers drought tolerance in Arabidopsis thaliana. *J. Exp. Bot.* 67, 421–434. doi: 10.1093/jxb/erv477

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. *Genomics Proteomics Bioinf.* 8, 77–80. doi: 10.1016/s1672-0229(10)60008-3

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics-a J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118

Yuan, G., He, S., Bian, S., Han, X., Liu, K., Cong, P., et al. (2020). Genome-wide identification and expression analysis of major latex protein (MLP) family genes in the apple (Malus domestica Borkh.) genome. *Gene* 733, 144275. doi: 10.1016/j.gene.2019.144275

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., et al. (2022). ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Briefings Bioinf.* 23 (4), 1–12. doi: 10.1093/bib/bbac222

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 December 2024
DOI 10.3389/fpls.2024.1515944

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Gang Nie,
Sichuan Agricultural University, China
Shuhua Wei,
Ningxia Academy of Agriculture and Forestry Sciences, China

*CORRESPONDENCE
Xuemin Wang
wangxuemin@caas.cn
Jun Tang
tangjun@caas.cn

RECEIVED 23 October 2024
ACCEPTED 18 November 2024
PUBLISHED 16 December 2024

CITATION
Ma X, Liang Q, Han Y, Fan L, Yi D, Ma L, Tang J and Wang X (2024) Integrated transcriptomic, proteomic and metabolomic analyses revealing the roles of amino acid and sucrose metabolism in augmenting drought tolerance in *Agropyron mongolicum*. *Front. Plant Sci.* 15:1515944. doi: 10.3389/fpls.2024.1515944

Integrated transcriptomic, proteomic and metabolomic analyses revealing the roles of amino acid and sucrose metabolism in augmenting drought tolerance in *Agropyron mongolicum*

Xiaoran Ma[1], Qingwei Liang[2], Yusi Han[1], Lu Fan[1], Dengxia Yi[1], Lin Ma[1], Jun Tang[1*] and Xuemin Wang[1*]

[1]Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Chifeng Institute of Agriculture and Animal Husbandry Science, Chifeng, China

Drought, a major consequence of climate change, initiates molecular interactions among genes, proteins, and metabolites. *Agropyron mongolicum* a high-quality perennial grass species, exhibits robust drought resistance. However, the molecular mechanism underlying this resistance remaining largely unexplored. In this study, we performed an integrated analysis of the transcriptome, proteome, and metabolome of *A. mongolicum* under optimal and drought stress conditions. This combined analysis highlighted the pivotal role of transporters in responding to drought stress. Moreover, metabolite profiling indicated that arginine and proline metabolism, as well as the pentose phosphate pathway, are significantly involved in the drought response of *A. mongolicum*. Additionally, the integrated analysis suggested that proline metabolism and the pentose phosphate pathway are key elements of the drought resistance strategy in *A. mongolicum* plants. In summary, our research elucidates the drought adaptation mechanisms of *A. mongolicum* and identifies potential candidate genes for further study.

KEYWORDS

***Agropyron mongolicum*, drought stress, proteomics, transcriptomics, metabolomics**

1 Introduction

In the context of agricultural development, drought serves as a pivotal constraint, hindering plant growth, diminishing crop yields, and limiting the extent of cultivable land. Arid and semi-arid regions, which represent about 36% of the global land area and 43% of the world's arable land, are particularly susceptible to this issue. As a result, global water scarcity has emerged as a significant factor contributing to the decline in crop yields (Du et al., 2017; Han et al., 2017; Zurbriggen et al., 2010). To survive under such conditions, plants have developed a range of defense mechanisms that facilitate adaptation to arid environments. The study of these adaptive mechanisms is essential for advancements in agriculture, crop breeding, and the enhancement of varietal resilience (Alvarez et al., 2008).

Recent studies on the drought resistance of plant stems and roots have underscored the importance of several key agronomic traits. Firstly, the length of roots during the seedling stage is identified as a critical factor in improving crop yield under drought stress (Ahmed et al., 2019). Additionally, in situations where subsoil moisture is limited, enhancing root mass and density can significantly improve yield components by facilitating more efficient water uptake (Fang et al., 2017). Furthermore, the total root length is another trait that plays a role in drought resistance, as it affects the soil distribution of roots and the amount of water they can absorb (Wasaya et al., 2018). Moreover, the survival or vigor of seedlings is a practical indicator for assessing drought tolerance in controlled laboratory conditions. Lastly, early growth, or vigor, especially concerning the size of leaves and stems at the early stages of plant development, is a crucial crop trait associated with enhanced water utilization, greater biomass, and higher grain yields (Zhao et al., 2019).

Agropyron mongolicum Keng, a diploid allogamous species (2n = 2x = 14, PP), belongs to the genus *Agropyron Gaertn* and is renowned as a high-quality perennial forage grass. It is predominantly distributed across regions in China, such as Hebei, Shanxi, Gansu, Ningxia, and Inner Mongolia (Boyer, 1982). This species flourishes in sandy and desert grasslands and frequently acts as a companion species in areas prone to desertification (Vij and Tyagi, 2010). *A. mongolicum* is primarily cultivated in arid and semi-arid regions characterized by water scarcity, where drought stress markedly constrains its growth and productivity. In 1990, the species was officially designated as a wild-type cultivated variety by the National Forage Variety Appraisal Committee.

Agropyron mongolicum exhibits high seed-setting and germination rates, robust vitality and adaptability, early greening in spring, and late senescence in autumn. These attributes render it an ideal forage option for areas with scarce resources, especially during the winter and spring seasons. The species boasts a high nutritional content, tender shoots and roots, and is characterized by good palatability and resistance to cold, drought, adverse soil conditions, and sandstorms. These characteristics make *A. mongolicum* particularly suitable for enhancing natural grasslands, serving as a windbreak and for sand stabilization, as well as for establishing artificial grasslands. Despite the substantial research on its agronomic properties, there has been limited exploration into the molecular basis of its drought resistance. Considering its outstanding drought tolerance, *A. mongolicum* shows potential as a genetic resource for the development of stress-resistant crops through breeding programs (Che and Li, 2007; Esfahanian et al., 2017; Zurbriggen et al., 2010).

Drought stress markedly impacts the growth and productivity of *A. mongolicum*, triggering alterations in its intrinsic molecular responses. An integrated multi-omics strategy yields profound insights into essential biological processes, metabolic pathways, and regulatory networks within plants (Amiour et al., 2012; Bittencourt et al., 2022; Chin et al., 2020; Leao et al., 2022; Li et al., 2019; Moreno et al., 2021; Remmers et al., 2018; Shu et al., 2022; Srivastava et al., 2013; Wang et al., 2019; Yun et al., 2019). Advances in sequencing technologies have facilitated comprehensive analyses of gene expression patterns, providing robust tools for molecular diagnosis and classification (Xia et al., 2023).

Multi-omics technologies have been extensively employed to elucidate the mechanisms of stress resistance in diverse plants and crops, encompassing responses to water stress, salt stress, disease resistance, and drought stress. These studies have pinpointed genes and metabolites associated with pivotal traits, as well as variations in the accumulation of bioactive components. Findings indicate that various plants and crops mount unique responses to different abiotic stresses. For instance, multi-omics high-throughput technologies have uncovered that E3-ubiquitin ligase proteins play a role in regulating abiotic stress responses in rice (Li et al., 2023; Melo et al., 2021). In wheat, proteomic and metabolomic approaches have been utilized to investigate the drought resistance mechanisms of two spring wheat varieties, Bahar (drought-resistant) and Kavir (drought-sensitive). Metabolomic analysis revealed marked alterations in primary metabolites, including amino acids, sugars, and organic acids, under drought stress. Specifically, the Bahar variety exhibited accumulation of branched-chain amino acids, lysine, proline, aromatic amino acids, arginine, and methionine, with the branched-chain pathway involved in tryptophan accumulation activated, thus contributing to auxin production. Conversely, the Kavir variety's metabolome exhibited fewer affected pathways, with purine metabolism being one of only two significantly impacted under stress conditions (Amirbakhtiar et al., 2019; Francki et al., 2016; Guo et al., 2020b; Michaletti et al., 2018; Saini et al., 2022; Shewry et al., 2017).

In the present study, we performed an integrated omics analysis of the transcriptome, proteome, and metabolome in the shoots and roots of *A. mongolicum* under both normal and drought stress conditions to elucidate its drought tolerance mechanisms. Differential expression analysis identified substantial changes in transcripts, proteins, and metabolites among various plant tissues in response to drought stress. Our comprehensive multi-omics investigation highlighted the pivotal roles of arginine and proline metabolism, as well as the pentose phosphate pathway, in the drought resistance of *A. mongolicum*. These findings offer new perspectives on the genes that could potentially enhance plant drought tolerance.

2 Materials and methods

2.1 Plant material

The *A. mongolicum* was gifted by Inner Mongolia Agricultural University. The seeds were initially surface-sterilized with 2% sodium hypochlorite for 45 minutes, then soaked overnight in distilled water. Germination took place on a moist blotting sheet within a petri dish at 28°C, under a 14-hour light/10-hour dark photoperiod. Seven days after germination, the seedlings were transferred to a hydroponic system with 1/2-strength Murashige and Skoog (MS) medium as the nutrient source, and were cultivated for an additional 14 days. The 14-day-old seedlings were subjected to a 20% (w/v) PEG 6000 solution to simulate drought stress. Following a 24-hour treatment period, both treated (drought-stressed) and untreated (control) seedlings were harvested, flash-frozen in liquid nitrogen, and stored at -80°C for subsequent analysis. The growth and sample collection were conducted across three independent biological replicates. These samples were utilized for subsequent transcriptomic, proteomic, and metabolomic analyses.

2.2 Transcriptome analysis

Total RNA was extracted using ethanol precipitation and the CTAB-PBIOZOL method. The quality and quantity of the RNA samples were evaluated using a Qubit fluorometer and a Qsep400 high-throughput bio-fragment analyzer (Met ware Biotechnology Co., Ltd. Wuhan, China). The library preparation followed standard Illumina protocols, and sequencing was carried out on the Illumina Nova-seq 6000 platform by a commercial service provider. The raw reads were filtered using Fastq to remove reads with adapters (Patel and Jain, 2012). Owing to the absence of a reference genome, transcriptome assembly was conducted *de novo*. Novel gene prediction was performed using String Tie, which applies network flow algorithms and optional *de novo* transcript assembly to splice transcripts. Gene expression was quantified using feature counts to calculate gene alignment statistics, followed by the computation of FPKM (Fragments Per Kilobase Million mapped reads) values based on gene length. Differential gene expression (log2 fold change ≥ +1 or ≤ -1 and FDR ≤ 0.05) was analyzed using edgeR (Robinson et al., 2010).

2.3 Protein sample preparation and proteomic analysis

The protein extraction and nano LC-MS/MS analysis were performed following the method described by (Singh et al., 2023). Total protein was extracted by suspending 1 g of ground tissue in an extraction buffer containing 1% SDS, 100 mM Tris-HCl, 7 M urea, 2 M thiourea, 1 mM PMSF, and 2 mM EDTA. The mixture was shaken, ultrasonicated on ice for 10 minutes, and the centrifuged to obtain the protein solution. Four times the volume of chilled acetone was added, and the proteins were precipitated overnight at -20°C, followed by centrifugation at 4°C to collect the precipitate. The precipitate was washed with cold acetone and dissolved in 8 M urea. Protein concentration was determined using a BCA kit according to the manufacturer's instructions.

For tryptic digestion, equal number of proteins were taken from each sample. The proteins were reduced with 10 mM DTT for 45 minutes at 37°C and alkylated with 50 mM iodoacetamide (IAM) for 15 minutes in the dark at room temperature. The protein precipitate was collected using 4 times the volume of chilled acetone precipitation at -20°C for 2 hours and resuspended in 200 µL of 25 mM ammonium bicarbonate solution. Protein was digested overnight at 37°C with 3 µL of trypsin (Promega). After digestion, peptides were desalted using a C18 cartridge, drying using a vacuum concentrator, and redissolved in 0.1% (v/V) formic acid.

LC-MS/MS analysis was performed on an Orbitrap Astral MS system coupled to a Thermo Scientific™ Vanquish™ Neo UHPLC system, Samples were injected via an autosampler and trapped on a PepMap Neo Trap Cartridge column (300 µm ×5 mm, 5 µm), then separated on an Easy-Spray™ PepMap™ Neo UHPLC column (150 µm × 15 cm, 2 µm) over a 22-minute gradient. For DIA experiments, the Orbitrap Astral MS was set to a full MS resolution of 240,000 at 200 m/z with a scan range of 380-980 m/z. The full MS AGC was set to 500%. Fragment ion scans were recorded at a resolution of 80,000 and Maximum injection time (ms) of 3 ms using 299 windows of 2-Th scanning from 380-980 m/z. Ions were fragmented using HCD with a normalized collision energy (NCE) of 25%.

MS raw data were analyzed using DIA-NN (v1.8.1) with a library-free method. Three databases were used to create a spectra library utilizing deep learning algorithms from neural networks: the MWXS-24-694-a_*Agropyron_mongolicum*_Keng. blast. pep. fasta (containing 25,403 sequences)、MWXS-24-694-a_*Agropyron_mongolicum*_Keng.angel.pep.fasta (containing 337 sequences)、and iRT2.fasta (containing 1 sequences).of. The Match Between Runs (MBR) option was employed to generate the spectral library from the DIA data, which was then reanalyze using this library. The Flase discovery rate (FDR) for search results was set to < 1% at both the protein and precursor ion levels. The remaining identifications were used for further quantitative analysis. Differentially expressed proteins (with a log2 fold change ≥ +1 or ≤ -1 and FDR ≤ 0.05) were identified using edgeR.

2.4 Metabolite profiling using GC-MS

The samples were subjected to vacuum freeze-drying using a lyophilizer (Scientz-100F) and then ground into powder form using a grinder (MM 400, Retsch) at 30 Hz for 1.5 minutes. Fifty milligrams of powdered sample were weighed using an electronic balance (MS105Dµ), followed by the addition of and then 1200 µL of a -20°C pre-cooled 70% methanolic aqueous extract solution (adjusted to maintain the ratio of 1200 µL extract per 50 mg sample). internal standard was added, the mixture was vortexed every 30 minutes for 30 seconds, repeating this process 6 times. After centrifugating at 12,000 rpm for 3 minutes, the supernatant was collected, filtered through a 0.22 µm microporous membrane,

and stored in injection vial for UPLC-MS/MS analysis. Metabolites levels were normalized using ribitol as an internal standard, and differential expression analysis (log2fold change ≥ +1 or ≤ - 1 and FDR ≤ 0.05) was perform educing edgeR (Robinson et al., 2010).

2.5 Integration of transcriptome, proteome and metabolome data

Based on the results of Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, pathways shared between differentially expressed genes (DEGs) and proteins (DEPs), DEGs and differentially expressed metabolites (DEMs), or DEPs and DEMs were selected to assess their roles in *A. mongolicum* under drought conditions.

3 Results

To elucidate the regulatory mechanisms of *A. mongolicum* under drought stress, we conducted transcriptomic, proteomic, and metabolomic analyses on drought-stressed seedlings. Drought stress negatively impacts multiple biological processes in plants, such as growth and physiological metabolism. Plants counteract drought stress by engaging physiological functions that mitigate, alleviate, or repair damage. Proline (Pro), a compound ubiquitous in plants, serves as a pivotal marker for evaluating drought tolerance, with elevated Pro levels being associated with enhanced drought resilience. Consequently, we conducted physiological assays to assess the drought response characteristics of *A. mongolicum* shoots. Our findings revealed a marked increase in proline activity 24 hours after the onset of drought stress (Supplementary Figure 1).

3.1 Transcriptional dynamics of *A. mongolicum* under drought stress

To elucidate the transcriptional changes elicited by drought stress, RNA sequencing was performed on the roots and shoots of 14-day-old seedlings that were either treated or untreated with PEG6000 for 24 hours. A total of 104.24 Gb of clean data was generated from the 12 samples, with each sample exceeding 7 Gb and a Q30 score greater than 94% (Supplementary Table 1). The differential gene expression analysis, conducted using DESeq2, involved P-value adjustments according to the Benjamini & Hochberg method. Our analysis revealed 7661upregulated and 5346 downregulated genes in the shoots (fold change ≥2 and q-value ≤0.05), as well as 9,592 upregulated and 6,215 downregulated genes in the roots (Figure 1A, Supplementary Tables 2, 3). Hierarchical clustering analysis uncovered distinct expression profiles between the shoots and roots. Furthermore, to assess the expression patterns of differentially expressed genes (DEGs) under varying treatment durations, we aggregated the DEGs from the comparison groups, performed hierarchical clustering, normalized the data using Z-score, and constructed a heatmap (Figure 1B). The heatmap revealed that the 0 h *vs* 24 h comparison groups exhibited opposite trends.

Gene Ontology (GO) enrichment analysis revealed that both upregulated and downregulated genes in the shoots were significantly associated with metabolic pathways, including those involved in redox enzyme activity, α-linolenic acid metabolism, and cytokinin metabolism (Figure 2A). Similarly, in the roots, pathways related to eicosanoid metabolism and cytokinin metabolism were prominently enriched (Figure 2B). Utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, we identified key pathways that were enriched under drought stress conditions. Specifically, in the shoots, genes exhibiting changes in expression were primarily enriched in pathways such as secondary metabolite biosynthesis, glycerophospholipid metabolism, and pentose and glucuronate interconversions (Figure 2C). In the roots, the enriched pathways encompassed nucleotide sugar biosynthesis, glycerophospholipid metabolism, and pentose and glucuronate interconversions (Figure 2D).

In a similar vein, we identified 841 differentially expressed transcription factors (TFs) belonging to 26 TF families under drought conditions. Notably, TF families such as *C2H2*, *AP2/ERF*, *bZIP*, *bHLH*, and *NAC* were highly expressed in both shoots and roots, highlighting their pivotal roles in regulating the plant's response to drought stress (Figure 3A). Moreover, the KEGG enrichment circle plot analysis revealed that differentially expressed genes (DEGs) in both shoot and root tissues were primarily enriched in pathways including KEGG pathway ko1100 (Metabolic pathways), ko01110 (Biosynthesis of secondary metabolites), ko04626 (Plant-pathogen interaction), and ko03010 (Ribosome), suggesting a significant association between amino acid metabolism and oxidative phosphorylation (Figure 3B). The comprehensive analysis indicates that DEGs are particularly enriched in amino acid biosynthesis and carbohydrate metabolism pathways.

3.2 Proteomic analysis of *A. mongolicum* in response to drought stress

To augment the insights gained from the global transcriptome analysis, we undertook a high-throughput proteomic study. This effort yielded a dataset comprising 95,877 peptides, which facilitated the identification and quantification of 7,286 distinct proteins (Supplementary Figure 2). Within this dataset, we identified 1,179 differentially expressed proteins (DEPs), with 773 in the shoots and 406 in the roots, based on a stringent criterion of a fold change ≥2 and a q-value <0.05 (Figure 4A). Hierarchical clustering analyses elucidated the differential protein expression patterns between shoots and roots, revealing distinct protein profiles for each tissue (Figure 4B). Gene Ontology (GO) enrichment analysis indicated that, under drought stress conditions, the shoots exhibited upregulated pathways such as cold acclimation, positive regulation of the response to water deprivation, and peptidyl-threonine dephosphorylation (Figure 5A). Conversely, downregulated pathways in the shoots pertained to photosynthesis, light harvesting in photosystem I, and the metabolic

10.3389/fpls.2024.1515944

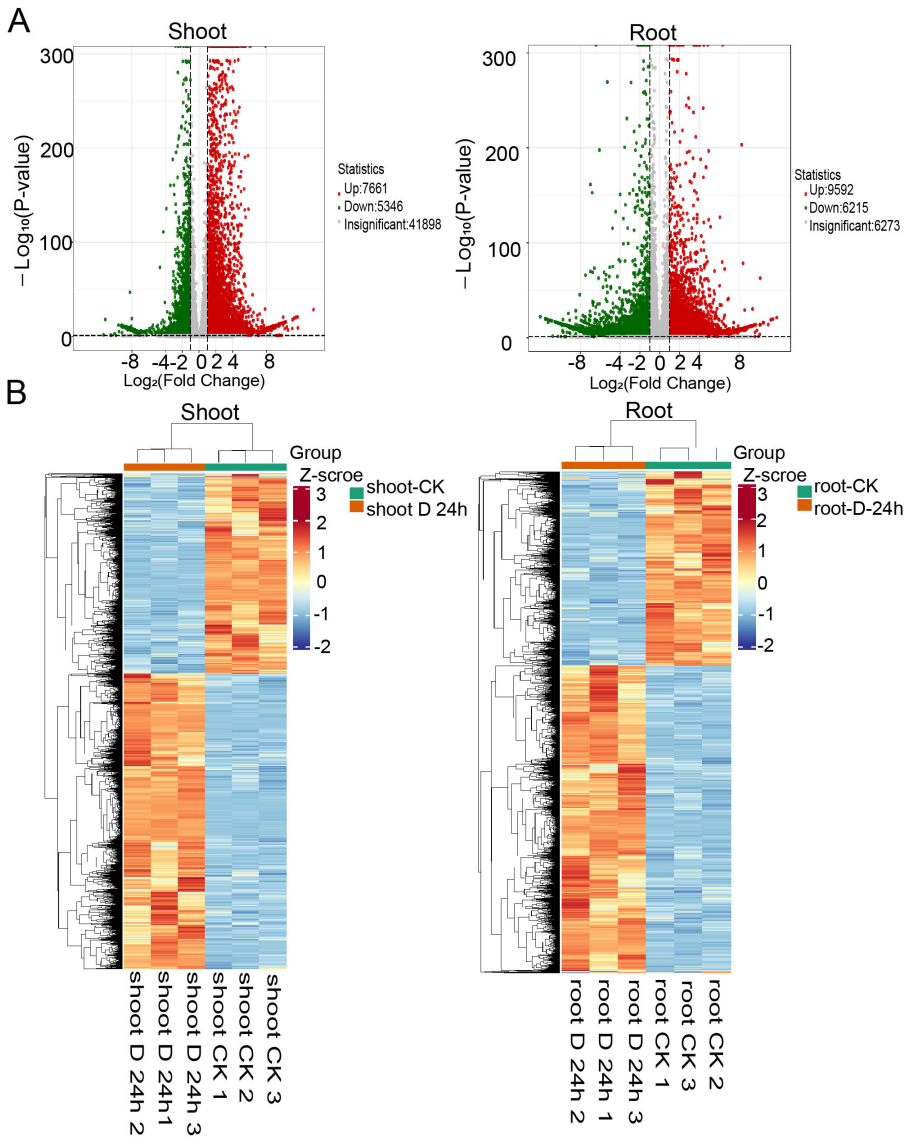


FIGURE 1
Differential expression genes identified under PEG treatment. **(A)** Venn diagrams showing the differentially expressed genes identified in PEG treatment in shoots (left) and roots (right). **(B)** Heatmaps displaying the normalized read counts of differentially expressed genes in PEG treatment in shoots (left) and roots (right).

processes of galacturonan and pectin (Figure 5B). In roots, upregulated pathways were associated with responses to hydrogen peroxide, as well as the degradation of cell wall polysaccharides and hemicellulose (Figure 5C). In contrast, downregulated pathways in the roots involved activities such as fructose-bisphosphate aldolase, sucrose: sucrose 1F-fructosyltransferase, and hexose alcohol dehydrogenase (Figure 5D). These observations suggest that carbohydrate metabolism is profoundly impacted by drought stress, with alterations in both shoot and root tissues (Figures 5A–D).

To delve into the interplay between differentially expressed proteins (DEPs), amino acid biosynthesis, and carbohydrate metabolism, we conducted a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. In the shoots, the upregulated DEPs were predominantly enriched in pathways such as isoflavonoid biosynthesis, arginine and proline metabolism, degradation of valine, leucine, and isoleucine, as well as starch and sucrose metabolism (Figure 6A). Conversely, the downregulated DEPs pertained to pathways including phenylpropanoid biosynthesis, plant hormone signal transduction, and ATP-dependent chromatin remodeling (Figure 6B). In the roots, the upregulated DEPs were implicated in the degradation of valine, leucine, isoleucine, tryptophan, and lysine (Figure 6C), while the downregulated DEPs were associated with pathways such as photosynthesis-antenna proteins, valine, leucine, and isoleucine degradation, histidine metabolism, glycolysis/gluconeogenesis, and carbon metabolism (Figure 6D). The analysis highlights a significant enrichment of DEPs in the pathways of amino acid metabolism and carbohydrate metabolism, suggesting a crucial role for these pathways in the plant's response to drought stress.

10.3389/fpls.2024.1515944

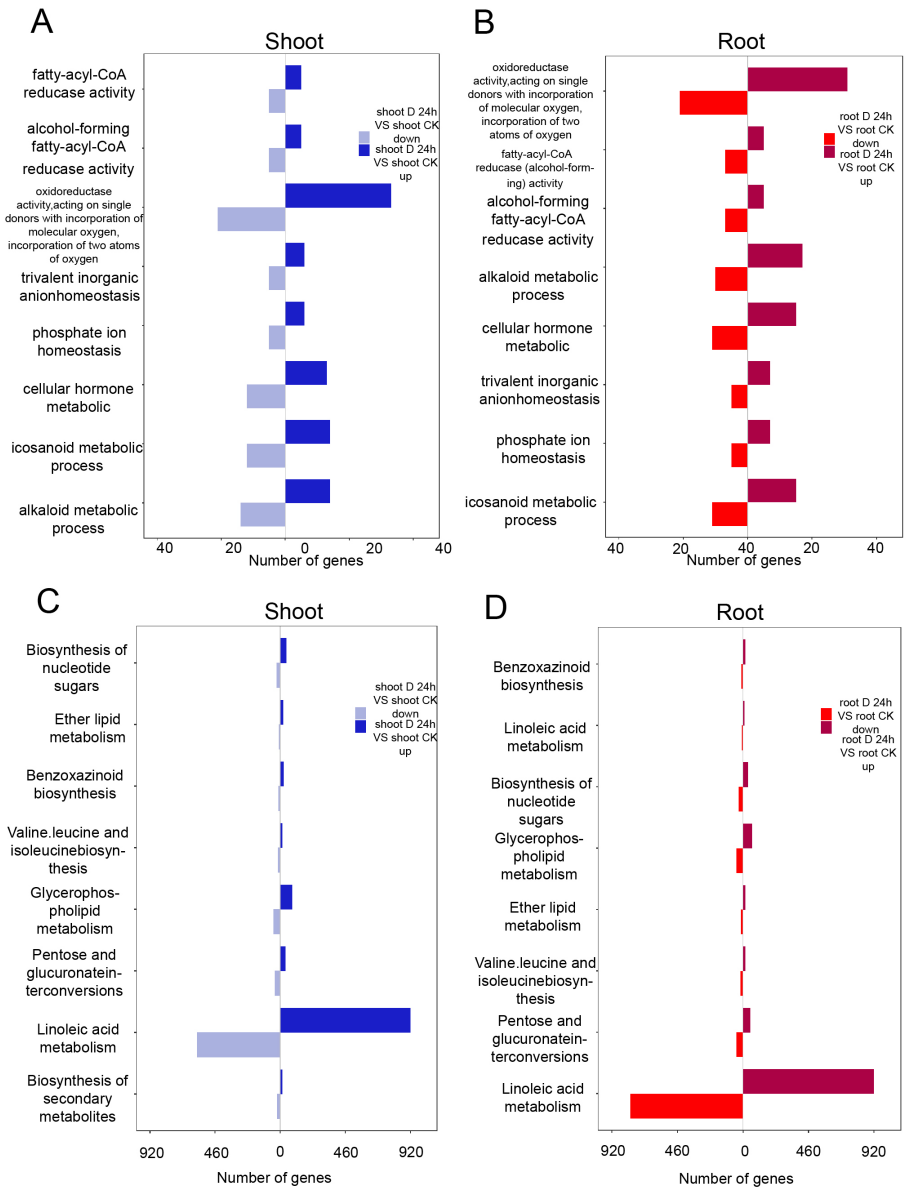


FIGURE 2
Differential expression genes identified under PEG treatment. **(A, B)** GO analysis of the differentially expressed genes in shoots **(A)** and roots **(B)**. **(C, D)** KEGG analysis of the differentially expressed genes in shoots **(C)** and roots **(D)**.

3.3 Integrated analysis of DEGs and DEPs

To explore the relationship between transcriptomic and proteomic responses to drought stress, we combined mRNA expression data from the transcriptome with protein abundance data from the proteome (Figure 7A). A comparative Gene Ontology (GO) enrichment analysis of differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) uncovered critical biological pathways affected by drought stress. In the shoots, both DEGs and DEPs exhibited enrichment in pathways related to the catabolic process of L-phenylalanine, cold acclimation, and the biogenesis of plant-type primary and secondary cell walls. For cellular components (CC), an enrichment in the chloroplast thylakoid membrane was observed, and in terms of molecular functions (MF), the analysis emphasized the pathways involving cellulose synthase (UDP-forming) and cellulose synthase activity.

In roots, both differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) were found to be enriched in biological processes (BP) such as cellulose biosynthesis and the plant-type primary cell wall pathway. The cellular component (CC) analysis highlighted the plasma membrane-anchored component pathway, while the molecular function (MF) analysis showed enrichment in pathways related to secondary active sulfate transporter activity. These findings imply that the identified DEGs and DEPs may be involved in regulating transport proteins, cell wall synthesis, and the production of

10.3389/fpls.2024.1515944

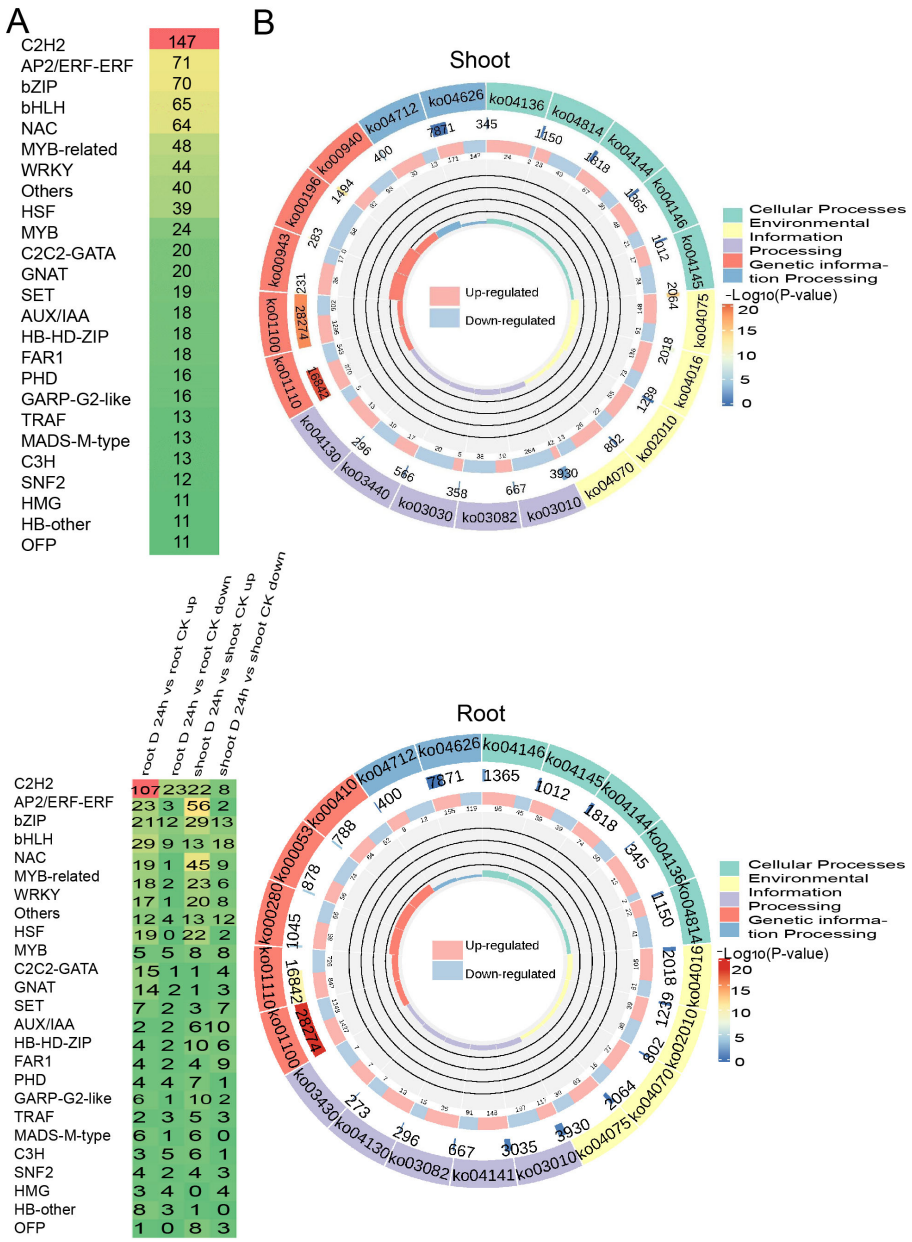


FIGURE 3
Differentially expressed transcription factors (DE-TFs) identified under PEG treatment. **(A)** The distribution of differentially expressed transcription factors (DE-TFs) under drought conditions in both shoot and root, with a family-wise breakdown of the up-regulated DE-TFs in response to drought stress. **(B)** KEGG Enrichment Circle Plot of Differential Genes expressed genes under drought conditions in the shoot (above) and root (below).

metabolic products (Figure 7B). Notably, under drought stress, there was significant enrichment of pathways such as cold acclimation, plant-type primary cell wall process, secondary cell wall biogenesis, and nitrate assimilation in both DEGs and DEPs, emphasizing the significance of carbohydrate and amino acid metabolism in the plant's response to drought.

Further Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of DEPs and DEGs indicated that in the shoots, the primary enrichment occurred in pathways related to flavonoid and flavanol biosynthesis, starch and sucrose metabolism, glycerophospholipid metabolism, and the photosynthesis-antenna proteins pathway. In contrast, the roots showed enrichment in pathways such as photosynthesis-antenna proteins, carbon metabolism, nitrogen metabolism, cofactor biosynthesis, fatty acid metabolism, and phenylpropanoid biosynthesis. These results underscore the importance of starch and sucrose metabolism, carbon metabolism, nitrogen metabolism, and phenylpropanoid biosynthesis in the plant's response to drought stress, with a notable differential expression of genes and proteins that are integral to carbohydrate metabolism and amino acid metabolism (Figure 7C).

10.3389/fpls.2024.1515944

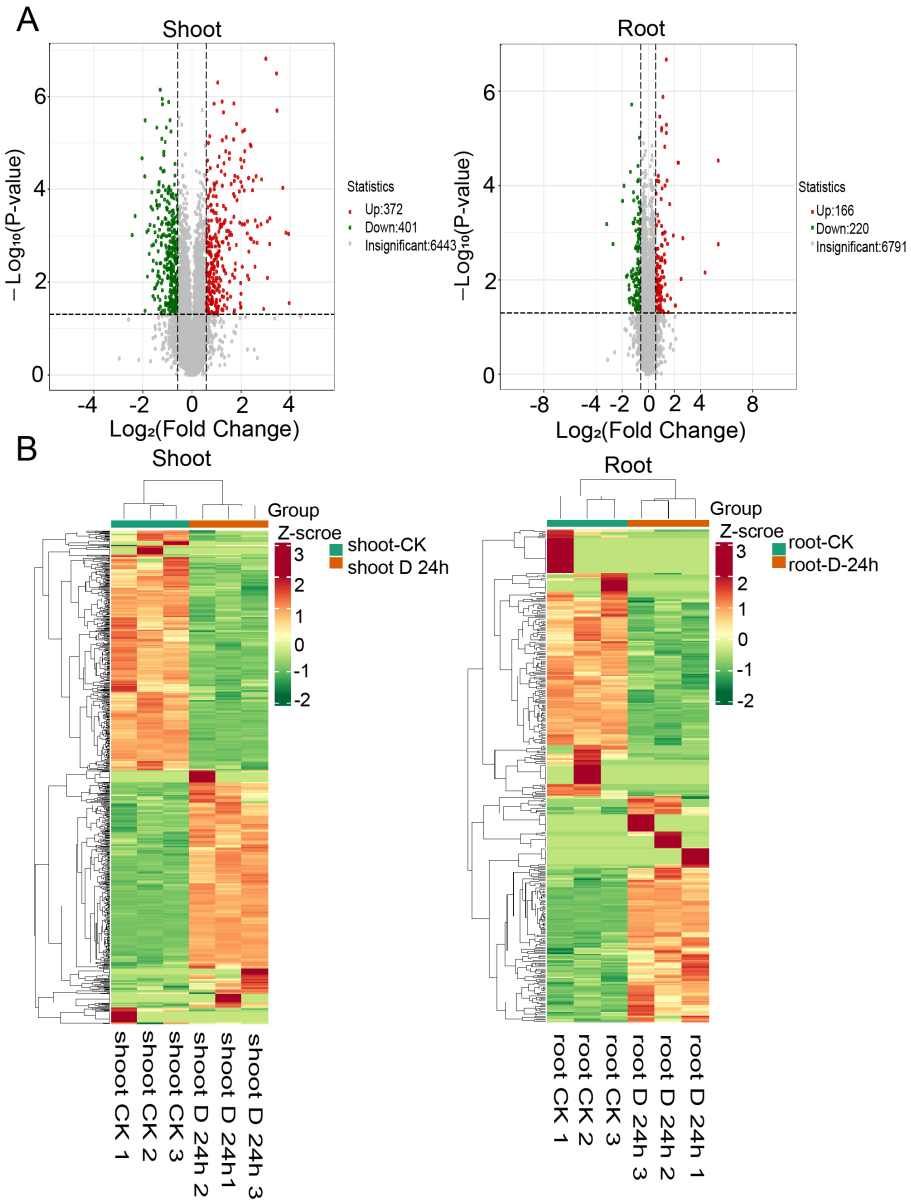


FIGURE 4
Proteome response of shoot and root under drought stress conditions. **(A)** Volcano plot of differentially expressed proteins (log2 fold change≥1 or ≤1 and P-value ≤ 0.05) under drought (DS) and control (CK) conditions in shoots (left) and roots (right) **(B)** Heatmaps displaying the normalized read counts of differentially expressed proteins in PEG treatment in shoots (left) and roots (right).

3.4 Metabolite analysis under drought stress

To connect transcriptomic and proteomic findings with metabolic pathways, we investigated the changes of major metabolites accumulation in shoots and roots of *A. mongolicum* under drought conditions. Employing ultra-high-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS), we generated detailed metabolite profiles. Principal component analysis (PCA) revealed that the expression profiles of differentially expressed metabolites (DEMs) were distinct across the samples, with the exception of the shoot-CK and root-CK samples (Figure 8A). The shoot-D-24h and root-D-24h samples were distinctly separated from the quality control (QC) group based on principal component 1 (PC1), which accounted for 57.93% of the variance. A heatmap depicted the differential regulation of various metabolite categories under drought conditions (Figure 8B). Furthermore, a venn diagram showed that out of the DEMs, 121 were shared between shoots and roots, with 230 unique to shoots and 303 unique to roots (Figure 8C). The volcano plot in illustrated that there were 241 upregulated DEMs and 110 downregulated DEMs in the shoots, whereas in the roots, there were 110 upregulated DEMs and 323 downregulated DEMs (Figure 9A). KEGG pathway enrichment analysis showed that the upregulated DEMs in shoots were involved in pathways such as

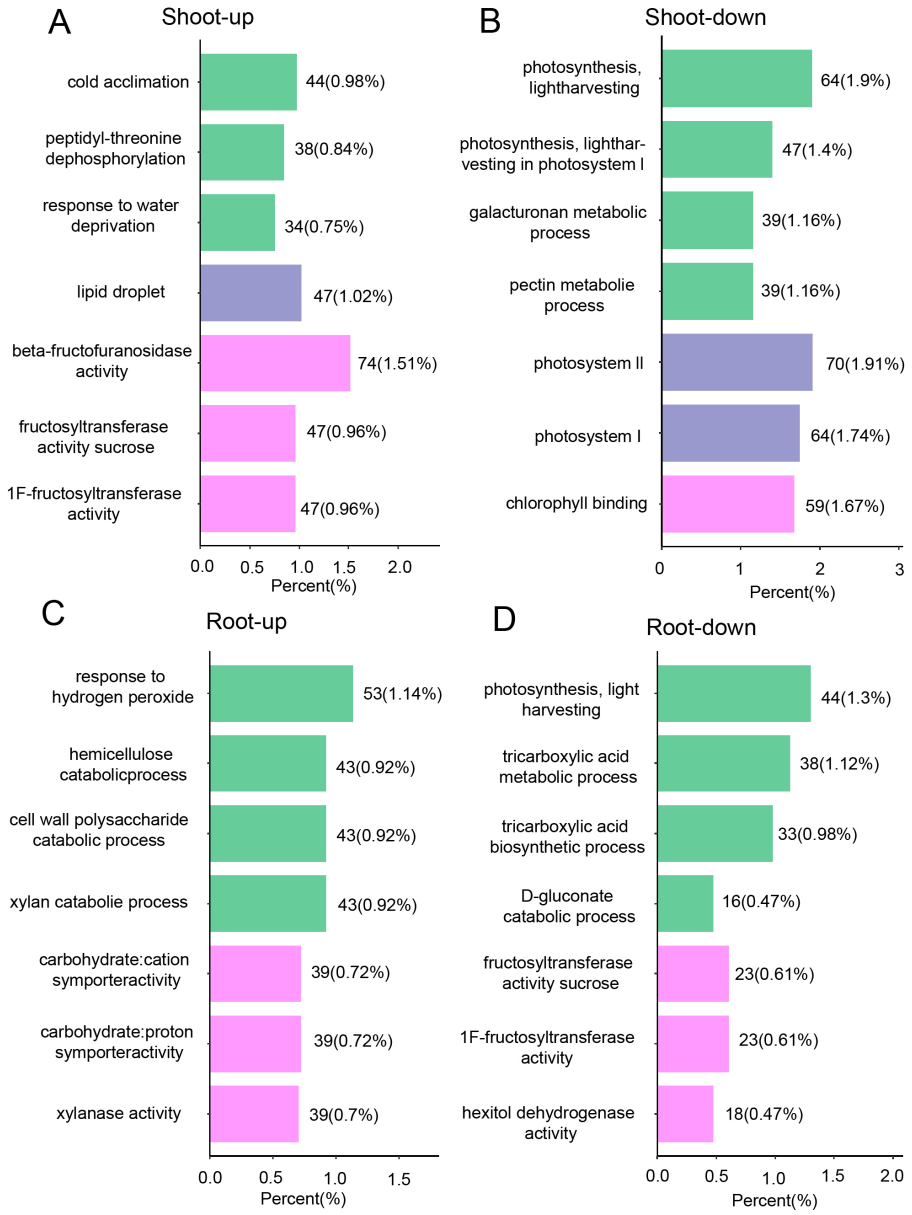


FIGURE 5
Proteome response of shoot and root under drought stress conditions. GO analysis of differentially expressed genes that are up-regulated **(A)** and down-regulated **(B)** in shoot, as well as those that are up-regulated **(C)** and down-regulated **(D)** in root. The green purple, and pink terms present biological process (BP), cellular component, and the molecular function (MF) category, according to GO analysis, respectively.

phenylalanine metabolism, amino acid biosynthesis, ABC transporters, and carotenoid biosynthesis. In roots, upregulated DEMs were linked to glycolysis/gluconeogenesis, galactose metabolism and linoleic acid metabolism, while downregulated DEMs were associated with propanoate metabolism, ABC transporters, and pantothenate and CoA biosynthesis (Figure 9B). After analyzing samples from CK and drought-treated groups (Shoot-D-24h and Root-D-24h), significantly differential metabolites were identified. In the CK and Shoot-D-24h samples, major upregulated metabolites included lipids (e.g., Beta-Hydroxypalmitic Acid), amino acids and derivatives (e.g., asn-pro-lys, aspl-pro, leu-Pro), and alkaloids (e.g., demissine, 6-Methylnicotinamide, alanine betaine) (Supplementary Table 4). For the CK and Root-D-24h samples, significantly downregulated metabolites included phenolic acids (e.g., 5-O-Galloyl-D-hamamelose*, 2-O-P-Coumaroylhydroxycitric Acid, 1,2-O-Diferuloylglycerol*), flavonoids (e.g., Tricetin 3'-glucuronide*, Tricin-7-O-(2''-Malonyl) rhamnoside) and others (e.g., 4-Hydroxybenzaldehyde, Sorbitol-6-phosphate, Maltotriose) (Supplementary Table 5). Our findings suggested that metabolites in shoots were largely associated with amino acid metabolism, whereas those in roots were primarily linked to carbohydrate

10.3389/fpls.2024.1515944

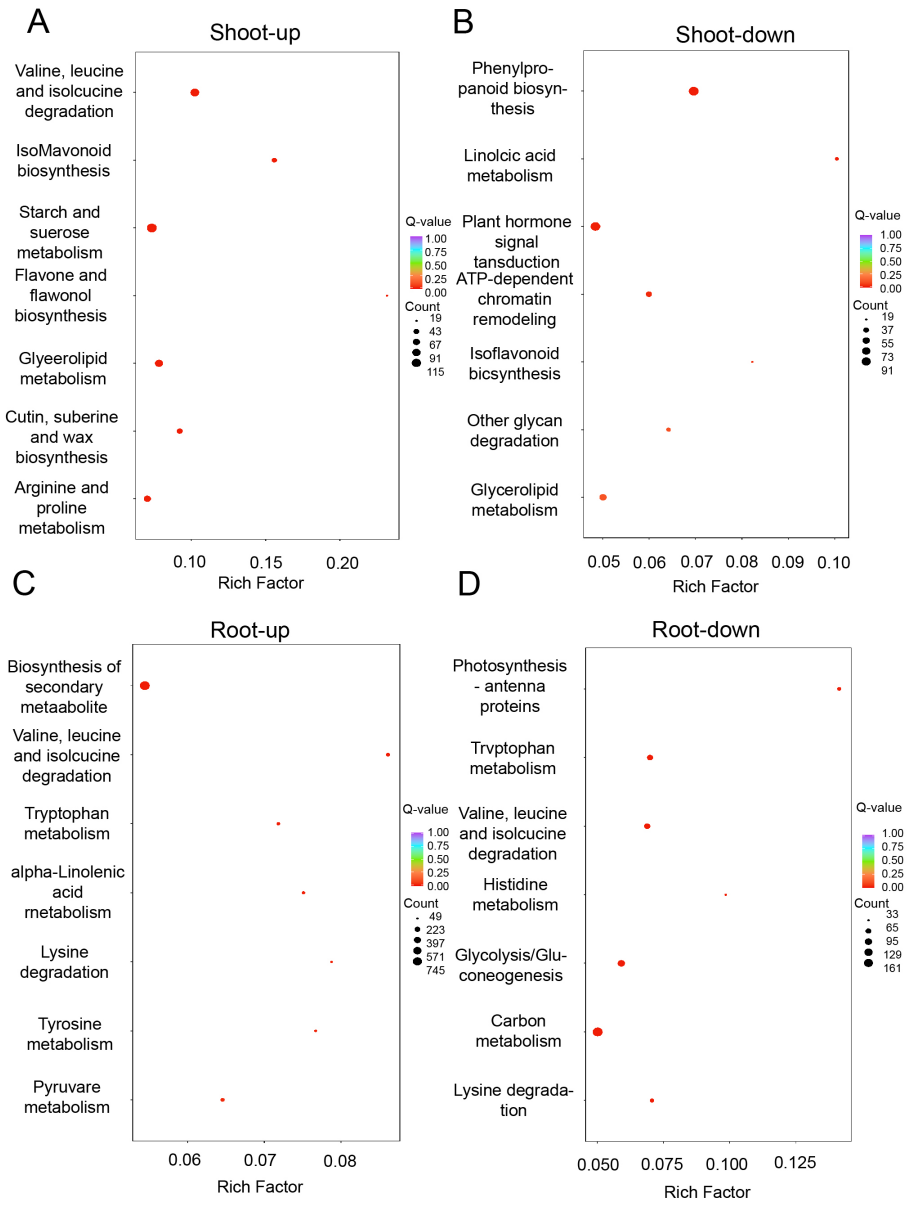


FIGURE 6
Proteome response of shoot and root under drought stress conditions. KEGG analysis of differentially expressed genes that are up-regulated **(A)** and down-regulated **(B)** in shoot, as well as those that are up-regulated **(C)** and down-regulated **(D)** in root.

metabolism. Overall, the results indicate that metabolites are primarily associated with carbohydrate metabolism, amino acid metabolism.

3.5 Integrated analysis of transcriptomics, proteomics, and metabolomics

To gain a deeper understanding of the interplay among transcriptomics, proteomics, and metabolomics datasets, an integrated multi-omics analysis was conducted using a functional relatedness framework. The differentially expressed genes (DEGs, 26,131), differentially expressed proteins (DEPs, 1,179), and differentially expressed metabolites (DEMs, 775) identified in both shoots and roots were concurrently mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Supplementary Tables 6–8), which allowed for the elucidation of their functional interconnections.

A KEGG pathway enrichment analysis was performed, and the bubble chart revealed that in the shoots, the most prominently enriched pathways included the degradation of valine, leucine, and isoleucine, galactose metabolism, arginine and proline metabolism, arginine biosynthesis, tryptophan metabolism, and phenylalanine metabolism. In contrast, the roots exhibited significant enrichment in pathways such as starch and sucrose metabolism, the pentose phosphate pathway, phenylalanine metabolism, carbon metabolism, and glycolysis/gluconeogenesis. Specifically, the interplay among genes, proteins, and metabolites was analyzed for

10.3389/fpls.2024.1515944

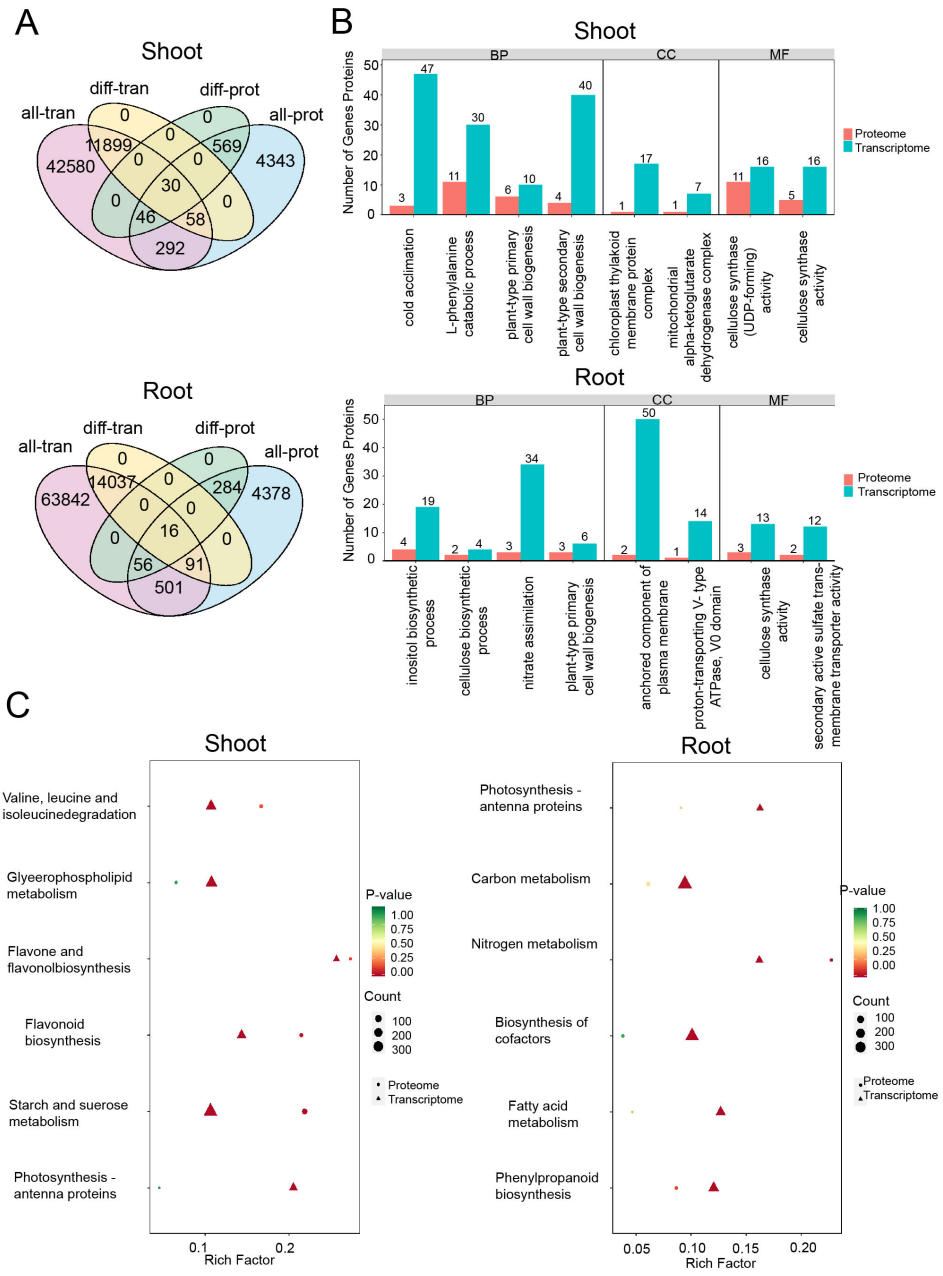


FIGURE 7
DEGs/DEPs identified in drought conditions. **(A)** Correlated genes and proteins expressed in drought conditions. all_ proteins/transcripts present the numbers of total proteins/transcripts identified under PEG treatments. diff_ proteins/transcripts present the numbers of differentially expressed proteins/transcripts identified in the treatments. **(B)** GO analysis of the differentially expressed genes in shoots (above) and roots (below). **(C)** KEGG analysis of the differentially expressed genes in shoots (left) and roots (right).

the arginine and proline metabolism pathway and the pentose phosphate pathway (Figure 10). The findings suggested that under drought conditions, amino acid metabolism is predominantly enriched in the shoots, whereas carbohydrate metabolism is mainly enriched in the roots.

Moreover, the KEGG pathway maps illustrated that the arginine and proline metabolism pathway showed significant differences in the expression of related genes and proteins, including L-Arginine-P, Agmatine, Putrescine, and S-Adenosyl-L-methionine. In response to drought stress, the pentose phosphate pathway exhibited considerable changes in metabolite expression, with notable alterations observed in metabolites such as β-D-Fructose-1,6-bisphosphate, D-Erythrose 4-phosphate, D-Ribose 5-phosphate, and D-Glucose 6-phosphate (Supplementary Figure 3).

3.6 Differential genes under drought stress

This study employs integrated transcriptomic, proteomic, and metabolomic analysis techniques to systematically investigate the

10.3389/fpls.2024.1515944

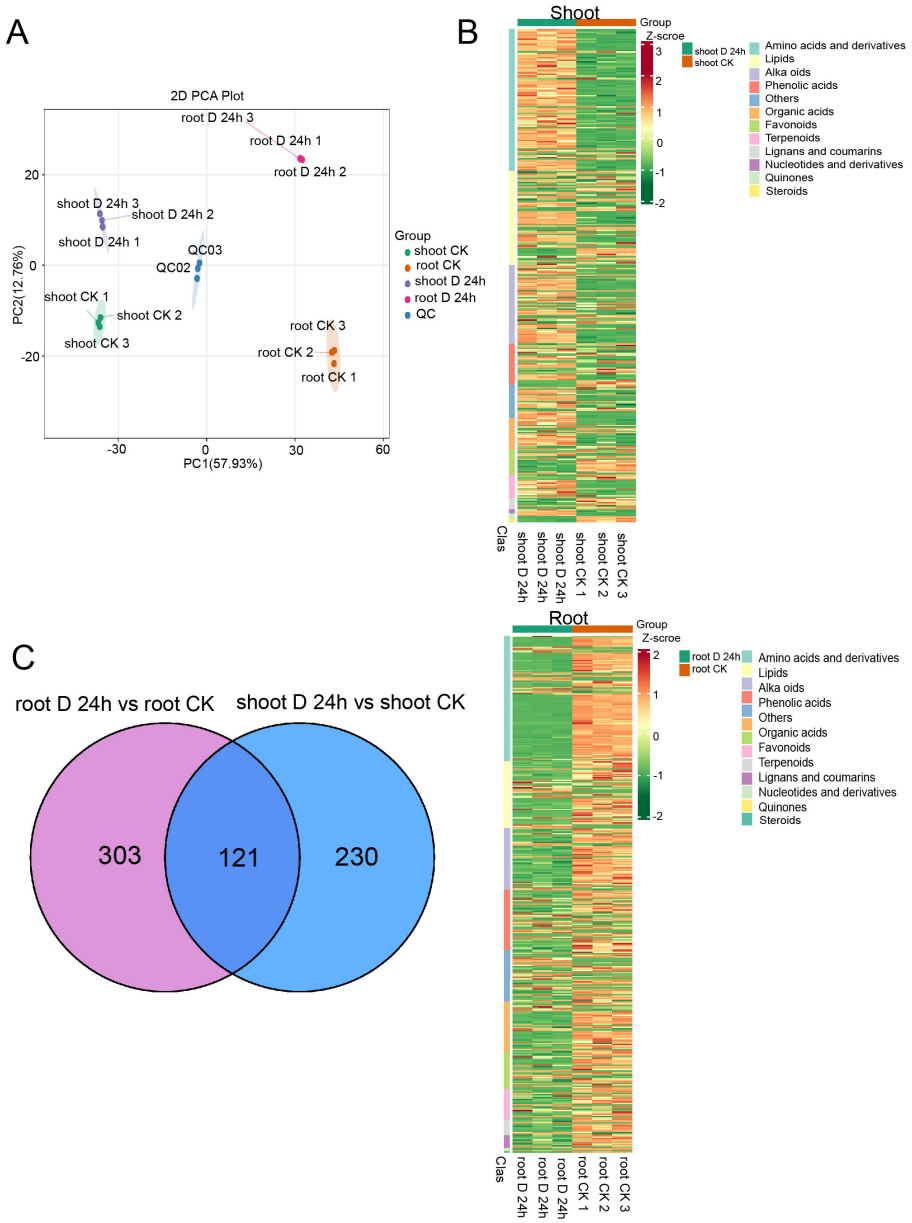


FIGURE 8
Metabolite response of shoot and root under drought stress conditions. **(A)** Principal components analysis (PCA) of the differentially expressed metabolites. **(B)** Heatmaps displaying the normalized read counts of differentially expressed proteins in PEG treatment in shoots (above) and roots (below). **(C)** Venn diagrams showing the differentially expressed metabolites identified in PEG treatment in shoots (right) and roots (left).

regulatory pathways of *A. mongolicum* under drought stress. Initially, in the proteomic screening, the GO database was analyzed for biological processes, cellular components, and molecular functions, identifying 16 highly expressed differentially expressed proteins (DEPs) associated with drought resistance. By querying the TAIR (https://www.arabidopsis.org/) website, the homologous gene families for these proteins were identified, including 14779 (*ASNS3*), 18687 (*ALDH7*), 18892 (*ALDH7B4*), 19487 (*AtUGT85A2*), 20718 (*TA*), 22603 (*SLC52A*), 17999 (*UDP-glucose*), 23189 (*AtXTH13*), 25241 (*AtOEP16-2*), 25473 (*GAPDH*), 25717 (*MBF1*), 994 (*ABCB*), 2543 (*LKR*), 3137 (*P53*), 8113 (*P5CS1*), and 17705 (*AtCIMS*). Following a combined analysis of transcriptomic, proteomic, and metabolomic data, four genes were selected: *ALDH7B4*, *ASNS3*, *P5CS1*, and *LKR*. The study revealed that each of these genes has corresponding differentially expressed genes (DEGs) and differentially expressed metabolites (DEMs). To validate the RNA-seq results for *ASNS3, P5CS1, LKR*, and *ALDH7B4* under drought stress, we performed an expression validation analysis using qRT-PCR. The results indicated that *ASNS3* expression was significantly upregulated at 24 h post-drought treatment relative to the control, while the expression levels of *P5CS1, LKR*, and *ALDH7B4* remained unchanged (Supplementary Figure 4). The KEGG Orthology (KO) numbers associated with these genes are Ko00280, Ko00310, Ko00330,

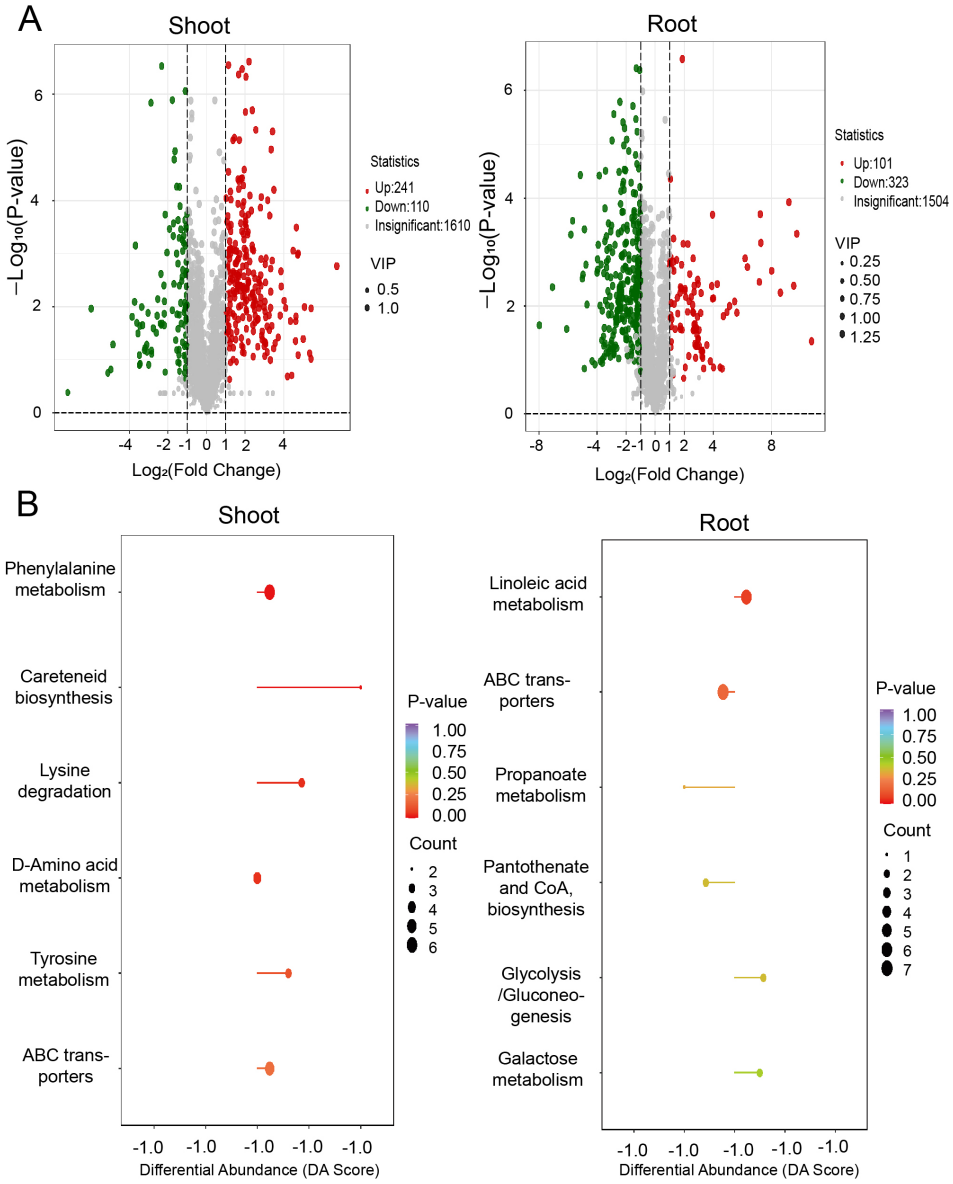


FIGURE 9
Metabolite response of shoot and root under drought stress conditions. **(A)** Venn diagrams showing the differentially expressed metabolites identified in PEG treatment in shoots (left) and roots (right). **(B)** KEGG analysis of the differentially expressed metabolites in shoots (left) and roots (right).

Ko00380, and Ko01230 (Supplementary Figure 5 and Supplementary Table 9). Additionally, the study discovered that *ASNS3*, *P5CS1*, and *LKR* are primarily enriched in amino acid biosynthesis pathways and their derivatives, whereas *ALDH7B4* is mainly associated with the pentose phosphate pathway.

4 Discussion

Drought is a major environmental stressor affecting plant growth. Therefore, elucidating the mechanisms of drought stress tolerance in *A. mongolicum* is important to enhance plant stress resistance. Multi-omics analysis has become established as an important analytical tool for understanding biochemical processes. The variations in amino acids profiles under distinct water stress conditions mirrored those found in *Lotus japonicus* (Sanchez et al., 2012) and maize plants (Alvarez et al., 2008). Proline levels are associated with severe water shortages in numerous plant species (Witt et al., 2012), with variations being genotype-specific and contingent upon the degree of water stress (Bowne et al., 2012). As a recognized compatible solute, prolines crucial for osmotic adjustments; It shields cellular structures during water stress periods and plays a significant role in neutralizing reactive

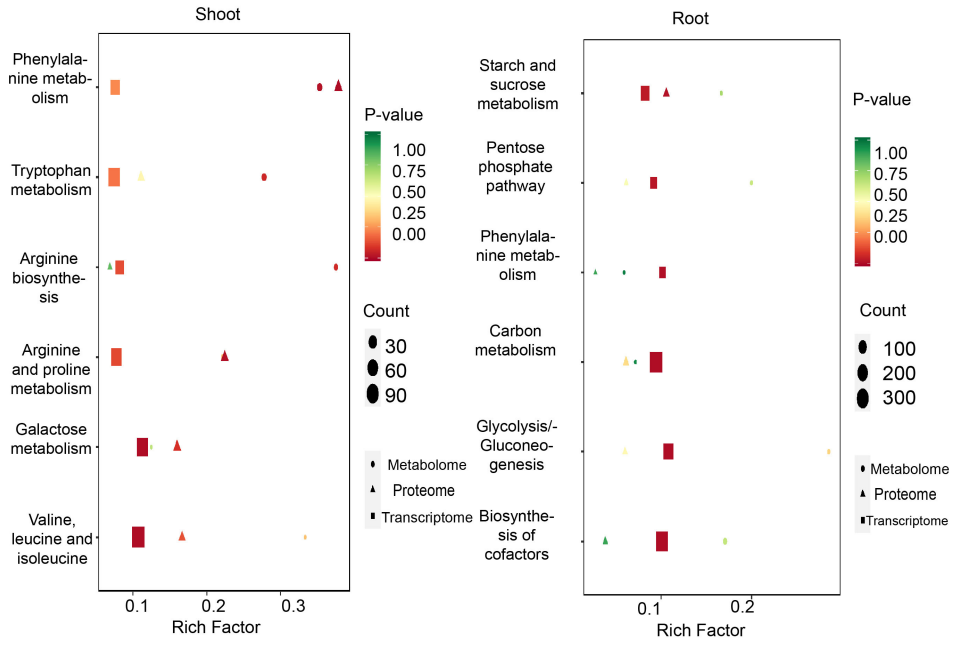


FIGURE 10
DEGs, DEPs and DEMs identified in drought conditions. KEGG analysis of the DEGs、DEPs and DEMs in shoots (left) and roots (right).

oxygen species (ROS),thereby mitigating the negative impacts of drought stress on plant metabolism (Zadehbagheri et al., 2014).

4.1 Arginine and proline as determinative amino acids in drought stress

Amino acids are pivotal in modulating plant stress resilience to stress by affecting osmotic balance, ion movement, stomatal regulation, and chemical equilibrium. Proline, known for its role as an osmotic regulator, can enhances plant tolerance and defend cells against various abiotic stresses (Jia et al., 2016). In addition to functioning as an effective Osmo protectant, proline is involved in multiple biochemical processes, especially under challenging environmental conditions. It also plays essential roles in neutralizing free radicals, regulating cellular redox status, sequestering metal ions, and triggering protective responses within plants (Raza et al., 2023). Proline is synthesized from glutamate, a precursor, by the enzymes pyrroline-5-carboxylate synthase (*P5CS*) and pyrroline-5-carboxylate reductase (*P5CR*) (Hu et al., 1992). Inducing *P5CS* expression to elevate proline concentrations enables *Oryza sativa* to tolerate increased salt and water stress (Su and Wu, 2004). While arginine, an essential amino acid, serves diverse function in plants, providing crucial nitrogen and supporting cellular processes that enhance resistance to various abiotic stresses (Wang et al., 2024). Arginine has been demonstrated to stimulate the synthesis of proline (Bokhary et al., 2020).

Recent studies have highlighted the roles of urea cycle intermediates, including ornithine, aspartate, arginine, and citrulline, in the mechanisms that confer plant tolerance to abiotic stress (Blume et al., 2019; Kalamaki et al., 2009; Shi et al., 2013; Song et al., 2020). In peanut plants, aspartate, a precursor to arginine, also showed a decrease under most abiotic stress conditions. The significance of urea cycle intermediates, including ornithine, aspartate, arginine and citrulline, in enhancing plant resilience to abiotic stress has been increasingly recognized, leading to increased production of other amino acids (Matysiak et al., 2020; Song et al., 2020). Exogenous application of arginine has been shown to increase both fresh and dry weight of shoots and roots in maize plants, and pre-treatment with arginine has similarly effect on stressed sunflower plants. The stimulation effect of arginine on water deficit-stressed plants may be attributed to its role as an essential amino acid that promotes plant growth (Faraji and Sepehri, 2019). Arginine is crucial for plant growth and stress resistance. In some barley varieties, arginine can significantly enhance the morphological growth characteristics. Its application to *Helianthus annuus* L. results in increased branch and root length, as well as a greater number of leaves. Furthermore, pre-treatment with arginine can mitigate the effects of drought on wheat plant growth. Under drought stress, arginine treatment can lead to increased height, tiller number, leaves number, and flag leaf area in barley plants. These effects may be attributed to the conversion of L-arginine to proline and nitric oxide, and crucial for plants to counteract drought stress (Jarzyniak and Jasinski, 2014; Ramadan et al., 2019).

Drought stress triggers hyperosmotic stress in plant cells, negatively impacting growth, development, and productivity (Zhang et al., 2022). To counteract these effects, plants have various mechanisms such as enhancing photosynthesis and water use efficiency, or stimulating the synthesis and accumulation of osmoprotectant small molecules and antioxidant enzymes. Increased proline or soluble sugar accumulation under stress conditions can help maintain cellular osmotic potential (Sun et al., 2003).

After joint analysis, the genes regulated by amino acid and pentose phosphate metabolic pathways include *ASNS3, P5CS1*, and *LKY*, among which *ASNS3* has the highest expression abundance. Therefore, further studying the role of *Asn* gene in plant abiotic stress response is of great significance for improving plant stress tolerance and *sustainable agricultural production.*es are vital for plant adaptation to abiotic stress (Kausar et al., 2022; Kosova et al., 2021, Kosova et al., 2018; Nawae et al., 2020). Amino acids are key molecules for plant growth and development and in response to abiotic stress. Among them, asparagine (*Asn*) has the highest expression level in plants and plays a crucial role in plant stress tolerance. *Asn* is involved in multiple processes in plants, such as nitrogen metabolism, protein synthesis, and storage (Iqbal et al., 2022). Under non-biological stress conditions, including drought stress, plants accumulate *Asn* as a compatible solute (Altman, 2003; Qu et al., 2019), Therefore, further studying the role of *Asn* gene in plant abiotic stress response is of great significance for improving plant stress tolerance and sustainable agricultural production. Under drought stress, ASN and SNP treatment enhances drought resilience in plants by increasing osmolyte concentration, relative water content, and leaf water potential, promoting osmolyte synthesis. Our research shows that *A. mongolicum* resists drought stress by accumulating more *Asn*. Akin et al (Akin and Kaya, 2024). research found that the synergistic effect of *Asn* and *SNP* on osmolyte synthesis and water status is more pronounced, suggesting their potential as a complementary strategy for improving plant drought resistance. The stronger synergistic effect of *Asn* and *SNP* on osmolyte production and water status indicates their potential as a complimentary approach to enhancing plant drought resistance. It is likely that *Asn* and NO interact with each other and the plant in certain ways that increase the plant's tolerance to drought, which might account for the synergistic action of these two compounds in enhancing the water relations of cotton plants during drought stress. *Asn* plays a role in osmoregulation and supports the maintenance of cellular turgor under drought stress (Schwendner et al., 2018; Yadav et al., 2019).

A variety of plant species harness the key function of *Asn* in response to abiotic stress. Under stress conditions, increased activities of asparagine synthetase (*AS*) and glutamate dehydrogenase help counteract NH^{4+} accumulation. Asparagine (*Asn*), a product of *AS*, plays a crucial role in nitrogen storage and transport due to its stability and high N:C ratio (Lam et al., 1996). In energy-limited conditions such as salinity, where glutamine synthetase (GS) and glutamate synthase (GOGAT) are inhibited and *AS* is activated, *Asn* is involved in nitrogen recycling and flow in plant cells, promoting nitrogen assimilation into *Asn*, which is rich in nitrogen and suitable for long-distance transport or long-term storage. Consequently, the upregulation of *OsAS1* enhances salt tolerance and grain yield in rice under salt stress conditions. Furthermore, *Asn*, the main N-transporter in alfalfa, accumulates in nodules and has been associated with N feedback, which inhibits symbiotic N_2 fixation (King and Purcell, 2005; Larrainzar et al., 2009; Serraj et al., 2001). Studies have shown lower transpiration may alter the long-distance transport of nitrogen compounds, leading to an accumulation of nitrogen compounds in nodules under water stress. However, recent studies have not found any accumulation of nitrogen compounds in pea nodules under artificial reduction of plant transpiration, indicating that transpiration affects the long-distance transport of metabolites and its effects on the drought-induced inhibition of symbiotic nitrogen fixation (Aldasoro et al., 2019). In our study, we found that under drought stress, the accumulation of *Asn* in *A. mongolicum* significantly increased and the abundance of *Asn* is about five times that of other genes, playing an important regulatory role in the biosynthesis metabolism pathway of amino acids. *Asn* can participate in regulating plants' responses to drought stress, helping plants better adapt to drought environments.

Moreover, the variable number of genes encoding *ASNS* among plant species, such as the single gene identified in alfalfa (Shi et al., 1997), rice (Nakano et al., 2000), soybean (Yamagata et al., 1998), and asparagus (Davies and King, 1993). Multiple *ASNS* genes have been found in pea (Tsai and Coruzzi, 1990),sunflower (Herrera-Rodríguez et al., 2002, Herrera-Rodríguez et al., 2004), soybean (Hughes et al., 1997), barley (Moller et al., 2003) and *Arabidopsis*. Some *ASNS* genes (e.g. pea *AS1*, *AtASN1*, sunflower *HAS1*, *HAS1.1*) are negatively regulated by light and sugars, being primarily expressed in dark-grown plants (Herrera-Rodríguez et al., 2004). RNA profiling of wheat identified three putative *ASNS* genes that were up-regulated in response to drought stress (Mohammadi et al., 2007). The result is the same as the experimental results. Under severe drought conditions, *ASNS*, which accumulates in ageing leaves, suggesting drought-induced senescence (Begona Herrera-Rodriguez et al., 2006; Eason et al., 2000). In a recent study, metabolites belonging to the aspartate pathway (including *Asn*, Ser, and Met) were reported as biomarkers for yield gap-based drought tolerance, accurately predicting more than 94% of drought tolerance in wheat (Yadav et al., 2019).

Drought responses in plants are closely related to the enzymes *P5CS* and *ProDH* involved in proline metabolism, which are influenced by soil moisture. H_2O_2production is also related to soil moisture across various stages of development, emphasizing the importance of considering seedlings and soil moisture conditions when studying drought stress. These results highlight the significant influence of soil moisture treatment on plant drought stress and provide insights into the underlying mechanisms. H_2O_2, as a ROS, plays a crucial role in intracellular communication and participates in plant adaptation to specific conditions. Oxidative stress caused by unfavorable environmental conditions, such as heat or drought, results in excessive ROS accumulation and triggers protective mechanisms, including proline accumulation through the upregulation of *P5CS* activity and the decrease of *ProDH* activity. Thus, H_2O_2 is involved in proline metabolism as a regulatory signaling molecule (Ben Rejeb et al., 2014; Uchida et al., 2002; Wen et al., 2013; Yang et al., 2009). This conclusion is consistent with the experimental results.

Drought resistance mechanisms in plants are diverse, with increasing root length and osmotic pressure regulatory substances

being common strategies. as a crucial osmotic regulatory substance, proline plays a pivotal role in lowering the water potential of plant cells, strengthening their water absorption and retention capabilities, and thus safeguarding cellular osmotic balance and subcellular structural stability (Trovato et al., 2008). *P5CS* genes, a key role in plant proline biosynthesis, have been reported in a variety of plant species, such as A. thaliana, Cajanus cajan, N. benthamiana, and Oryza sativa (Guo et al., 2024; Igarashi et al., 1997; Ku et al., 2011; Su and Wu, 2004; Surekha et al., 2014) P5CS, a key enzyme in the proline biosynthesis pathway, not only enhances root growth but also plays a crucial role in plant resistance to drought. The electrical conductivity of cell leakage provides an indirect assessment of cell membrane damage (Bajji et al., 2002). *OE-SpP5CS* in *A. thaliana* exhibited a lower degree of membrane damage compared to the wild type (WT). Transgenic rice with *P5CS* increased their proline content under water shortage conditions and demonstrated an increased fresh shoot weight of 50-95% after PEG treatments (Su and Wu, 2004). The overexpression of *SpP5CS* in *A. thaliana* augments the proline content to enhance drought stress resistance. Drawing upon the results from heterologous expression in A. thaliana and qPCR in S. purpurea, it can be concluded that SpP5CS enhances the drought resistance of plants by boosting proline production.

LKR is a key enzyme in the lysine metabolism pathway, involved in the conversion of lysine to alpha-amino hexanoic acid, which is part of the saccharopine pathway. Lin et al (Lin et al., 2023). study identified proline as a crucial metabolite in grapevines' responses to water stress. The study also discovered that the expression levels of genes associated with proline synthesis, such as *P5CS* and *LKR/SDH*, are upregulated during drought stress. Furthermore, the research suggests that grapevines bolster their drought resistance through the production of proline, a process mediated by *P5CS* and *LKR/SDH* enzymes (Degu et al., 2019; Hochberg et al., 2013)

4.2 Auxiliary pentose phosphate pathway under drought stress

In this study, several genes screened are involved in the glycolysis and pentose phosphate pathways. The pentose phosphate pathway (PPP) is closely linked to glycolysis and contributes to numerous metabolic pathways, contributing to numerous metabolic pathways. Sugars are vital participants in alleviating plant tolerance to abiotic stress (Keunen et al., 2013). Carbohydrate metabolism serves as a primary pathway in regulating cellular carbon and energy demands during drought stress, particularly through the accumulation of water-soluble sugars, which is of great significance for the induction of drought resistance in rice. Research on PEG-induced drought responses have demonstrated alterations in carbohydrate metabolism across various plant species, including lilac, pitaya, lentil, soybean, perennial ryegrass, alfalfa, sorghum and tomato (Abdel-Ghany et al., 2020; Cao et al., 2018; Foti et al., 2021; Guo et al., 2020a; Li et al., 2015; Siddiqui et al., 2021; Wang et al., 2021; Zhang and Shi, 2018; Zhou et al., 2022).

As knows *ALDH7B4* is primarily enriched in the pentose phosphate pathway. Expression of plant *ALDH7B4* is responsive to turgor and can protect cells from oxidative stress. In *Arabidopsis thaliana*, the *ALDH7B4* protein is induced by ABA, osmotic, and wound stress, as well as transgenic overexpression, leading to osmotic and oxidative stress tolerance (Hou and Bartels, 2015). Ectopic expression of the soybean *ALDH7* gene in tobacco and *A.thaliana* reduces MDA levels and sensitivity to hydrogen peroxide and methyl viologen, while rice *ALDH7 T-DNA* insertion mutants show increased sensitivity to various stresses (Kirch et al., 2004; Kotchoni et al., 2006; Missihoun et al., 2014). These findings collectively indicate that *ALDH7B4* plays a crucial role in plant adaptation and tolerance to abiotic stress (Shin et al., 2009; Suzuki et al., 2016).

5 Conclusion

This study conducted an analysis of transcriptome, proteome, and metabolome data from the shoots and roots of *A. mongolicum* seedlings to elucidate the underlying mechanisms of drought response and tolerance. The study found that arginine, proline, and Pentose Phosphate pathway metabolites play significant roles in the drought resistance of A. mongolicum. The comprehensive analysis of the integrated data revealed a synchronized response of genes, proteins, and metabolites associated with amino acid and pentose phosphate metabolic pathways in A. mongolicum under drought stress. This integrated approach is expected to enhance detailed investigations into the mechanisms of abiotic stress tolerance and response in A. mongolicum and potentially in other plant 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 authors.

Author contributions

XM: Writing – original draft; Methodology. QL: Writing – review & editing, Investigation. YH: Writing – review & editing, Investigation. LF: Writing – review & editing, Formal analysis. DY: Writing – review & editing, Supervision, Validation. LM: Writing – review & editing, Supervision, Validation. JT: Writing – review & editing, Validation. XM: Writing – review & editing, Validation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Inner Mongolia Seed Industry Science

and technology innovation major demonstration project (2022JBGS0014), the earmarked fund for China Agriculture Research System (CARS-34) and the National Crop Germplasm Resources Center (NCGRC-63).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2024.1515944/full#supplementary-material

SUPPLEMENTARY FIGURE 1
After 24 hours of drought treatment, the proline content in the seedlings of *A. mongolicum* significantly increased.

SUPPLEMENTARY FIGURE 2
Protein level identification results.

SUPPLEMENTARY FIGURE 3
Under drought conditions, The KEGG pathway maps for Arginine and proline metabolism(left) and Pentose phosphate metabolism(right).

SUPPLEMENTARY FIGURE 4
Expression analysis of ASNS3, P5CS1 LKR and ALDH7B4 genes during drought treatment ing to qRT-PCR.

SUPPLEMENTARY FIGURE 5
correlation-network diagram.

References

Abdel-Ghany, S. E., Ullah, F., Ben-Hur, A., and Reddy, A. S. N. (2020). Transcriptome analysis of drought-resistant and drought-sensitive sorghum (Sorghum bicolor) genotypes in response to PEG-induced drought stress. *Int. J. Mol. Sci.* 21, 772. doi: 10.3390/ijms21030772

Ahmed, H. G. M. D., Sajjad, M., Li, M., Azmat, M. A., Rizwan, M., Maqsood, R. H., et al. (2019). Selection criteria for drought-tolerant bread wheat genotypes at seedling stage. *Sustainability.* 11, 11-17. doi: 10.3390/su11092584

Akin, S., and Kaya, C. (2024). Asparagine and nitric oxide jointly enhance antioxidant capacity and nitrogen metabolism to improve drought resistance in cotton: Evidence from long-term field trials. *Food Energy Secur.* 13. doi: 10.1002/fes3.502

Aldasoro, J., Larrainzar, E., and Arrese-Igor, C. (2019). Application of anti-transpirants temporarily alleviates the inhibition of symbiotic nitrogen fixation in drought-stressed pea plants. *Agric. Water Management.* 213, 193–199. doi: 10.1016/j.agwat.2018.10.014

Altman, A. (2003). From plant tissue culture to biotechnology: Scientific revolutions, abiotic stress tolerance, and forestry. *In Vitro Cell. Dev. Biology-Plant* 39, 75–84. doi: 10.1079/ivp2002379

Alvarez, S., Marsh, E. L., Schroeder, S. G., and Schachtman, D. P. (2008). Metabolomic and proteomic changes in the xylem sap of maize under drought. *Plant Cell Environment.* 31, 325–340. doi: 10.1111/j.1365-3040.2007.01770.x

Amiour, N., Imbaud, S., Clément, G., Agier, N., Zivy, M., Valot, B., et al. (2012). The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. *J. Exp. Botany.* 63, 5017–5033. doi: 10.1093/jxb/ers186

Amirbakhtiar, N., Ismaili, A., Ghaffari, M. R., Firouzabadi, F. N., and Shobbar, Z. S. (2019). Transcriptome response of roots to salt stress in a salinity-tolerant bread wheat cultivar. *PloS One* 14. doi: 10.1371/journal.pone.0213305

Bajji, M., Kinet, J. M., and Lutts, S. (2002). The use of the electrolyte leakage method for assessing cell membrane stability as a water stress tolerance test in durum wheat. *Plant Growth Regulation.* 36, 61–70. doi: 10.1023/a:1014732714549

Begona Herrera-Rodriguez, M., Maldonado, J. M., and Perez-Vicente, R. (2006). Role of asparagine and asparagine synthetase genes in sunflower (Helianthus annuus) germination and natural senescence. *J. Plant Physiol.* 163, 1061–1070. doi: 10.1016/j.jplph.2005.10.012

Ben Rejeb, K., Abdelly, C., and Savoure, A. (2014). How reactive oxygen species and proline face stress together. *Plant Physiol. Biochem.* 80, 278–284. doi: 10.1016/j.plaphy.2014.04.007

Bittencourt, C. B., da Silva, T. L. C., Neto, J. C. R., Vieira, L. R., Leao, A. P., Ribeiro, J. A. D., et al. (2022). Insights from a multi-omics integration (MOI) study in oil palm (Elaeis guineensis jacq.) response to abiotic stresses: part one-salinity. *Plants-Basel.* 11, 19. doi: 10.3390/plants11131755

Blume, C., Ost, J., Muehlenbruch, M., Peterhaensel, C., and Laxa, M. (2019). Low CO2 induces urea cycle intermediate accumulation in Arabidopsis thaliana. *PloS One* 14. doi: 10.1371/journal.pone.0210342

Bokhary, S. U.F., Wang, L., Zheng, Y., and Jin, P. (2020). Pre-storage hot water treatment enhances chilling tolerance of zucchini (Cucurbita pepo L.) squash by regulating arginine metabolism. *Postharvest Biol. Technology.* 166. doi: 10.1016/j.postharvbio.2020.111229

Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., et al. (2012). Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Mol. Plant* 5, 418–429. doi: 10.1093/mp/ssr114

Boyer, J. S. (1982). Plant productivity and environment (crop genetic improvement). *Science.* 218, 443–448. doi: 10.1126/science.218.4571.443

Cao, X., Zhong, C., Zhu, C., Zhu, L., Zhang, J., Wu, L., et al. (2018). Ammonium uptake and metabolism alleviate PEG-induced water stress in rice seedlings. *Plant Physiol. Biochem.* 132, 128–137. doi: 10.1016/j.plaphy.2018.08.041

Che, Y. H., and Li, L. H. (2007). Genetic diversity of prolamines in Agropyron mongolicum Keng indigenous to northern China. *Genet. Resour. Crop Evolution.* 54, 1145–1151. doi: 10.1007/s10722-006-9006-7

Chin, E. L., Ramsey, J. S., Mishchuk, D. O., Saha, S., Foster, E., Chavez, J. D., et al. (2020). Longitudinal Transcriptomic, Proteomic, and Metabolomic Analyses ofCitrus sinensis (L.) Osbeck Graft-Inoculated with "Candidatus Liberibacter asiaticus". *J. Proteome Res.* 19, 719–732. doi: 10.1021/acs.jproteome.9b00616

Davies, K. M., and King, G. A. (1993). Isolation and characterization of a cDNA clone for a harvest-induced asparagine synthetase from Asparagus officinalis L. *Plant Physiol.* 102, 1337–1340. doi: 10.1104/pp.102.4.1337

Degu, A., Hochberg, U., Wong, D. C. J., Alberti, G., Lazarovitch, N., Peterlunger, E., et al. (2019). Swift metabolite changes and leaf shedding are milestones in the acclimation process of grapevine under prolonged water stress. *BMC Plant Biol.* 19. doi: 10.1186/s12870-019-1652-y

Du, J. C., Li, X. Q., Li, T. T., Yu, D. Y., and Han, B. (2017). Genome-wide transcriptome profiling provides overwintering mechanism of *Agropyron mongolicum*. *BMC Plant Biol.* 17, 13. doi: 10.1186/s12870-017-1086-3

Eason, J. R., Johnston, J. W., de Vré, L., Sinclair, B. K., and King, G. A. (2000). Amino acid metabolism in senescing Sandersonia aurantiaca flowers: cloning and characterization of asparagine synthetase and glutamine synthetase cDNAs. *Aust. J. Plant Physiol.* 27, 389–396. doi: 10.1071/PP99200

Esfahanian, E., Nejadhashemi, A. P., Abouali, M., Adhikari, U., Zhang, Z., et al. (2017). Development and evaluation of a comprehensive drought index. *J. Environ. Management.* 185, 31–43. doi: 10.1016/j.jenvman.2016.10.050

Fang, Y., Du, Y., Wang, J., Wu, A., Qiao, S., Xu, B., et al. (2017). Moderate drought stress affected root growth and grain yield in old, modern and newly released cultivars of winter wheat. *Front. Plant Science.* 8. doi: 10.3389/fpls.2017.00672

Faraji, J., and Sepehri, A. (2019).). Ameliorative effects of TiO2nanoparticles and sodium nitroprusside on seed germination and seedling growth of wheat under PEG-stimulated drought stress. *J. Seed Science.* 41, 309–317. doi: 10.1590/2317-1545v41n3213139

Foti, C., Kalampokis, I. F., Aliferis, K. A., and Pavli, O. I. (2021). Metabolic responses of two contrasting lentil genotypes to PEG-induced drought stress. *Agronomy-Basel.* 11, 1190. doi: 10.3390/agronomy11061190

Francki, M. G., Hayton, S., Gummer, J. P. A., Rawlinson, C., and Trengove, R. D. (2016). Metabolomic profiling and genomic analysis of wheat aneuploid lines to identify genes controlling biochemical pathways in mature grain. *Plant Biotechnol. J.* 14, 649–660. doi: 10.1111/pbi.12410

Guo, C.-J., Zhang, T., Leng, Q., Zhou, X., Zhong, J., and Liu, J.-L. (2024). Dynamic Arabidopsis P5CS filament facilitates substrate channelling. *Nat. Plants.* 10, 880-889. doi: 10.1038/s41477-024-01697-w

Guo, T., Tian, C., Chen, C., Duan, Z., Zhu, Q., and Sun, L. Z. (2020a). Growth and carbohydrate dynamic of perennial ryegrass seedlings during PEG-simulated drought and subsequent recovery. *Plant Physiol. Biochem.* 154, 85–93. doi: 10.1016/j.plaphy.2020.06.008

Guo, W., Xin, M., Wang, Z., Yao, Y., Hu, Z., Song, W., et al. (2020b). Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nat. Commun.* 11, 1-12. doi: 10.1038/s41467-020-18738-5

Han, H. M., Liu, W. H., Lu, Y. Q., Zhang, J. P., Yang, X. M., Li, X. Q., et al. (2017). Isolation and application of P genome-specific DNA sequences of *Agropyron* Gaertn. in Triticeae. *Planta.* 245, 425–437. doi: 10.1007/s00425-016-2616-1

Herrera-Rodríguez, M. B., Carrasco-Ballesteros, S., Maldonado, J. M., Pineda, M., Aguilar, M., and Pérez-Vicente, R. (2002).). Three genes showing distinct regulatory patterns encode the asparagine synthetase of sunflower (Helianthus annuus). *New Phytologist.* 155, 33–45. doi: 10.1046/j.1469-8137.2002.00437.x

Herrera-Rodríguez, M. B., Maldonado, J. M., and Pérez-Vicente, R. (2004). Light and metabolic regulation of HAS1, HAS1.1and HAS2, three asparagine synthetase genes in Helianthus annuus. *Plant Physiol. Biochem.* 42, 511–518. doi: 10.1016/j.plaphy.2004.05.001

Hochberg, U., Degu, A., Toubiana, D., Gendler, T., Nikoloski, Z., Rachmilevitch, S., et al. (2013). Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. *BMC Plant Biol.* 13, 183. doi: 10.1186/1471-2229-13-184

Hou, Q., and Bartels, D. (2015). Comparative study of the aldehyde dehydrogenase (ALDH) gene superfamily in the glycophyte *Arabidopsis thaliana* and *Eutrema halophytes*. *Ann. Bot.* 115, 465–479. doi: 10.1093/aob/mcu152

Hu, C. A., Delauney, A. J., and Verma, D. P. (1992). A bifunctional enzyme (delta 1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Nat. Acad. Sci. U. S. A.* 89, 9354–9358. doi: 10.1073/pnas.89.19.9354

Hughes, C. A., Beard, H. S., and Matthews, B. F. (1997). Molecular cloning and expression of two cDNAs encoding asparagine synthetase in soybean. *Plant Mol. Biol.* 33, 301–311. doi: 10.1023/a:1005784202450

Igarashi, Y., Yoshiba, Y., Sanada, Y., YamaguchiShinozaki, K., Wada, K., and Shinozaki, K. (1997). Characterization of the gene for Delta(1)-pyrroline-5-carboxylate synthetase and correlation between the expression of the gene and salt tolerance in Oryza sativa L. *Plant Mol. Biol.* . 33, 857–865. doi: 10.1023/a:1005702408601

Iqbal, A., Gui, H., Wang, X., Zhang, H., Zhang, X., and Song, M. (2022). Genome-wide expression analysis reveals involvement of asparagine synthetase family in cotton development and nitrogen metabolism. *BMC Plant Biol.* 22, 122. doi: 10.1186/s12870-022-03454-7

Jarzyniak, K. M., and Jasinski, M. (2014). Membrane transporters and drought resistance - a complex issue. *Front. Plant Science.* 5. doi: 10.3389/fpls.2014.00687

Jia, X., Sun, C. S., Zuo, Y. C., Li, G. Y., Li, G. B., Ren, L. Y., et al. (2016). Integrating transcriptomics and metabolomics to characterise the response of Astragalus membranaceus Bge. var. mongolicus (Bge.) to progressive drought stress. *BMC Genomics* 17, 16. doi: 10.1186/s12864-016-2554-0

Kalamaki, M. S., Merkouropoulos, G., and Kanellis, A. K. (2009). Can ornithine accumulation modulate abiotic stress tolerance in Arabidopsis. *Plant Signaling Behav.* 4, 1099–1101. doi: 10.4161/psb.4.11.9873

Kausar, R., Wang, X., and Komatsu, S. (2022). Crop proteomics under abiotic stress: from data to insights. *Plants-Basel.* 11, 2877. doi: 10.3390/plants11212877

Keunen, E., Peshev, D., Vangronsveld, J., Van den Ende, W., and Cuypers, A. (2013). Plant sugars are crucial players in the oxidative challenge during abiotic stress: extending the traditional concept. *Plant Cell Environment.* 36, 1242–1255. doi: 10.1111/pce.12061

King, C. A., and Purcell, L. C. (2005). Inhibition of N2 fixation in soybean is associated with elevated ureides and amino acids. *Plant Physiol.* 137, 1389–1396. doi: 10.1104/pp.104.056317

Kirch, H. H., Bartels, D., Wei, Y. L., Schnable, P. S., and Wood, A. J. (2004). The ALDH gene superfamily of Arabidopsis. *Trends Plant Science.* 9, 371–377. doi: 10.1016/j.tplants.2004.06.004

Kosova, K., Vitamvas, P., Prasil, I. T., Klima, M., and Renaut, J. (2021). Plant proteoforms under environmental stress: functional proteins arising from a single gene. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.793113

Kosova, K., Vitamvas, P., Urban, M. O., Prasil, I. T., and Renaut, J. (2018). Plant abiotic stress proteomics: the major factors determining alterations in cellular proteome. *Front. Plant Science.* 9. doi: 10.3389/fpls.2018.00122

Kotchoni, S. O., Kuhns, C., Ditzer, A., Kirch, H.-H., and Bartels, D. (2006). Over-expression of different aldehyde dehydrogenase genes in Arabidopsis thaliana confers tolerance to abiotic stress and protects plants against lipid peroxidation and oxidative stress. *Plant Cell Environment.* 29, 1033–1048. doi: 10.1111/j.1365-3040.2005.01458.x

Ku, H.-M., Hu, C.-C., Chang, H.-J., Lin, Y.-T., Jan, F.-J., and Chen, C.-T. (2011).). Analysis by virus induced gene silencing of the expression of two proline biosynthetic pathway genes in Nicotiana benthamiana under stress conditions. *Plant Physiol. Biochem.* 49, 1147–1154. doi: 10.1016/j.plaphy.2011.07.003

Lam, H. M., Coschigano, K. T., Oliveira, I. C., MeloOliveira, R., and Coruzzi, G. M. (1996). The molecular-genetics of nitrogen assimilation into amino acids in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 569–593. doi: 10.1146/annurev.arplant.47.1.569

Larrainzar, E., Wienkoop, S., Scherling, C., Kempa, S., Ladrera, R., Arrese-Igor, C., et al. (2009). Carbon metabolism and bacteroid functioning are involved in the regulation of nitrogen fixation in medicago truncatula under drought and recovery. *Mol. Plant-Microbe Interact.* 22, 1565–1576. doi: 10.1094/mpmi-22-12-1565

Leao, A. P., Bittencourt, C. B., da Silva, T. L. C., Neto, J. C. R., Braga, I. D., Vieira, L. R., et al. (2022). Insights from a multi-omics integration (MOI) study in oil palm (Elaeis guineensis jacq.) response to abiotic stresses: part two-drought. *Plants-Basel.* 11, 18. doi: 10.3390/plants11202786

Li, Z., Jing, W., Peng, Y., Zhang, X. Q., Ma, X., Huang, L. K., et al. (2015). Spermine alleviates drought stress in white clover with different resistance by influencing carbohydrate metabolism and dehydrins synthesis. *PloS One* 10, e0120708. doi: 10.1371/journal.pone.0120708

Li, H., Qamar, M. T., Yang, L., Liang, J., You, J., and Wang, L. (2023). Current progress, applications and challenges of multi-omics approaches in sesame genetic improvement. *Int. J. Mol. Sci.* 24. doi: 10.3390/ijms24043105

Li, T., Yun, Z., Wu, Q., Qu, H., Duan, X., and Jiang, Y. (2019). Combination of transcriptomic, proteomic, and metabolomic analysis reveals the ripening mechanism of banana pulp. *Biomolecules.* 9, 523. doi: 10.3390/biom9100523

Lin, Y. L., Liu, S. Y., Fang, X., Ren, Y. H., You, Z. J., Xia, J. X., et al. (2023). The physiology of drought stress in two grapevine cultivars: Photosynthesis, antioxidant system, and osmotic regulation responses. *Physiologia Plantarum* 175. doi: 10.1111/ppl.14005

Matysiak, K., Kierzek, R., Siatkowski, I., Kowalska, J., Krawczyk, R., and Miziniak, W. (2020). Effect of exogenous application of amino acids L-arginine and glycine on maize under temperature stress. *Agronomy-Basel* 10, 769. doi: 10.3390/agronomy10060769

Melo, F. V., Oliveira, M. M., Saibo, N. J. M., and Lourenco, T. F. (2021). Modulation of abiotic stress responses in rice by E3-ubiquitin ligases: A promising way to develop stress-tolerant crops. *Front. Plant Sci.* 12: doi: 10.3389/fpls.2021.640193

Michaletti, A., Naghavi, M. R., Toorchi, M., Zolla, L., and Rinalducci, S. (2018). Metabolomics and proteomics reveal drought-stress responses of leaf tissues from spring-wheat. *Sci. Rep.* 8. doi: 10.1038/s41598-018-24012-y

Missihoun, T. D., Hou, Q., Mertens, D., and Bartels, D. (2014). Sequence and functional analyses of the aldehyde dehydrogenase 7B4 gene promoter in Arabidopsis thaliana and selected Brassicaceae: regulation patterns in response to wounding and osmotic stress. *Planta.* 239, 1281–1298. doi: 10.1007/s00425-014-2051-0

Mohammadi, M., Kav, N. N. V., and Deyholos, M. K. (2007). Transcriptional profiling of hexaploid wheat (Triticum aestivum L.) roots identifies novel, dehydration-responsive genes. *Plant Cell Environ.* . 30, 630–645. doi: 10.1111/j.1365-3040.2007.01645.x

Moller, M. G., Taylor, C., Rasmussen, S. K., and Holm, P. B. (2003). Molecular cloning and characterisation of two genes encoding asparagine synthetase in barley (*Hordeum vulgare* L.). *Biochim. Et Biophys. Acta-Gene Structure Expression* . 1628, 123–132. doi: 10.1016/s0167-4781(03)00137-4

Moreno, J. C., Martinez-Jaime, S., Kosmacz, M., Sokolowska, E. M., Schulz, P., Fischer, A., et al. (2021). A multi-OMICs approach sheds light on the higher yield phenotype and enhanced abiotic stress tolerance in tobacco lines expressing the carrot lycopene β-cyclase1 gene. *Front. Plant Science.* 12. doi: 10.3389/fpls.2021.624365

Nakano, K., Suzuki, T., Hayakawa, T., and Yamaya, T. (2000). Organ and cellular localization of asparagine synthetase in rice plants. *Plant Cell Physiol.* 41, 874–880. doi: 10.1093/pcp/pcd006

Nawae, W., Shearman, J. R., Tangphatsornruang, S., Punpee, P., Yoocha, T., Sangsrakru, D., et al. (2020). Differential expression between drought-tolerant and drought-sensitive sugarcane under mild and moderate water stress as revealed by a comparative analysis of leaf transcriptome. *Peerj.* 8, E9608. doi: 10.7717/peerj.9608

Patel, R. K., and Jain, M. (2012). NGS QC toolkit: A toolkit for quality control of next generation sequencing data. *PloS One* 7. doi: 10.1371/journal.pone.0030619

Qu, C., Hao, B., Xu, X., Wang, Y., Yang, C., Xu, Z., et al. (2019). Functional research on three presumed asparagine synthetase family members in poplar. *Genes* 10. doi: 10.3390/genes10050326

Ramadan, A. A., Abd Elhamid, E. M., and Sadak, M. S. (2019). Comparative study for the effect of arginine and sodium nitroprusside on sunflower plants grown under salinity stress conditions. *Bull. Natl. Res. Centre.* 43. doi: 10.1186/s42269-019-0156-0

Raza, A., Charagh, S., Abbas, S., Hassan, M. U., Saeed, F., Haider, S., et al. (2023). Assessment of proline function in higher plants under extreme temperatures. *Plant Biol.* 25, 379–395. doi: 10.1111/plb.13510

Remmers, I. M., D'Adamo, S., Martens, D. E., de Vos, R. C. H., Mumm, R., America, A. H. P., et al. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom Phaeodactylum tricornutum during nitrogen limitation. *Algal Research-Biomass Biofuels Bioproducts.* 35, 33–49. doi: 10.1016/j.algal.2018.08.012

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Saini, D. K., Srivast, P., Pal, N., and Gupta, P. K. (2022). Meta-QTLs, ortho-meta-QTLs and candidate genes for grain yield and associated traits in wheat (Triticum aestivum L.). *Theor. Appl. Genet.* 135, 1049–1081. doi: 10.1007/s00122-021-04018-3

Sanchez, D. H., Schwabe, F., Erban, A., Udvardi, M. K., and Kopka, J. (2012).). Comparative metabolomics of drought acclimation in model and forage legumes. *Plant Cell Environment.* 35, 136–149. doi: 10.1111/j.1365-3040.2011.02423.x

Schwendner, P., Bohmeier, M., Rettberg, P., Beblo-Vranesevic, K., Gaboyer, F., Moissl-Eichinger, C., et al. (2018). Beyond Chloride Brines: Variable Metabolomic Responses in the Anaerobic Organism Yersinia intermedia MASE-LG-1 to NaCl and MgSO4 at Identical Water Activity. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.00335

Serraj, R., Vadez, V., and Sinclair, T. R. (2001). Feedback regulation of symbiotic N2 fixation under drought stress. *Agronomie.* 21, 621–626. doi: 10.1051/agro:2001153

Shewry, P. R., Corol, D. I., Jones, H. D., Beale, M. H., and Ward, J. L. (2017). Defining genetic and chemical diversity in wheat grain by 1H-NMR spectroscopy of polar metabolites. *Mol. Nutr. Food Res.* 61, 1600807. doi: 10.1002/mnfr.201600807

Shi, L. F., Twary, S. N., Yoshioka, H., Gregerson, R. G., Miller, S. S., Samac, D. A., et al. (1997). Nitrogen assimilation in alfalfa: Isolation and characterization of an asparagine synthetase gene showing enhanced expression in root nodules and dark-adapted leaves. *Plant Cell.* 9, 1339–1356. doi: 10.1105/tpc.9.8.1339

Shi, H., Ye, T., Chen, F., Cheng, Z., Wang, Y., Yang, P., et al. (2013). Manipulation of arginase expression modulates abiotic stress tolerance in Arabidopsis: effect on arginine metabolism and ROS accumulation. *J. Exp. Botany.* 64, 1367–1379. doi: 10.1093/jxb/ers400

Shin, J. H., Kim, S. R., and An, G. (2009). Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol.* 149, 905–915. doi: 10.1104/pp.108.130716

Shu, J., Ma, X., Ma, H., Huang, Q., Zhang, Y., Guan, M., et al. (2022). Transcriptomic, proteomic, metabolomic, and functional genomic approaches of Brassica napus L. during salt stress. *PloS One* 17, e0262587. doi: 10.1371/journal.pone.0262587

Siddiqui, M. H., Khan, M. N., Mukherjee, S., Alamri, S., Basahi, R. A., Al-Amri, A. A., et al. (2021). Hydrogen sulfide (H2S) and potassium (K+) synergistically induce drought stress tolerance through regulation of H+-ATPase activity, sugar metabolism, and antioxidative defense in tomato seedlings. *Plant Cell Rep.* 40, 1543–1564. doi: 10.1007/s00299-021-02731-3

Singh, V., Gupta, K., Singh, S., Jain, M., and Garg, R. (2023). Unravelling the molecular mechanism underlying drought stress response in chickpea via integrated multi-omics analysis. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1156606

Song, Q., Joshi, M., DiPiazza, J., and Joshi, V. (2020). Functional relevance of citrulline in the vegetative tissues of watermelon during abiotic stresses. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00512

Srivastava, V., Obudulu, O., Bygdell, J., Lofstedt, T., Ryden, P., Nilsson, R., et al. (2013). OnPLS integration of transcriptomic, proteomic and metabolomic data shows multi-level oxidative stress responses in the cambium of transgenic hipI-superoxide dismutase Populus plants. *BMC Genomics* 14, 1755. doi: 10.1016/j.algal.2018.08.012

Su, J., and Wu, R. (2004). Stress-inducible synthesis of proline in transgenic rice confers faster growth under stress conditions than that with constitutive synthesis. *Plant Science.* 166, 941–948. doi: 10.1016/j.plantsci.2003.12.004

Sun, C. X., Palmqvist, S., Olsson, H., Borén, M., Ahlandsberg, S., and Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. *Plant Cell.* 15, 2076–2092. doi: 10.1105/tpc.014597

Surekha, C., Kumari, K. N., Aruna, L. V., Suneetha, G., Arundhati, A., and Kishor, P. B. K. (2014). Expression of the Vigna aconitifolia P5CSF129A gene in transgenic pigeonpea enhances proline accumulation and salt tolerance. *Plant Cell Tissue Organ Culture.* 116, 27–36. doi: 10.1007/s11240-013-0378-z

Suzuki, N., Bassil, E., Hamilton, J. S., Inupakutika, M. A., Zandalinas, S. I., Tripathy, D., et al. (2016). ABA is required for plant acclimation to a combination of salt and heat stress. *PloS One* 11. doi: 10.1371/journal.pone.0147625

Trovato, M., Mattioli, R., and Costantino, P. (2008). Multiple roles of proline in plant stress tolerance and development. *Rendiconti Lincei-Scienze Fisiche E Naturali.* 19, 325–346. doi: 10.1007/s12210-008-0022-8

Tsai, F. Y., and Coruzzi, G. M. (1990). Dark-induced and organ-specific expression of two asparagine synthetase genes in Pisum sativum. *EMBO J.* 9, 323–332. doi: 10.1002/j.1460-2075.1990.tb08114.x

Uchida, A., Jagendorf, A. T., Hibino, T., Takabe, T., and Takabe, T. (2002). Effects of hydrogen peroxide and nitric oxide on both salt and heat stress tolerance in rice. *Plant Science.* 163, 515–523. doi: 10.1016/s0168-9452(02)00159-0

Vij, S., and Tyagi, A. K. J. P. B. J. (2010). Emerging trends in the functional genomics of the abiotic stress response in crop plants. *Plant Biotechnol. J.* 5, 361–380. doi: 10.1111/j.1467-7652.2007.00239.x

Wang, L., Liu, L., Huang, A., Zhang, H., and Zheng, Y. (2024). The metabolism of amino acids, AsA and abscisic acid induced by strigolactone participates in chilling tolerance in postharvest zucchini fruit. *Front. Plant Science.* 15. doi: 10.3389/fpls.2024.1402521

Wang, A., Ma, C., Ma, H., Qiu, Z., and Wen, X. (2021). Physiological and proteomic responses of pitaya to PEG-induced drought stress. *Agriculture-Basel.* 11, 632. doi: 10.3390/agriculture11070632

Wang, Z., Shi, H., Yu, S., Zhou, W., Li, J., Liu, S., et al. (2019). Comprehensive transcriptomics, proteomics, and metabolomics analyses of the mechanisms regulating tiller production in low-tillering wheat. *Theor. Appl. Genet.* 132, 2181–2193. doi: 10.1007/s00122-019-03345-w

Wasaya, A., Zhang, X., Fang, Q., and Yan, Z. (2018). Root phenotyping for drought tolerance: A review. *Agronomy-Basel.* doi: 10.3390/agronomy8110241

Wen, J. F., Gong, M., Liu, Y., Hu, J. L., and Deng, M. H. (2013). Effect of hydrogen peroxide on growth and activity of some enzymes involved in proline metabolism of sweet corn seedlings under copper stress. *Scientia Horticulturae.* 164, 366–371. doi: 10.1016/j.scienta.2013.09.031

Witt, S., Galicia, L., Lisec, J., Cairns, J., Tiessen, A., Araus, J. L., et al. (2012). Metabolic and phenotypic responses of greenhouse-grown maize hybrids to experimentally controlled drought stress. *Mol. Plant* 5, 401–417. doi: 10.1093/mp/ssr102

Xia, H., Zhang, Z., Luo, C., Wei, K., Li, X., Mu, X., et al. (2023). MultiPrime: A reliable and efficient tool for targeted next-generation sequencing. *Imeta.* 2. doi: 10.1002/imt2.143

Yadav, A. K., Carroll, A. J., Estavillo, G. M., Rebetzke, G. J., and Pogson, B. J. (2019). Wheat drought tolerance in the field is predicted by amino acid responses to glasshouse-imposed drought. *J. Exp. Botany.* 70, 4931–4947. doi: 10.1093/jxb/erz224

Yamagata, H., Nakajima, A., Bowler, C., and Iwasaki, T. (1998). Molecular cloning and characterization of a cDNA encoding asparagine synthetase from soybean (Glycine max L.) cell cultures. *Bioscience Biotechnol. Biochem.* 62, 148–150. doi: 10.1271/bbb.62.148

Yang, S.-L., Lan, S.-S., and Gong, M. (2009). Hydrogen peroxide-induced proline and metabolic pathway of its accumulation in maize seedlings. *J. Plant Physiol.* 166, 1694–1699. doi: 10.1016/j.jplph.2009.04.006

Yun, Z., Li, T., Gao, H., Zhu, H., Gupta, V. K., Jiang, Y., et al. (2019). Integrated transcriptomic, proteomic, and metabolomics analysis reveals peel ripening of harvested banana under natural condition. *Biomolecules.* 9, 167. doi: 10.3390/biom9050167

Zadehbagheri, M., Azarpanah, A., and Javanmardi, S. (2014). Proline metabolite transport an efficient approach in corn yield improvement as response to drought conditions. *American-Eurasian J. Agric. Environ. Sci.* 14, 476–485. doi: 10.5829/idosi.aejaes.2013.13.12.12284

Zhang, C., and Shi, S. (2018). Physiological and proteomic responses of contrasting alfalfa (Medicago sativa L.) varieties to PEG-induced osmotic stress. *Front. Plant Science.* 9. doi: 10.3389/fpls.2018.00242

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

Zhao, Z., Rebetzke, G. J., Zheng, B., Chapman, S. C., and Wang, E. (2019). Modelling impact of early vigour on wheat yield in dryland regions. *J. Exp. Botany.* 70, 2535–2548. doi: 10.1093/jxb/erz069

Zhou, Y., Li, H., Chen, H., Yang, X., Yu, T., Wang, Y., et al. (2022).). Proteomic investigation of molecular mechanisms in response to PEG-induced drought stress in soybean roots. *Plants-Basel.* 11, 1173. doi: 10.3390/plants11091173

Zurbriggen, M. D., Hajirezaei, M. R., and Carrillo, N. J. B. G. E. R. (2010). Engineering the future. Development of transgenic plants with enhanced tolerance to adverse environments. *Science.* 27, 33–56. doi: 10.1126/science.218.4571.443

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 December 2024
DOI 10.3389/fpls.2024.1520786

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Lei Feng,
South China Normal University, China
Guolong Li,
Shaanxi University of Chinese Medicine, China
Chen Xiao Dan,
Northwest University, China

*CORRESPONDENCE
Haiying Lei
xiaoleikkxx@163.com
Tianzeng Niu
tznius@126.com
Ake Liu
akeliu@126.com

RECEIVED 31 October 2024
ACCEPTED 26 November 2024
PUBLISHED 23 December 2024

CITATION
Li J, Wang X, Lu J, Song H, Lei H, Niu T and Liu A (2024) Genome-wide identification and expression analysis of the WRKY gene family in *Sophora flavescens* during tissue development and salt stress.
Front. Plant Sci. 15:1520786.
doi: 10.3389/fpls.2024.1520786

Genome-wide identification and expression analysis of the WRKY gene family in *Sophora flavescens* during tissue development and salt stress

Jin Li[1], Xi Wang[1], Junjie Lu[2], Huifang Song[1], Haiying Lei[1*], Tianzeng Niu[1*] and Ake Liu[1*]

[1]Department of Life Sciences, Changzhi University, Changzhi, China, [2]College of Life Science, Nanyang Normal University, Nanyang, China

Sophora flavescens is a traditional Chinese medicinal herb rich in various bioactive secondary metabolites, such as alkaloids and flavonoids, and exhibits remarkable resistance to abiotic stress. The WRKY transcription factor (TF) family is one of the largest plant-specific TF families and plays a crucial role in plant growth, development, and responses to abiotic stress. However, a comprehensive genome-wide analysis of the WRKY gene family in *S. flavescens* has not yet been conducted. In this study, we identified 69 *SfWRKY* genes from the *S. flavescens* genome and classified them into seven distinct subfamilies based on phylogenetic analysis. Transposed duplications and dispersed duplications were found to be the primary driving forces behind the expansion of the SfWRKY family. Additionally, several *cis*-acting elements related to the stress response and hormone signaling were discovered within the promoter regions of *SfWRKYs*. Transcriptomic analyses across five tissues (leaves, flowers, pods, roots, and stems) revealed that genes exhibiting high expression levels in specific tissues generally showed high expression across all the examined tissues. Coexpression network constructed based on metabolomic and transcriptomic analyses of root and pod development indicated that *SfWRKY29* may play a significant role in regulating the biosynthesis of secondary metabolites during tissue development. The RT-qPCR results of gene expression analysis revealed that several *SfWRKY* genes were significantly induced in response to the accumulation of secondary metabolites or salt stress. Our study systematically analyzed WRKY TFs in *S. flavescens*, which provides valuable reference data for further studies on the key roles of *SfWRKY* genes in growth development as well as their responses under salt stress conditions.

KEYWORDS

***Sophora flavescens*, WRKY gene family, phylogenetic analysis, tissue development, salt stress**

1 Introduction

The WRKY family is one of the largest transcription factor (TF) families, which has been reported extensively involved in regulating the biosynthesis of secondary metabolites, developmental processes and stress responses (Goyal et al., 2022; Song et al., 2023). Since the first WRKY protein structure was found in sweet potato (Ipomoea batatas *L.*) (Ishiguro and Nakamura, 1994), WRKY TFs have been extensively identified in many plants, such as *Polygonum cuspidatum, Artemisia annua, Fagopyrum tataricum* etc (Bao et al., 2018; Lv et al., 2020; Chen et al., 2021). The WRKY proteins are named for their highly conserved WRKY domain (~60 amino acids), which contains a conserved motif (WRKYGQK) located at the N-terminus, followed closely by a zinc finger motif (Eulgem et al., 2000; Cheng et al., 2019). The WRKY region consists of four lines with a β-fold composition of Zn^{2+} that coordinates with Cys/His residues to form a zinc finger structure (Yamasaki et al., 2005; Duan et al., 2007).

In terms of phylogeny, WRKY proteins are classified into three groups (I-III) based on the number of WRKY domains and the type of zinc finger structure. Group I proteins contain two WRKY domains and a C_2H_2 ($CX_{4-5}CX_{22-23}HXH$) zinc finger motif. In contrast, both Group II and Group III proteins possess only a single WRKY domain along with either a C_2H_2 or a modified C_2HC ($CX_7CX_{23}HXC$) zinc finger motif (Goyal et al., 2022; Long et al., 2023). Furthermore, Group II can be subdivided into five distinct subgroups (IIa~IIe). The amino acid sequences of WRKY proteins specifically bind to the W-box *cis*-regulatory element (TTGACT/C) within target gene promoters, thereby inducing their expression. This interaction plays a crucial role in regulating plant secondary metabolite synthesis, growth and development, as well as responses to biotic and abiotic stresses (Wani et al., 2021; Zhang et al., 2018).

Increasing evidence suggested that WRKYs serve as important regulatory foundations for plant growth and development (Jiang et al., 2017; Wang et al., 2023). Under short-day conditions, WRKY12 and WRKY13 were involved in regulating the flowering time of *Arabidopsis thaliana* (Li et al., 2016; Ma et al., 2020). In rice, OsWRKY78 plays a role in stem elongation and seed development (Zhang et al., 2011). WRKY26, WRKY45, and WRKY75 participate alongside ethylene in inhibiting the growth of primary roots and lateral roots during shade-avoidance response (Rosado et al., 2022). Following treatment with As + Fe, the expression of rice *OsWRKY71* increases, promoting root system development while also participating in the regulation of gibberellin synthesis pathways (Mirza and Gupta, 2024). In both *A. thaliana* and rapeseed, *WRKY70* was primarily expressed in leaves where it plays a significant role in leaf senescence (Ülker et al., 2007; Liu et al., 2023). Furthermore, it has been demonstrated that WRKY TFs possess functions that regulate the biosynthesis of terpenoids, alkaloids, flavonoids, etc. AaWRKY1 positively regulates artemisinin biosynthesis by promoting the expression of *DBR2*, *CYP71AV1*, and *ADS* genes within *A. annua* (Han et al., 2014). VqWRKY31 activates salicylic acid defense signals, which alter the accumulation of quercetin, flavonoids, and proanthocyanidins (Yin et al., 2022). In *Coptis chinensis*, CcWRKY7, CcWRKY29 and CcWRKY32 may regulate berberine alkaloid biosynthesis (Huang et al., 2023). PeWRKY30 serves as a key factor for flavonoid biosynthesis in passion fruit (Ma et al., 2024).

Previous studies have demonstrated that WRKY TFs play a crucial role in plant defensive responses to environmental stress (Jiang et al., 2017; Goyal et al., 2022; Feng et al., 2023). The overexpression of *SlWRKY8* in tomato significantly enhances its resistance to pathogen infection and positively regulates responses to drought and salt stress (Gao et al., 2019). Furthermore, in tomato, SlWRKY57 functions as a negative regulator in the response to salt stress by directly inhibiting the transcription of salt-responsive genes (*SlRD29B* and *SlDREB2*) as well as ion homeostasis genes (*SlSOS1*) (Ma et al., 2023). Under low temperature and drought conditions, the overexpression of *PoWRKY1* in *A. thaliana* has been shown to improve seed germination activity and promote root growth in transgenic plants (Wei et al., 2021). In addition, under drought and salt stress, the overexpression of *MfWRKY40* facilitates taproot elongation, enhances osmoregulation, and improves tolerance in *A. thaliana* plants (Huang et al., 2022). In wheat, the TaWRKY plays a role in regulating the response to aluminum and manganese ion stress (Luo et al., 2024). Tomato WRKY23 can enhance the salt and osmotic stress tolerance of transgenic *Arabidopsis* by modulating the ethylene and auxin pathways (Singh et al., 2023). Overexpression of *TaWRKY17* can significantly enhance the salt tolerance of wheat (Yu et al., 2023). Therefore, WRKY TFs can regulate the growth and development and environmental adaptation from multiple dimensions.

Sophora flavescens (Kusen), a Chinese herbal medicine, is widely used in the treatment of inflammation, solid tumors, and analgesic effect (Lee et al., 2018; Cheng et al., 2022). The main active ingredients in *S. flavescens* are alkaloids and flavonoids (Dong et al., 2021), which are involved in treating diseases such as hepatitis, tumors, and diabetes. Comparative genomics analysis can provide us with an efficient way to identify members of certain gene family and conduct studies on their potential functions (Liu et al., 2024). Given the significant contribution of the WRKY gene family to plant stress tolerance and secondary metabolite biosynthesis, in this study, we conducted phylogenetic analysis, sequence characters, tissue-specific expressions of *WRKY* genes in *S. flavescens*. Our findings will provide insight into the mechanism of environmental adaptability and secondary metabolite biosynthesis in *S. flavescens*, and also provide reference information for molecular breeding.

2 Materials and method

2.1 *WRKY* gene identification and sequence retrieval in *S. flavescens*

The genome and protein sequences of *S. flavescens* and *Sophora moorcroftiana* were obtained from previous studies (Qu et al., 2023; Yin et al., 2023). The HMM configuration file for WRKY domain (PF03106) was downloaded from the Pfam database (El-Gebali et al., 2019). Candidate WRKY members coding in the genomes of *S. flavescens* and *S. moorcroftiana* were identified using HMMER

(v3.2.1) software, with an *E*-value threshold set at 10^{-2} (Prakash et al., 2017). Only sequences containing the WRKY domain were considered as members of the WRKY family. To ensure the completeness of SfWRKY repertoire, we also examined the assembled novel transcripts obtained from the transcriptome assembly in section 2.5 for member identification. Furthermore, the *WRKY* genes in *A. thaliana* were derived from previous study (Abdullah-Zawawi et al., 2021).

2.2 Physicochemical properties of WRKY TFs

The *WRKY* genes of *S. flavescens* were named according to their relative positions on the chromosomes. To investigate their protein properties, we utilized the ProtParam program (ExPASy tools, http://web.expasy.org/protparam/) to estimate the molecular weights (MWs) and theoretical isoelectric points (pI).

2.3 Phylogenetic analysis of the WRKY family

To construct the phylogenetic tree of the WRKY members, we utilized the ClustalW program in MEGA 7 (v7.0.26) software (Kumar et al., 2016) to perform multiple sequence alignments of the WRKY domain regions from above mentioned three species. Subsequently, we employed the Neighbor-Joining (NJ) method within MEGA 7 to build the phylogenetic tree, selecting Poisson model for amino acid substitution and applying pairwise deletion for gap treatment. To ensure the accuracy, we assessed the support for each relative branch through 1000 bootstrap replicates.

2.4 Prediction of gene duplications and *cis*-acting regulatory elements

The DupGen_finder software (Qiao et al., 2019) was employed to conduct analysis of gene duplication patterns in *S. flavescens*. It includes whole genome duplication (WGD), tandem duplication (TD), transposed duplication (TRD), proximal duplication (PD), and dispersed duplication (DSD). The non synonymous substitution rate (*Ka*) and synonymous substitution rate (*Ks*) for TD gene pairs were calculated using KaKs_calculator3 (Zhang, 2022). The YN (Yang and Nielsen) model (Yang and Nielsen, 2000) was selected to compute the *Ka*/*Ks* ratio, which serves as an indicator of selective pressure on duplicated gene pairs. The gene density and intergenomic syntenic block analysis were conducted following the methods of Feng et al. (2024). Based on the genomic annotation information, TBtools software was utilized to obtain the upstream two kb genome sequence of *SfWRKY* genes from start codon. Subsequently, potential *cis*-regulatory elements were predicted and identified using default parameters via the PlantCARE website (Lescot, 2002).

2.5 Analysis of *SfWRKY* expression profiles based on transcriptome sequencing

To investigate the expression patterns of the *SfWRKY* genes in different tissues and growth stages, we collected four tissues (stem, flower, leaf, and root) of *S. flavescens* cultivated for over five years at Changzhi International Shennong Traditional Chinese Medicine Cultural Expo Park (Shangdang District, Changzhi City, Shanxi Province) on July 12, 2021. Additionally, we sampled pods at six different developmental stages between July 12 and August 6, 2021, with sampling conducted every five days. The roots of *S. flavescens* at eight distinct developmental stages, which were sown in September 2022, were collected on the 20th day of each month from April to November 2023 in Dianshang County (Lucheng District, Changzhi City, Shanxi Province). Each sample comprised three biological replicates collected in 50 mL centrifuge tubes and immediately frozen in liquid nitrogen before being stored at -80°C for further analysis. The experimental methods and analytical approaches for transcriptomics (stem, flower, leaf, and root; pod and root development) and metabolomics (pod and root development) were adapted from Zhong et al. (2024). The raw data from RNA-seq samples were archived in the NCBI database under accession number PRJNA1136989. The co-expression network of transcriptomes and metabolomes for pods and roots was constructed using R (v 4.2.2) with the WGCNA (v1.71) (Langfelder and Horvath, 2008) following the methods of Liu et al. (2024). The networks with inter-gene weight values greater than 0.3 were visualized using Cytoscape (v.3.8.2) (Otasek et al., 2019).

2.6 Expression analysis of *SfWRKY* genes based on RT-qPCR

The roots of *S. flavescens* cultivated in Dianshang County were collected for real-time quantitative PCR (RT-qPCR) analysis. The plants included two different cultivation years, C and S represent the *S. flavescens* sowed in 2024 and 2022, respectively (sowing occurs every April), with samples collected in July 2024. For salt stress treatment, one-year-old seedlings of *S. flavescens* were irrigated with 250 mM NaCl solution (Salt), while the normal condition (NC) irrigated with equivalent distilled water. After treatment for 14 days, leaves were collected for RT-qPCR analysis. Each sample consisted of three biological replicates were collected into 50 mL centrifuge tubes and immediately frozen in liquid nitrogen before being stored at -80°C.

Total RNA was extracted from the tissues of *S. flavescens* using the polysaccharide polyphenol total RNA extraction kit (Beijing GeneBetter, China). The integrity of the RNA was confirmed with a Nanodrop 2000 spectrophotometer (Thermo Fisher, USA). Reverse transcription was performed using the HiScript II Q RT SuperMix kit (Takara, Takara Biomedical Technology (Beijing) Co., Ltd.), followed by qPCR utilizing the SYBR qPCR Master Mix kit (Vazyme, Nanjing, China). Specific primers were designed using Primer5 software (v5.00) (Supplementary Table S1). *EF-1α* was

utilized as the reference gene. All reactions were performed in triplicate, and the relative expression levels of genes were calculated using the $2^{-\triangle\triangle Ct}$ method. Statistical significance was assessed using Student's *t* test.

3 Results

3.1 Sixty-nine *SfWRKY* genes identified in *S. flavescens* genome

We totally identified 69 *SfWRKY* genes based on genomic and transcriptomic data, which were named *SfWRKY01-SfWRKY69* according to their positions on the chromosomes (Figure 1; Supplementary Table S2). The *69 SfWRKY* genes encoded proteins of varying sizes, ranging from the largest protein (SfWRKY12) with a molecular weight (MW) of 83.5 kD and composed of 757 amino acids, to the smallest protein (SfWRKY56) with an MW of 18.8 kD and containing 167 amino acids. The theoretical isoelectric point (pI) ranges from 4.8 (SfWRKY05) to 9.78 (SfWRKY18), indicating that different SfWRKY proteins perform various functions under different microenvironments. The 69 *SfWRKY* genes are unevenly distributed across nine chromosomes of *S. flavescens* (Figure 1), with the majority located on chromosome 4 (Chr4, 15 genes) and Chr6 (12 genes), followed by Chr7 with 10 genes, and the least on Chr9 with only three genes.

Through correlation analysis between the *WRKY* gene numbers from the previous studies and their corresponding genome size, the former was not linearly correlated with the latter (Pearson's correlation coefficient (r) = -0.0440, $p > 0.05$; Supplementary Figure S1). The number of *SfWRKY* genes was lower than those of *A. thaliana* (72), *S. moorcroftiana* (83), *Medicago sativa* (91), and *Arachis hypogaea* (158). However, it is relatively higher compared to other plants, such as *Panicum miliaceum* (32), *Platycodon grandiflorus* (42), and *Dendrobioum catenatum* (62). This indicates that there is no significant correlation between the size of a species' genome and the size of the *WRKY* gene family.

3.2 Multiple sequence alignment, phylogenetic analysis and classification of *SfWRKY* genes

To further understand the evolutionary diversity of the *SfWRKY* genes, we constructed an NJ tree for the WRKY family members of *S. flavescens*, *S. moorcroftiana*, and *A. thaliana* based on their WRKY domains (Figure 2). Accordingly, the SfWRKY members were classified into three well defined groups. The 69 *SfWRKY* genes were unevenly distributed among the three groups, with 16 members in Group I, 44 members in Group II, and nine members in Group III (Figure 2B). The majority of *S. flavescens* and *S. moorcroftiana* WRKY members exhibit a one-to-one clustering pattern on the evolutionary tree, whereas the *Arabidopsis* WRKY members tend to form a cohesive cluster. Members in Group I have two WRKY domains located in N-terminal and C-terminal regions. Group II has largest member number, which is further subdivided into five subfamilies (Figure 2C, IIa-IIe). Groups IIa and IIb tend to cluster into one branch, while groups IId and IIe tend to cluster together. In Group III, four closely located members (within 0.32 Mb) on the chromosome of *S. moorcroftiana* tend to cluster together, possibly originating from tandem duplication events. Among these groups or subfamilies, the member number in *S. flavescens* and *S. moorcroftiana* was similar (Figures 2B, C).

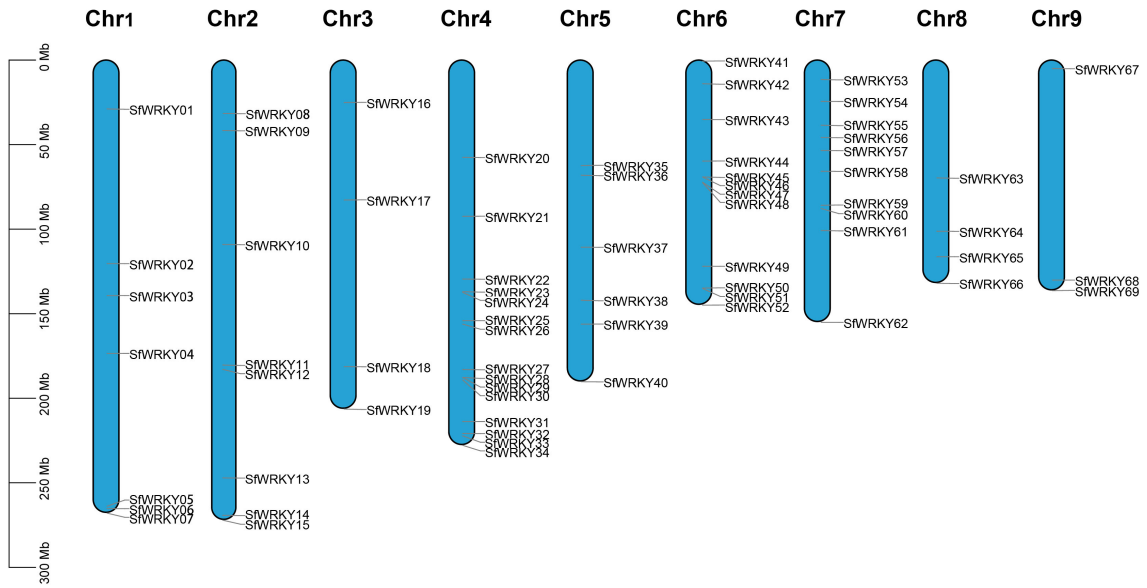


FIGURE 1
Chromosome distribution of *SfWRKY* genes. The chromosomal position of each *SfWRKY* gene was mapped according to the genome annotation file of *S. flavescens*. The chromosome number is labeled at the top of each chromosome. The scale is in mega bases (Mb) on the left.

10.3389/fpls.2024.1520786

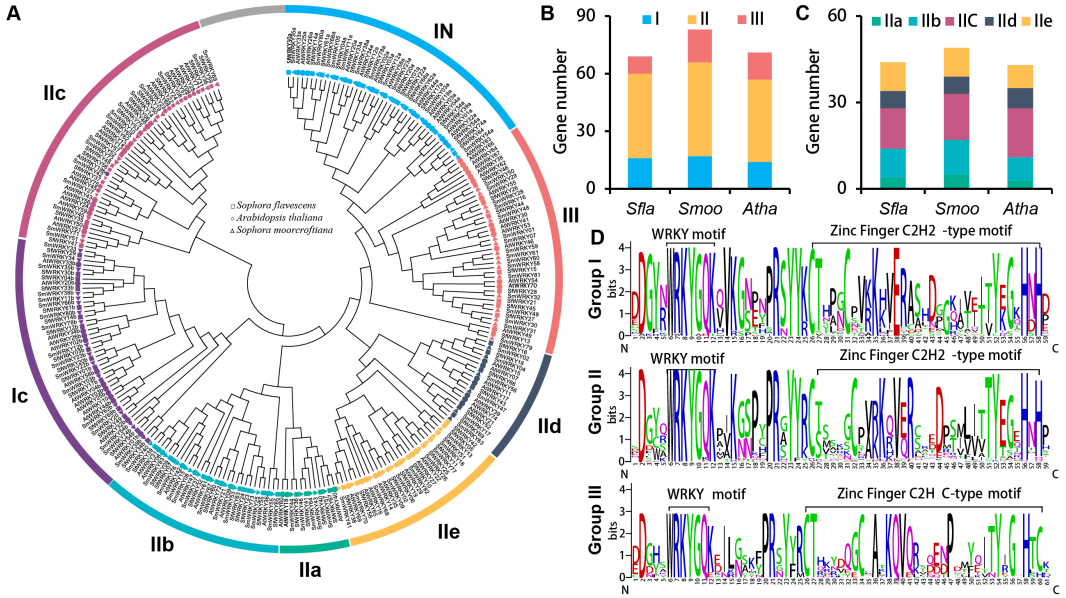


FIGURE 2
Phylogenetic analysis of *SfWRKY* genes. **(A)** Phylogenetic tree of WRKY proteins of *S. flavescens*, *S. moorcroftiana*, and *A. thaliana*. Different shapes (square, circle, and triangle) labeled at the end of each branch are used to mark the corresponding species, while different colors represent different groups. **(B)** The gene number comparison of 3 WRKY groups (I, II and III) among the *S. flavescens*, *S. moorcroftiana* and *A. thaliana*; **(C)** The member number comparison of five subfamilies in group II (IIa, IIb, IIc, IId, and IIe) among the *S. flavescens*, *S. moorcroftiana* and *A. thaliana*; **(D)** The comparative analysis of the conserved motif for the three groups of SfWRKYs based on multiple sequence alignment. The overall height of each stack represents the conservation degree of the sequence at the position. The letter at the top indicates the amino acid residue that occurs most frequently at that position.

Subsequently, multiple sequence alignments were conducted for the WRKY domains of the three SfWRKY groups (Figure 2D). The results revealed that the WRKY domains were highly conserved across all categories, containing a heptapeptide domain (WRKYGQK) and a zinc-finger domain (C2H2 or C2HC). Additionally, there were six WRKY domain variants, five of which were WRKYGEK (SfWRKY56, and SfWRKY61). The other three variants included WRKYGKK (SfWRKY35 and SfWRKY47), WKKYAQT (SfWRKY29), and WRVKGQE (SfWRKY28). The gene structure and motif distribution analyses showed members from same group or subfamily showed similar characteristics (Figure 3), which also provided evidence for the classification based on only phylogenetic analysis.

3.3 Gene duplication and collinearity of *WRKY* genes in *S. flavescens*

The DupGen_finder was used for analysis of *SfWRKY* gene duplication pattern, which included one pair of genes derived from WGD (such as *SfWRKY04* and *SfWRKY33*), 8 pairs from TRD (such as *SfWRKY20* and *SfWRKY33*), and 10 pairs of DSD (such as *SfWRKY61* and *SfWRKY20*) events (Figure 4A). Among these TRD and DSD gene pairs, it is generally presenting one gene corresponding to multiple duplicate genes rather than one corresponding to one, which may be related to multiple rounds of duplication occurred in *S. flavescens* genome. For example, the TRD genes corresponding to *SfWRKY33* are *SfWRKY03*, *SfWRKY20*, *SfWRKY30*, *SfWRKY61*, and *SfWRKY68*, and the DSD genes corresponding to *SfWRKY53* are *SfWRKY06*, *SfWRKY31*, and *SfWRKY48*. We further calculated the *Ka*, *Ks*, and *Ka*/*Ks* values for the duplicated *SfWRKY* gene pairs, and the results showed that the *Ka*/*Ks* values ranged from 0.09 to 0.39 (< 1), indicating that these genes underwent strong purifying selection during evolution (Table 1). This implies that gene duplications may contribute to the diversification and expansion of the SfWRKY gene family, while these duplicated genes are subject to strong functional constraints.

We also conducted a collinearity analysis among the three species, *S. flavescens*, *S. moorcroftiana*, and *A. thaliana*. The results showed that 16 *SfWRKY* genes from *S. flavescens* are present in collinear blocks with 11 counterparts from *A. thaliana*, and 24 *SfWRKY* genes from *S. flavescens* are present in collinear blocks with 12 counterparts from *S. moorcroftiana* (Figure 4B). It was observed that collinear genes tended to cluster into the same groups in phylogenetic tree (Figure 2), implying that they have a common evolutionary origin.

3.4 *Cis*-regulatory element distribution of *SfWRKY* genes

To understand the potential roles of the SfWRKY family members in plant growth and development, response to plant hormones, and environmental stresses, we analyzed the distribution of *cis*-elements in the upstream promoter regions of the *SfWRKY* genes (Figure 5). Totally, 18 plant growth and development related *cis*-elements were identified. The most

10.3389/fpls.2024.1520786

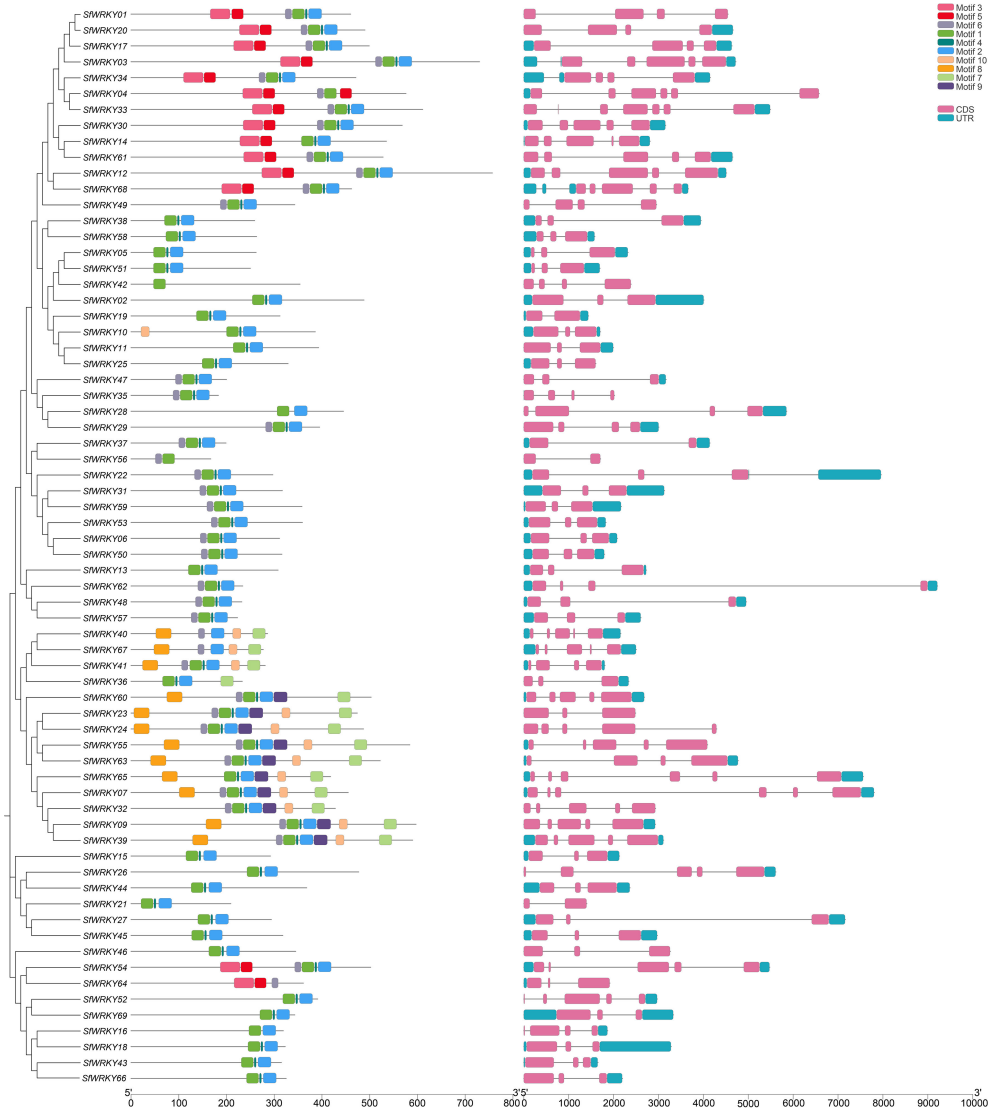


FIGURE 3
The motif distribution and gene structure of *SfWRKY* genes. The distribution of MEME motifs among the 69 WRKY members of *S. flavescens* was analyzed to unveil conserved motifs shared by related proteins. Additionally, the gene structure of these 69 *SfWRKY* genes was examined, with gray lines representing intron regions within the gene sequence. Different colors were employed to denote distinct motifs or gene regions, as depicted in the upper right corner of the figure.

abundant elements were the Box-4 (ATTAAT) and the G-box (TACGTG), which were involved in light responsiveness. They accounted for 30% and 20% of the total elements identified in this category. *SfWRKY51* contains the highest number of such elements (13). Other elements were also observed, such as circadian control and tissue-specific motifs like the GT1-motif (photosynthetic reaction regulation), TCT-motif (light-responsive elements), and GATA-motif (plant development). Regarding the phytohormone responsive elements, the ABRE element involved in ABA responsiveness, the TCA-element involved in salicylic acid responsiveness, the TGACG-motif involved in MeJA-responsiveness and the ERE element involved in ethylene signal regulation were identified abundant in upstream regions of the *SfWRKY* genes. These four pattern elements represent more than 45% of the total hormone-responsive elements. In *SfWRKY67*, the number of ABRE elements is the highest (8), while in *SfWRKY11*, the number of ERE elements is the highest (7). In addition, other elements were also found, like the as-1 element involved in SA and oxidative stress responsiveness, and the CGTCA-motif involved in MeJA-responsiveness in this category.

In the abiotic and biotic stress category, different elements associated with stress responses, such as oxidation, defense, drought, wounding, heat, and low temperature, were observed. After our analysis in *S. flavescens*, the largest part of the elements belonging to the abiotic and biotic stress category corresponds to two general stress-responsive motifs, namely the MYB (CCAAT box) and MYC (CACATG box) binding sites, representing 30% and 20% of the total identified *cis*-elements, respectively. In addition, other stress-specific *cis*-elements were identified. Several of them were responsive to wounding and pathogens—including the

10.3389/fpls.2024.1520786

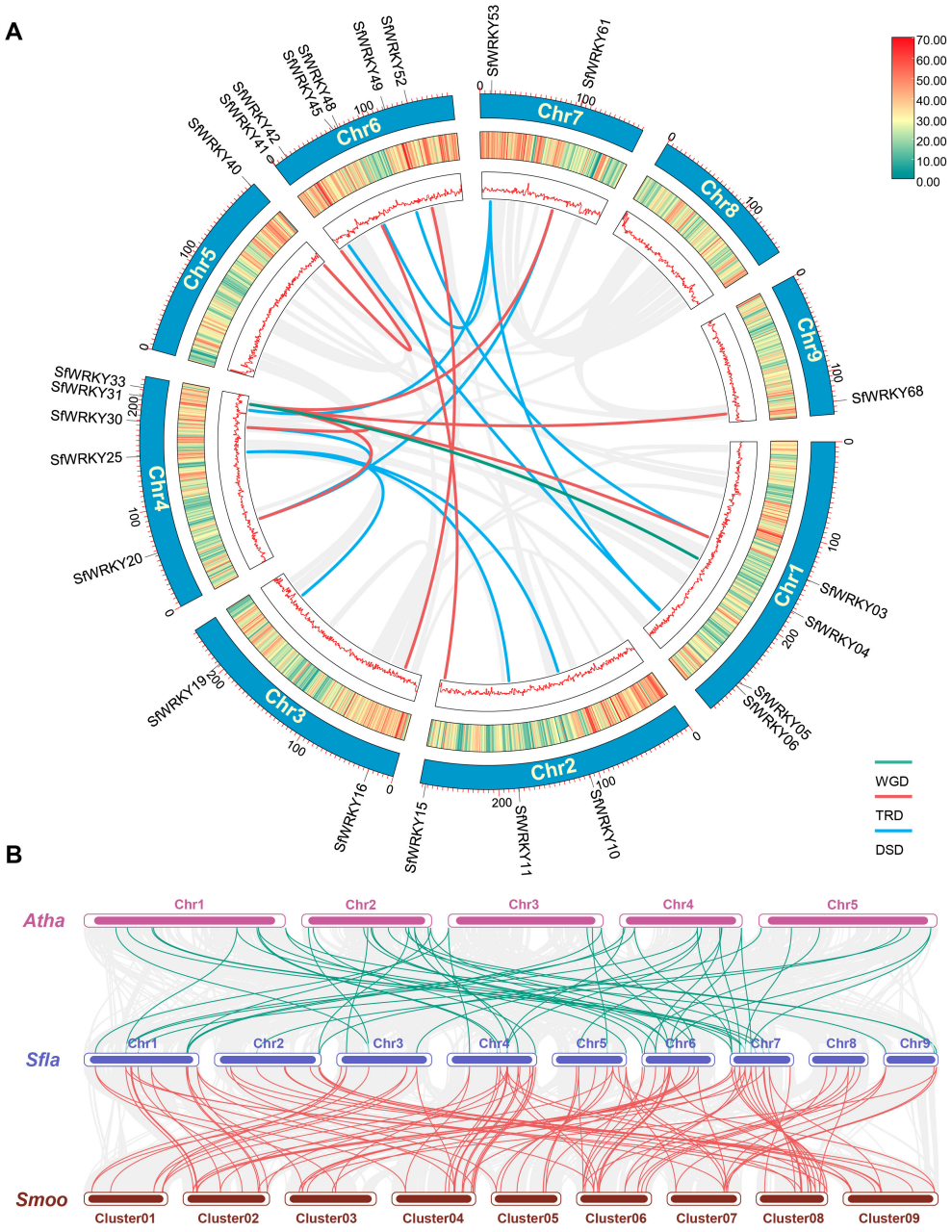


FIGURE 4
Gene duplication patterns and syntenic analyses of *WRKY* genes in *S. flavescens*, *S. moorcroftiana*, and *A. thaliana*. **(A)** The chromosome locations of duplicated *SfWRKY* genes resulted from WGD, TRD and DSD events on the Circos diagram. The different colors (red to yellow to green) in the middle track indicated high to low gene density (the number of genes per bin determined by sliding window analysis). **(B)** The intergenomic syntenic blocks identified among *A. thaliana* (Atha), *S. flavescens* (Sfla), and *S. moorcroftiana* (Smoo). The collinear blocks are represented by gray lines, while the *WRKY* gene collinearity is highlighted by green and red lines.

WRKY-box (W-box), the TC-rich repeats, and the wound responsive motif (WUN-motif). Temperature-related elements, such as the stress responsive element (STRE) and the low-temperature responsive (LTR) motifs; drought related elements, such as the MYB BINDING SITE (MBS) and the Dehydration -responsive element (DRE)-core; and anaerobic conditions like the Anaerobic response element (ARE) motif (Figure 5). Among them, *SfWRKY37* contains the highest number of such elements, followed by *SfWRKY60*.

3.5 Expression profiles of *SfWRKY* genes among different tissues

To investigate the expression patterns of SfWRKY family members among different tissues, we analyzed their expression levels in four different tissues, including leaves, flowers, roots, and stems, as well as different development stages of pods and roots. The results showed significant variations in the expression patterns of different *SfWRKY* genes across various tissues (Figure 6). Notably,

10.3389/fpls.2024.1520786

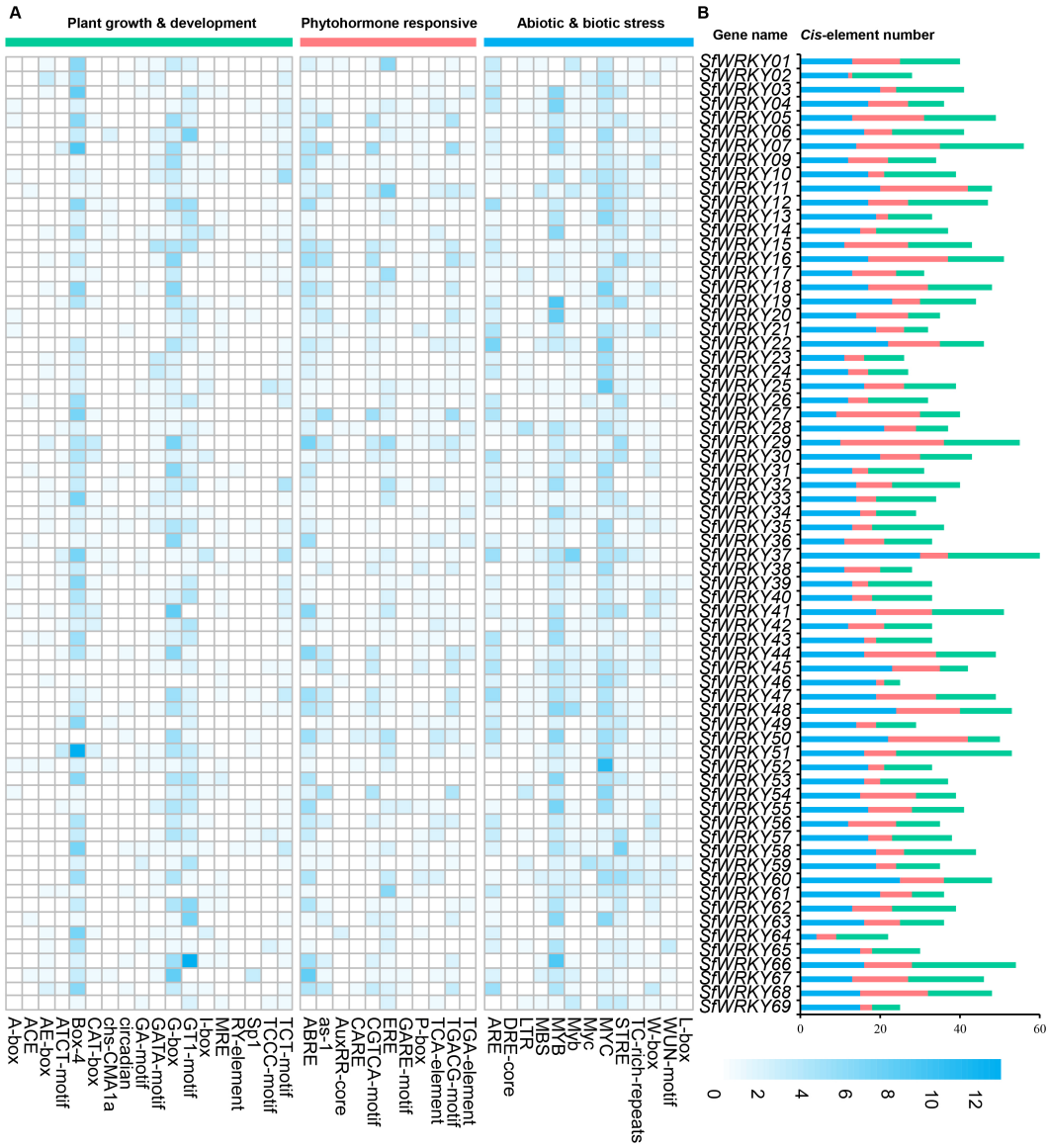


FIGURE 5
Analysis of the *cis*-elements located in the promoter regions of *SfWRKY* genes. **(A)** The different intensity colors indicate the numbers of different *cis*-elements in the promoter regions of *SfWRKY* genes (upstream 2kb region). **(B)** The different colored histograms represent the sum of the *cis*-acting elements in each category.

SfWRKY52 and *SfWRKY66* exhibited consistently high expression levels throughout all examined tissues and developmental stages. We can classify the expression patterns of *SfWRKY* genes among different tissues mainly into three categories based on the tissue expression profiles: the first category includes genes with high expression levels, such as *SfWRKY52*, *SfWRKY66*, *SfWRKY43*, and *SfWRKY52*, which are highly expressed among the root, stem, leaf, and flower tissues, and across all the six developmental stages of the pod tissue. The second category includes low-expressed genes among many tissues, such as *SfWRKY28*, *SfWRKY68*, and *SfWRKY47*, among all the four tissues; *SfWRKY12* and *SfWRKY28* across the pod tissues; and *SfWRKY12*, *SfWRKY28*, and *SfWRKY21* across the root tissues. The third category includes genes that do not express, such as *SfWRKY61*, *SfWRKY63*, and *SfWRKY62*, which are not expressed in any of the five tissues. It is noteworthy that the genes exhibiting high expression levels in one tissue or nearly ubiquitous high expression across all tissues, implying their predominant involvement in plant growth and development.

The gene coexpression networks were constructed based on transcriptome and metabolome data obtained from developmental stages of *S. flavescens* pods and roots. The results revealed that *SfWRKY29* and *SfWRKY41* were identified as hub-genes among the pod network (Figure 6B), indicating their pivotal roles in regulating the biosynthesis of secondary metabolites during pod development. IFR (2'-hydroxyisoflavanone reductase) and IFS (2-hydroxyisoflavanone synthase) emerged as key enzymes within the highly interconnected metabolic pathways. We further analyzed the coexpression network constructed based on the transcriptomics and metabolomics of *S. flavescens* root development, revealing *SfWRKY29* as a central hub with the highest connectivity to other

10.3389/fpls.2024.1520786

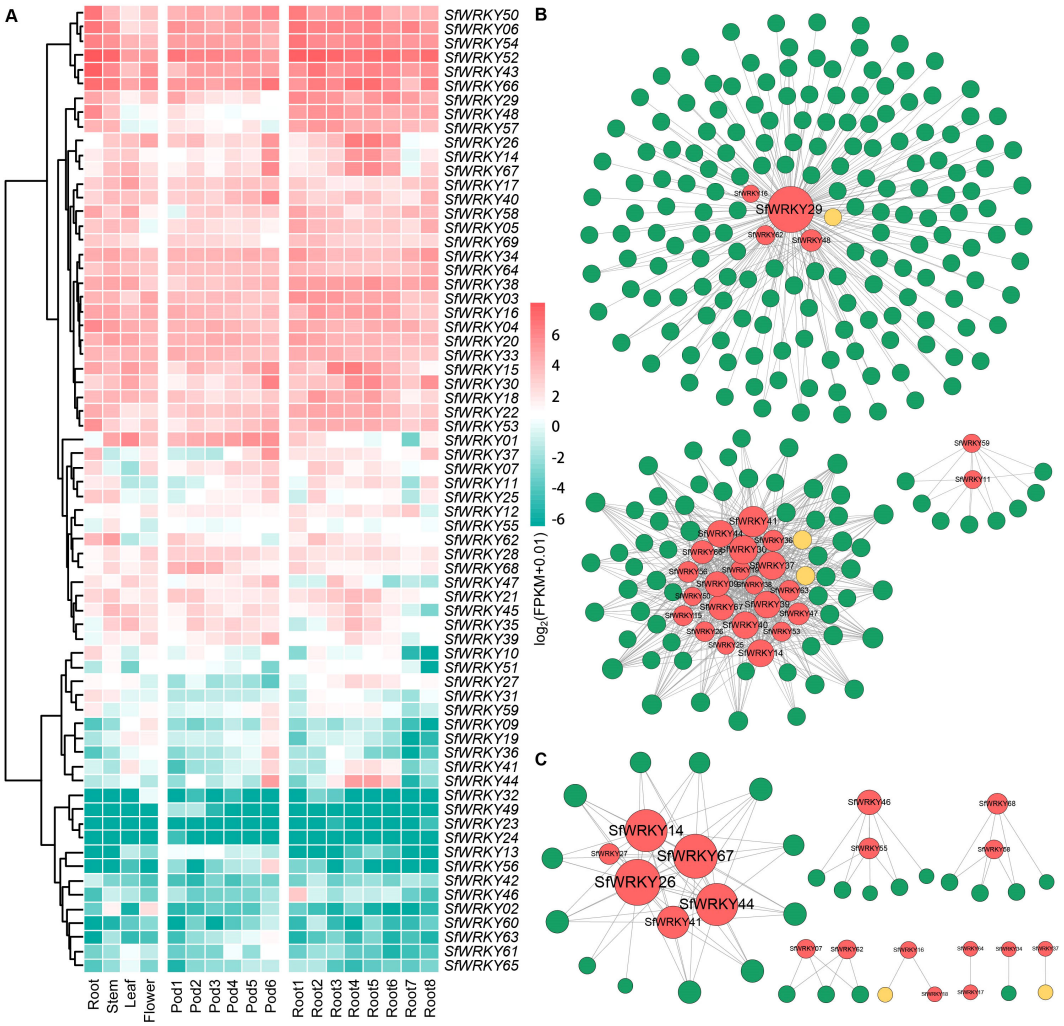


FIGURE 6
Expression patterns of *SfWRKY* genes among different tissues. **(A)** Expression profiles of the *SfWRKY* genes among different tissues or developmental stages. **(B)** Coexpression network of *SfWRKY* genes reconstructed based on transcriptome and metabolome during the pod development. **(C)** Coexpression network of *SfWRKY* genes reconstructed based on transcriptome and metabolome during the root development.

genes (Figure 6C), indicating its crucial involvement in regulating biosynthesis of secondary metabolites during root development. Notably, only I2'H (isoflavone 2'-hydroxylase) and I3'H (isoflavone 3'-hydroxylase) were identified as pivotal enzymes operating within this metabolic pathway.

3.6 Expression analysis of *SfWRKY* genes in *S. flavescens* roots with different cultivated years based on RT-qPCR

As we know, the roots of *S. flavescens* can accumulate secondary metabolites such as flavonoids and alkaloids as they grow, and the accumulation increases with the growing years (Lei et al., 2021). Excessive accumulation often leads to autotoxicity, which inhibits plant growth or causes continuous monocropping obstacle. Given the outstanding performance of *WRKY* genes in resistance to adverse conditions, we used RT-qPCR technology to analyze the expression patterns of WRKY family members in the roots of *S. flavescens* with cultivated years. As shown in Figure 7, among the 15 selected *SfWRKY* genes, all showed significantly higher expression in the SR (roots from *S. flavescens* sowed two years ago) than in the CR (roots from *S. flavescens* sowed in current years), with the greatest expression difference being observed in *SfWRKY44* (upregulated by more than 335 times), followed by *SfWRKY41* and *SfWRKY39*. Other *SfWRKY* genes also showed varying degrees of significant upregulation. This result implies that *WRKY* genes in *S. flavescens* play an important role in the response to the accumulation of secondary metabolites.

3.7 Expression analysis of *WRKY* genes in leaves of *S. flavescens* under salt stress based on RT-qPCR

To investigate the expression patterns of WRKY family members in the leaf tissues of *S. flavescens* under salt stress, we used RT-qPCR to analyze the expression patterns of 9 *SfWRKY*

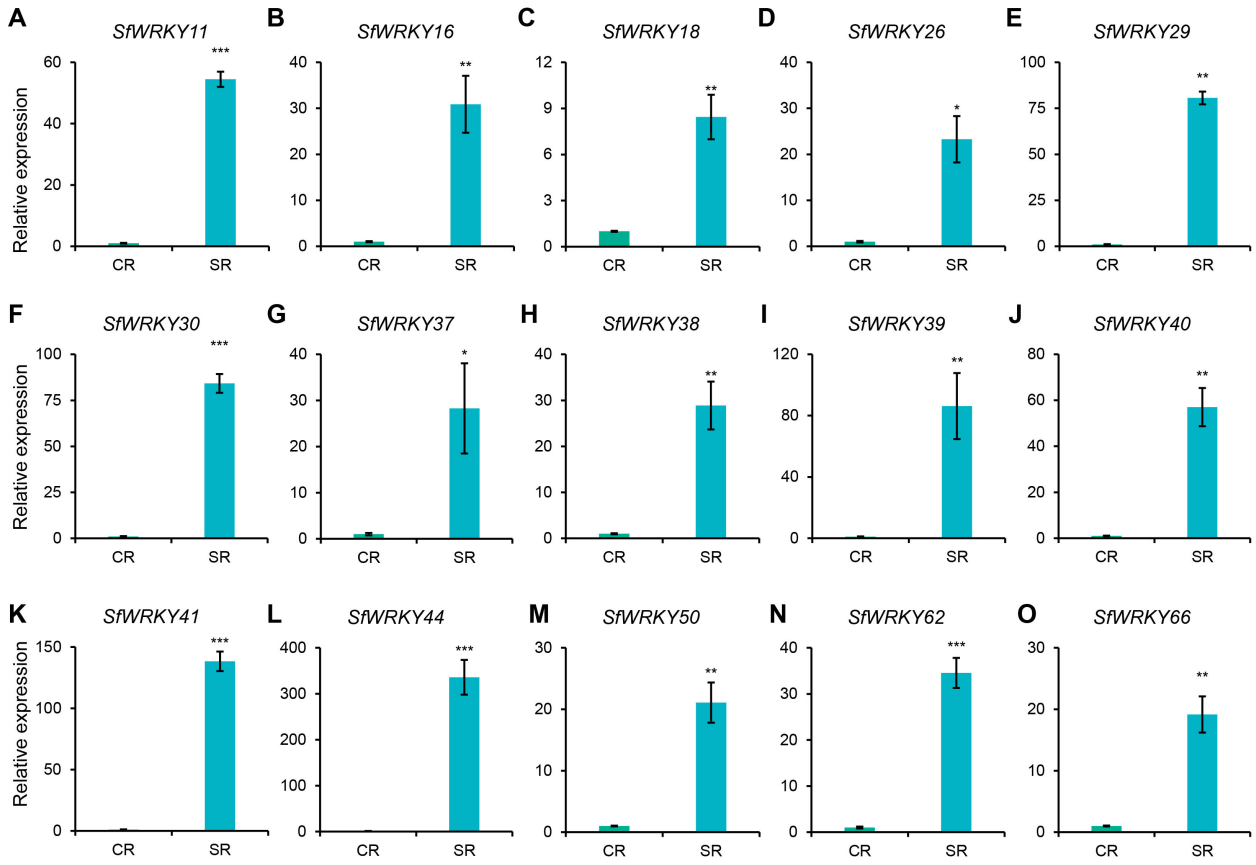


FIGURE 7
Expression analysis of *SfWRKY* genes in S. flavescens roots with different cultivated years by RT-qPCR. Panels **(A–O)** represent the relative expression of *SfWRKY11*, *SfWRKY16*, *SfWRKY18*, *SfWRKY26*, *SfWRKY29*, *SfWRKY30*, *SfWRKY37*, *SfWRKY38*, *SfWRKY39*, *SfWRKY40*, *SfWRKY41*, *SfWRKY44*, *SfWRKY50*, *SfWRKY62*, and *SfWRKY66* genes, respectively. Each sample included three biological replicates. The symbols *, **, and *** above the error bar indicate a statistically significant difference between CR (roots from S. flavescens sowed in current years) and SR (roots from S. flavescens sowed two years ago) samples at $p < 0.05$, $p < 0.01$, and $p < 0.001$ significance level.

genes (Figure 8). The results showed that seven *SfWRKY* genes (*SfWRKY18*, *SfWRKY20*, *SfWRKY35*, *SfWRKY52*, *SfWRKY55*, *SfWRKY66*, and *SfWRKY67*) were downregulated under salt stress, while two genes (*SfWRKY03* and *SfWRKY50*) were upregulated. The downregulation of *SfWRKY55* was the most significant, while the *SfWRKY03* was upregulated by 1.8 times when exposed to salt stress. Their significant expression differences under salt stress may suggest that they are involved in the response of *S. flavescens* to salt stress.

4 Discussion

S. flavescens is an herbaceous plant whose roots are often used in traditional Chinese medicine. It contains various active ingredients, including alkaloids, flavonoids, and others, which have shown significant effects in treating inflammation, fever, cancer, and skin diseases (He et al., 2015; Gao et al., 2024). *S. flavescens* also has a wide suitable growth range, including in barren gullies, shrublands, or fields. The plants exhibited robust stress resistance, demonstrating practical cold tolerance, heat resistance, and saline-alkali tolerance. It has been reported that the members of the WRKY family participate in various stress regulation networks in plants (Jiang et al., 2017; Goyal et al., 2022). Recently, studies were mainly focused on the analysis of effective components in *S. flavescens* and their related therapeutic mechanisms (Kong et al., 2024; Lin et al., 2024), but the regulatory roles of WRKY TFs in the growth and development of *S. flavescens*, especially in the regulation of stress conditions, are not clear. In this study, we conducted a comprehensive analysis of the 69 *S. flavescens* WRKY members based on the recently published genome of *S. flavescens* (Qu et al., 2023), suggesting the evolutionary diversity and complexity of the *SfWRKY* gene family. It has a significantly lower count than the 74 *WRKYs* in *Arabidopsis* and the 109 *WRKYs* in *Oryza sativa* (Supplementary Table S3) (Abdullah-Zawawi et al., 2021) but higher than those in *Nelumbo nucifera* (65), *Ananas comosus* (54), and *Glycyrrhiza uralensis* (52) (Xie et al., 2018; Li et al., 2019; Xiao et al., 2024).

Our evolutionary analysis classified the *SfWRKY* genes into three main groups, with the second group having the highest number of members, further divided into five subgroups, consistent with classifications in other plants (Eulgem et al., 2000; Rushton et al., 2010; Liu et al., 2020). Following the well-defined classification, the SfWRKY gene family was subdivided into three

10.3389/fpls.2024.1520786

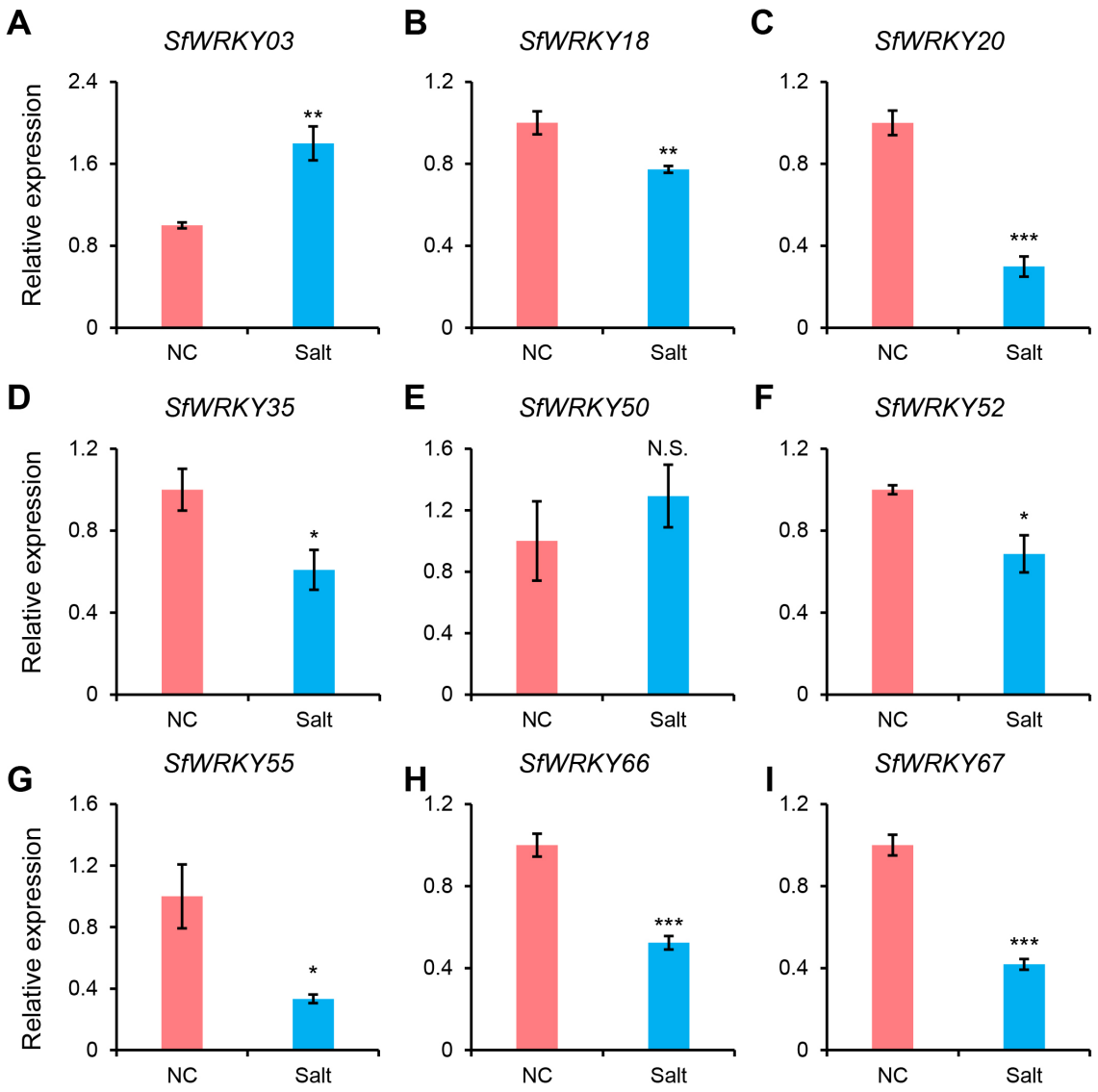


FIGURE 8
Expression analysis of *SfWRKY* genes in the leaves of S. flavescens under salt stress by RT-qPCR. Panels **(A–I)** represent the relative expression of *SfWRKY03*, *SfWRKY18*, *SfWRKY20*, *SfWRKY35*, *SfWRKY50*, *SfWRKY52*, *SfWRKY55*, *SfWRKY66*, and *SfWRKY67* genes, respectively. Each sample contain three biological replicates. The symbols *, **, and *** above the error bar indicate a statistically significant difference between the NC (normal condition) and Salt (250 mM NaCl) treatments at $p < 0.05$, $p < 0.01$, and $p < 0.001$ significance level (Student's t test). N.S., not significant.

groups (Figure 2A). Notably, Group II contains the highest number of *SfWRKY* genes, which may be attributed to gene duplication. In Group II, a total of 11 gene duplications were identified (Table 1). It also revealed extensive collinearity between *S. flavescens*, *A. thaliana*, and *S. moorcroftiana*, suggesting a common ancestor before the divergence of these lineages. The variation in collinearity between *S. flavescens* and *S. moorcroftiana*, compared to *A. thaliana*, is consistent with their evolutionary relationships. The presence of conserved motifs in *SfWRKY* genes supports their functional conservation across different plants. Gene and genome duplications have long been considered as a fundamental source of evolutionary innovation, offering an expanded molecular reservoir for the adaptive evolution of key pathways, plant development, and ecological transitions (Panchy et al., 2016). Previous studies suggest that the expansion of the WRKY gene family was mainly due to tandem and segmental duplication events (Chen et al., 2019). Gene duplication analysis suggests that TRD and DSD have greatly contributed to the expansion of SfWRKY gene family (Table 1). Selection pressure analysis showed that *SfWRKYs* have undergone purifying selection, offering an explanation for the observed differences. The *Ka*/*Ks* ratio calculations for all inferred duplicated genes were less than one. Therefore, these gene pairs may have experienced negative selection after duplication, with limited functional divergence, indicating stable changes in amino acid sequences and subfuncitionalization during the evolution of *S. flavescens* (Cusack and Wolfe, 2007).

WRKY TFs are crucial in the regulation of gene expression, as they specifically bind to the W-box motif located within the promoter

TABLE 1 Synonymous and nonsynonymous substitution rates for the duplicated gene pairs among *S. flavescens WRKY* genes.

Duplicate gene 1	Duplicate gene 2	Duplication pattern	*Ka*	*Ks*	*Ka*/*Ks*
SfWRKY04	*SfWRKY33*	WGD	0.127513	0.416783	0.305945
SfWRKY11	*SfWRKY25*	DSD	0.186373	0.531341	0.350759
SfWRKY41	*SfWRKY40*	TRD	0.330525	1.78668	0.184994
SfWRKY19	*SfWRKY25*	DSD	0.563535	2.83681	0.198651
SfWRKY06	*SfWRKY53*	DSD	0.452041	2.89165	0.156326
SfWRKY48	*SfWRKY53*	DSD	0.625776	3.7787	0.165606
SfWRKY16	*SfWRKY52*	TRD	0.526062	3.89375	0.135104
SfWRKY15	*SfWRKY45*	TRD	0.630472	3.92101	0.160793
SfWRKY31	*SfWRKY53*	DSD	0.538184	3.99948	0.134563
SfWRKY10	*SfWRKY25*	DSD	0.523031	4.0695	0.128525
SfWRKY68	*SfWRKY33*	TRD	0.625876	4.20783	0.148741
SfWRKY20	*SfWRKY33*	TRD	0.497031	4.21263	0.117986
SfWRKY61	*SfWRKY33*	TRD	0.576182	4.22011	0.136532
SfWRKY03	*SfWRKY33*	TRD	0.513669	4.30337	0.119364
SfWRKY30	*SfWRKY33*	TRD	0.561847	4.31147	0.130314

regions of target genes (Jiang et al., 2017). Similar to other plant species, the majority of SfWRKY proteins possess a WRKYGQK domain. Nevertheless, multiple variants of the *SfWRKY* gene have been identified (Figure 5); for instance, the WRKYGKK variant in soybean failed to effectively bind to the W-box (Zhou et al., 2008). The similarities in characteristics between SfWRKY35 and SfWRKY47 in *S. flavescens* require further investigation.

The *cis*-acting elements within gene promoters are crucial for understanding gene regulation, as they interact with transcription factors (Hernandez-Garcia and Finer, 2014). The promoters of *SfWRKY* genes contain various *cis*-acting elements that are closely associated with stress responses, plant hormone signaling, and plant growth and development. This indicates their significant roles in the response to both biotic and abiotic stresses. Previous study indicates that the expression of *WRKY* genes in specific tissues significantly influences plant growth and development (Wang et al., 2023). In this study, we observed significant differences in the expression levels of various *SfWRKY* genes across the leaves, flowers, pods, roots, and stems of *S. flavescens*. Among these, *SfWRKY52* and *SfWRKY66* exhibited consistently high expression levels across all the tissues and developmental stages (Figure 6A), suggesting their crucial roles involved in plant growth and development. Furthermore, through a coexpression network analysis of transcriptomic and metabolomic data from the pods and roots of *S. flavescens*, we identified *SfWRKY29* as a core gene. This finding implies its key regulatory role in biosynthesis during pod and root development.

The *WRKY* gene family plays a crucial role in regulating plant responses to various abiotic and biotic stresses (Jiang et al., 2017). To investigate the response of *SfWRKY* genes under salt stress, we analyzed their expression levels. The results indicated that eight *SfWRKY* genes exhibited differential expression under salt stress conditions. Overexpression of the *GmWRKY34* gene in *A. thaliana* significantly enhanced the plants' salt tolerance (Zhou et al., 2015), while *CdWRKY2* was found to negatively regulate lateral root growth under salt stress (Shao et al., 2023). Therefore, we hypothesize that these eight *SfWRKY* genes may be involved in the regulation of leaf responses to salt stress in *S. flavescens*. Of course, more studies are required to provide functional validation of *SfWRKY* genes through molecular biology techniques.

5 Conclusion

Our study conducted a comprehensive analysis of the SfWRKY gene family in *S. flavescens* through bioinformatic methods. A total of 69 *SfWRKY* genes were identified and classified into seven subfamilies (I, IIa, IIb, IIc, IId, IIe, and III), and characterize the physicochemical properties, chromosomal locations, phylogenetic relationships, synteny features, gene structures and *cis*-regulatory elements were characterized. Coexpression analysis of the transcriptomes and metabolomes from different tissues or different stages, it was found that *SfWRKY29* exhibited the highest connectivity with other genes, indicating that it plays a crucial role in regulating the biosynthesis of secondary metabolites. The RT-qPCR results of gene expression analysis revealed that some *SfWRKY* genes of *S. flavescens* were significantly induced in response to the accumulation of secondary metabolites or salt stress. Our study would lay a foundation for understanding the roles of *WRKY* genes in the growth and development of *S. flavescens* as well as their molecular mechanisms under abiotic stress.

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

JL: Data curation, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. XW: Data curation, Funding acquisition, Writing – review & editing. JJL: Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft. HS: Data curation, Formal analysis, Writing – original draft. HL: Conceptualization, Funding acquisition, Writing – review & editing, Investigation. TN: Conceptualization, Investigation, Writing – review & editing, Methodology, Supervision. AL: Conceptualization, Methodology, Supervision, Writing – review & editing, Formal analysis, Funding acquisition, Visualization.

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 Research project supported by the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2022L515), Shanxi scholarship council of China (2021-153), and Shanxi Basic Research Foundation (Free exploration) project (20210302124232, 202203021221235). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1520786/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Relationship between WRKY numbers in reported species and their corresponding genome sizes.

SUPPLEMENTARY TABLE 1
Primer sequences used for RT-qPCR.

SUPPLEMENTARY TABLE 2
Detailed information of the *S. flavescens WRKY* gene family.

References

Abdullah-Zawawi, M.-R., Ahmad-Nizammuddin, N.-F., Govender, N., Harun, S., Mohd-Assaad, N., and Mohamed-Hussein, Z.-A. (2021). Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in *Arabidopsis* and rice. *Sci. Rep.* 11, 19678. doi: 10.1038/s41598-021-99206-y

Bao, W., Wang, X., Chen, M., Chai, T., and Wang, H. (2018). A WRKY transcription factor, PcWRKY33, from Polygonum cuspidatum reduces salt tolerance in transgenic *Arabidopsis thaliana*. *Plant Cell Rep.* 37, 1033–1048. doi: 10.1007/s00299-018-2289-2

Chen, X., Li, C., Wang, H., and Guo, Z. (2019). WRKY transcription factors: evolution, binding, and action. *Phytopathol. Res.* 1, 13. doi: 10.1186/s42483-019-0022-x

Chen, T., Li, Y., Xie, L., Hao, X., Liu, H., Qin, W., et al. (2021). AaWRKY17, a positive regulator of artemisinin biosynthesis, is involved in resistance to Pseudomonas syringae in *Artemisia annua*. *Hortic. Res.* 8, 217. doi: 10.1038/s41438-021-00652-6

Cheng, W., Liu, D., Guo, M., Li, H., and Wang, Q. (2022). Sophoraflavanone G suppresses the progression of triple-negative breast cancer via the inactivation of EGFR-PI3K-AKT signaling. *Drug Dev. Res.* 83, 1138–1151. doi: 10.1002/ddr.21938

Cheng, X., Zhao, Y., Jiang, Q., Yang, J., Zhao, W., Taylor, I. A., et al. (2019). Structural basis of dimerization and dual W-box DNA recognition by rice WRKY domain. *Nucleic Acids Res.* 47, 4308–4318. doi: 10.1093/nar/gkz113

Cusack, B. P., and Wolfe, K. H. (2007). When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* 23, 270–272. doi: 10.1016/j.tig.2007.03.010

Dong, Y., Jia, G., Hu, J., Liu, H., Wu, T., Yang, S., et al. (2021). Determination of alkaloids and flavonoids in *Sophora flavescens* by UHPLC-Q-TOF/MS. *J. Anal. Methods Chem.* 2021, 1–13. doi: 10.1155/2021/9915027

Duan, M. R., Nan, J., Liang, Y. H., Mao, P., Lu, L., Li, L., et al. (2007). DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein. *Nucleic Acids Res.* 35, 1145–1154. doi: 10.1093/nar/gkm001

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–d432. doi: 10.1093/nar/gky995

Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/s1360-1385(00)01600-9

Feng, L., Li, Y., Zhou, Y.-L., Meng, G.-H., Ji, Z.-L., Lin, W.-H., et al. (2023). Integrative transcriptomic and proteomic analyses reveal a positive role of BES1 in salt tolerance in *Arabidopsis*. *Front. Plant Sci.* 14, 1034393. doi: 10.3389/fpls.2023.1034393

Feng, L., Teng, F., Li, N., Zhang, J.-C., Zhang, B.-J., Tsai, S.-N., et al. (2024). A reference-grade genome of the xerophyte *Ammopiptanthus mongolicus* sheds light on its evolution history in legumes and drought-tolerance mechanisms. *Plant Commun.* 5, 100891. doi: 10.1016/j.xplc.2024.100891

Gao, Y. F., Liu, J. K., Yang, F. M., Zhang, G. Y., Wang, D., Zhang, L., et al. (2019). The WRKY transcription factor WRKY8 promotes resistance to pathogen infection and mediates drought and salt stress tolerance in Solanum lycopersicum. *Physiol. Plant.* 168, 98–117. doi: 10.1111/ppl.12978

Gao, B. B., Wang, L., Li, L. Z., Fei, Z. Q., Wang, Y. Y., Zou, X. M., et al. (2024). Beneficial effects of oxymatrine from Sophora flavescens on alleviating Ulcerative colitis by improving inflammation and ferroptosis. *J. Ethnopharmacol.* 332, 118385. doi: 10.1016/j.jep.2024.118385

Goyal, P., Devi, R., Verma, B., Hussain, S., Arora, P., Tabassum, R., et al. (2022). WRKY transcription factors: evolution, regulation, and functional diversity in plants. *Protoplasma* 260, 331–348. doi: 10.1007/s00709-022-01794-7

Han, J., Wang, H., Lundgren, A., and Brodelius, P. E. (2014). Effects of overexpression of AaWRKY1 on artemisinin biosynthesis in transgenic *Artemisia annua* plants. *Phytochemistry* 102, 89–96. doi: 10.1016/j.phytochem.2014.02.011

He, X., Fang, J., Huang, L., Wang, J., and Huang, X. (2015). *Sophora flavescens* Ait.: Traditional usage, phytochemistry and pharmacology of an important traditional Chinese medicine. *J. Ethnopharmacol.* 172, 10–29. doi: 10.1016/j.jep.2015.06.010

Hernandez-Garcia, C. M., and Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* 217-218, 109–119. doi: 10.1016/j.plantsci.2013.12.007

Huang, X., Jia, A., Huang, T., Wang, L., Yang, G., and Zhao, W. (2023). Genomic profiling of WRKY transcription factors and functional analysis of CcWRKY7, CcWRKY29, and CcWRKY32 related to protoberberine alkaloids biosynthesis in *Coptis chinensis* Franch. *Front. Genet.* 14, 1151645. doi: 10.3389/fgene.2023.1151645

Huang, Z., Wang, J., Li, Y., Song, L., Chen, D. E., Liu, L., et al. (2022). A WRKY protein, MfWRKY40, of resurrection plant *Myrothamnus flabellifolia* plays a positive role in regulating tolerance to drought and salinity stresses of *Arabidopsis*. *Int. J. Mol. Sci.* 23, 8145. doi: 10.3390/ijms23158145

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and β-amylase from sweet potato. *Mol. Gen. Genet.* 244, 563–571. doi: 10.1007/bf00282746

Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., and Zhang, J. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59, 86–101. doi: 10.1111/jipb.12513

Kong, S., Liao, Q., Liu, Y., Luo, Y., Fu, S., Lin, L., et al. (2024). Prenylated flavonoids in *Sophora flavescens*: A systematic review of their phytochemistry and pharmacology. *Am. J. Chin. Med.* 52, 1087–1135. doi: 10.1142/s0192415x24500447

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

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

Lee, J., Jung, J., Son, S. H., Kim, H. B., Noh, Y. H., Min, S. R., et al. (2018). Profiling of the major phenolic compounds and their biosynthesis genes in *Sophora flavescens* aiton. *Sci. World J.* 2018, 6218430. doi: 10.1155/2018/6218430

Lei, H., Niu, T., Song, H., Bai, B., Han, P., Wang, Z., et al. (2021). Comparative transcriptome profiling reveals differentially expressed genes involved in flavonoid biosynthesis between biennial and triennial *Sophora flavescens*. *Ind. Crop Prod.* 161, 113217. doi: 10.1016/j.indcrop.2020.113217

Lescot, M. (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, 325–327. doi: 10.1093/nar/30.1.325

Li, W., Wang, H., and Yu, D. (2016). *Arabidopsis* WRKY transcription factors WRKY12 and WRKY13 oppositely regulate flowering under short-day conditions. *Mol. Plant* 9, 1492–1503. doi: 10.1016/j.molp.2016.08.003

Li, J., Xiong, Y., Li, Y., Ye, S., Yin, Q., Gao, S., et al. (2019). Comprehensive analysis and functional studies of WRKY transcription factors in *Nelumbo nucifera*. *Int. J. Mol. Sci.* 20, 5006. doi: 10.3390/ijms20205006

Lin, C. F., Lin, M. H., Hung, C. F., Alshetaili, A., Tsai, Y. F., Jhong, C. L., et al. (2024). The anti-inflammatory activity of flavonoids and alkaloids from *Sophora flavescens* alleviates psoriasiform lesions: Prenylation and methoxylation beneficially enhance bioactivity and skin targeting. *Phytother. Res.* 38, 1951–1970. doi: 10.1002/ptr.8140

Liu, T., Li, Y., Wang, C., Zhang, D., Liu, J., He, M., et al. (2023). Brassica napus transcription factor Bna.A07.WRKY70 negatively regulates leaf senescence in *Arabidopsis thaliana*. *Plants* 12, 347. doi: 10.3390/plants12020347

Liu, A., Liu, C., Lei, H., Wang, Z., Zhang, M., Yan, X., et al. (2020). Phylogenetic analysis and transcriptional profiling of *WRKY* genes in sunflower (*Helianthus annuus* L.): Genetic diversity and their responses to different biotic and abiotic stresses. *Ind. Crop Prod.* 148, 112268. doi: 10.1016/j.indcrop.2020.112268

Liu, A., Lu, J., Song, H., Wang, X., Wang, M., Lei, Z., et al. (2024). Comparative genomics and transcriptomics analysis of the bHLH gene family indicate their roles in regulating flavonoid biosynthesis in *Sophora flavescens*. *Front. Plant Sci.* 15, 1445488. doi: 10.3389/fpls.2024.1445488

Long, L., Gu, L., Wang, S., Cai, H., Wu, J., Wang, J., et al. (2023). Progress in the understanding of WRKY transcription factors in woody plants. *Int. J. Biol. Macromol.* 242, 124379. doi: 10.1016/j.ijbiomac.2023.124379

Luo, D., Xian, C., Zhang, W., Qin, Y., Li, Q., Usman, M., et al. (2024). Physiological and transcriptomic analyses reveal commonalities and specificities in wheat in response to aluminum and manganese. *Curr. Issues Mol. Biol.* 46, 367–397. doi: 10.3390/cimb46010024

Lv, B., Wu, Q., Wang, A., Li, Q., Dong, Q., Yang, J., et al. (2020). A WRKY transcription factor, FtWRKY46, from Tartary buckwheat improves salt tolerance in transgenic *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 147, 43–53. doi: 10.1016/j.plaphy.2019.12.004

Ma, J., Li, C., Sun, L., Ma, X., Qiao, H., Zhao, W., et al. (2023). The SlWRKY57-SlVQ21/SlVQ16 module regulates salt stress in tomato. *J. Integr. Plant Biol.* 65, 2437–2455. doi: 10.1111/jipb.13562

Ma, Z., Li, W., Wang, H., and Yu, D. (2020). WRKY transcription factors WRKY12 and WRKY13 interact with SPL10 to modulate age-mediated flowering. *J. Integr. Plant Biol.* 62, 1659–1673. doi: 10.1111/jipb.12946

Ma, F., Zhou, H., Yang, H., Huang, D., Xing, W., Wu, B., et al. (2024). WRKY transcription factors in passion fruit analysis reveals key PeWRKYs involved in abiotic stress and flavonoid biosynthesis. *Int. J. Biol. Macromol.* 256, 128063. doi: 10.1016/j.ijbiomac.2023.128063

Mirza, Z., and Gupta, M. (2024). Iron reprogrammes the root system architecture by regulating OsWRKY71 in arsenic-stressed rice (*Oryza sativa* L.). *Plant Mol. Biol.* 114, 11. doi: 10.1007/s11103-024-01420-5

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 20, 185. doi: 10.1186/s13059-019-1758-4

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

Prakash, A., Jeffryes, M., Bateman, A., and Finn, R. D. (2017). The HMMER web server for protein sequence similarity search. *Curr. Protoc. Bioinf.* 60, 3.15.11–13.15.23. doi: 10.1002/cpbi.40

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2

Qu, Z., Wang, W., and Adelson, D. L. (2023). Chromosomal level genome assembly of medicinal plant *Sophora flavescens*. *Sci. Data.* 10, 572. doi: 10.1038/s41597-023-02490-8

Rosado, D., Ackermann, A., Spassibojko, O., Rossi, M., and Pedmale, U. V. (2022). WRKY transcription factors and ethylene signaling modify root growth during the shade-avoidance response. *Plant Physiol.* 188, 1294–1311. doi: 10.1093/plphys/kiab493

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

Shao, A., Xu, X., Amombo, E., Wang, W., Fan, S., Yin, Y., et al. (2023). CdWRKY2 transcription factor modulates salt oversensitivity in Bermudagrass [*Cynodon dactylon* (L.) Pers.]. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1164534

Singh, D., Debnath, P., Sane, A. P., and Sane, V. A. (2023). Tomato (Solanum lycopersicum) WRKY23 enhances salt and osmotic stress tolerance by modulating the ethylene and auxin pathways in transgenic Arabidopsis. *Plant Physiol. Biochem.* 195, 330–340. doi: 10.1016/j.plaphy.2023.01.002

Song, H., Cao, Y., Zhao, L., Zhang, J., and Li, S. (2023). Review: WRKY transcription factors: Understanding the functional divergence. *Plant Sci.* 334, 111770. doi: 10.1016/j.plantsci.2023.111770

Ülker, B., Shahid Mukhtar, M., and Somssich, I. E. (2007). The WRKY70 transcription factor of *Arabidopsis influences* both the plant senescence and defense signaling pathways. *Planta* 226, 125–137. doi: 10.1007/s00425-006-0474-y

Wang, H., Chen, W., Xu, Z., Chen, M., and Yu, D. (2023). Functions of WRKYs in plant growth and development. *Trends Plant Sci.* 28, 630–645. doi: 10.1016/j.tplants.2022.12.012

Wani, S., Anand, S., Singh, B., Bohra, A., and Joshi, R. (2021). WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Rep.* 40, 1071–1085. doi: 10.1007/s00299-021-02691-8

Wei, Z., Ye, J., Zhou, Z., Chen, G., Meng, F., and Liu, Y. (2021). Isolation and characterization of PoWRKY, an abiotic stress-related WRKY transcription factor from *Polygonatum odoratum*. *Physiol. Mol. Biol. Plants* 27, 1–9. doi: 10.1007/s12298-020-00924-w

Xiao, J., Gao, P., Yan, B., Zhao, Y., Nan, T., Kang, C., et al. (2024). Genome-wide identification of *Glycyrrhiza uralensis* Fisch. *WRKY* gene family and expression analysis under salt stress. *Plant Stress* 13, 100520. doi: 10.1016/j.stress.2024.100520

Xie, T., Chen, C., Li, C., Liu, J., Liu, C., and He, Y. (2018). Genome-wide investigation of *WRKY* gene family in pineapple: evolution and expression profiles during development and stress. *BMC Genomics* 19, 490. doi: 10.1186/s12864-018-4880-x

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., et al. (2005). Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell.* 17, 944–956. doi: 10.1105/tpc.104.026435

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous nubstitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43. doi: 10.1093/oxfordjournals.molbev.a026236

Yin, W., Wang, X., Liu, H., Wang, Y., Van Nocker, S., Tu, M., et al. (2022). Overexpression of VqWRKY31 enhances powdery mildew resistance in grapevine by promoting salicylic acid signaling and specific metabolite synthesis. *Horticult. Res.* 9, uhab064. doi: 10.1093/hr/uhab064

Yin, X., Yang, D., Liu, Y., Yang, S., Zhang, R., Sun, X., et al. (2023). *Sophora moorcroftiana* genome analysis suggests association between sucrose metabolism and drought adaptation. *Plant Physiol.* 191, 844–848. doi: 10.1093/plphys/kiac558

Yu, Y., Wu, Y., and He, L. (2023). A wheat WRKY transcription factor TaWRKY17 enhances tolerance to salt stress in transgenic Arabidopsis and wheat plant. *Plant Mol. Biol.* 113, 171–191. doi: 10.1007/s11103-023-01381-1

Zhang, M., Chen, Y., Nie, L., Jin, X., Liao, W., Zhao, S., et al. (2018). Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis*. *Sci Rep.* 8, 5197. doi: 10.1038/s41598-018-23558-1

Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genom. Proteom. Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002

Zhang, C., Xu, Y., Lu, Y., Yu, H., Gu, M., and Liu, Q.-Q. (2011). The WRKY transcription factor OsWRKY78 regulates stem elongation and seed development in rice. *Planta* 234, 541–554. doi: 10.1007/s00425-011-1423-y

Zhong, Y., Xie, Y., Zhang, D., Li, G., and Yu, J. (2024). Integrated metabolomic and transcriptomic analysis of metabolic diversity and biosynthesis of glucosinolates and flavonoids in various cultivars of radish microgreens. *Food Biosci.* 59, 104055. doi: 10.1016/j.fbio.2024.104055

Zhou, Q. Y., Tian, A. G., Zou, H. F., Xie, Z. M., Lei, G., Huang, J., et al. (2008). Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* 6, 486–503. doi: 10.1111/j.1467-7652.2008.00336.x

Zhou, L., Wang, N., Gong, S., Lu, R., Li, Y., and Li, X. (2015). Overexpression of a cotton (*Gossypium hirsutum*) *WRKY* gene, GhWRKY34, in *Arabidopsis* enhances salt-tolerance of the transgenic plants. *Plant Physiol. Bioch.* 96, 311–320. doi: 10.1016/j.plaphy.2015.08.016

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2025
DOI 10.3389/fpls.2025.1534085

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Zhipeng Liu,
Lanzhou University, China
Yunwei Zhang,
China Agricultural University, China
Piyi Xing,
Shandong Agricultural University, China

*CORRESPONDENCE
Jiajun Yan
51037529@qq.com
Shiqie Bai
baishiqie@126.com

RECEIVED 25 November 2024
ACCEPTED 16 January 2025
PUBLISHED 07 February 2025

CITATION
Li X, Song D, Li M, Li D, You M, Peng Y, Yan J and Bai S (2025) An initial exploration of core collection construction and DNA fingerprinting in *Elymus sibiricus* L. using SNP markers.
Front. Plant Sci. 16:1534085.
doi: 10.3389/fpls.2025.1534085

An initial exploration of core collection construction and DNA fingerprinting in *Elymus sibiricus* L. using SNP markers

Xinrui Li[1,2,3], Daping Song[1], Mingfeng Li[3], Daxu Li[3], Minghong You[3], Yan Peng[2], Jiajun Yan[1*] and Shiqie Bai[1*]

[1]School of Life Science and Engineering, Southwest University of Science and Technology, Mianyang, China, [2]College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, China, [3]Institute of Herbaceous Plants, Sichuan Academy of Grassland Science, Chengdu, China

Elymus sibiricus L., an excellent forage and ecological restoration grass, plays a key role in grassland ecological construction and the sustainable development of animal husbandry. In China, the wild germplasm resources of *E. sibiricus* are abundant, and they are shaped by similar and contrasting climatic conditions to form distinct populations, which enrich the genetic diversity of *E. sibiricus*. To more comprehensively aggregate *E. sibiricus* germplasm resources at a lower cost and to more accurately utilize its genetic variation, this study conducted a preliminary exploration of core germplasm collections and fingerprinting of *E. sibiricus* using single nucleotide polymorphism (SNP) markers. By combining multiple evaluation measures with weighted processing, we successfully identified 36 materials from 90 wild *E. sibiricus* samples to serve as a core collection. Genetic diversity assessments, allele evaluations, and principal component analyses of the 36 core germplasm samples all indicate that these 36 samples accurately and comprehensively represent the genetic diversity of all 90 *E. sibiricus* germplasm accessions. Additionally, we identified 290 SNP loci from among the high-quality SNP loci generated by whole-genome sequencing of the 90 *E. sibiricus* samples as candidate markers. Of these, 52 SNP loci were selected as core markers for DNA fingerprinting of *E. sibiricus*. Using kompetitive allele-specific PCR (KASP) technology, we also performed population origin identification for 60 wild *E. sibiricus* germplasm accessions based on these core markers. The core SNP markers screened in this study were able to accurately distinguish between *E. sibiricus* germplasms from the Qinghai–Tibet Plateau and those from elsewhere. This study not only provides a reference for the continued collection and identification of *E. sibiricus* germplasm resources but also offers a scientific basis for their conservation and utilization.

KEYWORDS

***Elymus sibiricus* L., SNP, core collection, DNA fingerprinting, KASP, population identification**

1 Introduction

Elymus sibiricus L., the Elymus L. genus type species, is a perennial, self-pollinating, and allotetraploid grass widely distributed across Eurasia that plays a crucial role in meadow steppe and meadow communities, often as a dominant species (Yan et al., 2007b; Li et al., 2021). *Elymus sibiricus* has strong adaptability, and high forage quality. It naturally occurs across various habitats, including alpine meadows at elevations ranging from 1,500 to 4,900 meters, forest clearings, shrublands, slopes, and gravel beds along river valleys (Peng et al., 2022; Yen and Yang, 2022; Chen et al., 2023). *Elymus sibiricus* has abundant wild resources within China, where it mainly occurs in Northeast China, North China, Northwest China, and the Qinghai–Tibet Plateau (Li et al., 2021). *E. sibiricus* exhibits excellent cold resistance, with its seedlings able to withstand temperatures as low as -4°C, and it can safely overwinter under extreme low temperatures ranging from -40°C to -30°C (Yan et al., 2010). Chen and He (2004) found that even at an altitude of 3200 m in the Qinghai Lake region, *E. sibiricus* cv.duoye maintains high yields of forage and seeds. In North China, *E. sibiricus* typically returns to green in April, and in the Northwest region, it starts to germinate when the daily average temperature reaches above 0°C, and enters the greening period when the daily average temperature rises above 4°C (Li et al., 2000; Abulaiti et al., 2008). As a meso-xerophytic plant, *E. sibiricus* exhibits strong drought resistance during the germination and seedling stages, outperforming Agropyron cristatum (L.) Gaertn, and other congeneric species such as *E. dahuricus* Turcz., *E. nutans* Griseb., *E. excelsus* Turcz., and *E. tangutorum* (Nevski) Hand.-Maz in terms of drought tolerance (Yu et al., 2011; Chen et al., 2016; He et al., 2023). Liu et al. (2022) found that under drought stress, apoplastic barrier in the endodermis could maintain the balanced growth of *E. sibiricus*, which contributes to drought tolerance of *E. sibiricus*. The antioxidant defense system, key metabolic substances, specific transcription factor families, and genes related to signal transduction are also considered important factors in *E. sibiricus*' strong drought resistance (Li et al., 2020; Yang et al., 2020; Yu et al., 2022; An et al., 2024). Additionally, *E. sibiricus* has a certain tolerance to salt stress, which makes it suitable for the restoration and improvement of saline-alkali lands (Yang et al., 2015). Based on these characteristics, *E. sibiricus* is extensively used for the revegetation of degraded grasslands, soil stabilization, and forage production in the Qinghai–Tibet Plateau and other high-altitude areas of western China (Xie et al., 2015; Bai and Yan, 2020). Thus far, *E. sibiricus* is among the few native grass species that have achieved large-scale seed production and commercial utilization in the Qinghai–Tibet Plateau region owing to its high seed yield potential (Yan et al., 2024). Previously, research on the germplasm resources of *E. sibiricus* has primarily focused on resource evaluation and genetic diversity assessment based on phenotypic characteristics and second-generation molecular markers such as simple sequence repeats (SSRs) and sequence-related amplified polymorphisms (SRAPs) (Ma et al., 2008; Yan et al., 2008; Zeng et al., 2022; Zhang et al., 2024). In contrast, Guo et al. (2016) used morphological traits and SSR markers to identify four *E. sibiricus* cultivars from the Northwest Plateau of Sichuan, China. Additionally, Xie et al. (2015) utilized start codon targeted polymorphisms (SCoT) markers to identify 69 different cultivars and wild accessions of *E. sibiricus*. However, there have been no reports identifying the population origins for wild *E. sibiricus* germplasm accessions. *E. sibiricus* germplasm from different habitats have experienced varying climate types, leading to significant phenotypic and genetic differences among wild eco-geographic populations, which provides rich genetic resources and diverse selection bases for *E. sibiricus* germplasm improvement (Zhang et al., 2022; Xiong et al., 2024b; Yan et al., 2024). Characterizing and analyzing wild *E. sibiricus* germplasm greatly facilitates the preservation and utilization of these resources. However, methods based on morphological traits and the use of second-generation molecular markers to identify the origins of wild *E. sibiricus* germplasm are both time-consuming and inefficient. Moreover, the morphological variation in wild *E. sibiricus* germplasm is extensive, and its traits are often influenced by environmental changes, dramatically increasing the difficulty of source identification and hindering the collection and utilization of these germplasm resources (Yan et al., 2007a; Li et al., 2023a). Therefore, establishing an economical, efficient, and accurate method for identifying the population source of *E. sibiricus* germplasm is crucial for advancing the research and conservation of its genetic resources, origin identification, and sustainable development of the *E. sibiricus* industry.

Recently, a high-quality *E. sibiricus* reference genome (6.53 GB) of 'Chuancao No. 2', a nationally approved variety in China, has been released, which provides a foundation for population genomics studies of *E. sibiricus*. Yan et al. (2024) performed whole-genome resequencing of 90 wild *E. sibiricus* germplasm samples from various habitat types and thus identified 80,148,422 high-quality SNP loci by alignment and comparison with the 'Chuancao No. 2' *E. sibiricus* reference genome. Based on these SNP loci, the study not only highlighted the rich genetic diversity of wild *E. sibiricus* germplasm but also was able to divide the 90 wild *E. sibiricus* germplasm samples into four distinct groups, Qinghai–Tibet Plateau (QTP), Northwest China (NW), North China (NC), and Northeast China (NE) groups, providing important genetic evidence for the population classification of *E. sibiricus* resources. Compared with molecular markers such as SSRs, inter-simple sequence repeat (ISSRs), and SRAPs, SNP markers are distributed more evenly across the entire genome, and they are characterized by higher density, greater polymorphism, and more stable inheritance (Rafalski, 2002; Varshney et al., 2009; Li et al., 2023c). SNPs can be readily adapted to automated genotyping methods and can be identified via high-throughput automated detection (Zhang et al., 2020). The KBiosciences Kompetitive allele-specific PCR (KASPar) system is one of the most ideal high-throughput SNP genotyping platforms given its high accuracy and low cost (Semagn et al., 2014), and it has been widely utilized in studies on wheat (Grewal et al., 2020), rice (Steele et al., 2024), cotton (Islam et al., 2015), cucumber (Kahveci et al., 2021), and broccoli (Shen et al., 2020, 2021). Thus, constructing a core collection based on SNP markers is more accurate and effective (Van Inghelandt et al., 2010; Dou et al., 2023). A core collection is characterized by its heterogeneity,

diversity, representativeness, and practicality, as it is not merely a simple molecular genetic snapshot of an entire species but rather a representative subset that retains most of the genetic information of the original broader germplasm (Brown, 1989; Gu et al., 2023). The construction of a core collection not only provides strong support for genetic breeding, conservation biology, and systematic studies, but also has high academic and practical value in promoting germplasm resource exchange and utilization as well as gene bank management more generally (Gautam et al., 2004; Gu et al., 2023; Aribi, 2024). Lee et al. (2020) selected 67 pumpkin accessions as a core collection from among 610 pumpkin (Cucurbita moschata) germplasm accessions based on 2,071 high-quality SNPs. Similarly, Ketema et al. (2020) identified 94 accessions as a core collection from among 357 Ethiopian cowpea (Vigna unguiculata [L.] Walp.) germplasm samples based on genetic diversity analysis using SNP markers. These core accessions captured the full genetic diversity of the 357 Ethiopian cowpea germplasm samples, providing an empirical foundation for the collection, conservation, and utilization of Ethiopian cowpea germplasm. Girma et al. (2020) selected 387 sorghum (Sorghum bicolor L.) accessions as a core collection from 1,628 sorghum germplasms based on SNP markers, providing important resources for subsequent sorghum breeding, genomic research, and genetic studies.

Additionally, SNP markers are one of the marker types recommended by the International Union for the Protection of New Varieties of Plants (UPOV) in the BMT Molecular Testing Guidelines for constructing DNA fingerprint databases (Button, 2008). DNA fingerprint markers consist of a small number of highly representative markers that can be used to distinguish between different individuals or groups within the same species, and because of their advantages of being convenient and enabling rapid identification, as well as their accurate and stable results, they are widely used in the study of crop germplasm resource diversity and in variety identification (Karihaloo, 2015; Luo et al., 2023). Fan et al. (2021) selected 24 SNP loci with high polymorphism information content and strong sequence conservation to construct a DNA fingerprint map for tea plant (Camellia sinensis [L.] O. Kuntze) varieties, and it was used to accurately distinguish all 103 tested tea plant germplasm samples. Tian et al. (2021) constructed an SNP-DNA fingerprint database containing more than 20,000 maize (Zea mays L.) samples based on 200 core SNP loci; this database can thus be used in variety authentication, purity determination, and the protection of plant variety rights. Wang et al. (2021a) used SNP markers to construct a DNA fingerprint map for 216 cigar tobacco (Nicotiana tabacum L.) germplasm resources, providing a scientific basis for the selection and identification of high-quality cigar tobacco germplasm resources. However, the application of SNPs in the study of *E. sibiricus* has primarily focused on molecular marker-assisted breeding and inferring the demographic history of populations. For instance, Zhang et al. (2019) employed specific-locus amplified fragment sequencing (SLAF-seq) technology to successfully construct a high-density genetic linkage map for *E. sibiricus* and identify QTLs and candidate genes associated with seed traits. Moreover, they successfully identified genes related to the adaptation of *E. sibiricus* to high-altitude climatic conditions and important agronomic traits through Genetic Selection and Genome-Wide Association Analysis (GWAS) (Zhang et al., 2022). Xiong et al. (2024b) based on SNP data from the pan-chloroplast genome, inferred that the ancestors of *E. sibiricus* originated from the Qinghai-Tibet Plateau and underwent a complex migration history. Han et al. (2022) explored the factors influencing the geographic distribution pattern and genetic spatial structure of *E. sibiricus* on the Qinghai-Tibet Plateau using SNP markers. To date, there have been no reports on the construction of a core germplasm collection and DNA fingerprinting map for *E. sibiricus* based on SNP markers.

In this study, we thus constructed a core collection of *E. sibiricus* based on high-quality SNP markers generated from whole-genome resequencing of 90 wild *E. sibiricus* accessions (Yan et al., 2024). We further conducted genetic diversity analyses on these core accessions to evaluate their representativeness. Additionally, to accurately distinguish among different *E. sibiricus* germplasm sources and identify the population sources of wild germplasm accessions, we explored the construction of an *E. sibiricus* DNA fingerprint map based on SNP markers. Ultimately, a set of 52 core SNP markers was selected for the construction of the DNA fingerprint map. Subsequently, we utilized the KASPar platform to genotype 60 representative wild *E. sibiricus* accessions in order to differentiate them and identify their population sources, thereby validating the accuracy and effectiveness of the characterized core SNP markers. These findings not only enhance the conservation and utilization of *E. sibiricus* germplasm resources but also provide scientific evidence and data references for their continued collection and identification.

2 Materials and methods

2.1 Plant materials and DNA extraction

Nine *E. sibiricus* samples (Supplementary Table 1) were selected for Sanger sequencing to validate the accuracy of the core SNP loci. Sixty wild *E. sibiricus* samples (Supplementary Table 2) were also selected for KASP genotyping to evaluate the effectiveness of the 31 KASP markers. All plant materials were provided by the Sichuan Academy of Grassland Sciences (SAG). The seeds of these *E. sibiricus* samples were germinated in nutrient bowls (16 cm in diameter and 16 cm in height) filled with mixed soil (soil:nutrient soil:vermiculite, 3:4:1.5 [*V*:*V*:*V*]). The nutrient bowls were then placed in a growth chamber under controlled conditions (day/night cycle 16/8 h, 20/15°C; 60 ± 5% relative humidity; 400 μE·m−2 ·s −1 PPFD). For each germplasm accession, when the plants reached the seedling stage, young leaves were collected from each plant and stored at -80°C for subsequent DNA extraction.

The leaf samples were ground with approximately 4-mm-diameter steel beads in a Sceintz-48 tissue grinder (SCIENTZ, Ningbo, China) after sufficient chilling in liquid nitrogen. Genomic DNA was extracted from the plant samples using the TIANGEN Kit (TIANGEN BIOTECH, Co., Ltd., Beijing, China) according to the manufacturer's instructions. The concentration

and purity of the DNA were measured using a NanoDrop2000 UV spectrophotometer (Thermo Scientific, Waltham, MA, USA).

2.2 Data collection

The genomic data of *E. sibiricus* 'Chuancao No. 2' and the whole-genome sequencing data of 90 wild *E. sibiricus* samples (Supplementary Table 1) were downloaded from the National Genomics Data Center (NGDC) (BioProject accession number PRJCA029280). All SNP markers used in this analysis were derived from 80,148,422 high-quality SNP loci identified in the resequencing data of these 90 *E. sibiricus* germplasm samples (Yan et al., 2024).

2.3 Core collection development

To establish a core germplasm collection of *E. sibiricus* accessions, Core Hunter II software (Thachuk et al., 2009) was used to analyze and evaluate the representativeness of 90 *E. sibiricus* accessions. Using Core Hunter II software, we applied a weighted approach combining multiple evaluation measures (modified Roger's distance and Shannon's Diversity Index) to screen the number of core accessions for proportions of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 of the total germplasm collection, which was based on the parameters recommended on the Core Hunter website (https://github.com/cropinformatics/CoreHunter). The final core collection size was determined by evaluating allele coverage (CV) and comparing the expected heterozygosity (H_e), Shannon–Weaver index, Nei's gene diversity index, and polymorphism information content (PIC) between the core collection and the entire germplasm collection.

2.4 SNP selection

To select SNP markers with high marker quality, strong representativeness, high discriminative power, a uniform distribution across the genome, and high specificity for fingerprinting analysis, we followed five criteria (outlined below), based on the research of Sun (2017) and Wang et al. (2021a): (1) specificity, (2) uniform marker distribution and high marker quality, (3) substantial PIC values, (4) Hardy–Weinberg equilibrium, (5) uniqueness. (1) To ensure that the selected SNP markers have high specificity, we extracted 200 bp sequences upstream and downstream of each SNP, totaling 401 bp. We then performed BLAST alignments and retained only those SNPs that uniquely aligned to the reference genome. (2) Based on the premise that the SNP markers were uniformly distributed across the 14 chromosomes of *E. sibiricus*, we retained SNPs with no missing genotype data and discarded those with a minor allele frequency (MAF) below 20%. (3) The PIC values of SNP loci were calculated by a Perl script, and loci with PIC values less than 0.35 were discarded. (4) Hardy–Weinberg equilibrium was tested using VCFtools software (v.0.1.13) (Danecek et al., 2011) with the parameters –max-missing 1 –maf 0.2 –hwe 0.01, and the loci with a *p*-values greater than 0.01 were retained. (5) A Perl script was used to identify SNP loci that did not have mutations in other loci within 100 bp before and after labeling. Thus, we ultimately selected high-quality SNP markers for subsequent fingerprinting analysis.

2.5 Fingerprint construction and *generation* of 2D *barcodes*

The screened SNP loci were used to construct a DNA fingerprint map using RStudio (R version 4.4.0). To facilitate the viewing of genotype information for each germplasm accession, 2D barcodes were generated for each accession using the online software Caoliaoerweima (http://cli.im/), which can provide the genotype of each germplasm accession after the barcode is scanned.

2.6 Verification of SNP locus authenticity by sanger sequencing

The specific primers for 52 core SNP loci were designed using Primer Premier5 software (Lalitha, 2000). Parameters for primer design were as follows: GC content, 44%–72%, melting temperature (Tm), 56-68°C, and length, 19-25 bp, and the primers were synthesized by Sangon Bioengineering Co., Ltd. (Shanghai, China) (Supplementary Table 3). DNA from nine *E. sibiricus* samples, selected from the 90 materials, was used as the template for PCR amplification. The total volume of the PCR mixture was 20 µl, containing 1 µl Phanta Max Super-Fidelity DNA polymerase (Vazyme, Nanjing, China), 10 µl 2X Phanta Max buffer, 1 µl dNTP (10 mM each), 2 µL of genomic DNA, 0.8 µl of primer mix (containing 10 µM of primer F and 10 µM of primer R), and 5.2 µL of ddH_2O. The PCR reaction conditions were as follows: initial denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min, with a final extension at 72°C for 5 min. All PCR-amplified products were separated by agarose gel electrophoresis, and the target fragments were extracted and recovered under a Tanon-3500 Gel Imaging System (Tanon Science & Technology Co., Ltd., Shanghai, China). The specific procedure is as follows: 5 ml of PCR product were subjected to 1% agarose gel electrophoresis at 150 V and 100 mA. After observing for 10-20 minutes, the target PCR band was excised from the gel, recovered. Subsequently, the target fragments were sequenced using the 3730XL sequencer (Thermo Fisher Scientific), and the sequencing data were analyzed using SnapGene (V6.0.2) (Li et al., 2023b) and SeqMan (v7) software (Swindell and Plasterer, 1997).

2.7 Kompetitive allele-specific PCR genotyping

For each SNP locus retained after screening, 200 bp sequences upstream and downstream of the SNP were extracted, and KASP

primers were designed and developed by Genepioneer Biotechnologies (Nanjing, China). The online platform Primer3Plus (https://www.primer3plus.com/) was utilized for the design of KASP markers, wherein for each KASP target site, one common reverse primer and two allele-specific forward primers were designed based on the flanking sequences around the variant position (SNP). The primer design parameters are as follows: GC content between 30% and 60%, with an optimal GC content of 45%, Tm value between 55 and 61°C, and the size of the PCR product not larger than 120 bp. Primers were appended with tails compatible with the standard FAM or VIC labels (FAM tail: 5′-GAAGGTGACCAAGTTCATGCT-3′; VIC tail: 5′-GAAGGTCGGAGTCAACGGATT-3′), with a targeted SNP positioned at the 3′ end. Kompetitive Allele-Specific PCR (KASP) assays were performed using the CFX Connect™ Real-Time System (Bio-Rad, Hercules, CA, USA). The newly synthesized primers were diluted to 10 μM using TE buffer (pH 8.0) and mixed according to a ratio of forward genotyping primer 1:forward genotyping primer 2:reverse universal primer of 1:1:3 (*V*:*V*:*V*) to serve as the primer mix. DNA samples were diluted to match the concentration of the lowest concentration sample in the batch, and each 5-μL reaction mixture contained 1.25 μL of the diluted DNA sample. Finally, the total amplification reaction volume was 5 μL in each well of a 96-well plate and consisted of 2.5 μL of 2× KASP Master mix (JasonGen Biological Technology Co., Ltd, Beijing, China), 1.25 μL of primer mix, and 1.25 μL of DNA sample as described previously. The 96-well PCR reaction plates are subjected to sealing, shaking, and centrifugation to ensure thorough and even mixing of the reaction system. After centrifugation, the PCR reaction is carried out using the following cycling program: activation at 95°C for 10 min, 10 touchdown cycles of 95°C for 20 s and 61–55°C for 60 s (decreasing by 0.6°C each cycle), and then 27 cycles of denaturation and annealing/elongation at 95°C for 20 s and 55°C for 60 s. Fluorescent signals of the reactions were detected and genotyping data were analyzed using the Omega Fluorostar scanner (BMG Labtech, Ortenberg, Germany) and KlusterCaller (v2.22.0.5) software (Bansal et al., 2021).

2.8 Data analysis

Based on the population VCF file, we organized the genotyping information of selected SNP loci in 90 samples using the Notepad– (version 2.2) (https://gitee.com/cxasm/notepad–) text editor. Subsequently, the organized results were imported into PowerMarker (version 3.25) software to calculate genetic diversity indices, including Polymorphic Information Content (PIC), heterozygosity rate, Minor Allele Frequency (MAF), and Nei's gene diversity index (Liu and Muse, 2005). Principal component analysis (PCA) was performed using Tassel (v5.1) (Bradbury et al., 2007), and based on the clustering results of the samples, the graphical representations were created using R language and ggplot2 package (Pryszcz and Gabaldón, 2016). The distance matrix between individuals was calculated based on SNPs using MEGA X (v10.2.6) (Kumar et al., 2018). The phylogenetic tree was constructed using IQ-TREE2 software (v2.2.0) based on the neighbor-joining method, with 1,000 bootstrap replicates (Minh et al., 2020). The specific parameter settings were as follows: -s *.phy –seqtype DNA -T 10 -B 1000 -m NJ –boot-trees. Based on the filtered SNPs, population structure was analyzed using Admixture software (v1.3.0) with the parameters -C 0.01 -s time –cv -j4 (Alexander et al., 2009). The number of subpopulations (K value) was pre-set to range from 1 to 10 for clustering, and cross-validation was performed on the clustering results. The optimal number of clusters was determined based on the minimum cross-validation error rate.

3 Result

3.1 Screening of core germplasm collection

In this study, a comprehensive evaluation and screening of 90 accessions of *E. sibiricus* germplasm resources were conducted according to weighted values of modified Roger's distance (weight 0.7) and Shannon's diversity index (weight 0.3). Through an assessment of coverage, 36 accessions (representing 40% of the germplasm resources) were ultimately identified as representative core germplasm accessions for *E. sibiricus* (Figure 1A; Supplementary Table 4, Supplementary Data 1). Further, we analyzed the genetic diversity of these 36 core accessions. The observed heterozygosity (H_o) ranged from 0.028 to 1.000 (mean = 0.044), while the expected heterozygosity (H_e) ranged from 0.095 to 0.500 (mean = 0.293). The Nei diversity index ranged from 0.097 to 0.509. Additionally, the average values of the Shannon–Wiener index and PIC were 0.459 and 0.243, respectively. These genetic diversity metrics of the core germplasm collection were found to be similar to those of the original full set of 90 accessions of *E. sibiricus* germplasm accessions ($P < 0.05$), indicating that the core germplasm collection effectively represents the overall genetic diversity (Supplementary Table 5). The allele frequency assessment revealed that the ten potential genotypes derived from the four nucleotides A, C, G, and T (as well as their dinucleotide combinations) were consistently distributed across all germplasm and the core germplasm (Figure 1B; Supplementary Table 6). Furthermore, the comparison of MAF values between the 36 core accessions and the entire germplasm pool indicated that the highest proportion of MAF values fell within the range of 0.05–0.10. The distribution of MAF values in the range of 0.20–0.50 was consistent between the core germplasm collection and the original full germplasm collection (Supplementary Figure 1). Finally, PCA of all germplasm materials and the core germplasm indicated that the core germplasm collection aligned well with the distribution plot of all *E. sibiricus* materials (Figure 1C). These results demonstrate that the 36 core accessions accurately capture the genetic diversity of all 90 *E. sibiricus* germplasm accessions.

10.3389/fpls.2025.1534085

3.2 SNP fingerprint construction

3.2.1 The screening and evaluation of SNP markers

Based on resequencing data from 90 *E. sibiricus* germplasm accessions derived from four populations (NW, NE, NC, and QTP), with an average sequencing depth of approximately 11×, a total of 80,148,422 high-quality SNPs were identified for fingerprint profiling marker screening (Yan et al., 2024). Considering the subsequent application of the selected SNP loci for KASP genotyping, the primary screening criterion was set to ensure specificity by requiring each SNP marker to align uniquely to the reference genome within a sequence window of 200 bp upstream and downstream (totaling 401 bp). Furthermore, based on other key criteria, such as the distribution of SNP markers across the genome, the missing base call rate of 0, the MAF greater than 0.2, and the PIC values greater than 0.35, a total of 290 SNP loci were ultimately selected as candidate markers (Supplementary Table 7).

The genetic diversity analysis of the 290 candidate markers revealed that H_o ranged from 0.333 to 0.633, with an average value of 0.495; the MAF ranged from 0.340 to 0.50, with an average value of 0.415. Additionally, the average values for Nei's genetic diversity index, the Shannon–Wiener index, and PIC were 0.485, 0.675, and 0.366, respectively (Figures 2A–D; Supplementary Table 8). These results indicated that the 290 candidate markers exhibited high levels of polymorphism. The accuracy and effectiveness of the candidate markers were assessed through population structure analysis of the 90 wild *E. sibiricus* germplasms using the 290 candidate markers. Through the construction of a phylogenetic tree, it was found that the 290 candidate loci could effectively distinguish *E. sibiricus* germplasms from the QTP population from those of non-QTP populations (including MW, NC, and NE populations) (Figure 3A). However, there was a certain degree of error in identifying *E. sibiricus* germplasms from the NW, NC, and NE populations. For example, although 15 *E. sibiricus* germplasms from the NW population clustered together, 11 *E. sibiricus* germplasms from the NW population were scattered among the NE and NC populations. Furthermore, some *E. sibiricus* germplasms from the NC and NE populations were also clustered together. PCA also indicated a clear overlap among the *E. sibiricus* accessions from the NC, NE, and NW populations (Figure 3B).

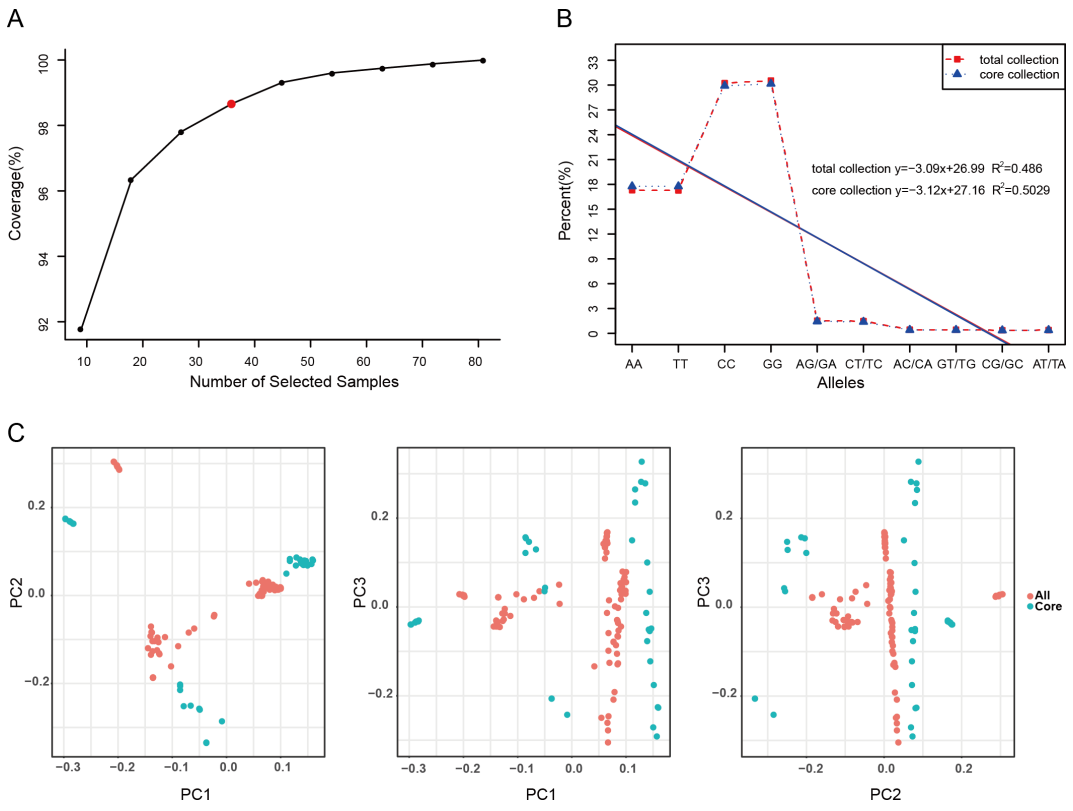


FIGURE 1
Screening and evaluation of the *Elymus sibiricus* core collection. **(A)** The evaluation of allele coverage for 90 *E. sibiricus* samples. The red dot in the graph represents the *E. sibiricus* core collection of 36 accessions. **(B)** The frequency distribution of genotypes in all 90 *E. sibiricus* samples and the core collection of 36 accessions. The ten potential genotypes are shown along the *x*-axis, and the *y*-axis shows the proportion of each genotype. The red dashed line indicates the genotype distribution of all germplasm resources, while the blue dashed line indicates the genotype distribution of the selected core germplasm accessions. The red solid line represents the fitted curve based on the genotype distribution of all germplasm resources, and the blue solid line represents the fitted curve based on the genotype distribution of the selected core germplasm. The upper right corner of the graph shows the linear equation fit to the data, and R^2 is the coefficient of determination. **(C)** The principal component analysis (PCA) plot of 90 *E. sibiricus* samples and the selected core collection. Each dot represents a sample, the red dots represent all 90 *E. sibiricus* samples, and the blue dots represent the core collection of 36 accessions.

10.3389/fpls.2025.1534085

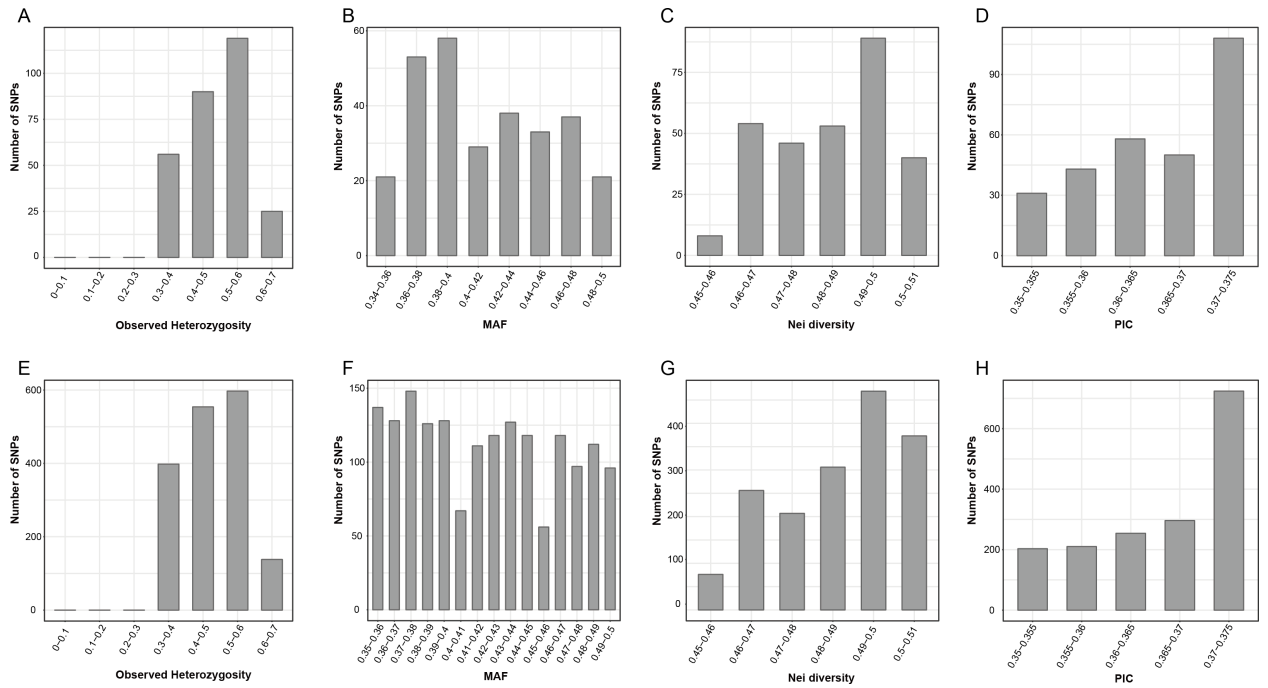


FIGURE 2
Population genetic analysis of *Elymus sibiricus* accessions based on SNP loci. **(A)** Observed heterozygosity (H_O), **(B)** minor allele frequency (MAF), **(C)** Nei's diversity index, and **(D)** polymorphism information content (PIC) values of 90 *E. sibiricus* samples based on a set of 290 candidate SNP markers. **(E)** H_O, **(F)** MAF, **(G)** Nei's diversity index, and **(H)** PIC values of 90 *E. sibiricus* samples based on a set of 338 candidate SNP markers (FP338).

Given that the classification of the NE, NW, and NC populations of *E. sibiricus* germplasms is based on whole-genome SNP data and that there is geographic overlap among these three populations, some materials from the NE and NC populations and NW and NE populations were obtained from locations that are geographically extremely close to each other (Supplementary Figure 2). Therefore, we hypothesize that it is challenging to completely and accurately distinguish accessions from the NE, NC, and NW populations from each other using a simple set of SNP loci combinations alone. To test this hypothesis, we abandoned the specificity criteria for SNP locus selection and constructed a set of candidate markers comprising 338 SNPs (FP338) with higher genetic diversity (Figures 2E–H; Supplementary Tables 8, 9). Using the FP338 loci, a phylogenetic tree and PCA were conducted for all 90 *E. sibiricus* germplasm accessions (Figures 3C, D). Using this subset of SNPs, it was still not possible to completely distinguish the materials from the NE, NC, and NW populations, with seven accessions from the NW population, six from the NE population, and two from the NC population still failing to be accurately classified. Notably, in the classification results using the FP338 candidate markers, seven germplasms from the QTP population (SAG-XZ18018, SAG-XZ18002, SAG-XZ18004, SAG-XZ18014, SAG-XZ18006, SAG-XZ18007, and SAG-XZ18012) were clustered together. Additionally, the genetic distances among these seven germplasms were also closer in the clustering results based on the 290 SNP loci, which was consistent with the results of the delineation of QTP populations accessions based on genome-wide SNPs, indicating that these candidate loci effectively reflect the genetic similarities among these germplasms.

3.2.2 Screening of core SNP markers and construction of DNA fingerprints

Despite the limitations of the 290 candidate SNPs for use in identifying the population location of some *E. sibiricus* germplasm accessions from NE, NC, and NW populations, they could still be used to accurately predict the source of germplasm from the QTP population, as well as distinguish differences among individual germplasm accessions. In order to rapidly and economically distinguish *E. sibiricus* germplasm sources, 52 core SNPs were ultimately selected after screening for loci with high PIC and MAF values from a pool of 290 candidate markers, and ensuring that these SNP loci were evenly distributed across the 14 chromosomes of *E. sibiricus*, with the genotypes C/C, A/A, T/T, and G/G represented by yellow, green, blue, and purple, respectively; additionally, missing sites are represented in gray, and heterozygous sites are shown in white (Figure 4A; Supplementary Data 2). Using this set of core SNP loci, pairwise comparisons were conducted for the 90 *E. sibiricus* samples, and the results successfully distinguished each *E. sibiricus* germplasm accession (Figure 4B). Furthermore, through the analysis of the population structure of the 90 *E. sibiricus* samples, it was possible to accurately differentiate germplasms from the QTP population from those obtained from non-QTP populations, which indicated that the 52 core loci indeed effectively represent the 290 candidate SNPs (Figure 5). Genotyping data for the 52 core SNPs of the 90 *E. sibiricus* germplasms were encoded using the online software Caoliaoerweima (http://cli.im/), and 2D barcode fingerprints were generated for each germplasm (Supplementary Data 3).

3.2.3 Design of KASP primers and identification of *E. sibiricus* germplasm resources

The accuracy of analysis based on the 52 core SNP loci was further evaluated via PCR and Sanger sequencing for nine samples (SAG-NM18033, SAG-HB18004, SAG-XJ18028, SAG-XJ18013, SAG-NM18050, SAG-SC18007, SAG-GS18003, SAG-XZ18007, and PI598784) that were randomly selected from among the 90 samples. We found that the verification results for some SNP loci deviated from those obtained by next-generation sequencing, leading to reduced accuracy and reduced specificity of the loci. Ultimately, only 31 out of the 52 core SNPs could be successfully converted into KASP markers (Supplementary Table 10).

To evaluate the effectiveness of these 31 KASP markers in classifying *E. sibiricus* germplasms, 16 samples were randomly selected from among the 90 germplasm accessions and subjected to KASP analysis along with 44 other *E. sibiricus* germplasm accessions from the NC, NW, NE, and QTP populations (Supplementary Table 2). We observed that 27 KASP markers demonstrated a high discriminatory power, which were able to achieve distinct genotyping results in 60 samples (Supplementary Data 4). However, there were four KASP markers (KASP-SNP14, KASP-SNP20, KASP-SNP28, KASP-SNP18) with relatively poor typing results. In some samples, the homozygous or heterozygous status of the corresponding SNP sites could not be accurately identified, which might be related to the specificity of these loci in different samples. A genetic distance matrix and a phylogenetic tree were constructed based on the genotyping results (Supplementary Table 11). Thus, accessions from the QTP population were separated by smaller genetic distances and were clearly clustered together in the phylogenetic tree (Supplementary Figure 3). However, three samples from outside the QTP population (NC-test12, NC-test9, and NE-test6) were mixed into this cluster, such that the genetic distance between NC-test9 and QTP-test4 was 0.0556, that between NE-test6 and QTP-test5 was 0.1429, and that between NC-test12 and QTP-test8 was 0. This unexpected clustering may be related to the failure of some SNP loci to serve as functional KASP markers. Furthermore, there remained a lack of clarity in the population source identification of germplasms from the NW, NC, and NE populations. For example, two samples from the NC population (NC-test4 and SAG-NM18037) and four samples from the NE population (NE-test9, NE-test10, NE-test5, and PI598775) clustered with most of the samples from the NW population. Additionally, three samples from the NE population (NE-test11, NE-test8, and NE-test1) were also clustered with eight

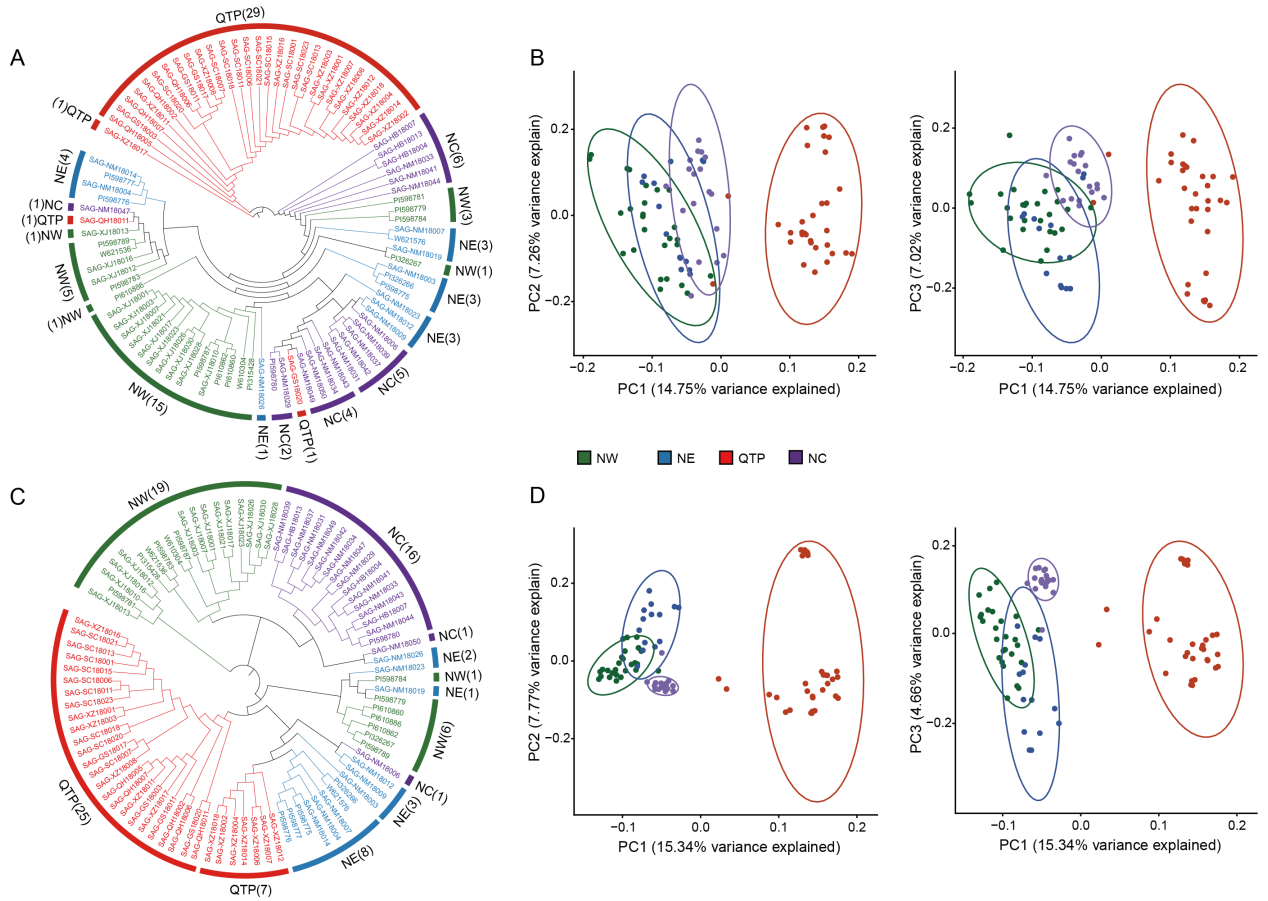


FIGURE 3
Population structure analysis of 90 *E. sibiricus* samples based on the selected SNP loci. **(A, B)** The phylogenetic tree and principal component analysis (PCA), respectively, of 90 *E. sibiricus* samples based on a set of 290 candidate SNP markers. **(C, D)** The phylogenetic tree and PCA, respectively, of 90 *E. sibiricus* samples based on a set of 338 candidate SNP markers (FP338).

10.3389/fpls.2025.1534085

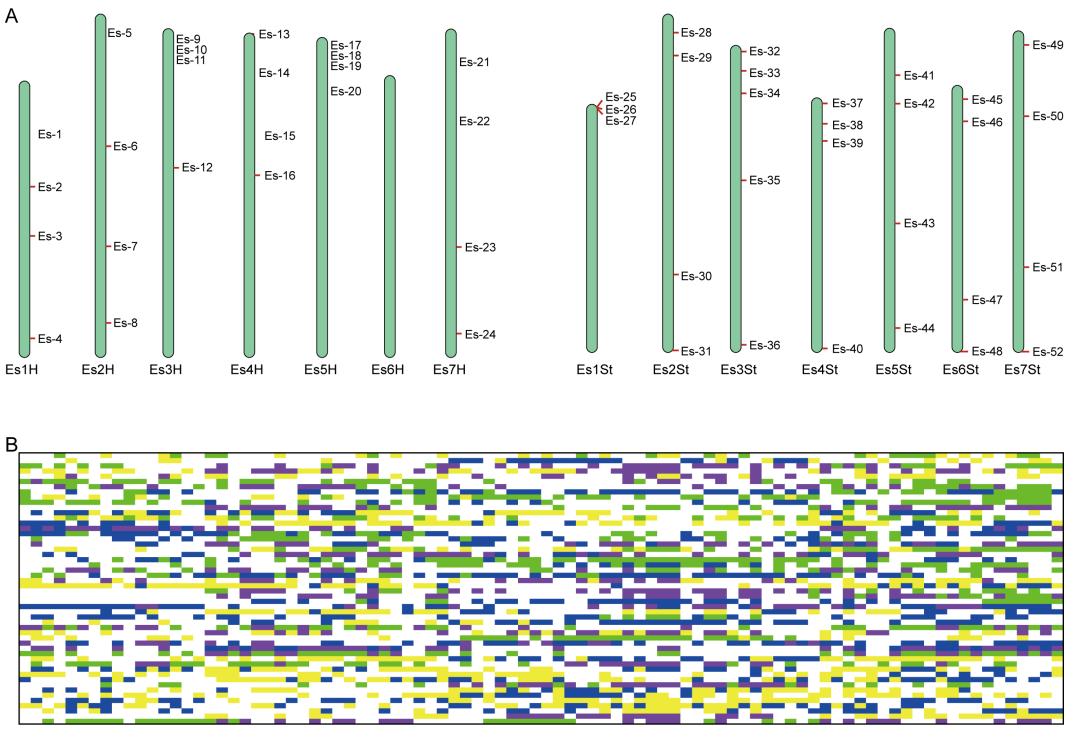


FIGURE 4
Fingerprint analysis of 90 *Elymus sibiricus* samples. **(A)** The distribution of 52 core SNP markers on 14 *E. sibiricus* chromosomes. **(B)** DNA fingerprint composed of 52 core SNP markers. Each row corresponds to a SNP marker, and each column corresponds to a *E. sibiricus* sample. The genotypes C/C, A/A, T/T, and G/G are represented in yellow, green, blue, and purple, respectively. Missing sites are represented in gray, and heterozygous sites are shown in white.

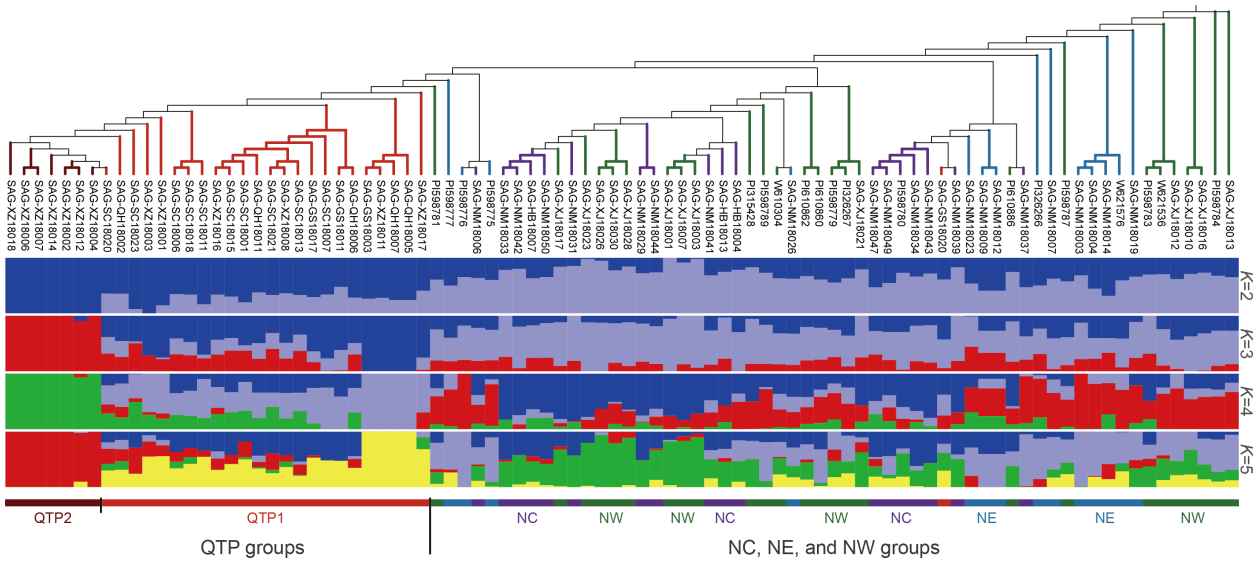


FIGURE 5
Population structure analysis of 90 *Elymus sibiricus* samples based on the 52 core SNP markers. From top to bottom, the figure shows a phylogenetic tree of 90 *E. sibiricus* samples based on 52 core SNP markers, the sample names, and the population structure of 90 *E. sibiricus* germplasm resources for different numbers of subpopulations *K*.

germplasms from the NC population. These results again highlight the limitations of using a small number of SNP locus combinations to differentiate *E. sibiricus* germplasm accessions with similar genetic backgrounds.

4 Discussion

As a widely distributed species across the Eurasian continent, the wild germplasm resources of *E. sibiricus* are extremely rich and abundant (Ma et al., 2012). Influenced by climatic factors in different habitats, the genetic variation among *E. sibiricus* accessions exhibits distinct regional characteristics (Li et al., 2023e); for instance, the phenology of heading and the presence of small spines at the base of the stems and leaf sheaths in wild *E. sibiricus* are significantly associated with high altitude (Li et al., 2023a). The phenotype characterized by the appearance of downy hair on the basal leaf sheaths during the seedling stage is found to have an extremely significant positive correlation with both longitude and latitude, while it demonstrates an extremely significant negative correlation with altitude, annual mean temperature, and annual average rainfall (Li et al., 2023d). Genetic variation in these traits, which manifest clear regional characteristics, may represent adaptations of *E. sibiricus* germplasm under different environmental conditions, not only enriching the genetic diversity of *E. sibiricus* but also promoting population differentiation and providing an important genetic resource foundation for the development and breeding of new *E. sibiricus* varieties with superior traits (Zhang et al., 2022; Xiong et al., 2024a). Furthermore, determining how to scientifically identify, preserve, and protect these wild *E. sibiricus* germplasm resources is a prerequisite for their effective utilization. The genetic diversity of multiple important crops, including sorghum (Cuevas et al., 2017), sweet potato (Su et al., 2017), pumpkin (Lee et al., 2020), cowpea (Ketema et al., 2020), radish (Xing et al., 2024), and wheat (Soleimani et al., 2020), has been preserved as much as possible through the construction of a core germplasm collection based on SNP markers. However, in the construction of a core germplasm collection for *E. sibiricus*, the focus has primarily remained on phenotypic traits and the use of SSRs as markers. Yan et al. (2017) used SSR markers to identify 47 accessions as the core germplasm collection from among 148 samples of *E. sibiricus*. Zeng et al. (2022) identified five accessions as the core germplasm collection for breeding by evaluating nine agronomic traits of 76 *E. sibiricus* germplasm accessions. The observation of plant phenotypic traits is one of the more traditional and intuitive methods of core germplasm identification. However, most of the phenotypic traits in plants are quantitative traits affected by multiple minor-effect genes and are therefore often susceptible to environmental influences (Wang et al., 2021a; Mackay and Anholt, 2024). Based on literature reports and field observations, we found that even the same *E. sibiricus* germplasm accession could exhibit significant phenotypic differences under different environmental conditions (Zhang, 2020; Jia, 2021). Additionally, there is substantial variation in the phenotypes of *E. sibiricus* across different growth years. Specifically, the agronomic traits of *E. sibiricus* are generally optimal during the second and third years of growth after plants are established, showing considerable differences from those in the first year, indicating the need for a longer period for phenotypic characterization of *E. sibiricus* germplasm accessions (Zhou et al., 2000; Chen, 2013; Qi et al., 2023). These factors present a substantial hindrance in the construction of a core germplasm bank for *E. sibiricus* using phenotypic traits. Additionally, although SSR markers are widely used in the construction of plant core germplasm collections, SNP markers offer several advantages: they are more abundant, have higher density, exhibit greater levels of polymorphism, and are more stable (Rafalski, 2002; Mammadov et al., 2012). Moreover, the identification and statistical analysis of SNP markers are even easier and more convenient through the utilization of the high-throughput automatic detection capabilities of the KASP genotyping platform (Smith and Maughan, 2015; Dipta et al., 2024). Therefore, we identified 36 *E. sibiricus* accessions as the core germplasm collection based on SNP markers, which can accurately represent the genetic diversity of the original 90 *E. sibiricus* germplasm accessions. This study represents the initial exploration of constructing a core germplasm resource for *E. sibiricus* based on a large number of SNP markers, providing a valuable reference for future collection and preservation of *E. sibiricus* germplasm resources. Concurrently, this core collection not only effectively protect the genetic diversity of *E. sibiricus* germplasm and prevent the loss of resources, but also serve as the basic material for breeding, providing important genetic resources for the breeding of new varieties.

Currently, the identification of wild germplasm resources of *E. sibiricus* mainly focuses on distinguishing *E. sibiricus* germplasm from congeneric *E. nutans* germplasm. Owing to the rich phenotypic variation in the wild germplasm of these two forage grasses, some of the wild germplasm of the two species are morphologically very similar, and the regions of their geographic distributions overlap. Li et al. (2023d) identified 990 *E. sibiricus* accessions and 246 *E. nutans* accessions from 1,723 wild *Elymus* germplasm resources by combining phenotypic trait analysis with flow cytometry. However, the population source identification of wild *E. sibiricus* germplasms has not been reported. The construction of fingerprints based on SNPs and the combination of KASP for typing identification can be utilized to accurately and efficiently differentiate among various germplasm sources within a species. This approach has been widely applied in the identification of varieties and population structure analysis in many species (Yang et al., 2019; Wang et al., 2021b, 2022; Yang et al., 2022). Therefore, in this study, based on the SNP markers obtained from the whole-genome sequencing data of 90 wild *E. sibiricus* accessions from the QTP, NC, NE and NW populations, we initially explored the feasibility of identifying the population origins of wild *E. sibiricus* germplasm accessions using SNP fingerprints combined with KASP genotyping. Unlike diploid species such as maize and cigar tobacco that have already established SNP fingerprints (Tian et al., 2021; Wang et al., 2021a), *E. sibiricus*, as an allotetraploid species, presents challenges in precise SNP genotyping due to the presence of homologous chromosomes between sub-genomes and a large number of repetitive sequences on the chromosomes, which result in many markers showing

multiple copy phenomena (Yan et al., 2024). Therefore, we prioritized the specificity of SNPs as the primary criterion for selecting candidate SNP sites for fingerprint construction, in order to screen SNP sites with single-copy characteristics, thus simplifying the complex polyploid genotyping to a diploid-like genotyping, and facilitating KASP detection. However, similar to findings in the construction of SNP fingerprints for allotetraploid upland cotton (*Gossypium hirsutum* L.), this screening criterion limits the number of available SNPs and genetic diversity (Sun, 2017). In this study, the selected core SNP markers were able to accurately discriminate germplasm from QTP and non-QTP populations, but there were obvious errors in discriminating germplasm from NC, NE and NW populations. Some of the accessions from these three populations were clustered together based on the genotyping results, which indicated that the selected SNP markers lacked sufficient discriminatory power for differentiating the germplasm from these three populations. We speculate that this result may be related to the similar genetic backgrounds of individuals from these three populations. The materials from geographically closer populations may exhibit more genetic homogeneity owing to both environmental similarities as well as gene flow between populations. Consequently, it is challenging to accurately identify their population source using only a few SNP loci. For the QTP population of *E. sibiricus*, there was not only a large geographic distance from the NE, NC, and NW populations, but also genetic differences likely shaped by the unique environmental conditions of the QTP, such as its high altitude, which may lead to significant genetic differentiation from the non-QTP populations. Accordingly, even a small number of SNP loci were sufficient to distinguish germplasms from the QTP population from those from other populations. The similar findings have been reported in studies on tobacco (Wang et al., 2021a) and honeysuckle (Li et al., 2023b) fingerprint maps. Additionally, we also infer that a more comprehensive collection and sequencing analysis of *E. sibiricus* germplasm from the regions corresponding to these three populations would obtain richer SNP information. This will not only facilitate more detailed population differentiation of accessions from these regions but also help construct more discriminative SNP fingerprints, thereby playing a crucial role in identifying the population origins of *E. sibiricus* accessions. It is noteworthy that Yan et al. (2024), based on SNP datasets obtained from whole-genome sequencing, found that even *E. sibiricus* germplasm with close geographical origins were classified into different populations. Therefore, compared to the SCoT molecular markers used in previous study to identify the geographical origin of *E. sibiricus* (Xie et al., 2015), the DNA fingerprints constructed based on SNP markers in this study can identify the population origin of *E. sibiricus* germplasm, which is more beneficial for analyzing the population structure of *E. sibiricus* germplasm. Furthermore, studies have indicated that the *E. sibiricus* germplasm from the QTP population possesses higher genetic diversity, and it is inferred that the Qinghai-Tibet Plateau is very likely the center of origin for *E. sibiricus* (Xiong et al., 2024b; Yan et al., 2024). Consequently, it is of significant importance for the conservation, rational utilization and molecular breeding of *E. sibiricus* germplasm resources to rapidly discern whether the *E. sibiricus* germplasm originates from the QTP population (QTP) through fingerprints.

5 Conclusion

In this study, we successfully constructed a core collection comprising 36 *E. sibiricus* samples by integrating published sequencing data from 90 wild *E. sibiricus* accessions. Additionally, 290 candidate SNP markers and 52 core SNP markers were identified for the development of *E. sibiricus* DNA fingerprints, according to a series of strict screening criteria and evaluation methods. Subsequently, utilizing KASP technology, we genotyped 60 wild *E. sibiricus* accessions using these core SNP markers. The core SNP makers were able to accurately identify the germplasms from the QTP population, but there were some errors in the population origin identification of *E. sibiricus* germplasm from the NC, NE, and NW populations. We speculate that this result is owing to the relatively similar genetic backgrounds of *E. sibiricus* from these three populations. Therefore, we advocate for a more comprehensive collection and sequencing analysis of *E. sibiricus* germplasm resources from the regions corresponding to these three populations, which would enrich the SNP database and facilitate more precise analyses of population structure and gene flow in these areas, as well as the selection of more effective DNA marker combinations. In summary, our study preliminarily validated the feasibility of constructing a core germplasm set and DNA fingerprint for *E. sibiricus* based on SNPs. This work not only underscores the reliability and precision of SNP markers but also lays a crucial foundation for future efforts in the collection, conservation, and utilization of *E. sibiricus* germplasm resources.

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

XL: Conceptualization, Data curation, Formal Analysis, Investigation, Software, Writing – original draft, Writing – review & editing. DS: Formal Analysis, Investigation, Methodology, Writing – original draft. ML: Investigation, Methodology, Writing – original draft. DL: Writing – original draft, Conceptualization, Resources. MY: Conceptualization, Resources, Writing – original draft. YP: Conceptualization, Writing – review & editing. JY: Conceptualization, Writing – review & editing, Project administration. SB: Data curation, Funding acquisition, Project administration, Resources, 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 financially supported by the Sichuan Science and Technology Program -Science and Technology Innovation Talent Project: Precise Identification and Breeding Application of Important Traits of Elymus

sibiricus, a Native Grass Species on the Qinghai-Tibet Plateau (2023JDRC0040); Research Project on Innovation and Development of Forestry and Grassland Science and Technology (2023132016); 2023 Major Innovation Platform Construction Special Project of the National Grass Industry Technology Innovation Center -Phenotypic and Genotypic Identification of Important Grass Germplasm Resources (CCPTZX2023W01); the Natural Science Foundation of Southwest University of Science and Technology (Grant No. 23zx7143). The Development and Utilization of Elymus Germplasm Resources (2023-RC-high level-5).

Acknowledgments

We thank the laboratory staff at the Sichuan Academy of Grassland Science and the College of Grassland Science and Technology, Sichuan Agricultural University, for their support.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1534085/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of minor allele frequency (MAF) values in the core collection and the original full germplasm collection. Intervals of MAF values corresponding to SNPs are shown along the *x*-axis; the *y*-axis represents the frequency of all SNPs in each interval.

SUPPLEMENTARY FIGURE 2
Geographic distribution of 90 Elymus *sibiricus* samples, both at a broad scale and locally.

SUPPLEMENTARY FIGURE 3
Clustering analysis dendrogram of 60 wild *Elymus sibiricus* germplasm accessions based on genotyping results for 31 KASP markers. The bands next to the dendrogram indicate the population origin of the samples, where red, purple, blue, and green represent *E. sibiricus* samples from the QTP, NC, NE, and NW regions, respectively.

References

Abulaiti, M., Salamu,, and Aji, R. (2008). Analysis of the climate conditions of growth of Elymus sibinicus grass. *Xinjiang Agric. Sci.* 45 (S1), 222–224. doi: CNKI:SUN:XJNX.0.2008-S1-076

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

An, Y., Wang, Q., Cui, Y., Liu, X., Wang, P., Zhou, Y., et al. (2024). Comparative physiological and transcriptomic analyses reveal genotype specific response to drought stress in Siberian wildrye (*Elymus sibiricus*). *Sci. Rep.* 14, 21060. doi: 10.21203/rs.3.rs-4251832/v1

Aribi, M. M. (2024). "Plant gene banks: conservation of genetic resources," in *Sustainable Utilization and Conservation of Plant Genetic Diversity*. Eds. J. M. Al-Khayri, S. M. Jain and S. Penna (Springer, Singapore), 753–775.

Bai, S. Q., and Yan, J. J. (2020). *Research and utilization of Elymus sibiricus Linn. germplasm resources* (Beijing, China: Science Press).

Bansal, M., Adamski, N. M., Toor, P. I., Kaur, S., Sharma, A., Srivastava, P., et al. (2021). A robust KASP marker for selection of four pairs of linked leaf rust and stripe rust resistance genes introgressed on chromosome arm 5DS from different wheat genomes. *Mol. Biol. Rep.* 48, 5209–5216. doi: 10.1007/s11033-021-06525-4

Bradbury, P. J., Zhang, Z. W., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Brown, A. H. D. (1989). Core collections: a practical approach to genetic resources management. *Genome* 31, 818–824. doi: 10.1139/g89-144

Button, P. (2008). The international union for the protection of new varieties of plants (upov) recommendations on variety denominations. *Acta Hortic.* 799, 191–200. doi: 10.17660/ActaHortic.2008.799.27

Chen, Y. (2013). *Study on photosynthetic traits and production performance of Elymus sibiricus in the Qinghai-Tibetan plateau*. Qinghai University, Xining.

Chen, S. M., Feng, J. J., Xiong, Y., Xiong, Y. L., Liu, Y. J., Zhao, J. M., et al. (2023). Evaluation and screening of wild *Elymus sibiricus* L. germplasm resources under salt stress. *Agronomy* 13, 2675. doi: 10.3390/agronomy13112675

Chen, G., and He, L. (2004). Evaluation of ecological adaptability and productivity of two species of *Elymus* in alpine region. *Pratacultural Sci.* 21, 4. doi: 10.3969/j.issn.1001-0629.2004.09.009

Chen, Y., Zhou, Q., Sun, J., and Tian, L. (2016). Comparative study on drought resistance of six native *Elymus* L.species seedlings. *J. Southwest Minzu University (Natural Sci. Edition)* 42, 6. doi: 10.11920/xnmdzk.2016.06.002

Cuevas, H. E., Rosa-Valentin, G., Hayes, C. M., Rooney, W. L., and Hoffmann, L. (2017). Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics* 18, 1–17. doi: 10.1186/s12864-016-3475-7

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Dipta, B., Sood, S., Mangal, V., Bhardwaj, V., Thakur, A. K., Kumar, V., et al. (2024). KASP: a high-throughput genotyping system and its applications in major crop plants for biotic and abiotic stress tolerance. *Mol. Biol. Rep.* 51, 508. doi: 10.1007/s11033-024-09455-z

Dou, T. Y., Wang, C. C., Ma, Y. L., Chen, Z. Y., Zhang, J., and Guo, G. G. (2023). CoreSNP: an efficient pipeline for core marker profile selection from genome-wide SNP datasets in crops. *BMC Plant Biol.* 23, 580. doi: 10.1186/s12870-023-04609-w

Fan, X. J., Yu, W. T., Cai, C. P., Lin, Y., Wang, Z. H., Fang, W. P., et al. (2021). Construction of molecular ID for tea cultivars by using of single-nucleotide polymorphism (SNP) markers. *Scientia Agricultura Sin.* 54 (8), 1751–1760. doi: 10.3864/j.issn.0578-1752.2021.08.014

Gautam, P. L., Singh, B. B., Saxena, S., and Sharma, R. K. (2004). "Collection, conservation and utilization of plant genetic resources," in *Plant Breeding: Mendelian to Molecular Approaches*. Eds. H. K. Jain and M. C. Kharkwal (Springer, Dordrecht), 657–690.

Girma, G., Nida, H., Tirfessa, A., Lule, D., Bejiga, T., Seyoum, A., et al. (2020). A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. *Plant Genome* 13, e20055. doi: 10.1002/tpg2.20055

Grewal, S., Hubbart-Edwards, S., Yang, C., Devi, U., Baker, L., Heath, J., et al. (2020). Rapid identification of homozygosity and site of wild relative introgressions in wheat through chromosome-specific KASP genotyping assays. *Plant Biotechnol. J.* 18, 743–755. doi: 10.1111/pbi.13241

Gu, R., Fan, S. H., Wei, S. P., Li, J. R., Zheng, S. H., and Liu, G. L. (2023). Developments on Core collections of plant genetic resources: do we know enough? *Forests* 14, 926. doi: 10.3390/f14050926

Guo, Z. H., Yang, H. Y., Zhang, C. L., Sun, M., Fu, K. X., Zhang, X. Q., et al. (2016). Identification of four *Elymus sibiricus* cultivar (strains) from northwest plateau of Sichuan using morphological traits and SSR markers. *Pratacult Sci.* 33, 1718–1727. doi: 10.11829/j.issn.1001-0629.2015-0664

Han, M., Zhang, J., Li, D., Sun, S., Zhang, C., Zhang, C., et al. (2022). Phylogeographical pattern and population evolution history of indigenous *elymus sibiricus* L. @ on qinghai-tibetan plateau. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.882601

He, Y., Hu, Y., Duan, H., Han, J., Liang, H., Chao, X., et al. (2023). Comprehensive evaluation of drought resistance of four *Elymus* varieties at seedling stage. *Chin. J. Grassland* 45, 77–87. doi: 10.16742/j.zgcdxb.20220303

Islam, M. S., Thyssen, G. N., Jenkins, J. N., and Fang, D. D. (2015). Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland cotton. *Plant Genome* 8, plantgenome2014.2007.0034. doi: 10.3835/plantgenome2014.07.0034

Jia, Z. Y. (2021). *Creation of new germplasm and association analysis of main agronomic traits in Elymus sibiricus L.* Inner Mongolia Agricultural University, Hohhot (IM.

Kahveci, E., Devran, Z., Özkaynak, E., Hong, Y., Studholme, D. J., and Tör, M. (2021). Genomic-assisted marker development suitable for *Cscvy-1* selection in cucumber breeding. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.691576

Karihaloo, J. L. (2015). "DNA fingerprinting techniques for plant identification," in *Plant Biology and Biotechnology*. Eds. B. Bahadur, M. Venkat Rajam, L. Sahijram and K. Krishnamurthy (Springer, New Delhi, India), 205–221.

Ketema, S., Tesfaye, B., Keneni, G., Amsalu Fenta, B., Assefa, E., Greliche, N., et al. (2020). DArTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp] germplasms. *PloS One* 15, e0239122. doi: 10.1371/journal.pone.0239122

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, 1547–1549. doi: 10.1093/molbev/msy096

Lalitha, S. (2000). Primer premier 5. *Biotech. Software Internet Rep.* 1, 270–272. doi: 10.1089/152791600459894

Lee, H. Y., Jang, S. Y., Yu, C. R., Kang, B. C., Chin, J. H., and Song, K. (2020). Population structure and genetic diversity of *Cucurbita moschata* based on genome-wide high-quality SNPs. *Plants* 10, 56. doi: 10.3390/plants10010056

Li, Q., Chang, F., and Dong, T. (2000). A preliminary study on flowering and seed production of 6 herbage grasses. *Grassland Prataculture.* 1, 41–43. doi: CNKI:SUN:NMCY.0.2000-01-010

Li, J., Chang, X., Huang, Q., Liu, P., Zhao, X., Li, F., et al. (2023b). Construction of SNP fingerprint and population genetic analysis of honeysuckle germplasm resources in China. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1080691

Li, X., Chen, S., Yan, J., Chen, L., Zhang, J., You, M., et al. (2021). Research progress on *Elymus sibiricus* Linn. germplasm resources. *J. Grassland Forage Sci.* 01), 6–17. doi: 10.3969/j.issn.2096-3971.2021.01.002

Li, M., Li, X., Li, Y., Chen, L., Wang, Y., Li, C., et al. (2023d). Identification of wild *Elymus sibiricus* germplasm resources and analysis of variation of dorsal hairs in basal leaf sheath. *Acta Agrestia Sin.* 31, 1026. doi: 10.11733/j.issn.1007-0435.2023.04.011

Li, J. W., Li, H., Liu, Z. W., Wang, Y. X., Chen, Y., Yang, N., et al. (2023c). Molecular markers in tea plant (*Camellia sinensis*): Applications to evolution, genetic identification, and molecular breeding. *Plant Physiol. Biochem.* 198, 107704. doi: 10.1016/j.plaphy.2023.107704

Li, X. L., Liu, T., Li, A. Q., Xiao, Y. H., Sun, K. P., and Feng, J. (2023e). Diversifying selection and climatic effects on major histocompatibility complex class II gene diversity in the greater horseshoe bat. *Evolutionary Appl.* 16, 688–704. doi: 10.1111/eva.13528

Li, C. Y., Wang, Y., Li, X. R., Li, Y. Z., Li, M. F., Chen, L. L., et al. (2023a). Morphological diversity and germplasm utilization potential of wild *Elymus sibiricus*. *Acta Prataculturae Sin.* 32, 67. doi: 10.11686/cyxb2022132

Li, M., Yang, J., Wang, X., Li, D., Zhang, C., Tian, Z., et al. (2020). Transcriptome profiles identify the common responsive genes to drought stress in two *Elymus* species. *J. Plant Physiol.* 250, 153183. doi: 10.1016/j.jplph.2020.153183

Liu, K. J., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, X., Wang, P., An, Y., Wang, C.-M., Hao, Y., Zhou, Y., et al. (2022). Endodermal apoplastic barriers are linked to osmotic tolerance in meso-xerophytic grass Elymus sibiricus. *Front. Plant Sci.* 13, 1007494. doi: 10.3389/fpls.2022.1007494

Luo, Z., Yao, Z., Yang, Y., Wang, Z., Zou, H., Zhang, X., et al. (2023). Genetic fingerprint construction and genetic diversity analysis of sweet potato (*Ipomoea batatas*) germplasm resources. *BMC Plant Biol.* 23, 355. doi: 10.1186/s12870-023-04329-1

Ma, X., Chen, S. Y., Zhang, X. Q., Bai, S. Q., and Zhang, C. B. (2012). Assessment of worldwide genetic diversity of Siberian wild rye (*Elymus sibiricus* L.) germplasm based on gliadin analysis. *Molecules* 17, 4424–4434. doi: 10.3390/molecules17044424

Ma, X., Zhang, X. Q., Zhou, Y. H., Bai, S. Q., and Liu, W. (2008). Assessing genetic diversity of *Elymus sibiricus* (Poaceae: *Triticeae*) populations from Qinghai-Tibet Plateau by ISSR markers. *Biochem. Systematics Ecol.* 36, 514–522. doi: 10.1016/j.bse.2008.03.003

Mackay, T. F. C., and Anholt, R. R. H. (2024). Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat. Rev. Genet.* 25, 639–657. doi: 10.1038/s41576-024-00711-3

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012, 728398. doi: 10.1155/2012/728398

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Peng, J., Sun, X. X., Shi, H. T., Chen, Y. J., Cai, X. R., Wang, X., et al. (2022). Nutritional value and molecular structure characteristics in different fractions of *Elymus sibiricus* from Qinghai-Tibet Plateau. *Chin. J. Anim. Nutr.* 34 (8), 5270–5282. doi: 10.3969/j.issn.1006-267x2022.08.048

Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44, e113–e113. doi: 10.1093/nar/gkw294

Qi, H., Liu, W., Liu, M., Liu, M., Liang, G., Zhang, Y., et al. (2023). Interannual differences in ear traits and spike type division of *E. sibiricus* L. wheat ears on the Qinghai-Tibet Plateau. *J. Nucl. Agric. Sci.* 37, 1751–1763. doi: 10.11869/j.issn.1000-8551.2023.09.1751

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi: 10.1016/S1369-5266(02)00240-6

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Shen, Y., Wang, J., Shaw, R. K., Yu, H., Sheng, X., Zhao, Z., et al. (2021). Development of GBTS and KASP panels for genetic diversity, population structure, and fingerprinting of a large collection of broccoli (*Brassica oleracea* L. var. italica) in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.655254

Shen, Y., Wang, J., Yu, H., Sheng, X., Zhao, Z., and Gu, H. (2020). Development of KASP markers for germplasm characterization and fingerprinting identification of broccoli in China. *Res. Square.* [Preprint]. doi: 10.21203/rs.3.rs-60988/v1

Smith, S. M., and Maughan, P. J. (2015). "SNP genotyping using KASPar assays," in *Plant genotyping: Methods and protocols*. Ed. J. Batley (Humana Press, New York, NY), 243–256.

Soleimani, B., Lehnert, H., Keilwagen, J., Plieske, J., Ordon, F., Naseri Rad, S., et al. (2020). Comparison between core set selection methods using different Illumina marker platforms: A case study of assessment of diversity in wheat. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01040

Steele, K., Quinton-Tulloch, M., Vyas, D., and Witcombe, J. (2024). Thousands of trait-specific KASP markers designed for diverse breeding applications in rice (*Oryza sativa*). *G3: Genes Genomes Genet.* 15 (1), jkae251. doi: 10.1093/g3journal/jkae251

Su, W., Wang, L., Lei, J., Chai, S., Liu, Y., Yang, Y., et al. (2017). Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. *PloS One* 12, e0172066. doi: 10.1371/journal.pone.0172066

Sun, Z. (2017). SNP fingerprint and Genome-wide association studies of important agronomic traits for upland cotton. Hebei Agricultural University, Baoding (HB.

Swindell, S. R., and Plasterer, T. N. (1997). SEQMAN: contig assembly. *Sequence Data Anal. Guidebook.* 70, 75–89. doi: 10.1385/0-89603-358-9:75

Thachuk, C., Crossa, J., Franco, J., Dreisigacker, S., Warburton, M., and Davenport, G. F. (2009). Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinf.* 10, 1–13. doi: 10.1186/1471-2105-10-243

Tian, H., Yang, Y., Wang, R., Fan, Y., Yi, H., Jiang, B., et al. (2021). Screening of 200 core SNPs and the construction of a systematic SNP-DNA standard fingerprint database with more than 20,000 maize varieties. *Agriculture* 11, 597. doi: 10.3390/agriculture11070597

Van Inghelandt, D., Melchinger, A. E., Lebreton, C., and Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* 120, 1289–1299. doi: 10.1007/s00122-009-1256-2

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

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, 1044–1057. doi: 10.1016/S2095-3119(21)63654-7

Wang, Y., Lv, H., Xiang, X., Yang, A., Feng, Q., Dai, P., et al. (2021a). Construction of a SNP fingerprinting database and population genetic analysis of cigar tobacco germplasm resources in China. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.618133

Wang, Y., Wu, X., Li, Y., Feng, Z., Mu, Z., Wang, J., et al. (2021b). Identification and validation of a core single-nucleotide polymorphism marker set for genetic diversity assessment, fingerprinting identification, and core collection development in bottle gourd. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.747940

Xie, W., Zhang, J., Zhao, X., Zhang, J., and Wang, Y. (2015). Siberian wild rye (*Elymus sibiricus* L.): Genetic diversity of germplasm determined using DNA fingerprinting and SCoT markers. *Biochem. Systematics Ecol.* 60, 186–192. doi: 10.1016/j.bse.2015.04.021

Xing, X., Hu, T., Wang, Y., Li, Y., Wang, W., Hu, H., et al. (2024). Construction of SNP fingerprints and genetic diversity analysis of radish (*Raphanus sativus* L.). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1329890

Xiong, Y., Li, D., Liu, T., Xiong, Y., Yu, Q., Lei, X., et al. (2024a). Extensive transcriptome data providing great efficacy in genetic research and adaptive gene discovery: a case study of *Elymus sibiricus* L.(Poaceae, Triticeae). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1457980

Xiong, Y., Xiong, Y. L., Jia, X. J., Zhao, J. M., Yan, L. J., Sha, L. N., et al. (2024b). Divergence in *Elymus sibiricus* is related to geography and climate oscillation: A new look from pan-chloroplast genome data. *J. Systematics Evol.* 62, 794–808. doi: 10.1111/jse.13020

Yan, J., Bai, S., Ma, X., Gan, Y., and Zhang, J. (2007b). Genetic diversity of *Elymus sibiricus* and its breeding in China. *Chin. Bull. Bot.* 02), 226–231. doi: 10.3969/j.issn.1674-3466.2007.02.015

Yan, J., Bai, S., Ma, X., Gan, Y., Zhang, C., You, M., et al. (2007a). Ear character diversity of native populations of *Elymus sibiricus* in the northwest plateau of Sichuan province. *Acta Prataculturae Sinica.* 16 (06), 99–106. doi: 10.3321/j.issn:1004-5759.2007.06.015

Yan, J., Bai, S., Ma, X., Zhang, X., Zhang, C., and You, M. (2008). Genetic diversity of *Elymus sibiricus* L. in the eastern Qinghai-Tibet Plateau of China detected by SRAP markers. *Acta Prataculturae Sinica.* 19 (01), 173–183. doi: 10.11686/cyxb20100124

Yan, J., Bai, S., Zhang, C., You, M., Li, D., and Zeng, Y. (2010). Ecological characteristics and morphological variation of wild *Elymus sibiricus* L. germplasm from Qinghai-Tibetan Plateau in China. *Chin. J. Grassland.* 32, 49–57. doi: CNKI:SUN:ZGCD.0.2010-04-008

Yan, J., Li, X., Wang, L., Li, D., Ji, C., Yang, Z., et al. (2024). A high-continuity and annotated reference genome of allotetraploid Siberian wildrye (*Elymus sibiricus* L., Poaceae: Triticeae). *bioRxiv* [Preprint]. doi: 10.1101/2024.04.17.589894

Yan, W., Ma, Y., Zhang, J., Wang, K., and Meng, Q. (2017). The analysis of genetic diversity and the construction of core collection for *Elymus sibiricus* L. *J. Grassland Forage Sci.* 38, 1–12. doi: 10.3969/j.issn.1673-8403.2017.01.001

Yang, G., Chen, S., Chen, L., Sun, K., Huang, C., Zhou, D., et al. (2019). Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* 12, 1–18. doi: 10.1186/s12284-019-0272-3

Yang, W., Liu, W., Ma, X., and Ma, L. (2020). Effects of ROS accumulation and antioxidant system in two different drought resistant *Elymus sibiricus* under drought stress. *Acta Agrestia Sin.* 28, 684. doi: 10.11733/j.issn.1007-0435.2020.03.012

Yang, Y., Lyu, M., Liu, J., Wu, J., Wang, Q., Xie, T., et al. (2022). Construction of an SNP fingerprinting database and population genetic analysis of 329 cauliflower cultivars. *BMC Plant Biol.* 22, 522. doi: 10.1186/s12870-022-03920-2

Yang, Y., Zhang, H., Zhou, H., Wang, W., Yin, H., Yao, B., et al. (2015). Effect of salt stress on physiological characteristics of alpine grassland pasture *Elymus sibiricus*. *Acta Agriculturae Boreali-occidentalis Sin.* 24, 156–162. doi: 10.7606/j.issn.1004-1389.2015.07.024

Yen, C., and Yang, J. L. (2022). *Biosystematics of Triticeae: Volume V. Genera: Campeiostachys, Elymus, Pascopyrum, Lophopyrum, Trichopyrum, Hordelymus, Festucopsis, Peridictyon, and Psammopyrum* (Germany: Springer Nature).

Yu, Q., Xiong, Y., Su, X., Xiong, Y., Dong, Z., Zhao, J., et al. (2022). Comparative metabolomic studies of siberian wildrye (*Elymus sibiricus* L.): a new look at the mechanism of plant drought resistance. *Int. J. Mol. Sci.* 24, 452. doi: 10.3390/ijms24010452

Yu, F., Yang, M., Gan, Y., Yang, S., Xiao, B., Zheng, Q., et al. (2011). Drought resistance evaluation of three grass species at germination stage from the Northwest Plateau of Sichuan Province. *Pratacultural Sci.* 28, 993–997. doi: CNKI:SUN:CYKX.0.2011-06-023

Zeng, H., Yue, J., Zhang, S., Yi, K., Wei, K., Yang, S., et al. (2022). Comprehensive evaluation of agronomic traits of 76 germplasm resources of *Elymus sibiricus*. *Acta Agrestia Sin.* 30, 3046. doi: 10.11733/j.issn.1007-0435.2022.11.021

Zhang, Z. (2020). *Construction of Genetic Linkage Map and QTL Localization for Seed Shattering Related Genes in Elymus sibiricus L* (Lanzhou (GS: Lanzhou University).

Zhang, Q., Li, S., Guo, S., and Tian, H. (2024). Morphology and SSR markers reveal the genetic diversity of *Elymus* species germplasm in Northwestern China. *Genet. Resour. Crop Evol.* 71, 2091–2103. doi: 10.1007/s10722-023-01768-5

Zhang, Z., Xie, W., Zhang, J., Wang, N., Zhao, Y., Wang, Y., et al. (2019). Construction of the first high-density genetic linkage map and identification of seed yield-related QTLs and candidate genes in *Elymus sibiricus*, an important forage grass in Qinghai-Tibet Plateau. *BMC Genomics* 20, 1–17. doi: 10.1186/s12864-019-6254-4

Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., et al. (2020). A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci. Rep.* 10, 5623. doi: 10.1038/s41598-020-62518-6

Zhang, Z., Zheng, Y., Zhang, J., Wang, N., Wang, Y., Liu, W., et al. (2022). High-altitude genetic selection and genome-wide association analysis of yield-related traits in *Elymus sibiricus* L. using SLAF sequencing. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.874409

Zhou, H., Yang, B., and Han, J. G. (2000). Studies on some structural characteristics of a community of grassland of *Elymus sibiricus* planted in different years. *Acta Agrestia Sin.* 8, 245. doi: 10.11733/j.issn.1007-0435.2000.04.002

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 April 2025
DOI 10.3389/fpls.2025.1518406

OPEN ACCESS

EDITED BY
Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY
Gograj Singh Singh Jat,
Division of Vegetable Science - IARI, India
Mamy Jayne Nelly Rajaofera,
Hainan Medical University, China

*CORRESPONDENCE
Yongdong Sun
sunyd@hist.edu.cn

†These authors have contributed equally to this work

RECEIVED 28 October 2024
ACCEPTED 01 April 2025
PUBLISHED 24 April 2025

CITATION
Zhao Z, Ao W, Luo W, Sun Y, Tokala VY, Liu J, Zhi S and Sun Y (2025) Identification and characterization of Csa-miR159s and their expression patterns under different abiotic stresses in cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* 16:1518406. doi: 10.3389/fpls.2025.1518406

Identification and characterization of Csa-miR159s and their expression patterns under different abiotic stresses in cucumber (*Cucumis sativus* L.)

Zhenxiang Zhao[1,2†], Wenhong Ao[1,2†], Weirong Luo[1,2], Yaoguang Sun[1,2], Vijay Yadav Tokala[3], Junjun Liu[1,2], Shenshen Zhi[1,2] and Yongdong Sun[1,2*]

[1]School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang, China, [2]Henan Province Engineering Research Center of Horticultural Plant Resource Utilization and Germplasm Enhancement, Henan Institute of Science and Technology, Xinxiang, China, [3]Horticulture Research and Extension, The Postharvest Education Foundation, La Pine, OR, United States

The miR159 gene family plays an essential role in plant growth and development, and stress response. Nevertheless, there are no reports defining its specific function in cucumber fruit expansion and response to abiotic stresses. In this study, we retrieved six Csa-miR159 sequences from the EnsemblPlants database, which were located on chromosome 1, chromosome 3, and chromosome 5 of cucumber, respectively. Phylogenetic analysis showed that Csa-miR159c/d/e/f belonged to one branch and Csa-miR159a/b to another. *C*is-acting regulatory elements (CREs) including light response elements, phytohormone response elements, stress response elements, regulatory elements associated with plant growth and development were distributed unevenly in the promoter regions of Csa-miR159s, which indicated that Csa-miR159s might mediate the stress response, and growth and development. Moreover, it was determined that *CsMYBs* were the target genes of Csa-miR159s through psRNA-Target prediction and qRT-PCR analysis. Further findings suggested that Csa-miR159b might negatively regulate cucumber fruit expansion by targeting *Cs1RMYB9*, *Cs1RMYB31*, *Cs2RMYB37* and *Cs2RMYB64*. Similarly, Csa-miR159d might negatively regulate cucumber fruit expansion by targeting *Cs2RMYB27* and *Cs2RMYB32*. In addition, the differential expression of Csa-miR159s suggested their potential response to abiotic stresses and plant phytohormones. This study would provide valuable information on the molecular characterization of Csa-miR159s and establish a foundation for further research on the mechanisms of Csa-miR159s in regulating fruit expansion and stress response.

KEYWORDS

cucumber (*Cucumis sativus*. L), Csa-miR159s, fruit expansion, abiotic stress, phytohormones

1 Introduction

Cucumber (*Cucumis sativus* L.) is a widely cultivated vegetable worldwide, with China being the largest producer and consumer. In 2023, China cultivated approximately 1.37 million hectares of cucumber, with a total yield of 80.21 million tons. This accounted for 60.40% of the global cucumber cultivation area and 82.01% of the total production (Food and Agriculture Organization of the United Nations, 2023). However, unfavorable environmental conditions such as inappropriate temperature, weak light, drought and salt stresses, have a significant impact on cucumber fruit expansion, leading to the decreased yield and quality. Therefore, it is important to understand the molecular mechanism of cucumber fruit expansion in order to improve yield and stress tolerance in cucumber.

MicroRNA (miRNA) is a class of highly conserved endogenous non-coding small RNA, regulating the expression of its target genes at both the levels of transcription and post-transcription by directly cleaving or inhibiting the translation of target mRNA (Jones-Rhoades et al., 2006). It plays crucial roles in various physiological and metabolic processes, such as plant growth and development, and stress response (Jeong and Green, 2013; Liu et al., 2007; Zhang et al., 2019). miR159 has been extensively studied in plants (Montes et al., 2014) and the studies reveal how it influences plant growth and development by targeting *MYB* family genes (Dubos et al., 2010). For instance, miR159-*GAMYB* pathway has been widely implicated in plant growth, stress response, and phytohormone signaling in various species such as *Arabidopsis* (Allen et al., 2010, 2007; Alonso-Peral et al., 2010), tomato (Zhang et al., 2020) and rice (Zhao et al., 2017). In Gloxinia (*Sinningia speciosa*), expression patterns of miR159 and *GAMYB* were negatively correlated during flower development (Li et al., 2013). In addition, some studies have demonstrated the important role of the miR159-*GAMYB* in fruit development. For instance, in tomato, overexpression of Sl-MIR159 led to the down-regulation of *SlGAMYB*, thereby inducing parthenocarpy and early fruit ripening (da Silva et al., 2017). Similarly, Sly-miR159-*SlGAMYB2* was also found to control fruit growth, as the inhibition of Sly-miR159 and overexpression of *SlGAMYB2* resulted in the larger fruit, while the loss of function of *SlGAMYB2* led to the smaller fruit (Zhao et al., 2022). In the case of grape, exogenous application of gibberellin (GA) promoted parthenocarpy, accompanied by the up-regulation of Vvi-miR159c and the down-regulation of *VvGAMYB* (Wang et al., 2018).

The miR159-*GAMYB* pathway is known to play a crucial role in the response to drought and salt stresses. Studies have shown that miR159 was induced by drought stress in plants such as *Arabidopsis* (Reyes and Chua, 2007), maize (Wei et al., 2009), wheat (Akdogan et al., 2016), barley (Hackenberg et al., 2015) and poplar (Fu et al., 2023). However, in potato, the expression level of miR159 decreased under drought treatment, while the expression level of *GAMYB*-like homologues increased (Pieczynski et al., 2013). *SlMYB33*, the target gene of Sly-miR159, was associated with the accumulation of proline and putrescine, which enhanced plant tolerance to drought stress (López-Galiano et al., 2019). Furthermore, it has been reported that miR159 can be induced by salt stress in *Arabidopsis* (Liu et al., 2008) and soybean (Li et al., 2023). Additionally, miR159-*GAMYB* plays a crucial role in some plant phytohormone signaling pathways, such as abscisic acid (ABA) (Reyes and Chua, 2007) and GA (Wang et al., 2017). For instance, in 'Zuijinxiang' grape, the expression level of VvimiR159 increased after GA treatment, while the expression level of *VvGAMYB* significantly decreased (Wang et al., 2018). In 'Rosario Bianco' grape, the expression of miR159 was up-regulated in the pulp after GA treatment, whereas the expression of miR159a/c was down-regulated in the pulp and pericarp (Han et al., 2014). Overall, the miR159-*GAMYB* pathway plays a role in response to abiotic stresses and plant phytohormones. In our previous study, differential expression of Csa-miR159b was observed between the ovary and expanded fruit using small RNA sequencing, which suggested that Csa-miR159b was involved in cucumber fruit expansion (Sun et al., 2019). However, there was a scarcity of studies on the functions of Csa-miR159s in relation to cucumber fruit expansion and stress response.

This study aims to characterize Csa-miR159s in cucumber and to investigate their roles in fruit expansion and stress response. In the present study, multiple sequence alignment, chromosomal location, secondary structure, phylogenetic relationship, *cis*-regulatory elements (CREs), and the target genes of Csa-miR159s were studied in detail. Additionally, expression profiles of Csa-miR159s were analyzed in the ovary and expanded fruit, and in response to different stresses and plant phytohormones. Our findings will provide valuable information for further functional analysis of Csa-miR159s in cucumber, and also provide references for improving cucumber yield and resilience.

2 Materials and methods

2.1 Identification of Csa-miR159s

A search for miR159 family members in cucumber was conducted using EnsemblPlants database (http://plants.ensembl.org/). The mature sequences of miR159s from various crop species (zucchini, watermelon, pumpkin, cucumber, melon, tomato, rice and *Arabidopsis*) were obtained from the PmiREN database (Guo et al., 2020). Multiple sequence alignments of Csa-miR159s were performed using ClustalW software (Thompson et al., 1994), and were used to generate a sequence logo diagram through the online website (https://weblogo.berkeley.edu/). TBtools software (Chen et al., 2023) was employed to visualize the distribution of Csa-miR159s on cucumber chromosomes. The RNA secondary structure of pre-MIR159s was predicted using the RNAstructure web server (http://rna.urmc.rochester.edu/RNAstructureWebServers/Predictl/Predictl.html). Mature sequences of miR159s were submitted to MEGA v5.1 software (Kumar et al., 2018) to construct phylogenetic relationships using the neighbor-joining (NJ) method with 1000 bootstrap replicates to assess branch

confidence. The 2000 bp promoter sequences upstream from the initiation codon of Csa-MIR159s were extracted from EnsemblPlants database. The putative *cis*-regulatory elements (CREs) were identified and analyzed using the PlantCARE tool (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002).

2.2 Prediction of target genes

psRNA-Target uses sequence complementarity and energy-based scoring to predict miRNA-target interactions. A score threshold of ≤5.0 was chosen based on established standards for psRNA-Target to ensure high confidence in predicted interactions. To predict the potential target relationships of Csa-miR159s and *CsMYBs*, their gene sequences were submitted to the psRNA-Target online website (https://www.zhaolab.org/psRNATarget/) (Liu et al., 2015), and target genes with a score ≤5.0 was selected and submitted to Cucurbit Genomics Database (CuGenDB) for further analysis.

2.3 Plant growth conditions and stress treatments

Seeds of cucumber (cv. Jinyou No. 1) were soaked in water at a temperature of 55°C for 15 min and then incubated at 28°C for 2 days to germinate. The germinated seeds were cultivated in a pot filled with a medium consisting of peat soil, perlite, and vermiculite in a 2:1:1 ratio, and placed in a climate-controlled chamber at a temperature of 28°C with a light period of 16 h and a dark period of 8 h. Cucumber seedlings at the three-leaf stage were transferred to the plastic greenhouse for continuous growth. Samples from ovary (on the day of anthesis), and expanded fruit (5 days after anthesis) were collected for gene expression analysis. For drought and NaCl stresses, cucumber seedlings at the three-leaf stage with a similar size and height were cultured into 40 L (113 cm × 73 cm × 5 cm) hydroponic pots. The control group was cultured in Hoagland nutrient solution. Drought stress was induced using a Hoagland nutrient solution containing 10% PEG-6000, while NaCl stress was induced using a Hoagland nutrient solution containing 150 mmol/L NaCl (Lu et al., 2022). Leaves were collected at 0, 3, 6, 12, and 24 h after treatment for gene expression analysis. For plant phytohormone treatments, the cucumber seedlings at the three-leaf stage were sprayed with 100 μmol L^{-1} ABA, 100 μmol L^{-1} salicylic acid (SA), 100 μmol L^{-1} jasmonic acid (JA), 50 μmol L^{-1} ethephon (ETH), 50 μmol L^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D), and 50 μmol L^{-1} GA, respectively, while the control condition was sprayed with double distilled water (Li et al., 2019). Plant phytohormone treatments were conducted once a day. After three consecutive days of treatment, cucumber leaves were collected for gene expression analysis. All treatments were performed with three biological replicates.

2.4 qRT-PCR analysis

RNA was isolated from various cucumber tissues using the TaKaRa MiniBEST Plant RNA Extraction Kit (TaKaRa, Dalian, China). Isolated RNA was stored at -80°C until further use to prevent degradation. The Mir-X miRNA First-Strand Synthesis Kit (TaKaRa, Dalian, China) was then utilized for first-strand complementary DNA (cDNA) synthesis. qRT-PCR was conducted using the TB Green® Premix Ex TaqTM II (Tli RnaseH Plus) (TaKaRa, Dalian, China), using *U6* snRNA as the endogenous control for Csa-miR159s, and *18S* as the endogenous control for *CsMYBs*. Stem-loop of mature Csa-miR159s was used for qRT-PCR. The specific primer sequences utilized in this study were provided in detail in Supplementary Table S1. Gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), each expression level was evaluated using three biological replicates.

3 Results

3.1 Identification of Csa-miR159s

Six Csa-miR159 sequences were identified from the EnsemblPlants database. The mature sequences of Csa-miR159s were 21-22 nt in length and highly conserved (Figure 1A). These sequences were mapped to chromosome 1 (Csa-miR159a), chromosome 3 (Csa-miR159b), and chromosome 5 (Csa-miR159c/d/e/f), respectively, based on their physical positions (Figure 1B). Although Csa-miR159c/d/e/f were located on the same chromosome, their mature sequences showed lower similarity. In contrast, Csa-miR159a and Csa-miR159b, which were located on the different chromosomes, shared higher similarity.

The prediction result of secondary structure showed that all pre-miR159s demonstrated a typical stem-loop structures (Figure 2). The number of sub-loops varied from 12 (Csa-miR159f) to 18 (Csa-miR159a), and the stem-loop folding free energy ranged from -102.5 kcal/mol (pre-miR159b) to -78.7 kcal/mol (Csa-miR159a/f) (Figure 2).

3.2 Phylogenetic relationship of miR159s

To better understand the evolutionary relationships among miR159s, we further analyzed the mature sequences from cucumber (Csa-miR159a/b/c/d/e/f), melon (Cme-miR159a/b), zucchini (Cma-miR159a/b/c/d), pumpkin (Cmo-miR159a/b), watermelon (Cla-miR159a/b), tomato (Sly-miR159a/b), rice (Osa-miR159a/b/c/d/e/f), and *Arabidopsis* (Ath-miR159a/b/c) (Figure 3). Twenty-seven miR159s were classified into two branches based on the evolutionary divergence. Csa-miR159c/d/e/f and Cma-

10.3389/fpls.2025.1518406

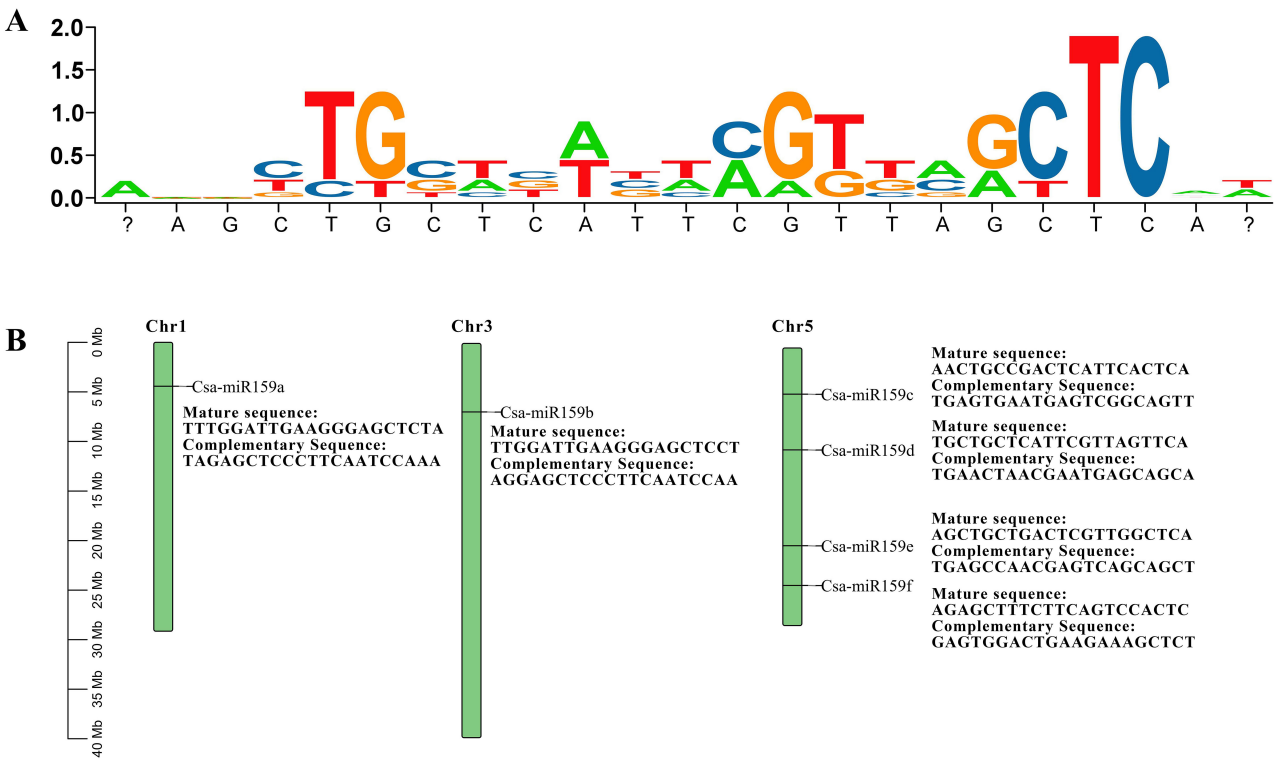


FIGURE 1
Mature sequences and chromosomal positions of Csa-miR159s. The ratio is measured in megabases (Mb). **(A)** Mature sequences; **(B)** Chromosomal positions.

miR159a/b belonged to one branch. Csa-miR159a/b were classified into another branch with remaining members. The phylogenetic tree revealed that cucumber miR159s share closer evolutionary relationships with those of zucchini.

3.3 Cis-regulatory elements analysis of Csa-miR159s

To investigate the potential functions of Csa-miR159s, we analyzed the CREs in the promoter regions. As shown in Figure 4, these CREs were grouped into four functional categories. The most abundant category was light response elements, which included Box 4, AAAC-motif, G-Box, TCT-motif, AE-box, GATA-motif, GT1-motif, I-box, GA-motif, TCCC-motif, ATCT-motif, chs-CMA2a, ATC-motif, and MRE. We also detected various phytohormone response elements, such as abscisic acid responsiveness (ABRE), gibberellin responsiveness (TATC-box and P-box), ethylene responsiveness (ERE), MeJA responsiveness (CGTCA-motif, TGACG-motif) and salicylic acid responsiveness (TCA-element, SARE). Furthermore, stress response elements were identified, including anaerobic induction (ARE), drought inducibility (MBS), heat induction (STRE), low temperature responsiveness (LTR), wound responsiveness (WUN-motif), and defense and stress responsiveness (TC-rich). Additionally, regulatory elements related to plant growth and development were also identified, such as zein metabolism regulation (O2-site), meristem expression (CAT-box), endosperm expression (GCN4-motif) and circadian control (circadian).

3.4 Prediction of target genes of Csa-miR159s

To investigate the regulatory mechanisms of Csa-miR159s, potential target genes were predicted using the psRNA-Target tool (Table 1). All target genes were named according to their subgroups and chromosomal positions from top to bottom (from *Cs1RMYB* to *Cs4RMYB*). The results revealed that *Cs2RMYB37*, *Cs2RMYB64*, *Cs1RMYB31*, *Cs1RMYB9* and *Cs2RMYB25* were recognized as the target genes of Csa-miR159a. Similarly, *Cs2RMYB37*, *Cs2RMYB64*, *Cs1RMYB31*, *Cs1RMYB9* and *Cs3RMYB1* were predicted as the target genes of Csa-miR159b. Additionally, *Cs2RMYB27* and *Cs2RMYB32* were found to be the target genes of Csa-miR159d. However, no target genes were detected for Csa-miR159c, Csa-miR159e and Csa-miR159f. Notably, all predicted target genes were classified as MYB or MYB-like transcription factors.

10.3389/fpls.2025.1518406

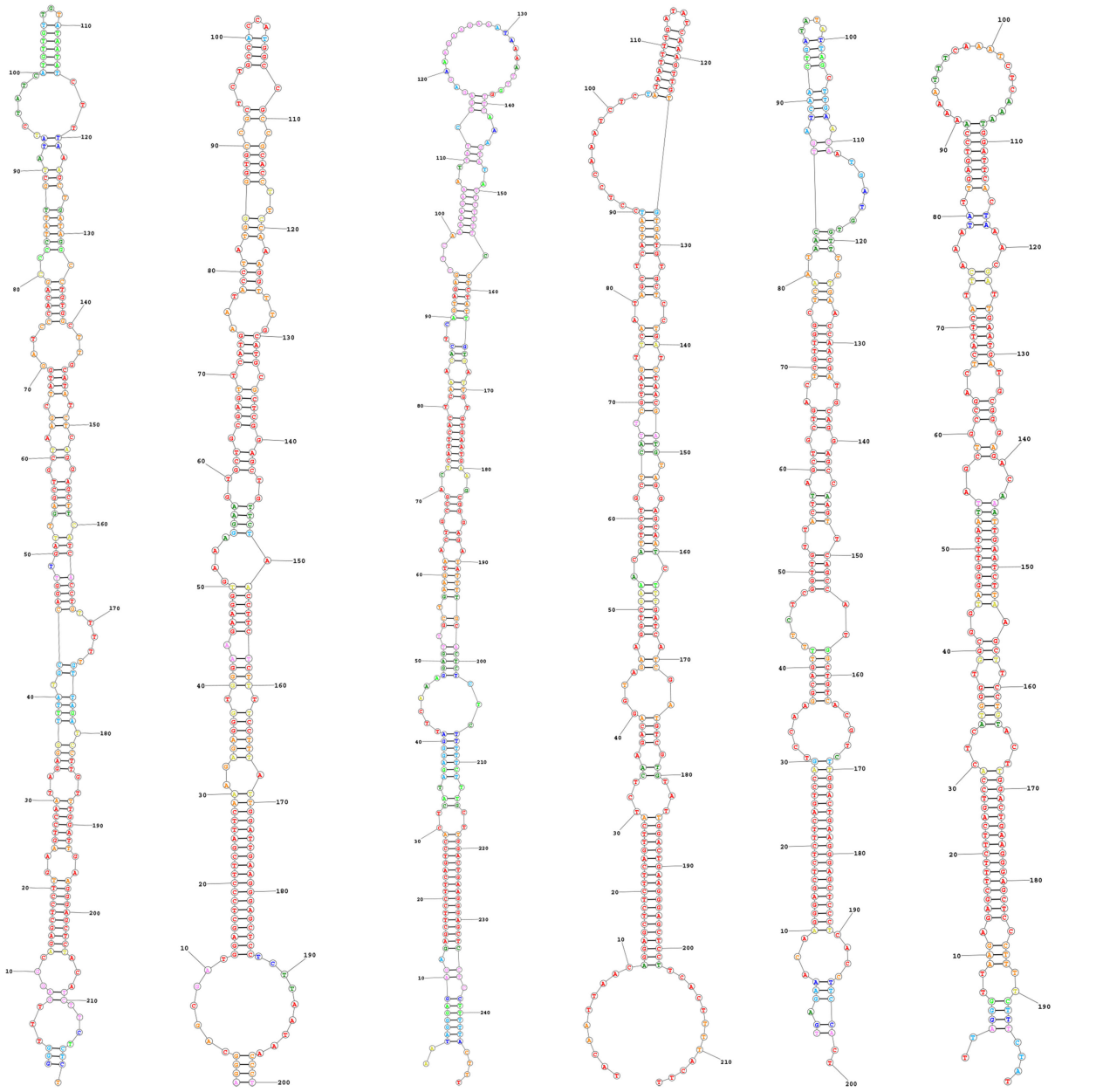


FIGURE 2
Secondary structures of Csa-miR159s.

3.5 Expression profiles of Csa-miR159s and their target genes

In this study, qRT-PCR was used to confirm the expression profiles of Csa-miR159s and their target genes in the ovary and expanded fruit (Figure 5). The expression of Csa-miR159b in the ovary was 2.44-fold higher than that in the expanded fruit, while the expression levels of Csa-miR159a/c/e/f in the ovary were lower than those in the expanded fruit. Notably, Csa-miR159d was only expressed in the ovary and was not detected in the expanded fruit. In terms of the target genes, *Cs1RMYB9*, *Cs1RMYB31*, *Cs2RMYB37*, *Cs2RMYB64*, *Cs2RMYB27* and *Cs2RMYB32* showed lower expression in the ovary compared to the expanded fruit, except for *Cs3RMYB1* which exhibited the opposite trend. These findings suggested that Csa-miR159b might negatively regulate cucumber fruit expansion by targeting *Cs1RMYB9*, *Cs1RMYB31*, *Cs2RMYB37* and *Cs2RMYB64*. Additionally, Csa-miR159d might negatively regulate cucumber fruit expansion by targeting

10.3389/fpls.2025.1518406

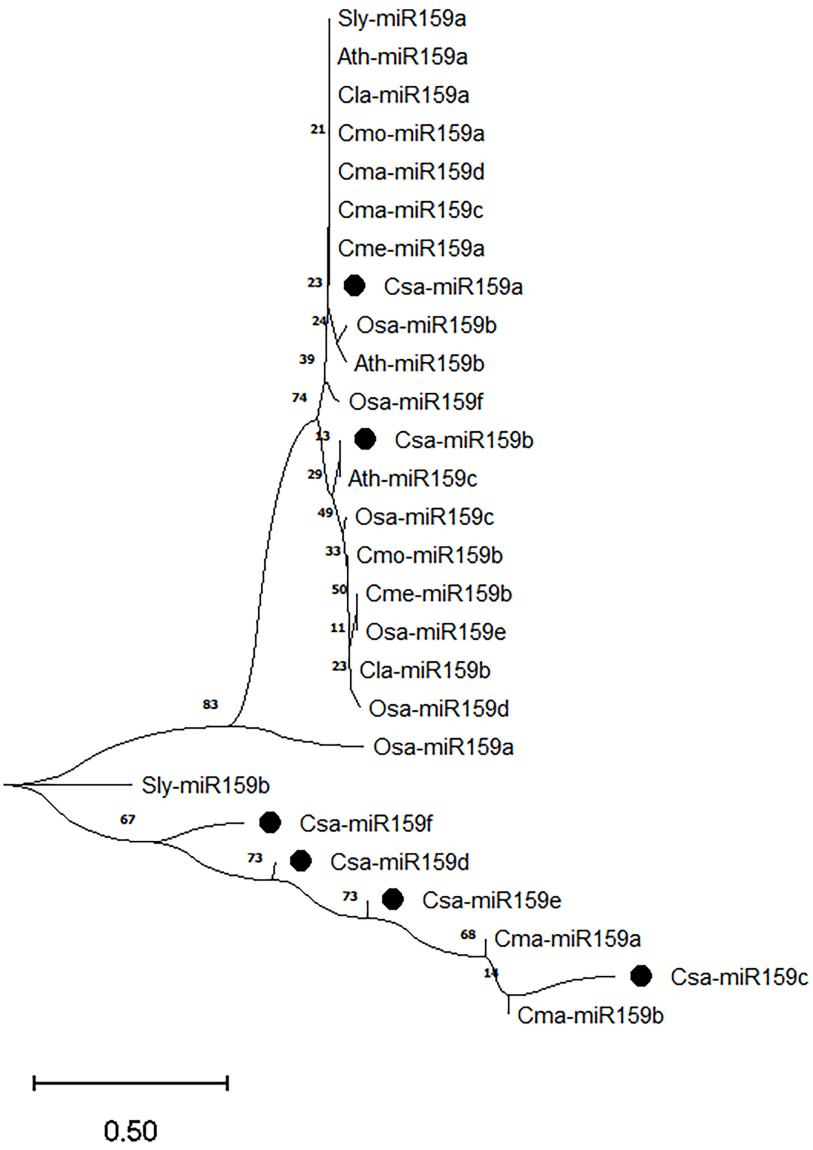


FIGURE 3
Phylogenetic analysis of miR159s from cucumber, melon, zucchini, pumpkin, watermelon, tomato, rice and *Arabidopsis*. Csa, cucumber; Cme, melon; Cma, zucchini; Cmo, pumpkin; Cla, watermelon; Sly, tomato; Osa, rice; Ath, *Arabidopsis*.

Cs2RMYB27 and *Cs2RMYB32*. Taken together, these results highlighted the tissue specificity and functional diversity of Csa-miR159s.

3.6 Csa-miR159s response to abiotic stresses and plant phytohormones

The study investigated the expression patterns of Csa-miR159s under different treatments including PEG, NaCl and plant phytohormones. Under PEG stress, the expression levels of Csa-miR159a/f increased gradually, reaching the top at 24 h after treatment. Specifically, Csa-miR159a was 24.4-fold higher than the control, and Csa-miR159f was 174.1-fold higher. Conversely, Csa-miR159c exhibited the increased expression, peaking at 12 h, and then decreasing to the lowest level at 24 h after treatment, 1.3-fold lower than the control. However, Csa-miR159b/d/e were significantly down-regulated under PEG stress. Compared to the control, their expression levels decreased by 76.9, 2.4 and 4.3-fold, respectively (Figure 6). Under NaCl stress, Csa-miR159a/d/e were all significantly up-regulated at 6 h, with increases of 88.5, 33.2 and 3.9-fold compared to the control. While Csa-miR159b was remarkably down-regulated from 3 h to 24 h. In contrast, the expression levels of Csa-miR159c initially decreased at 3 h, then increased at 6 h, and peaked at 24 h. Csa-miR159f showed the increased expression, peaking at 12 h, and then decreasing at 24 h (Figure 7). In relation to plant phytohormones, it was observed that Csa-miR159a/b/c were significantly induced by SA, ETH and 2,4-D. On the other hand, Csa-miR159d showed significant up-regulation in response to ABA, GA, SA, ETH, and 2,4-D. Furthermore, Csa-

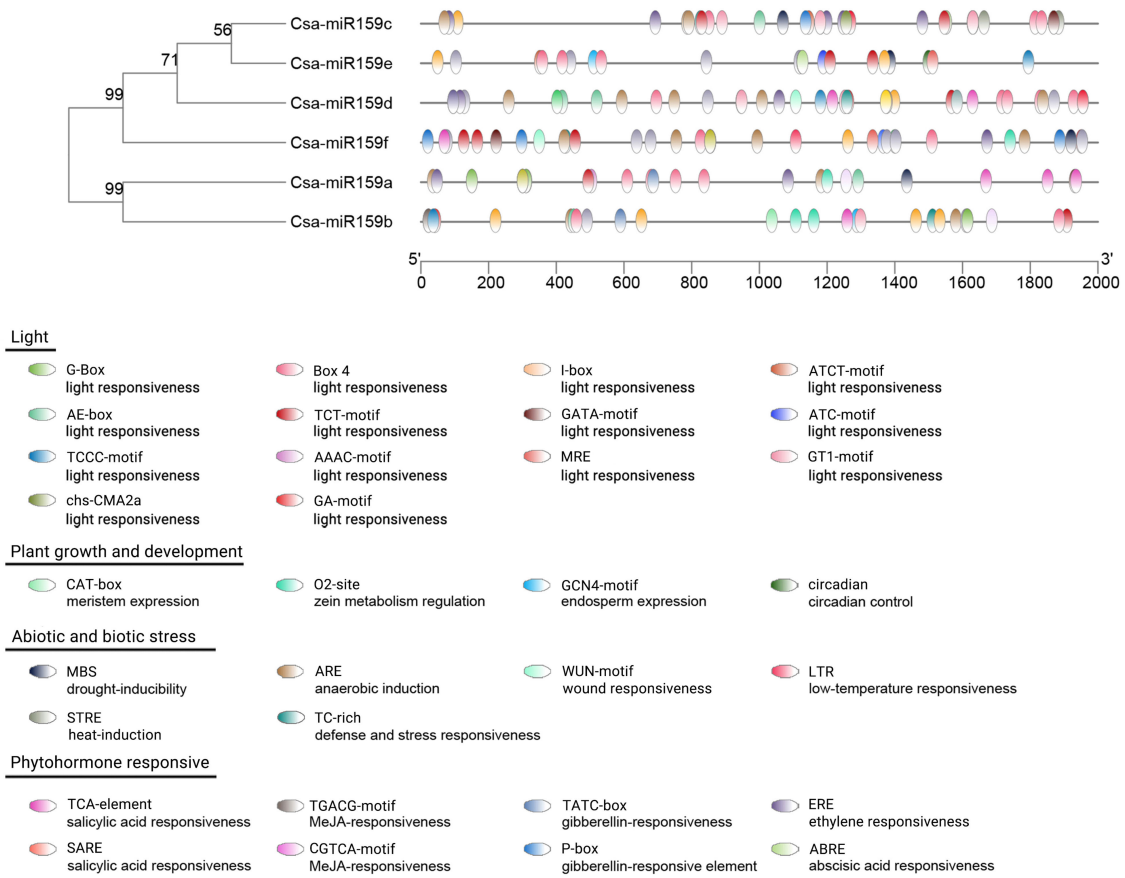


FIGURE 4
CREs in the promoter regions of Csa-miR159s. Different CRE was presented in the different color shape.

miR159e exhibited a remarkable down-regulation when exposed to ABA, GA, JA, ETH, and 2,4-D. Additionally, Csa-miR159f displayed significant up-regulation specifically in response to SA, ETH, and 2,4-D (Figure 8). These findings indicated that Csa-miR159s might be involved in plant stress response and phytohormone regulation.

4 Discussion

miRNA plays a crucial role in plant growth and development, and stress response by regulating the expression of their target genes. Among these miRNAs, miR159 has been extensively identified and characterized in numerous plant species, such as *Arabidopsis* (Palatnik et al., 2007), grape (Zhang et al., 2019), soybean (Li et al., 2023) and *Dendrobium officinale* (Hao and Zhang, 2022). However, there is limited research on the functions of Csa-miR159s regarding fruit expansion and abiotic stress response in cucumber. In this study, we identified six Csa-miR159s and their target genes. Csa-miR159s were unevenly distributed on chromosome 1, chromosome 3, and chromosome 5. Interestingly, Csa-miR159c/d/e/f, which had different mature sequences, were located on the same chromosome. Conversely, Csa-miR159a/b, which had higher sequence homology, were distributed on the different chromosomes. This suggested that the divergent chromosomal localization of Csa-miR159s could be a result of gene duplication events or evolutionary pressures that had caused their dispersion across different chromosomes, and this dispersion might have promoted the diversified functions of Csa-miR159s. Phylogenetic analysis revealed a close relationship between miR159s in cucumber and those in zucchini, possibly due to common evolutionary processes as the members of the *Cucurbitaceae* family.

The promoter region typically contains specific CREs with distinct functions. Analysis of CREs can offer insights into the potential functions of genes in the growth and development, and stress response. In our study, we identified stress-related elements in Csa-miR159s, including MBS, ARE, LTR, STRE and TC-rich, which suggested that Csa-miR159s might be associated with stress response in cucumber. Previous studies have demonstrated that up-regulated expression of miR159 enhanced stress tolerance in *Arabidopsis* (Reyes and Chua, 2007; Liu et al., 2008) and sweet potato (Yang et al., 2020). Conversely, some reports indicated that miR159 was down-regulated under salt stress and drought stress (Yang et al., 2020), and overexpression of miR159 increased stress sensitivity in rice (Wang et al., 2012) and potato (Pieczynski et al., 2013). In addition, Peng et al. (2018) observed that the expression of miR159 in rice was down-regulated after 3 h of salt stress treatment,

10.3389/fpls.2025.1518406

TABLE 1 Target genes prediction of Csa-miR159s.

miRNA_ACC	Rename	Target_Acc	Expectation	Target regions	Inhibition	Multiplicity	Description
Csa-miR159a	*Cs2RMYB37*	Csa4G022940.1	0.5	CDS:2507239-2508535	Cleavage	1	MYB-related transcription factor
	Cs2RMYB64	Csa7G043580.1	0.5	CDS:2401528-2403473	Cleavage	1	MYB transcription factor
	Cs1RMYB31	Csa6G105150.1	2.0	CDS:6908538-6908833	Cleavage	1	MYB-like transcription factor
	Cs1RMYB9	Csa2G035350.1	3.0	CDS:3526253-3527734	Translation	1	MYB transcription factor
	Cs2RMYB25	Csa3G264750.1	5.0	CDS:16264809-16266219	Cleavage	1	MYB family transcription factor
Csa-miR159b	*Cs2RMYB37*	Csa4G022940.1	0.5	CDS:2507239-2508535	Cleavage	11	MYB-related transcription factor
	Cs2RMYB64	Csa7G043580.1	0.5	CDS:2401528-2403473	Cleavage	1	MYB transcription factor
	Cs1RMYB31	Csa6G105150.1	2.0	CDS:6908538-6908833	Cleavage	1	MYB-like transcription factor
	Cs1RMYB9	Csa2G035350.1	3.0	CDS:3526253-3527734	Translation	1	MYB transcription factor
	Cs3RMYB1	Csa2G375240.1	5.0	CDS:18863762-18872002	Cleavage	1	Putative MYB transcription factor
Csa-miR159d	*Cs2RMYB27*	Csa3G386830.1	4.5	CDS:18944049-18946409	Cleavage	1	Putative MYB transcription factor
	Cs2RMYB32	Csa3G816030.1	5.0	CDS:31548636-31550625	Cleavage	1	MYB transcription factor

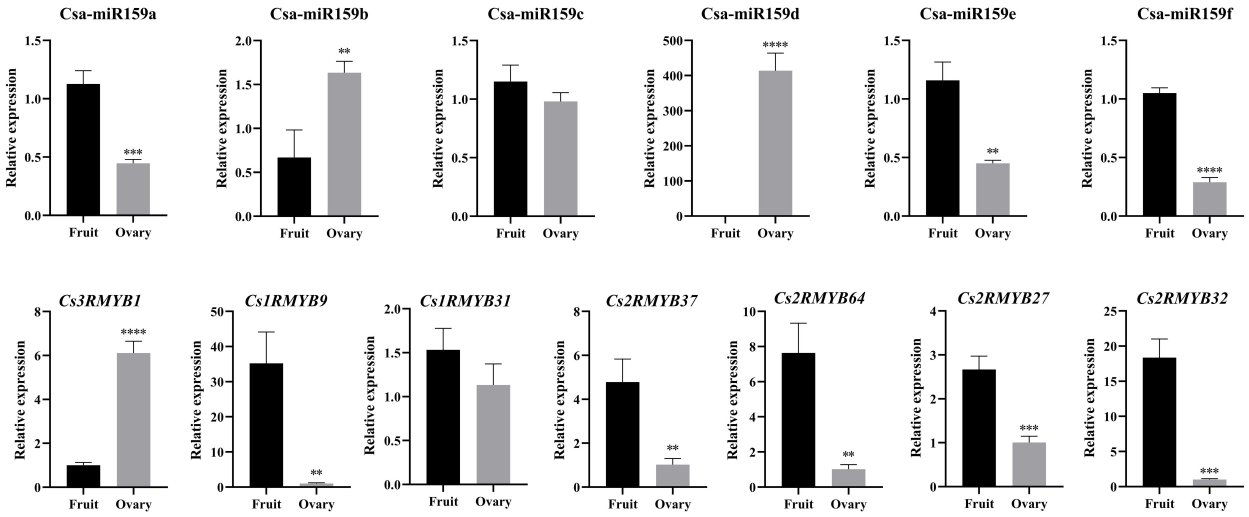


FIGURE 5
Relative expression of Csa-miR159s and *CsMYBs* in the ovary and the expanded fruit of cucumber by qRT-PCR. The X-axis indicated the tested tissue samples. Error bars represented ± standard deviation (SD) with three biological replicates. Different asterisks above the bars indicated significant differences. (** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

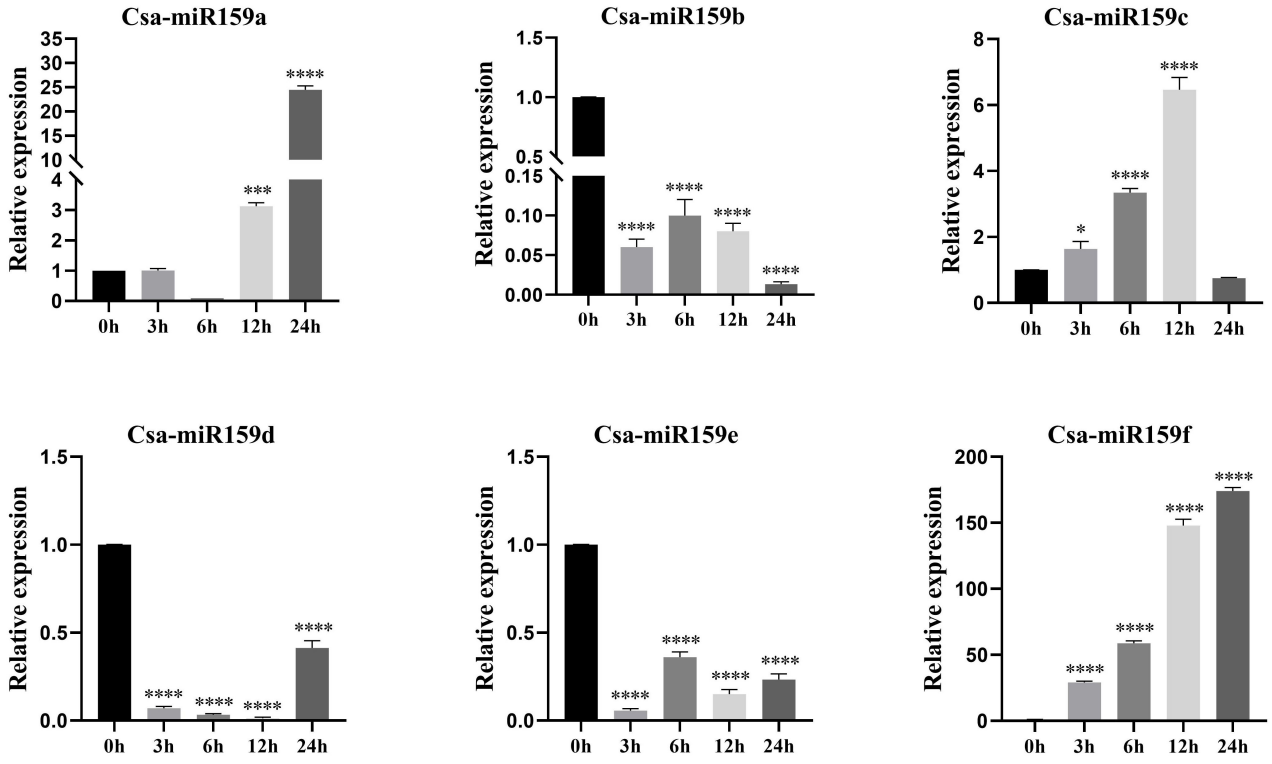


FIGURE 6
Relative expression of Csa-miR159s in response to 10% PEG-6000 treatment (0, 3, 6, 12 and 24 h). Error bars represented ± standard deviation (SD) with three biological replicates. Different asterisks above the bars indicated significant differences. (* $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$).

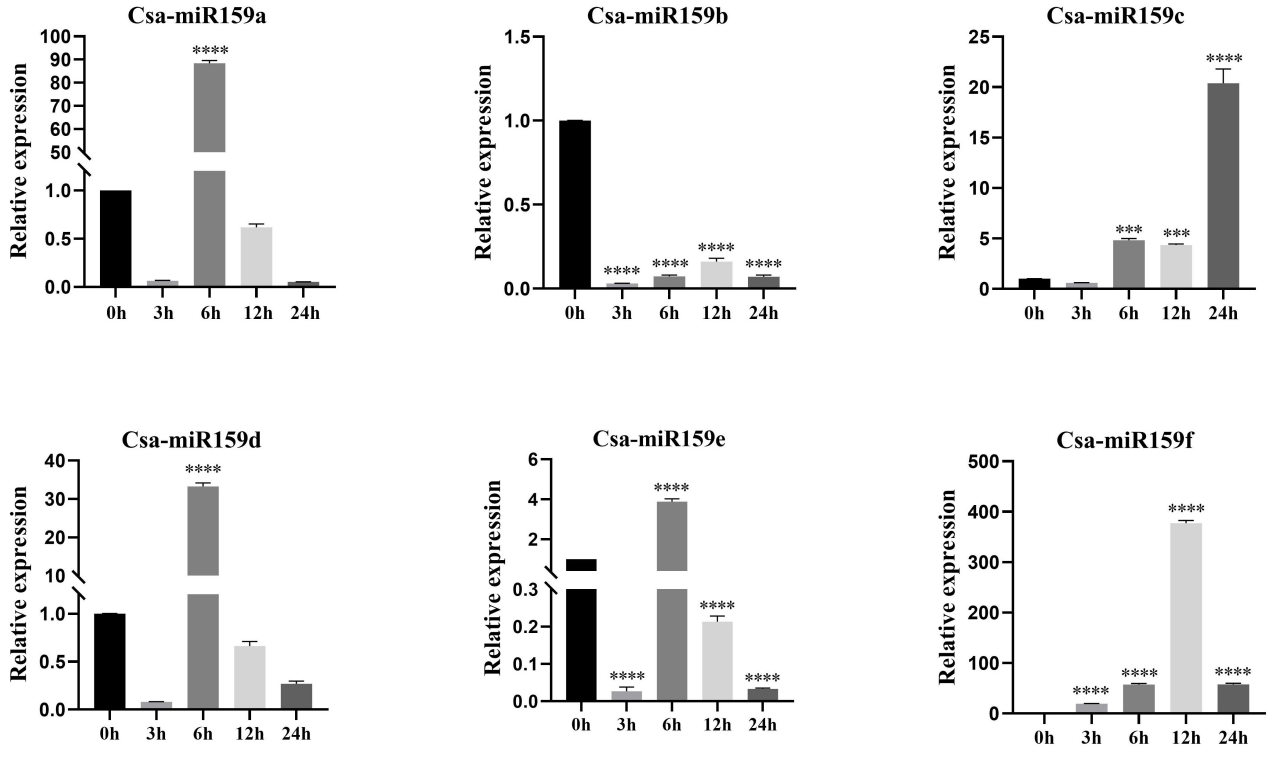


FIGURE 7
Relative expression of Csa-miR159s in response to 150 mmol/L NaCl treatment (0, 3, 6, 12 and 24 h). Error bars represented ± standard deviation (SD) with three biological replicates. Different asterisks above the bars indicated significant differences. (*** $p < 0.001$, **** $p < 0.0001$).

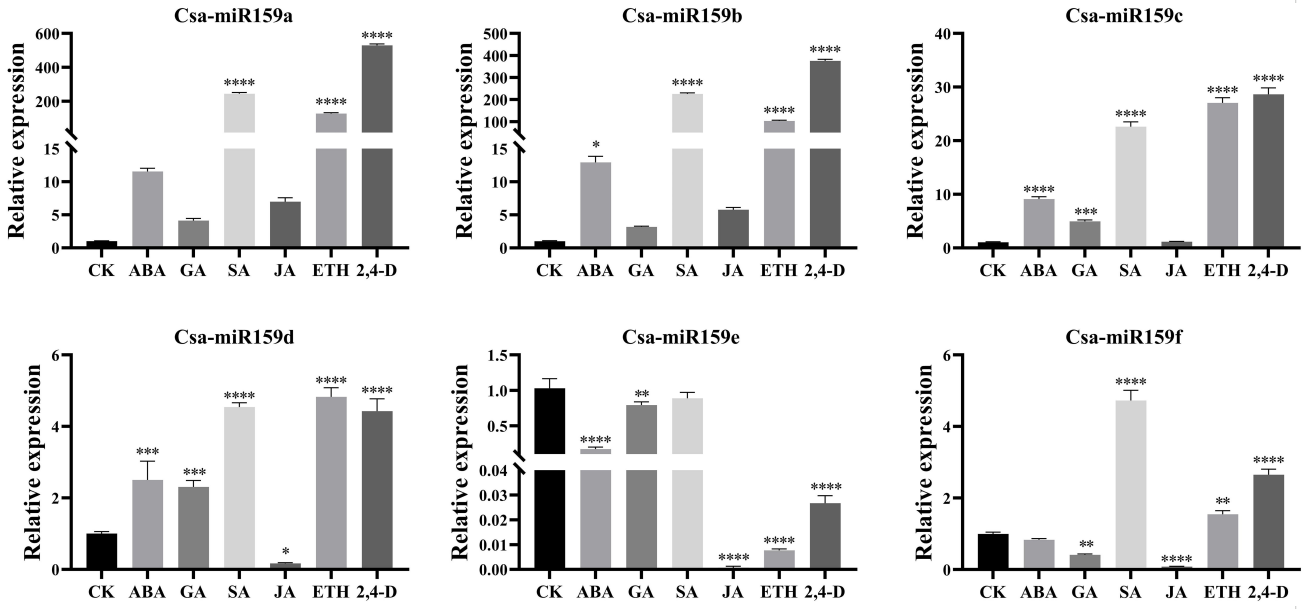


FIGURE 8
Relative expression of Csa-miR159s in response to different plant phytohormones, such as, ABA, GA, SA, JA, ETH and 2,4-D. Error bars represented ± standard deviation (SD) with three biological replicates. Different asterisks above the bars indicated significant differences. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

followed by up-regulation. Our results confirmed that Csa-miR159s could respond to PEG and NaCl stresses by qRT-PCR, and significant differences were observed in their expression patterns. Additional experimental validation is necessary to elucidate the transcriptional regulation mechanisms of Csa-miR159s under PEG and NaCl stresses.

Several studies have indicated that miR159 can respond to some plant phytohormones, including ABA (Reyes and Chua, 2007) and GA (Wang et al., 2017). For instance, the application of exogenous GA led to a significant decrease in the expression level of Fa-miR159a, while the expression level of Fa-miR159b remained unchanged in strawberry (Csukasi et al., 2012). Similarly, exogenous GA treatment resulted in the up-regulated expression of VvmiR159c during flowering, whereas VvmiR159a/b showed no significant changes in grape (Wang et al., 2018). In this study, CREs of six Csa-miR159s included various plant phytohormone response elements such as ABA, GA, SARE, ABRE, MeJA and ETH. Csa-miR159s exhibited distinct expression patterns under ABA, GA, SA, JA, ETH and 2,4-D treatments. This suggested that Csa-miR159s might play a significant role in plant phytohormone signaling pathways with varying response mechanisms.

Fruit development is a crucial stage in the life cycle of plants, encompassing a variety of intricate physiological and molecular processes. miR159 has been found to be crucial for ovule development and fruit set in tomato. Overexpression of Sly-miR159 caused abnormal ovule development, premature maturation, and seedless fruit in tomato (da Silva et al., 2017; Deng, 2020). Furthermore, Sly-miR159-*SlGAMYB2* pathway has been identified to regulate fruit morphology, whereby inhibition of Sly-miR159 led to larger fruit and a reduced length/width ratio (Zhao et al., 2022). Here, we discovered that the expression of Csa-miR159b/d was significantly higher in the ovary than that in the expanded fruit. Conversely, the expression levels of Csa-miR159a/c/e/f in the ovary were lower than those in the expanded fruit. These findings suggested that Csa-miR159s could be involved in fruit expansion in cucumber. miRNA regulates the growth and development in plants by inhibiting the expression of its target genes. In this study, *Cs2RMYB37*, *Cs2RMYB64*, *Cs1RMYB31*, *Cs1RMYB9* and *Cs3RMYB1* were predicted as the target genes of Csa-miR159b based on the PsRNA-Target results. While, *Cs2RMYB27* and *Cs2RMYB32* were found to be the target genes of Csa-miR159d. Furthermore, we found that *Cs1RMYB9*, *Cs1RMYB31*, *Cs2RMYB37*, *Cs2RMYB64*, *Cs2RMYB27* and *Cs2RMYB32* showed lower expression in the ovary compared to the expanded fruit by qRT-PCR. These findings suggested that Csa-miR159b might negatively regulate cucumber fruit expansion by targeting *Cs1RMYB9*, *Cs1RMYB31*, *Cs2RMYB37* and *Cs2RMYB64*. Similarly, Csa-miR159d might negatively regulate cucumber fruit expansion by targeting *Cs2RMYB27* and *Cs2RMYB32*. Taken together, our results suggested a vital role of Csa-miR159s in fruit expansion and stress response in cucumber. Further research is required to comprehend the functions of Csa-miR159s by performing a gain of function or loss of function assay.

5 Conclusions

In this study, six miR159 family members were identified in cucumber. Bioinformatics and expression profiles of Csa-miR159s were performed to discover their potential functions. The results

showed that Csa-miR159s played a crucial role in the response to PEG, NaCl and plant phytohormones. Additionally, it was found that Csa-miR159b/d might inhibit the cucumber fruit expansion by targeting their target genes. Our study provided a theoretical foundation for further investigation into the roles of Csa-miR159s under fruit expansion and abiotic stresses in cucumber.

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

ZZ: Formal Analysis, Investigation, Methodology, Writing – original draft. WA: Formal Analysis, Investigation, Methodology, Writing – original draft. WL: Data curation, Formal Analysis, Software, Writing – review & editing. YAS: Data curation, Formal Analysis, Software, Writing – review & editing. VT: Data curation, Formal Analysis, Software, Writing – review & editing. JL: Data curation, Formal Analysis, Software, Writing – review & editing. SZ: Data curation, Formal Analysis, Software, Writing – review & editing. YOS: Conceptualization, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the Key Science and Technology Project of Henan Province (No. 242102111125; No. 232102110177), Year 2024 Research Funding Program Based on Merit for Overseas Persons in Henan Province (No. 202404).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2025.1518406/full#supplementary-material

References

Akdogan, G., Tufekci, E. D., Uranbey, S., and Unver, T. (2016). miRNA-based drought regulation in wheat. *Funct. Integr. Genomics* 16, 221–233. doi: 10.1007/s10142-015-0452-1

Allen, R. S., Li, J., Alonso-Peral, M. M., White, R. G., Gubler, F., and Millar, A. A. (2010). MicroR159 regulation of most conserved targets in *Arabidopsis* has negligible phenotypic effects. *Silence.* 1, 18. doi: 10.1186/1758-907x-1-18

Allen, R. S., Li, J., Stahle, M. I., Dubroue, A., Gubler, F., and Millar, A. A. (2007). Genetic analysis reveals functional redundancy and the major target genes of the *Arabidopsis* miR159 family. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16371–16376. doi: 10.1073/pnas.0707653104

Alonso-Peral, M. M., Li, J., Li, Y., Allen, R. S., Schnippenkoetter, W., Ohms, S., et al. (2010). The MicroRNA159-regulated GAMYB-like genes inhibit growth and promote programmed cell death in *Arabidopsis*. *Plant Physiol.* 154, 757–771. doi: 10.1104/pp.110.160630

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A "one for all, all for one"bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Csukasi, F., Donaire, L., Casanal, A., Martinez-Priego, L., Botella, M. A., Medina-Escobar, N., et al. (2012). Two strawberry miR159 family members display developmental-specific expression patterns in the fruit receptacle and cooperatively regulate *Fa-GAMYB*. *New Phytol.* 195, 47–57. doi: 10.1111/j.1469-8137.2012.04134.x

da Silva, E. M., Silva, G., Bidoia, D. B., Azevedo, M. D., de Jesus, F. A., Pino, L. E., et al. (2017). microRNA159-targeted SIGAMYB transcription factors are required for fruit set in tomato. *Plant J.* 92, 95–109. doi: 10.1111/tpj.13637

Deng, Y. (2020). *Mechanism of Sly-miR159a regulating tomato fruit shape* (Shenzhen Univeristy: Shenzhen).

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

Food and Agriculture Organization of the United Nations (2023). Available online at: http://www.faoorg/faostat/ (Accessed 25 August 2023).

Fu, T., Wang, C., Yang, Y., Yang, X., Wang, J., Zhang, L., et al. (2023). Function identification of miR159a, a positive regulator during poplar resistance to drought stress. *Hortic. Res.* 10, 17. doi: 10.1093/hr/uhad221

Guo, Z., Kuang, Z., Wang, Y., Zhao, Y., Tao, Y., Cheng, C., et al. (2020). PmiREN: a comprehensive encyclopedia of plant miRNAs. *Nucleic Acids Res.* 48, D1114–D1121. doi: 10.1093/nar/gkz894

Hackenberg, M., Gustafson, P., Langridge, P., and Shi, B. (2015). Differential expression of microRNAs and other small RNAs in barley between water and drought conditions. *Plant Biotechnol. J.* 13, 2–13. doi: 10.1111/pbi.12220

Han, J., Fang, J., Wang, C., Yin, Y., Sun, X., Leng, X., et al. (2014). Grapevine microRNAs responsive to exogenous gibberellin. *BMC Genomics* 15, 17. doi: 10.1186/1471-2164-15-111

Hao, L., and Zhang, Y. (2022). Genome-wide analysis of miR159 gene family and predicted target genes associated with environmental stress in *dendrobium officinale*: A bioinformatics study. *Genes.* 13, 13. doi: 10.3390/genes13071221

Jeong, D. H., and Green, P. J. (2013). The role of rice microRNAs in abiotic stress responses. *J. Plant Biol.* 56, 187–197. doi: 10.1007/s12374-013-0213-4

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53. doi: 10.1146/annurev.arplant.57.032905.105218

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, 1547–1549. doi: 10.1093/molbev/msy096

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, 325–327. doi: 10.1093/nar/30.1.325

Li, X., Bian, H., Song, D., Ma, S., Han, N., Wang, J., 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., Miao, L., Huang, B., Gao, L., He, C., Yan, Y., et al. (2019). Genome-wide identification and characterization of cucumber BPC transcription factors and their responses to abiotic stresses and exogenous phytohormones. *Int. J. Mol. Sci.* 20, 5048. doi: 10.3390/ijms20205048

Li, B., Tao, P., Xu, F., He, P., and Wang, J. (2023). Function of Soybean miR159 family members in plant responses to low phosphorus, high salinity, and abscisic acid treatment. *Agron. (Basel)* 13, 16. doi: 10.3390/agronomy13071798

Liu, S., Li, J., Wu, J., Zhou, K., Zhou, H., Yang, J., et al. (2015). StarScan: a web server for scanning small RNA targets from degradome sequencing data. *Nucleic Acids Res.* 43, W480–W486. doi: 10.1093/nar/gkv524

Liu, P., Montgomery, T. A., Fahlgren, N., Kasschau, K. D., Nonogaki, H., and Carrington, J. C. (2007). Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. *Plant J.* 52, 133–146. doi: 10.1111/j.1365-313X.2007.03218.x

Liu, H., Tian, X., Li, Y., Wu, C., and Zheng, C. (2008). Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA.* 14, 836–843. doi: 10.1261/rna.895308

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, 402–408. doi: 10.1006/meth.2001.1262

López-Galiano, M. J., García-Robles, I., Gonzalez-Hernandez, A. I., Camañes, G., Vicedo, B., Real, M. D., et al. (2019). Expression of miR159 is altered in tomato plants undergoing drought stress. *Plants-Basel.* 8, 11. doi: 10.3390/plants8070201

Lu, L., Luo, W., Yu, W., Zhou, J., Wang, X., and Sun, Y. (2022). Identification and characterization of Csa-miR395s reveal their involvements in fruit expansion and abiotic stresses in cucumber. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.907364

Montes, R. A. C., Rosas-Cardenas, F. D., De Paoli, E., Accerbi, M., Rymarquis, L. A., Mahalingam, G., et al. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* 5, 15. doi: 10.1038/ncomms4722

Palatnik, J. F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., et al. (2007). Sequence and expression differences underlie functional specialization of *Arabidopsis* MicroRNAs miR159 and miR319. *Dev. Cell.* 13, 115–125. doi: 10.1016/j.devcel.2007.04.012

Peng, T., Wen, H., Zhao, Y., Wang, B., Jin, Y., Sun, H., et al. (2018). Identification and expression analysis of rice miRNA related to salt and drought stresses. *Acta Agric. Boreali-Sin.* 33, 20–27.

Pieczynski, M., Marczewski, W., Hennig, J., Dolata, J., Bielewicz, D., Piontek, P., et al. (2013). Down-regulation of CBP80 gene expression as a strategy to engineer a drought-tolerant potato. *Plant Biotechnol. J.* 11, 459–469. doi: 10.1111/pbi.12032

Reyes, J. L., and Chua, N. H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J.* 49, 592–606. doi: 10.1111/j.1365-313X.2006.02980.x

Sun, Y., Luo, W., Chang, H., Li, Z., Zhou, J., Li, X., et al. (2019). Identification of miRNAs and their target genes involved in cucumber fruit expansion using small RNA and degradome sequencing. *Biomolecules.* 9, 15. doi: 10.3390/biom9090483

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment throug sequence weighting, position-specific gap penalties and weight marix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Wang, C., Jogaiah, S., Zhang, W., Abdelrahman, M., and Fang, J. (2018). Spatio-temporal expression of miRNA159 family members and their *GAMYB* target gene during the modulation of gibberellin-induced grapevine parthenocarpy. *J. Exp. Bot.* 69, 3639–3650. doi: 10.1093/jxb/ery172

Wang, Y., Sun, F., Cao, H., Peng, H., Ni, Z., Sun, Q., et al. (2012). TamiR159 directed wheat TaGAMYB cleavage and its involvement in anther development and heat response. *PloS One* 7, 10. doi: 10.1371/journal.pone.0048445

Wang, M., Xie, Z., Sun, X., Li, X., Zhu, X., Wang, C., et al. (2017). Function analysis of miR159 and its target gene *VvGAMYB* in grape flower development. *Acta Hortic. Sin.* 44, 1061–1072. doi: 10.16420/j.issn.0513-353x.2016-0816

Wei, L., Zhang, D., Xiang, F., and Zhang, Z. (2009). Differentially expressed miRNAs potentially involved in the regulation of defense mechanism to drought stress in maize seedlings. *Int. J. Plant Sci.* 170, 979–989. doi: 10.1086/605122

Yang, Z., Zhu, P., Kang, H., Liu, L., Cao, Q., Sun, J., 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, 16. doi: 10.1186/s12864-020-6567-3

Zhang, W., Han, X., Zhu, X., Xie, Z., Jiu, S., Huang, Y., et al. (2019). Identification of the target genes of VvmiR159s and their regulation in response to GA in different tissues of grape berry. *Sci. Agric. Sin.* 52, 2858–2870.

Zhang, Y., Zhang, B., Yang, T., Zhang, J., Liu, B., Zhan, X., et al. (2020). The GAMYB-like gene *SlMYB33* mediates flowering and pollen development in tomato. *Hortic. Res.* 7, 133. doi: 10.1038/s41438-020-00366-1

Zhao, P., Wang, F., Deng, Y., Zhong, F., Tian, P., Lin, D., et al. (2022). Sly-miR159 regulates fruit morphology by modulating GA biosynthesis in tomato. *Plant Biotechnol. J.* 20, 833–845. doi: 10.1111/pbi.13762

Zhao, Y., Wen, H., Teotia, S., Du, Y., Zhang, J., Li, J., et al. (2017). Suppression of microRNA159 impacts multiple agronomic traits in rice (*Oryza sativa* L.). *BMC Plant Biol.* 17, 13. doi: 10.1186/s12870-017-1171-7

frontiers | Frontiers in Plant Science

TYPE Data Report
PUBLISHED 16 May 2025
DOI 10.3389/fpls.2025.1520832

OPEN ACCESS

EDITED BY
Changmian Ji,
Chinese Academy of Tropical Agricultural Sciences, China

REVIEWED BY
Li Xinrui,
Sichuan Agricultural University, China
Jiban Mitra,
Indian Council of Agricultural Research, India

*CORRESPONDENCE
Daiana A. Zhernova
zhernova.d@ya.ru
Alexey A. Dmitriev
alex_245@mail.ru

RECEIVED 31 October 2024
ACCEPTED 25 March 2025
PUBLISHED 16 May 2025

CITATION
Zhernova DA, Arkhipov AA, Rozhmina TA, Zhuchenko AA, Bolsheva NL, Sigova EA, Dvorianinova EM, Borkhert EV, Pushkova EN, Melnikova NV and Dmitriev AA (2025) Transcriptome map and genome annotation of flax line 3896.
Front. Plant Sci. 16:1520832.
doi: 10.3389/fpls.2025.1520832

Transcriptome map and genome annotation of flax line 3896

Daiana A. Zhernova[1*], Alexander A. Arkhipov[1], Tatiana A. Rozhmina[2], Alexander A. Zhuchenko[2,3], Nadezhda L. Bolsheva[1], Elizaveta A. Sigova[1,4], Ekaterina M. Dvorianinova[1,4], Elena V. Borkhert[1], Elena N. Pushkova[1], Nataliya V. Melnikova[1] and Alexey A. Dmitriev[1*]

[1]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia, [2]Federal Research Center for Bast Fiber Crops, Torzhok, Russia, [3]All-Russian Horticultural Institute for Breeding, Agrotechnology and Nursery, Moscow, Russia, [4]Moscow Institute of Physics and Technology, Moscow, Russia

KEYWORDS

flax, *Linum usitatissimum*, line 3896, linseed, transcriptome map, reference genome, genome annotation

1 Introduction

Flax (*Linum usitatissimum* L.) is valued and cultivated worldwide for its dual-purpose use as both a seed and fiber crop. It is one of the oldest domesticated and most widely used crops (Kvavadze et al., 2009). Flax genome has a diploid chromosome set 2x=2n=30 and a size ~450 Mb (Dvorianinova et al., 2022; You et al., 2023). Flax seeds are rich in omega-3 fatty acids, digestible proteins, dietary fiber, and lignans (Goyal et al., 2014). Consumption of linseed oil has beneficial effects on human health, reducing the risks of many disorders and promoting immunity (Mali et al., 2019; Saini et al., 2021; Al-Madhagy et al., 2023). In addition, flax seed oil is used for technical purposes and also serves as a highly nutritious feed for livestock (Xu et al., 2022; Yadav et al., 2024). Flax fiber goes into the production of eco-friendly textiles with high absorption capacity and composite materials (Asyraf et al., 2022; More, 2022).

Different flax varieties are grown for different purposes and vary considerably in their characteristics. Today, traditional breeding is being aided by biotechnology and molecular genetics to select individuals with the desired traits more quickly and efficiently due to high-quality genome assemblies and their annotations. Since it is the study of transcriptomes in various tissues and organs that allows us to establish the associations between a valuable characteristic and its causative region in the genome (Dmitriev et al., 2020; Guo et al., 2020).

To date, there is a great deal of disparate data on gene expression in different flax organs and tissues, under different growth conditions and at different stages of ontogeny. Much work was devoted to the study of the involvement of specific genes in important agronomic traits, resistance to biotic and abiotic stressors, the regulation of organogenesis in flax plants and other characteristics.

Agriculturally valuable traits include those related to yield and to the quality of oil and fiber. The yield and the quality of oil were the subject of great interest of many studies (Xie

et al., 2019; Miart et al., 2021; Gao et al., 2022; Jiang et al., 2022; Wang et al., 2022b; Dvorianinova et al., 2023b; Pushkova et al., 2024). Since the valuable product obtained from flax is fiber, the characteristics of phloem fibers were actively studied (Roach and Deyholos, 2007; Zhang and Deyholos, 2016; Gorshkov et al., 2017; Gorshkova et al., 2018; Gorshkov et al., 2019; Galinousky et al., 2020; Mokshina et al., 2020; Petrova et al., 2021; Guo et al., 2022; Mokshina et al., 2022; Yu et al., 2022; Bao et al., 2023; Gorshkova et al., 2023; Liu et al., 2023; Ibragimova and Mokshina, 2024). In addition, transcriptome analysis largely allowed the identification of genes associated with flax plant height (Guo et al., 2021), the length of the growing season, the time of flowering, and the duration of ripening (Gao et al., 2022; House et al., 2022).

Data on flax gene expression and co-expression under suboptimal environmental conditions allowed researches to assess the association of genes with resistance to pathogen infection (Galindo-González and Deyholos, 2016; Dmitriev et al., 2017; Wu et al., 2019b; Boba et al., 2021; He et al., 2022) and abiotic stressors (Yu et al., 2014; Dmitriev et al., 2016; Dash et al., 2017; Wu et al., 2018; Krasnov et al., 2019; Wu et al., 2019a; Huang et al., 2021; Wang et al., 2021; Soto-Cerda et al., 2022; Wang et al., 2022a; Danaeipour et al., 2023; Kostyn et al., 2023; Qiu et al., 2023; Wang et al., 2023; Zhang et al., 2024).

The regulation of organogenesis in flax plays an important role in understanding the development of valuable flax traits (Saha et al., 2021; Yuan et al., 2021; Qi et al., 2023; Zhao et al., 2023). The comparative study of expression profiles of linseed and fiber flax varieties identified genes associated with flax plant type, flax oil odor, and paleohistorical data (Sveinsson et al., 2014; Povkhova et al., 2021; Yang et al., 2022).

Several synthesis articles with annotations for genome assemblies of flax varieties were published: linseed CDC Bethune (Wang et al., 2012) and fiber flax YY5 (Sa et al., 2021). However, the CDC Bethune genome contains some errors because it was assembled only from Illumina reads, which did not allow researchers to resolve its complexity (Sa et al., 2021; Dvorianinova et al., 2023a). The YY5 genome was annotated with transcriptome data of a different variety for only five samples of mature flax plants: leaf, stem, root, flower, and fruit.

In the NCBI database, the reference genome of *L. usitatissimum* is currently represented by a high-quality assembly of line 3896 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030674075.2/, accessed on 12 October 2024) obtained by us earlier (Dvorianinova et al., 2023a). Line 3896 belongs to the group of linseed flax and is characterized by resistance to Fusarium wilt (Rozhmina and Loshakova, 2016; Dmitriev et al., 2017) and edaphic stressor (low acidity) (Rozhmina et al., 2020), high seed yield and oil content (our observations). In the present study, we complement previous studies of line 3896 with a transcriptome map and genome annotation, which were necessary to make further progress in the field of flax genome research. Our annotation was obtained with the use of RNA-Seq data, whose positive effect on the annotation result was previously shown (Salzberg, 2019; Gabriel et al., 2024). The study of flax genome organization and gene expression will allow the development of methods to obtain improved varieties with desired traits with high efficiency. The results of the study are of use for the identification of genes and polymorphisms responsible for valuable traits and development of modern breeding technologies: genome editing, marker-assisted and genomic selection.

2 Materials and methods

2.1 Plant material

Seeds of linseed line 3896 were provided by the Institute for Flax (Torzhok, Russia). To obtain transcriptome data, we collected a set of organs (Table 1; Figure 1) of line 3896 plants at different stages of vegetation under optimal growth conditions described in the next subsection.

2.2 Flax cultivation

Flax seeds were sterilized in 1% sodium hypochlorite and 96% ethanol, after which the seeds were washed twice with water and planted in 0.05% fungicide Maxim (Syngenta, Gaillon, France) solution in Petri dishes for 7 days. Seedlings were planted in the soil and continued to grow under greenhouse conditions at 20°C and ~50-70% relative humidity with regular watering.

For transcriptome analysis, a set of different organs/tissues at different development stages was collected. The plant parts and ontogenetic stages used in the study are listed in Table 1. Each of the samples was collected in two biological replicates. Moreover, each sample was a pool of organs from 2-10 different plants (except capsules and seeds), which is necessary to level out differences between samples and be able to capture trends common to the species (Takele Assefa et al., 2020). The flowers were marked with the date of the day it opened (day of flowering). Seeds from the same capsules were pooled. Capsules were not pooled. The age of the plants at the time of each collection and the collection conditions are shown in Table 1. Samples were collected in the middle of the day between 12 and 15 h under similar conditions to smooth the influence of circadian rhythms on gene expression profiles. Samples were collected in liquid nitrogen and stored at -70°C.

2.3 RNA isolation

Samples #1-6 and #8-18 were grinded using a TissueLyser II homogenizer (Qiagen, Hilden, Germany) with the addition of 3 ceramic beads for two minutes. The harder samples (#7 and #19-28) were homogenized using a disposable pestle inserted in a DeWALT DCD701D2 cordless drill/driver (DeWALT, Towson, MD, USA) at 1200-1500 rpm in 1.5 ml tubes in liquid nitrogen to a fine powder, without allowing the sample to thaw. RNA isolation from samples #1-18 was performed using the Quick-RNA Miniprep Kit (Zymo Research, Irvine, CA, USA). RNA isolation from capsule and seed

10.3389/fpls.2025.1520832

TABLE 1 Examined samples of flax line 3896 organs/tissues at different stages of ontogenesis.

#	ID	Title	Description
1	Lin_SAM_1 Lin_SAM_2	SAM	Shoot apical meristems (SAM) from the upper part of the shoot at 30 days after germination (DAG). Their diameter was about 0.5 mm. Pool from 10 different plants.
2	Lin_FAM_1 Lin_FAM_2	FAM	Floral apical meristems (FAM) from the upper part of the shoot at 30 DAG if the shoot reached the early bud stage. Pool from 5 different plants.
3	Lin_leaf_blade_top_1 Lin_leaf_blade_top_2	Leaf laminae of young leaves	Leaf blades of young leaves from the top of the shoot at 30 DAG. Pool from 5 different plants.
4	Lin_leaf_blade_top3_1 Lin_leaf_blade_top3_2	Leaf laminae of intermediate leaves	Intermediate leaf blades from the shoot at 3 cm from the top at 30 DAG. Pool from 5 different plants.
5	Lin_leaf_blade_top10_1 Lin_leaf_blade_top10_2	Leaf laminae of mature leaves	Mature leaf blades from the shoot at 10 cm from the top at 30 DAG. Pool from 5 different plants.
6	Lin_shoot_1-3_1 Lin_shoot_1-3_2	Stem fragments 1-3 cm from the top	Stem fragments 1-3 cm from the top at 30 DAG. Pool from 3 different plants.
7	Lin_shoot_9-10_1 Lin_shoot_9-10_2	Stem fragments 9-10 cm from the top	Stem fragments 9-10 cm from the top at 30 DAG. Pool from 2 different plants.
8	Lin_s_SAM_1 Lin_s_SAM_2	SAM of seedlings	SAM of seedlings from the upper part of the shoot between cotyledons at 5 DAG on Petri dishes. Pool from 6 different plants.
9	Lin_s_cotyledon_1 Lin_s_cotyledon_2	Cotyledons of seedling	Cotyledons of seedlings at 5 DAG on Petri dishes. There were no true leaves yet, just the cotyledons. Pool from 6 different plants.
10	Lin_s_hyp_1 Lin_s_hyp_2	Hypocotyls of seedlings	Hypocotyls of seedlings collected as the stem between cotyledons and roots at 5 DAG on Petri dishes. Pool from 6 different plants.
11	Lin_s_root_1 Lin_s_root_2	Roots of seedlings	Roots of seedlings at 5 DAG on Petri dishes. Pool from 6 different plants.
12	Lin_anther_1 Lin_anther_2	Anthers of mature flowers	Mature flower anthers (before opening) at 56 DAG. Pool from 6 different plants.
13	Lin_pistil_1 Lin_pistil_2	Carpels of mature flowers	Mature flower carpels (before pollination) at 56 DAG. Pool from 6 different plants.
14	Lin_filament_1 Lin_filament_2	Stamen filaments of mature flowers	Mature flower stamen filaments (before opening) at 56 DAG. Pool from 6 different plants.
15	Lin_petal_1 Lin_petal_2	Petals of mature flowers	Mature flower petals (before opening) at 56 DAG. Pool from 6 different plants.
16	Lin_sepal_1 Lin_sepal_2	Sepals of mature flowers	Mature flower sepals (before opening) at 56 DAG. Pool from 6 different plants.
17	Lin_flower_1 Lin_flower_2	Flowers without pedicels	Mature flowers without pedicels at 56 DAG. Pool from 4 different plants.
18	Lin_pedicel_1 Lin_pedicel_2	Pedicels of mature flowers	Pedicels of mature flowers at 56 DAG. Pool from 6 different plants.
19	Lin_capsule_3_1 Lin_capsule_3_2	Capsule at 3 DAF	Capsule without seeds at 3 days after flowering (DAF).
20	Lin_capsule_7_1 Lin_capsule_7_2	Capsule at 7 DAF	Capsule without seeds at 7 DAF.
21	Lin_capsule_14_1 Lin_capsule_14_2	Capsule at 14 DAF	Capsule without seeds at 14 DAF.
22	Lin_capsule_21_1 Lin_capsule_21_2	Capsule at 21 DAF	Capsule without seeds at 21 DAF.
23	Lin_capsule_28_1 Lin_capsule_28_2	Capsule at 28 DAF	Capsule without seeds at 28 DAF.
24	Lin_seed_3_1 Lin_seed_3_2	Seeds at 3 DAF	Seeds without capsule at 3 DAF.

(Continued)

TABLE 1 Continued

#	ID	Title	Description
25	Lin_seed_7_1 Lin_seed_7_2	Seeds at 7 DAF	Seeds without capsule at 7 DAF.
26	Lin_seed_14_1 Lin_seed_14_2	Seeds at 14 DAF	Seeds without capsule at 14 DAF.
27	Lin_seed_21_1 Lin_seed_21_2	Seeds at 21 DAF	Seeds without capsule at 21 DAF.
28	Lin_seed_28_1 Lin_seed_28_2	Seeds at 28 DAF	Seeds without capsule at 28 DAF.

samples (#19-28) was performed by CTAB with modifications described previously (Pushkova et al., 2024). After that, total RNA was additionally cleaned using the CleanRNA Standard kit (Evrogen, Moscow, Russia) according to the manufacturer's protocol with a DNAase I treatment step from the RNase-Free DNase Set (Qiagen). The quality of isolated RNA was checked by horizontal electrophoresis in 2% agarose gel and using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and concentrations were measured using a Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.4 cDNA library preparation and sequencing on the Illumina platform

The cDNA libraries were prepared using the QIAseq Stranded mRNA Select Kit (Qiagen) according to the manufacturer's protocol. The quality of the obtained cDNA libraries (agreement of the length of the obtained libraries with the expected one and the absence of adapter dimers) was assessed on a Qsep1-Plus capillary electrophoresis system (BiOptic, New Taipei City, Taiwan), and the concentration was evaluated on a Qubit 4 fluorometer (Thermo Fisher Scientific). The cDNA libraries (all samples from Table 1) were mixed equimolarly and sequenced on a NextSeq 2000 instrument (Illumina, San Diego, CA, USA) using the NextSeq 2000 P3 Reagents (100 Cycles) kit (Illumina) in 51 + 51 nucleotide format.

2.5 Genome assembly annotation

RNA-Seq reads were trimmed with fastp 0.23.4 using default parameters (Chen et al., 2018). Structural annotation for the flax line 3896 genome assembly (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030674075.2/, accessed on 12 October 2024) (Dvorianinova et al., 2023a) was performed using BRAKER3 3.0.8 (Gabriel et al., 2024). This process utilized our RNA-Seq transcriptome data in combination with known Viridiplantae sequences from OrthoDB protein database for gene prediction (Kuznetsov et al., 2023). Functional annotation of the predicted genes was carried out using a local version of InterProScan 5.69-101.0 to assign functional domains and predict gene functions (Jones et al., 2014). The completeness of the annotation was assessed using BUSCO 5.7.1 in protein mode (the eudicots_odb10 dataset). Default parameters of program settings were used in the data analysis.

2.6 Analysis of transcriptome data

Gene expression analysis was performed with PPline (Krasnov et al., 2015) with default parameters of program settings and included the following steps:

- Alignment of the trimmed RNA-Seq reads to the annotated in the present study reference genome of flax line 3896 using STAR 2.7.2b (Dobin et al., 2013).
- Quantification of gene expression: read overlaps with annotated genomic features were counted using featureCounts 1.6.0 (Liao et al., 2014).
- Calculation of the normalized gene expression values, in counts per million (CPM), using edgeR.
- Sample normalization using TMM to account for differences in library sizes and composition between samples.

2.7 Gene enrichment analysis

Custom Gene Ontology (GO) annotation for line 3896 was constructed using EggNOG-mapper (–tax_scope 33090) (emapper 2.1.12, eggNOG DB version: 5.0.2) for the longest proteins in the constructed gene annotation (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021). Differential gene expression analysis was conducted for a balanced subset of flax samples: mature leaf blades from the shoot at 10 cm from the top at 30 days after germination (DAG), stem fragments 1-3 cm from the top at 30 DAG, roots of seedlings at 5 DAG, hypocotyls of seedlings at 5 DAG, mature flower carpels before pollination at 56 DAG, capsule without seeds at 14 days after flowering (DAF), seeds without capsule at 14 DAF, shoot apical meristems (SAM) of seedlings at 5 DAG. Gene expression in each tissue was compared to that in the other tissues from the subset of flax samples. Genes with FDR (QLF) < 0.05 and logFC > 1.5 were selected for gene enrichment analysis.

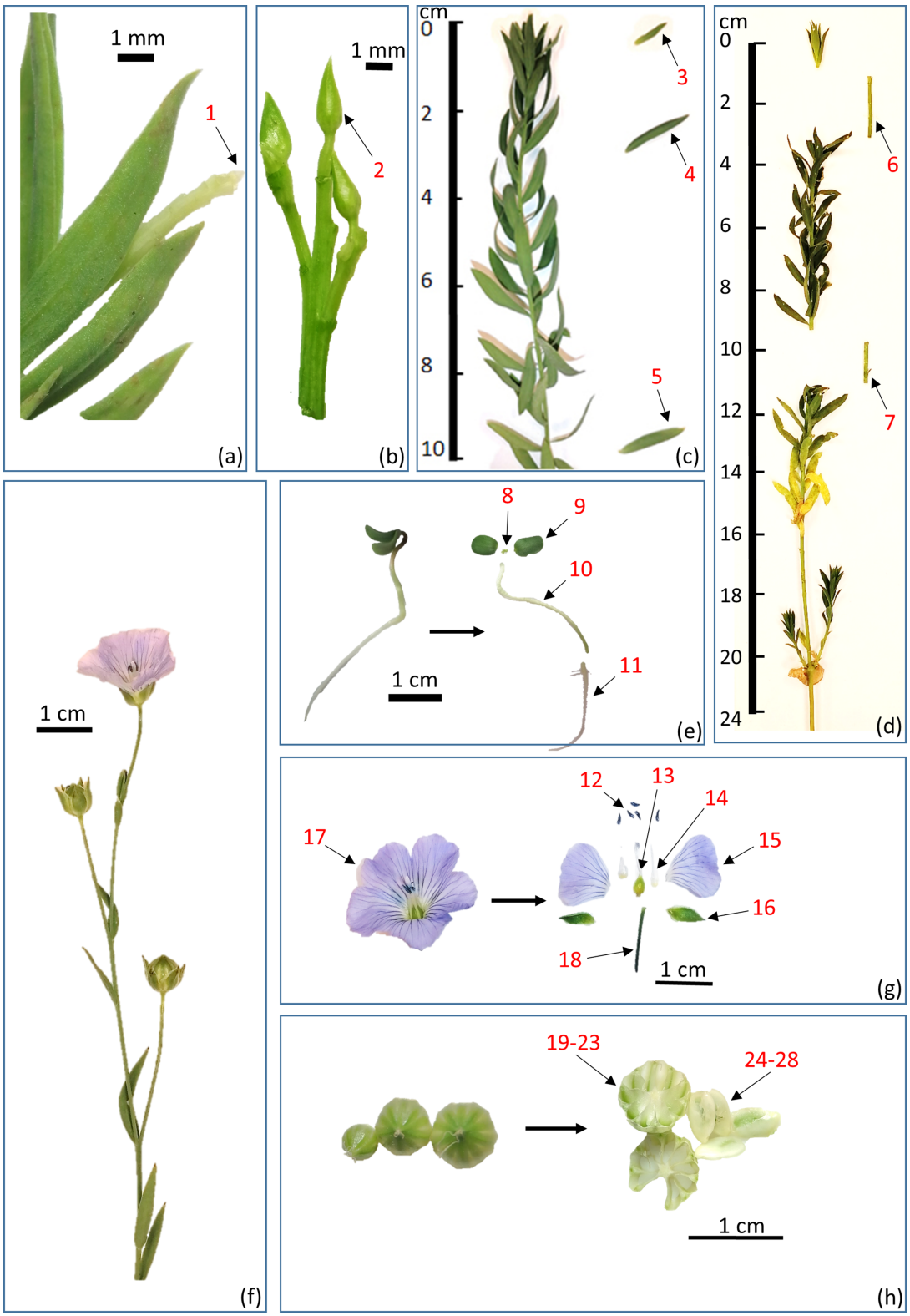


FIGURE 1
Flax organs/tissues studied by transcriptome analysis: **(a-d)** Apical parts of stem; **(e)** Seedling; **(f)** Mature plant with a flower and capsules; **(g)** Whole flower and its parts; **(h)** Capsules and seeds. 1 – SAM (shoot apical meristem); 2 – FAM (floral apical meristem); 3 – Leaf lamina of young leaf; 4 – Leaf lamina of intermediate leaf; 5 – Leaf lamina of mature leaf; 6 – Stem fragment 1-3 cm from the top; 7 – Stem fragment 9-10 cm from the top; 8 – SAM of seedling; 9 – Cotyledons of seedling; 10 – Hypocotyl of seedling; 11 – Root of seedling; 12 – Anthers of mature flower; 13 – Carpel of mature flower; 14 – Stamen filaments of mature flower; 15 – Petals of mature flower; 16 – Sepals of mature flower; 17 – Flower without pedicel; 18 – Pedicel; 19-23 – Capsule; 24-28 – Seeds. Numbers in Figure 1 corresponds to those in Table 1.

Gene enrichment was performed using the constructed annotation, the selected gene lists, and the topGO 2.54.0 package (Alexa and Rahnenfuhrer, 2023).

3 Preliminary data analysis

3.1 RNA-seq data characteristics

This article presents data of transcriptome analysis of 28 organs/tissues of flax line 3896: 7 samples of vegetative organs of actively growing individuals, 4 samples of seedlings, 7 samples of various parts of generative organs at flowering stage, and 10 samples of fruits and seeds at different stages of maturation (Table 1; Figure 1). We set out to cover all growth stages of flax and all key events in the development of valuable flax traits. From 5.4 to 20.5 million raw reads (51 + 51 bp) were obtained for each sample on the Illumina platform (two biological replicates were sequenced for each sample). The raw data were deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1172129.

After trimming, the reads were mapped to the genome assembly of line 3896, and on average ~95% of the reads for each sample were mapped (on average ~90% were uniquely mapped), confirming the high quality of the transcriptome data.

3.2 Genome annotation

We collected extensive data on gene expression in 28 organs/tissues of flax line 3896 at different development stages. Using the obtained transcriptome data and the Viridiplantae protein sequence database, we annotated the line 3896 genome assembly with BRAKER3: 39,055 genes and 45,154 transcripts were predicted, and 37,787 of these genes were annotated using InterProScan (Supplementary Data Sheet S1). The high completeness of annotation was achieved according to BUSCO (Benchmarking Universal Single-Copy Orthologs) – 95.6% (eudicots). Notably, 22.2% of all BUSCO were complete and single-copy and 73.4% were complete and duplicated. Such a high percentage of duplicated BUSCO was expected for an ancient tetraploid (Bolsheva et al., 2017).

3.3 Transcriptome map

Utilizing the obtained annotation of line 3896, we performed an analysis of our transcriptome data, which resulted in the identification of genes exhibiting tissue-specific and development stage-specific expression patterns within flax organs/tissues. This analysis led to the generation of a comprehensive transcriptome map for line 3896. To present the data in a convenient format for further analysis, we used PPline and RTrans (https://github.com/gskrasnov/RTrans, accessed on 17 October 2024). It was applied to evaluate the expression levels of the identified genes as read counts per million reads (CPM). Our transcriptome map is summarized in Supplementary Table S1 and presented as a heatmap (Supplementary Figure S1).

3.4 Gene pathway enrichment analysis

During the gene pathway enrichment analysis, the following was found out. For leaves collected at a distance of 10 cm from the apex, compared to the other organs/tissues, the GO terms were represented by the processes of photosynthesis, carbohydrate metabolism, plastid organization, electron transport chain, pigment synthesis, and transmembrane transport. These processes are characteristic of the main photosynthetic organ of the plant in the active phase (Müller and Munné-Bosch, 2021; Leister, 2023). In the stem fragment, located 1-3 cm from the top, the processes of vascular tissue histogenesis, vascular and phloem transport, stem morphogenesis, and response to auxin synthesized in the apical meristem predominated. This is logical, since we were dealing with an axial organ whose main function is the transport of metabolites, and the incision was made close to the site of differentiation (Yoshida et al., 2009; Kułak et al., 2023). It was determined that the major pathways in the seedling root included the processes of water and solute transport, root hair formation and growth, response to chemical and mechanical stimuli, and metabolism of auxin, other hormones, and secondary metabolites, which are the main processes occurring in the roots of vascular plants (Vissenberg et al., 2020; Li et al., 2021; Castillo-Jiménez et al., 2023). GO analysis of the hypocotyl transcriptome revealed differential expression of genes related to pathways of amino acid biosynthesis, as well as active regulation of biosynthetic processes, brassinosteroid metabolism, and cell growth. These findings indicate the presence of active development processes in the axial organs of seedlings (Favero et al., 2021). GO analysis for flax pistil demonstrated the representation of genes that are associated with the formation and development of generative structures, pollination, and pollen tube growth. Furthermore, it demonstrated representation of genes associated with active ion metabolism, which is necessary for directed pollen tube growth (Zhou et al., 2022). GO analysis for a capsule (14 DAF) indicated that the GO terms associated with the formation of secondary cell wall, synthesis of its components, and lignification prevailed. These processes enable the preparation of dry fruits for opening and seed dispersal (Seymour et al., 2013). Additionally, catabolism of organic compounds used for cell wall construction and seed maturation is active in ripening capsules. The differential expression pattern of flax seeds (14 DAF) was dominated by processes related to seed and fruit development and maturation, lipid storage and fatty acid synthesis, as well as abscisic acid (ABA) metabolism and regulation. Oil accumulation corresponds to the primary function of the seed as a reproductive organ, and ABA regulates its maturation (Sano and Marion-Poll, 2021; Dvorianinova et al., 2023b). The shoot apical meristem displayed a hallmark pattern indicative of actively dividing cells. The process of SAM is characterized by the macromolecule biosynthesis, ribosome assembly, translation, RNA processing,

DNA reparation, organelle formation, nucleosome assembly, and chromatin remodeling (Xue et al., 2020; Burian, 2021). Thus, the results of the gene enrichment analysis of the subset of flax samples in general looked logical and confirmed the adequacy of the obtained data. The results are presented in detail in Supplementary Table S2.

4 Conclusions

Flax is of great industrial and nutritional value and is therefore actively studied at the molecular-genetic level. To date, there are several flax genome assemblies, some with annotations, and many scattered gene expression data for different varieties. In this study, we used Illumina sequencing to obtain comprehensive transcriptome data for flax line 3896, whose genome is currently a reference for the species *Linum usitatissimum* L. in the NCBI database. Gene expression profiles were analyzed in 28 various flax organs/tissues at different stages of ontogenesis. With these data we were able to annotate the genome of line 3896 and generate a high-quality transcriptome map. The transcriptome map will allow the identification of genes that have a high expression level in a particular organ/tissue. Such genes may play a key role in the biological processes taking place in that organ/tissue. In addition, data on gene expression profiles during plant development can help to determine the most important time points at which the processes of interest occur. The transcriptome map also allows the determination of gene functions based not only on homology analysis, but also taking into account gene expression patterns in different organs/tissues. Thus, the transcriptome map and annotation presented in this work allow reaching a new level in the molecular-genetic studies of flax, the search for key genes responsible for the valuable traits, the development of new approaches in flax breeding and the creation of improved varieties.

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://www.ncbi.nlm.nih.gov/, PRJNA1172129.

Author contributions

DZ: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. AA: Investigation, Writing – original draft, Writing – review & editing. TR: Investigation, Writing – review & editing. AZ: Investigation, Writing – review & editing. NB: Investigation, Writing – review & editing. ES: Investigation, Writing – review & editing. ED: Investigation, Writing – review & editing. EB: Investigation, Writing – review & editing. EP: Investigation, Writing – review & editing. NM: Conceptualization, Investigation, Writing – review & editing. AD: Conceptualization, Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was financially supported by the Russian Science Foundation, grant number 24-64-00033, https://rscf.ru/project/24-64-00033/.

Acknowledgments

This work was performed using the equipment of the EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu_genome_ce.php).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1520832/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Expression heatmap based on $\log_2$(CPM/avg.CPM) of annotated genes in 28 organs/tissues of flax line 3896.

SUPPLEMENTARY TABLE S1
Expression levels of annotated genes in 28 organs/tissues of flax line 3896 as read counts per million reads (CPM).

SUPPLEMENTARY TABLE S2
Gene enrichment analysis for a balanced subset of flax samples: mature leaf blades from the shoot at 10 cm from the top at 30 DAG, stem fragments 1-3 cm from the top at 30 DAG, roots of seedlings at 5 DAG, hypocotyls of seedlings at 5 DAG, mature flower carpels before pollination at 56 DAG, capsule without seeds at 14 DAF, seeds without capsule at 14 DAF, SAM (shoot apical meristem) of seedlings at 5 DAG.

SUPPLEMENTARY DATA SHEET S1
Annotation of flax line 3896 genome assembly.

References

Alexa, A., and Rahnenfuhrer, J. (2023). *topGO: Enrichment Analysis for Gene Ontology (2.54.0)*. doi: 10.18129/B9.bioc.topGO

Al-Madhagy, S., Ashmawy, N. S., Mamdouh, A., Eldahshan, O. A., and Farag, M. A. (2023). A comprehensive review of the health benefits of flaxseed oil in relation to its chemical composition and comparison with other omega-3-rich oils. *Eur. J. Med. Res.* 28, 240. doi: 10.1186/s40001-023-01203-6

Asyraf, M. R. M., Ishak, M. R., Norrrahim, M. N. F., Amir, A. L., Nurazzi, N. M., Ilyas, R. A., et al. (2022). Potential of flax fiber reinforced biopolymer composites for cross-arm application in transmission tower: a review. *Fibers Polymers* 23, 853–877. doi: 10.1007/s12221-022-4383-x

Bao, Y., Zou, Y., Tian, R., Huang, X., Liu, L., Wang, B., et al. (2023). Transcriptome analysis of fiber development under high-temperature stress in flax (*Linum usitatissimum* L.). *Ind. Crops Prod.* 195, 116019. doi: 10.1016/j.indcrop.2022.116019

Boba, A., Kostyn, K., Kozak, B., Zalewski, I., Szopa, J., and Kulma, A. (2021). Transcriptomic profiling of susceptible and resistant flax seedlings after *Fusarium oxysporum* lini infection. *PloS One* 16, e0246052. doi: 10.1371/journal.pone.0246052

Bolsheva, N. L., Melnikova, N. V., Kirov, I. V., Speranskaya, A. S., Krinitsina, A. A., Dmitriev, A. A., et al. (2017). Evolution of blue-flowered species of genus *Linum* based on high-throughput sequencing of ribosomal RNA genes. *BMC Evolutionary Biol.* 17, 253. doi: 10.1186/s12862-017-1105-x

Burian, A. (2021). Does shoot apical meristem function as the germline in safeguarding against excess of mutations? *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.707740

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Castillo-Jiménez, A., Garay-Arroyo, A., de la Paz Sánchez, M., García-Ponce, B., Martínez-García, J. C., and Álvarez-Buylla, E. R. (2023). Extended discrete gene regulatory network model for the *Arabidopsis thaliana* root-hair cell fate. *bioRxiv*. doi: 10.1101/2023.11.15.567304

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Danaeipour, Z., Garoosi, G., Tohidfar, M., Bakhtiari Zadeh, M. R., and Mirjalili, M. H. (2023). Data mining of flax transcriptome profile under stress aluminum toxicity and zinc deficiency. *MGJ* 18, 21–32.

Dash, P. K., Rai, R., Mahato, A. K., Gaikwad, K., and Singh, N. K. (2017). Transcriptome landscape at different developmental stages of a drought tolerant cultivar of flax (*Linum usitatissimum*). *Front. Chem.* 5. doi: 10.3389/fchem.2017.00082

Dmitriev, A. A., Krasnov, G. S., Rozhmina, T. A., Novakovskiy, R. O., Snezhkina, A. V., Fedorova, M. S., et al. (2017). Differential gene expression in response to *Fusarium oxysporum* infection in resistant and susceptible genotypes of flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 17, 253. doi: 10.1186/s12870-017-1192-2

Dmitriev, A. A., Kudryavtseva, A. V., Krasnov, G. S., Koroban, N. V., Speranskaya, A. S., Krinitsina, A. A., et al. (2016). Gene expression profiling of flax (*Linum usitatissimum* L.) under edaphic stress. *BMC Plant Biol.* 16, 237. doi: 10.1186/s12870-016-0927-9

Dmitriev, A. A., Novakovskiy, R. O., Pushkova, E. N., Rozhmina, T. A., Zhuchenko, A. A., Bolsheva, N. L., et al. (2020). Transcriptomes of different tissues of flax (*Linum usitatissimum* L.) cultivars with diverse characteristics. *Front. Genet.* 11. doi: 10.3389/fgene.2020.565146

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dvorianinova, E. M., Bolsheva, N. L., Pushkova, E. N., Rozhmina, T. A., Zhuchenko, A. A., Novakovskiy, R. O., et al. (2022). Isolating *Linum usitatissimum* L. nuclear DNA enabled assembling high-quality genome. *Int. J. Mol. Sci.* 23, 13244. doi: 10.3390/ijms232113244

Dvorianinova, E., Pushkova, E., Bolsheva, N., Rozhmina, T., Zhernova, D., Sigova, E., et al. (2023a). Improving genome assembly of flax line 3896 with high-precision Illumina reads. *Russian J. Genet.* 59, S237–S240. doi: 10.1134/S102279542314003X

Dvorianinova, E. M., Zinovieva, O. L., Pushkova, E. N., Zhernova, D. A., Rozhmina, T. A., Povkhova, L. V., et al. (2023b). Key *FAD2*, *FAD3*, and *SAD* genes involved in the fatty acid synthesis in flax identified based on genomic and transcriptomic data. *Int. J. Mol. Sci.* 24, 14885. doi: 10.3390/ijms241914885

Favero, D. S., Lambolez, A., and Sugimoto, K. (2021). Molecular pathways regulating elongation of aerial plant organs: a focus on light, the circadian clock, and temperature. *Plant J.* 105, 392–420. doi: 10.1111/tpj.14996

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., et al. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* 34, 769–777. doi: 10.1101/gr.278090.123

Galindo-González, L., and Deyholos, M. K. (2016). RNA-seq transcriptome response of flax (*Linum usitatissimum* L.) to the pathogenic fungus *Fusarium oxysporum* f. sp. *lini*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01766

Galinousky, D., Padvitski, T., Mokshina, N., Gorshkov, O., Khotyleva, L., Gorshkova, T., et al. (2020). Expression of cellulose synthase-like genes in two phenotypically distinct flax (*Linum usitatissimum* L.) subspecies. *Genet. Resour. Crop Evol.* 67, 1821–1837. doi: 10.1007/s10722-020-00943-2

Gao, P., Qiu, S., Ma, X., Parkin, I. A., Xiang, D., and Datla, R. (2022). Spatiotemporal transcriptomic atlas of developing embryos and vegetative tissues in flax. *Plants* 11, 2031. doi: 10.3390/plants11152031

Gorshkov, O., Chernova, T., Mokshina, N., Gogoleva, N., Suslov, D., Tkachenko, A., et al. (2019). Intrusive growth of phloem fibers in flax stem: integrated analysis of miRNA and mRNA expression profiles. *Plants* 8, 47. doi: 10.3390/plants8020047

Gorshkov, O., Mokshina, N., Gorshkov, V., Chemikosova, S., Gogolev, Y., and Gorshkova, T. (2017). Transcriptome portrait of cellulose-enriched flax fibres at advanced stage of specialization. *Plant Mol. Biol.* 93, 431–449. doi: 10.1007/s11103-016-0571-7

Gorshkova, T., Chernova, T., Mokshina, N., Gorshkov, V., Kozlova, L., and Gorshkov, O. (2018). Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci. Rep.* 8, 14570. doi: 10.1038/s41598-018-32869-2

Gorshkova, T., Mokshina, N., Mitsuda, N., Gorshkov, O., and You, F. M. (2023). "Key stages of flax bast fiber development through the prism of transcriptomics," in *The flax genome*. Ed. B. Fofana (Springer, Cham), 149–198.

Goyal, A., Sharma, V., Upadhyay, N., Gill, S., and Sihag, M. (2014). Flax and flaxseed oil: an ancient medicine & modern functional food. *J. Food Sci. Technol.* 51, 1633–1653. doi: 10.1007/s13197-013-1247-9

Guo, D., Jiang, H., Yan, W., Yang, L., Ye, J., Wang, Y., et al. (2020). Resequencing 200 flax cultivated accessions identifies candidate genes related to seed size and weight and reveals signatures of artificial selection. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01682

Guo, D., Jiang, H., Ye, J., Zhang, A., Wang, Y., Gao, Y., et al. (2021). Transcriptome combined with population level validation reveals genomic loci controlling plant height in flax (*Linum usitatissimum* L.). *Ind. Crops Prod.* 172, 113998. doi: 10.1016/j.indcrop.2021.113998

Guo, Y., Wen, L., Chen, J., Pan, G., Wu, Z., Li, Z., et al. (2022). Comparative transcriptomic analysis identifies key cellulose synthase genes (*CESA*) and cellulose synthase-like genes (*CSL*) in fast growth period of flax stem (*Linum usitatissimum* l.). *J. Natural Fibers* 19, 10431–10446. doi: 10.1080/15440478.2021.1993510

He, R., Chang, Y., and Wang, J. (2022). Identification of genes responsible for stress resistance in *Fusarium oxysporum*-inoculated flax seedlings using weighted gene co-expression network analysis. *Eur. J. Plant Pathol.* 163, 513–528. doi: 10.1007/s10658-022-02497-8

House, M. A., Young, L. W., Robinson, S. J., and Booker, H. M. (2022). Transcriptomic analysis of early flowering signals in 'Royal'Flax. *Plants* 11, 860. doi: 10.3390/plants11070860

Huang, W.-G., Jiang, W.-D., Yao, Y.-B., Song, X.-X., Liu, Y., Chen, S., et al. (2021). Transcriptome profiling of flax (*Linum usttatissimum* L.) response to low potassium stress. *Acta Agronomica Sin.* 47, 1070–1081. doi: 10.3724/SP.J.1006.2021.04133

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Ibragimova, N., and Mokshina, N. (2024). Callose metabolism in flax fibers during gravity response: analysis of gene expression. *Russian J. Plant Physiol.* 71, 57. doi: 10.1134/S1021443724604567

Jiang, H., Guo, D., Liu, Y., Zhu, L., Xie, F., and Xie, L. (2022). RNA-Seq combined with population-level analysis reveals important candidate genes related to seed size in flax (*Linum usitatissimum* L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1015399

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kostyn, K., Boba, A., Kozak, B., Sztafrowski, D., Widuła, J., Szopa, J., et al. (2023). Transcriptome profiling of flax plants exposed to a low-frequency alternating electromagnetic field. *Front. Genet.* 14. doi: 10.3389/fgene.2023.1205469

Krasnov, G. S., Dmitriev, A. A., Kudryavtseva, A. V., Shargunov, A. V., Karpov, D. S., Uroshlev, L. A., et al. (2015). PPLine: an automated pipeline for SNP, SAP, and splice variant detection in the context of proteogenomics. *J. Proteome Res.* 14, 3729–3737. doi: 10.1021/acs.jproteome.5b00490

Krasnov, G. S., Dmitriev, A. A., Zyablitsin, A. V., Rozhmina, T. A., Zhuchenko, A. A., Kezimana, P., et al. (2019). Aluminum responsive genes in flax (*Linum usitatissimum* L.). *BioMed. Res. Int.* 2019, 5023125. doi: 10.1155/2019/5023125

Kułak, K., Wojciechowska, N., Samelak-Czajka, A., Jackowiak, P., and Bagniewska-Zadworna, A. (2023). How to explore what is hidden? A review of techniques for vascular tissue expression profile analysis. *Plant Methods* 19, 129. doi: 10.1186/s13007-023-01109-8

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., et al. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* 51, D445–D451. doi: 10.1093/nar/gkac998

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., et al. (2009). 30,000-year-old wild flax fibers. *Science* 325, 1359–1359. doi: 10.1126/science.1175404

Leister, D. (2023). Enhancing the light reactions of photosynthesis: Strategies, controversies, and perspectives. *Mol. Plant* 16, 4–22. doi: 10.1016/j.molp.2022.08.005

Li, H., Testerink, C., and Zhang, Y. (2021). How roots and shoots communicate through stressful times. *Trends Plant Sci.* 26, 940–952. doi: 10.1016/j.tplants.2021.03.005

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, D., Yuan, H., Yao, Y., Cheng, L., Tang, L., Kang, Q., et al. (2023). Genome-wide identification and expression pattern analysis of the *F5H* gene family in flax (*Linum usitatissimum* L.). *Agronomy* 13, 1108. doi: 10.3390/agronomy13041108

Mali, A. V., Padhye, S. B., Anant, S., Hegde, M. V., and Kadam, S. S. (2019). Anticancer and antimetastatic potential of enterolactone: Clinical, preclinical and mechanistic perspectives. *Eur. J. Pharmacol.* 852, 107–124. doi: 10.1016/j.ejphar.2019.02.022

Miart, F., Fontaine, J.-X., Mongelard, G., Wattier, C., Lequart, M., Bouton, S., et al. (2021). Integument-specific transcriptional regulation in the mid-stage of flax seed development influences the release of mucilage and the seed oil content. *Cells* 10, 2677. doi: 10.3390/cells10102677

Mokshina, N., Gorshkov, O., Galinousky, D., and Gorshkova, T. (2020). Genes with bast fiber-specific expression in flax plants-Molecular keys for targeted fiber crop improvement. *Ind. Crops Prod.* 152, 112549. doi: 10.1016/j.indcrop.2020.112549

Mokshina, N., Mikshina, P., and Gorshkova, T. (2022). Expression of cellulose synthase genes during the gravistimulation of flax (*Linum usitatissimum*) and poplar (*Populus alba× tremula*) plants. *Russian J. Bioorg. Chem.* 48, 529–542. doi: 10.1134/S106816202203013X

More, A. P. (2022). Flax fiber–based polymer composites: a review. *Advanced Composites Hybrid Mater.* 5, 1–20. doi: 10.1007/s42114-021-00246-9

Müller, M., and Munné-Bosch, S. (2021). Hormonal impact on photosynthesis and photoprotection in plants. *Plant Physiol.* 185, 1500–1522. doi: 10.1093/plphys/kiaa119

Petrova, N., Nazipova, A., Gorshkov, O., Mokshina, N., Patova, O., and Gorshkova, T. (2021). Gene expression patterns for proteins with lectin domains in flax stem tissues are related to deposition of distinct cell wall types. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.634594

Povkhova, L. V., Melnikova, N. V., Rozhmina, T. A., Novakovskiy, R. O., Pushkova, E. N., Dvorianinova, E. M., et al. (2021). Genes associated with the flax plant type (oil or fiber) identified based on genome and transcriptome sequencing data. *Plants* 10, 2616. doi: 10.3390/plants10122616

Pushkova, E., Dvorianinova, E., Povkhova, L., Rozhmina, T., Novakovskiy, R., Sigova, E., et al. (2024). Expression profiles of genes involved in lignan synthesis in developing flax seeds. *Russian J. Genet.* 60, 987–991. doi: 10.1134/S1022795424700443

Qi, F., Wang, F., Xiaoyang, C., Wang, Z., Lin, Y., Peng, Z., et al. (2023). Gene expression analysis of different organs and identification of AP2 transcription factors in flax (*Linum usitatissimum* L.). *Plants* 12, 3260. doi: 10.3390/plants12183260

Qiu, C., Qiu, H., Peng, D., Chen, J., Wang, Y., Stybayev, G., et al. (2023). The mechanisms underlying physiological and molecular responses to waterlogging in flax. *J. Natural Fibers* 20, 2198275. doi: 10.1080/15440478.2023.2198275

Roach, M. J., and Deyholos, M. K. (2007). Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. *Mol. Genet. Genomics* 278, 149–165. doi: 10.1007/s00438-007-0241-1

Rozhmina, T., and Loshakova, N. (2016). New sources of effective resistance genes to fusarium wilt in flax (*Linum usitatissimum* L.) depending on temperature. *Agric. Biol.* 51, 310–317. doi: 10.15389/agrobiology.2016.3.310eng

Rozhmina, T., Zhuchenko, A., Melnikova, N., and Smirnova, A. (2020). Resistance of flax gene pool samples to edaphic stress caused by low acidity. *Agric. Sci. Euro North East* 21, 133–140. doi: 10.30766/2072-9081.2020.21.2.133-140

Sa, R., Yi, L., Siqin, B., An, M., Bao, H., Song, X., et al. (2021). Chromosome-level genome assembly and annotation of the fiber flax (*Linum usitatissimum*) genome. *Front. Genet.* 12. doi: 10.3389/fgene.2021.735690

Saha, D., Shaw, A. K., Datta, S., and Mitra, J. (2021). Evolution and functional diversity of abiotic stress-responsive *NAC* transcription factor genes in *Linum usitatissimum* L. *Environ. Exp. Bot.* 188, 104512. doi: 10.1016/j.envexpbot.2021.104512

Saini, R. K., Prasad, P., Sreedhar, R. V., Akhilender Naidu, K., Shang, X., and Keum, Y.-S. (2021). Omega-3 polyunsaturated fatty acids (PUFAs): Emerging plant and microbial sources, oxidative stability, bioavailability, and health benefits — A review. *Antioxidants* 10, 1627. doi: 10.3390/antiox10101627

Salzberg, S. L. (2019). Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* 20, 92. doi: 10.1186/s13059-019-1715-2

Sano, N., and Marion-Poll, A. (2021). ABA metabolism and homeostasis in seed dormancy and germination. *Int. J. Mol. Sci.* 22, 5069. doi: 10.3390/ijms22105069

Seymour, G. B., Østergaard, L., Chapman, N. H., Knapp, S., and Martin, C. (2013). Fruit development and ripening. *Annu. Rev. Plant Biol.* 64, 219–241. doi: 10.1146/annurev-arplant-050312-120057

Soto-Cerda, B. J., Larama, G., Gajardo, H., Inostroza-Blancheteau, C., Cloutier, S., Fofana, B., et al. (2022). Integrating multi-locus genome-wide association studies with transcriptomic data to identify genetic loci underlying adult root trait responses to drought stress in flax (*Linum usitatissimum* L.). *Environ. Exp. Bot.* 202, 105019. doi: 10.1016/j.envexpbot.2022.105019

Sveinsson, S., McDill, J., Wong, G. K., Li, J., Li, X., Deyholos, M. K., et al. (2014). Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics. *Ann. Bot.* 113, 753–761. doi: 10.1093/aob/mct306

Takele Assefa, A., Vandesompele, J., and Thas, O. (2020). On the utility of RNA sample pooling to optimize cost and statistical power in RNA sequencing experiments. *BMC Genomics* 21, 312. doi: 10.1186/s12864-020-6721-y

Vissenberg, K., Claeijs, N., Balcerowicz, D., and Schoenaers, S. (2020). Hormonal regulation of root hair growth and responses to the environment in Arabidopsis. *J. Exp. Bot.* 71, 2412–2427. doi: 10.1093/jxb/eraa048

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., et al. (2012). The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. *Plant J.* 72, 461–473. doi: 10.1111/j.1365-313X.2012.05093.x

Wang, N., Lin, Y., Qi, F., Xiaoyang, C., Peng, Z., Yu, Y., et al. (2022a). Comprehensive analysis of differentially expressed genes and epigenetic modification-related expression variation induced by saline stress at seedling stage in fiber and oil flax, *Linum usitatissimum* L. *Plants* 11, 2053. doi: 10.3390/plants11152053

Wang, N., Qi, F., Wang, F., Lin, Y., Xiaoyang, C., Peng, Z., et al. (2023). Evaluation of differentially expressed genes in leaves vs. roots subjected to drought stress in flax (*Linum usitatissimum* L.). *Int. J. Mol. Sci.* 24, 12019. doi: 10.3390/ijms241512019

Wang, W., Wang, L., Wang, L., Tan, M., Ogutu, C. O., Yin, Z., et al. (2021). Transcriptome analysis and molecular mechanism of linseed (*Linum usitatissimum* L.) drought tolerance under repeated drought using single-molecule long-read sequencing. *BMC Genomics* 22, 109. doi: 10.1186/s12864-021-07416-5

Wang, S., Xu, J., Wang, Z., Li, Z., Yi, L., Yao, L., et al. (2022b). Gene screening for fatty acid synthesis of flax based on transcriptome sequencing. *Food Sci. Technol.* 42, e93721. doi: 10.1590/fst.93721

Wu, W., Nemri, A., Blackman, L. M., Catanzariti, A.-M., Sperschneider, J., Lawrence, G. J., et al. (2019b). Flax rust infection transcriptomics reveals a transcriptional profile that may be indicative for rust *Avr* genes. *PloS One* 14, e0226106. doi: 10.1371/journal.pone.0226106

Wu, J., Zhao, Q., Sun, D., Wu, G., Zhang, L., Yuan, H., et al. (2018). Transcriptome analysis of flax (*Linum usitatissimum* L.) undergoing osmotic stress. *Ind. Crops Prod.* 116, 215–223. doi: 10.1016/j.indcrop.2018.02.035

Wu, J., Zhao, Q., Wu, G., Yuan, H., Ma, Y., Lin, H., et al. (2019a). Comprehensive analysis of differentially expressed unigenes under NaCl stress in flax (*Linum usitatissimum* L.) using RNA-Seq. *Int. J. Mol. Sci.* 20, 369. doi: 10.3390/ijms20020369

Xie, D., Dai, Z., Yang, Z., Tang, Q., Deng, C., Xu, Y., et al. (2019). Combined genome-wide association analysis and transcriptome sequencing to identify candidate genes for flax seed fatty acid metabolism. *Plant Sci.* 286, 98–107. doi: 10.1016/j.plantsci.2019.06.004

Xu, L., Wei, Z., Guo, B., Bai, R., Liu, J., Li, Y., et al. (2022). Flaxseed meal and its application in animal husbandry: A review. *Agriculture* 12, 2027. doi: 10.3390/agriculture12122027

Xue, Z., Liu, L., and Zhang, C. (2020). Regulation of shoot apical meristem and axillary meristem development in plants. *Int. J. Mol. Sci.* 21, 2917. doi: 10.3390/ijms21082917

Yadav, S., Sharma, K., and Singh, D. (2024). "Modification methods of linseed oil and its applications in different product development," in *Linseed*. Eds. S. Langyan and A. Kumar. (Elsevier, Amsterdam), 219–225.

Yang, J.-W., Li, H.-K., Lin, Y.-F., Liu, W.-Y., and Wei, C.-Q. (2022). Transcriptome analysis of two flaxseed cultivars grown in Xinjiang and mining of differentially expressed genes related to flaxseed oil aroma. *Food Sci.* 43, 70–76. doi: 10.7506/spkx1002-6630-20201120-213

Yoshida, S., Iwamoto, K., Demura, T., and Fukuda, H. (2009). Comprehensive analysis of the regulatory roles of auxin in early transdifferentiation into xylem cells. *Plant Mol. Biol.* 70, 457–469. doi: 10.1007/s11103-009-9485-y

You, F. M., Moumen, I., Khan, N., and Cloutier, S. (2023). "Reference genome sequence of flax," in *The flax genome*. Eds. F. M. You and B. Fofana (Springer, Cham), 1–17.

Yu, Y., Huang, W., Chen, H., Wu, G., Yuan, H., Song, X., et al. (2014). Identification of differentially expressed genes in flax (*Linum usitatissimum* L.) under saline–alkaline stress by digital gene expression. *Gene* 549, 113–122. doi: 10.1016/j.gene.2014.07.053

Yu, Y., Yuan, H., Zhao, D., Yao, Y., Zhou, F., Cheng, L., et al. (2022). Transcriptome analysis of two flax varieties at different developmental stages reveals significant differential expression of genes related to stem fiber development. *J. Natural Fibers* 19, 1428–1442. doi: 10.1080/15440478.2020.1776662

Yuan, H., Guo, W., Zhao, L., Yu, Y., Chen, S., Tao, L., et al. (2021). Genome-wide identification and expression analysis of the WRKY transcription factor family in flax (*Linum usitatissimum* L.). *BMC Genomics* 22, 375. doi: 10.1186/s12864-021-07697-w

Zhang, N., and Deyholos, M. K. (2016). RNASeq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00950

Zhang, F., Liu, Y., Ma, J., Su, S., Chen, L., Cheng, Y., et al. (2024). Analyzing the diversity of MYB family response strategies to drought stress in different flax varieties based on transcriptome data. *Plants* 13, 710. doi: 10.3390/plants13050710

Zhao, L.-R., Li, W., Wang, L.-M., Qi, Y.-N., Li, W.-J., Xie, Y.-P., et al. (2023). Identification and relative expression pattern of *PLA1* gene family in flax. *Acta Agronomica Sin.* 49, 2949–2965. doi: 10.3724/SP.J.1006.2023.24224

Zhou, Z., Zheng, S., Haq, S. I. U., Zheng, D., and Qiu, Q. S. (2022). Regulation of pollen tube growth by cellular pH and ions. *J. Plant Physiol.* 277, 153792. doi: 10.1016/j.jplph.2022.153792

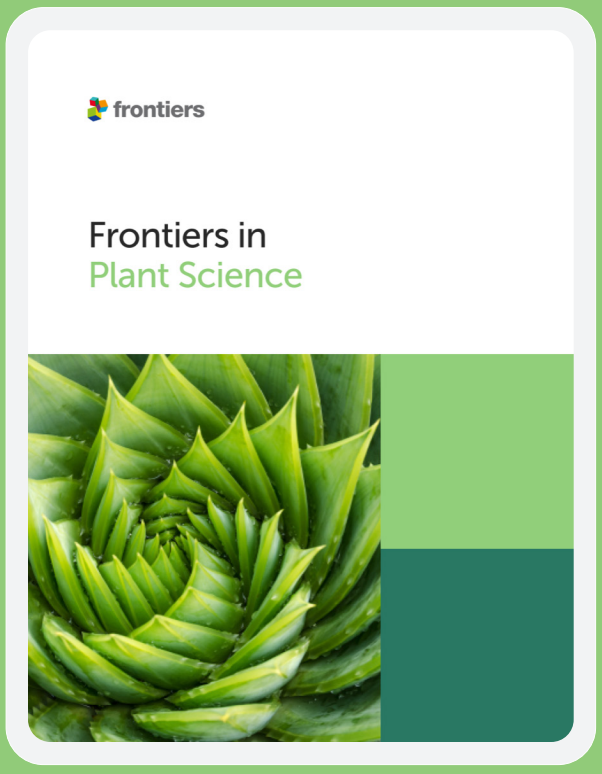
Frontiers in
Plant Science

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