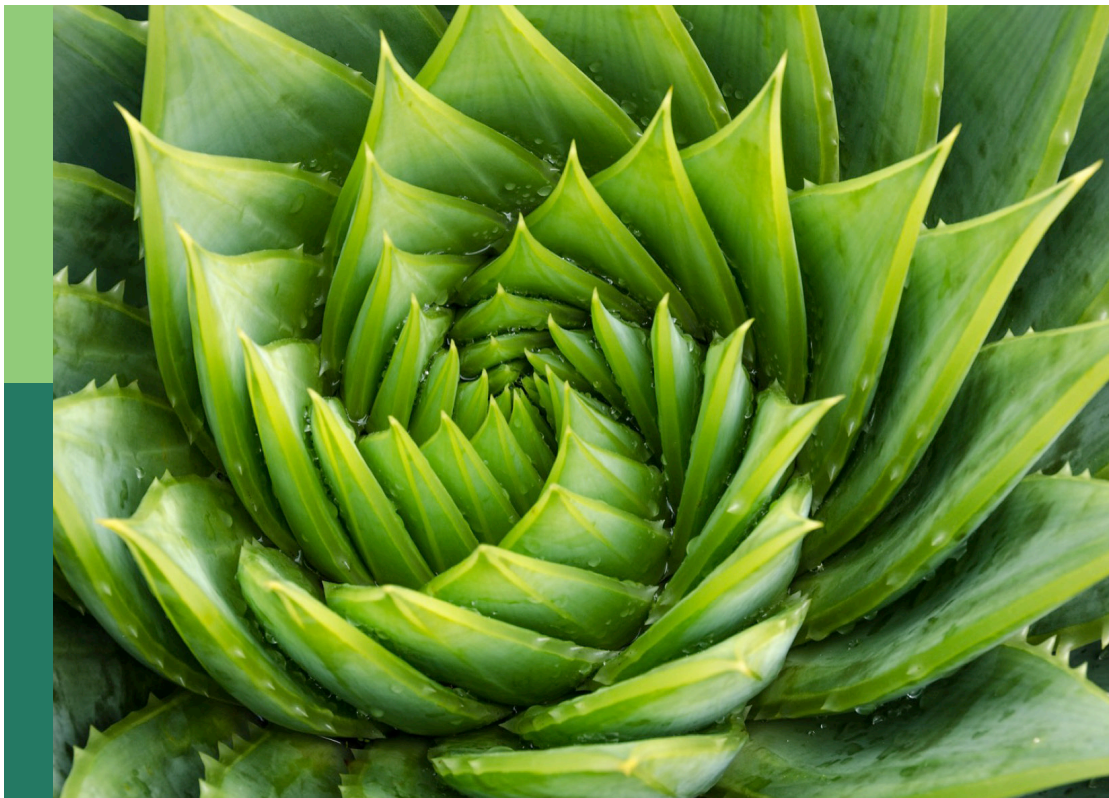
frontiers | Research Topics

Multi-omics strategies to analyze complex agronomic traits in plants

Edited by
Lin Chen and Guo-Fei Tan

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3789-3
DOI 10.3389/978-2-8325-3789-3

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 strategies to analyze complex agronomic traits in plants

Topic editors

Lin Chen — Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, China

Guo-Fei Tan — Guizhou Academy of Agricultural Sciences (CAAS), China

Citation

Chen, L., Tan, G.-F., eds. (2023). *Multi-omics strategies to analyze complex agronomic traits in plants*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3789-3

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 31 July 2023
DOI 10.3389/fpls.2023.1256629

OPEN ACCESS

EDITED AND REVIEWED BY
James Lloyd,
Stellenbosch University, South Africa

*CORRESPONDENCE
Lin Chen
chenlin@caas.cn

RECEIVED 11 July 2023
ACCEPTED 17 July 2023
PUBLISHED 31 July 2023

CITATION
Chen L and Tan G-F (2023) Editorial: Multi-omics strategies to analyze complex agronomic traits in plants. *Front. Plant Sci.* 14:1256629. doi: 10.3389/fpls.2023.1256629

Editorial: Multi-omics strategies to analyze complex agronomic traits in plants

Lin Chen[1]* and Guo-Fei Tan[2]

[1]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Institute of Horticulture, Guizhou Academy of Agricultural Sciences (GAAS), Guiyang, Guizhou, China

KEYWORDS

omics, genetic selection, complex traits, genetic improvement, crop breeding

Editorial on the Research Topic
Multi-omics strategies to analyze complex agronomic traits in plants

Complex traits in plants are usually controlled by multiple genes, and the genetic analysis of complex traits has always been a difficult problem in plant research. In the past decade, with the continuous development of omics technology, more and more omics technologies (genomics, transcriptome, metabolomics, proteomics, and others) have been applied to the genetic analysis of plant complex traits. Therefore, how to integrate these massive omics data has become a new research hotspot.

In the past year, a total of 21 manuscripts focusing on this topic were received, and 9 original research articles were selected for publication after rigorous peer review. These 9 articles reported research on seven different plants, namely, maize, celery, lettuce, apple, rapeseed, sweet cherry, and bamboo shoot. Among these articles, there are seven research articles focusing on the agronomic traits related to plant development, one research article focusing on heat stress related to plant response to abiotic stress, and one research article focusing on plant response to biotic stress. To the benefit of potential readers, the key points of the nine articles in this Research Topic are highlighted as follows:

The fleshy stem is the main product organ of stem lettuce. However, the molecular mechanism of the fleshy stem expansion in lettuce is still unknown. Huang et al. identified the key genes and metabolites during the fleshy stem expansion process by a comparative analysis of the transcriptome and metabolome. A total of 9,383 differentially expressed genes and 822 metabolites were found during this process. Moreover, these differentially expressed genes and metabolites were significantly enriched in the sugar synthesis, glycolysis, and plant hormone processes. These results provide important candidate genes and metabolites for further research into the molecular mechanisms of fleshy stem expansion.

The rapid decline in the quality of fresh bamboo shoots after receipt is one of the main reasons for the difficulty of their long-distance transportation and long-term preservation. Li, Z. et al. conducted morphological, physiological, transcriptomic, and microRNA-sequencing analyses to investigate the postharvest characteristics of fresh bamboo shoots. The differentially expressed genes and miRNAs were significantly enriched in structural

polysaccharide metabolism, starch and sucrose metabolism, and glycolysis pathways. Furthermore, a co-expression network of carbohydrate metabolism was constructed in this article. These results suggest that carbohydrate metabolism mainly affects the quality of postharvest bamboo shoots.

Gibberellin (GA) is an important hormone in the plant development process. Exogenous use of GA3 can improve the sweet cherry yield but decrease the fruit quality. Chen et al. identified the key genes and metabolites by integrating transcriptome and metabolome analysis in response to exogenous GA3 application in sweet cherry. They results showed that the content of ABA, JA, and IAA were significantly increased after the application of exogenous use of GA3. Moreover, they also found the WRKY transcription factors were more sensitive to the application of exogenous GA3. Their results provide a new insight into improving the yield and quality of sweet cherry.

Flowering time is one of the most important traits of maize and other plants for breeding. Wu et al. identified 82 significant SNPs and 117 candidate genes related to the maize flowering time by the GWAS method with an association panel consisting of 226 maize inbred lines. In addition, a total of 21 QTLs and 65 candidate genes were found to be related to maize flowering time by linkage analysis with an F2:3 population. By integrating GWAS, linkage analysis, and transcriptome analysis, 25 important candidate genes for maize flowering time were identified in this study. Their finding is novel and provides new gene resources for other researchers to study the genetic bases of maize flowering time.

Seminal roots play a key role for maize seedlings to obtain water and other nutrients. In this topic, Wang et al. identified three important candidate genes (Zm00001d021572, Zm00001d021579, and Zm00001d021861) by combining QTL analysis, which was conducted by using a maize–teosinte BC2F6 population, with transcriptome analysis. Moreover, they also found that Zm00001d021572 was selected during maize domestication. These results remind other researchers of the existence of excellent genetic resources in the wild germplasm resources of crops.

Plant height is the key element of the ideal plant architecture. The association mapping panel with 230 rapeseed accessions was used by Zhao et al. to understand the genetic basis of plant height in rapeseed. By integrating GWAS, haplotype analysis, and expression analysis, *BnaA01g09530D*, which encodes BRASSINOSTEROID-INSENSITIVE 2 and belongs to the GLYCOGEN SYNTHASE KINASE 3 (GSK3) family, was found to be significantly related to plant height in rapeseed. The results of this article provide an important candidate gene for the genetic improvement of rapeseed in line with the ideal plant architecture.

Besides plant height, the first branch height in rapeseed, which is another key element of the ideal plant architecture, has an important effect on yield and mechanized harvesting. Dong et al. found that 19 QTLs were related to the first branch height by linkage analysis, and 26 significant SNPs were found by GWAS. Among these hotspot regions, one major QTL located on Chr.A02 was found by GWAS and linkage analysis. In this region, it was confirmed by expression analysis and transgene analysis that *BnaA02g13010D*, which encodes a TCP transcription factor, can regulate the first branch height in rapeseed. These results provide valuable information for the genetic improvement of the first branch height in rapeseed.

Celery is a valuable edible crop and cannot endure high temperatures. Heat shock transcription factors (HSFs) are the most important transcription factors in plant response to heat stress and other abiotic stresses. Li, M. et al. identified 29 *AgHSFs* in celery and classified these genes into three classes. Based on the transcriptome data, the expression patterns of these genes were analyzed under heat stress. Among these genes, *AgHSFa6-1* acts as a positive regulator to improve the celery thermotolerance. These results provide a valuable gene resource for the improvement of celery to heat stress.

Alternaria blotch disease is one of the major fungal diseases in apple, which is caused by the *Alternaria alternata apple pathotype* (AAAP). To understand the molecular mechanism to AAAP in different varieties in apple, Liang et al. conducted allele-specific expression analysis and comparative transcriptomic analysis before and after AAAP inoculation. Based on this analysis, *MdFLS2* and its allele were found to be significantly related to resistance to AAAP in apple.

Based on the abovementioned articles published in this Research Topic, it can be seen that multi-omics techniques have greatly improved the efficiency of the genetic analysis of complex agronomic traits in plants, especially for some non-model plants. We hope that these research results can provide new ideas and assistance for the analysis of complex agronomic traits in plants. Finally, we are very grateful for the efforts of the journal editors, peer reviewers, and relevant authors. Without their efforts, this Research Topic would not appear in front of readers. We hope that readers can obtain valuable information from this topic to help them achieve success in the future.

Author contributions

LC: Formal Analysis, Funding acquisition, Validation, Writing – original draft, Writing – review & editing. G-FT: Validation, Writing – original draft, Writing – review & editing.

Funding

This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (No. 2022-YWF-ZYSQ-04), the

Ordos Science and Technology Plan (2022EEDSKJZDZX011), the major demonstration project of "The Open Competition" for Seed Industry Science and Technology Innovation in Inner Mongolia (No. 2022JBGS0016), the Guizhou Provincial Department of Science and Technology "Construction and Utilization of Horticultural Platform of Plant GenBank Creation" (No. Qiankehe Fuqi[2022] 005), and the Guizhou Modern Agriculture Research System (GZMARS)—Plateau Characteristic Vegetable Industry.

Acknowledgments

We greatly appreciate the efforts of the journal editors, peer reviewers, and authors.

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fpls.2022.1021161

OPEN ACCESS

EDITED BY
Guo-Fei Tan,
Guizhou Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Yogesh Kumar Ahlawat,
University of Florida, United States
Feng Que,
Nanjing Forestry University, China
Jian Zhang,
University of British Columbia-Okanagan Campus, Canada
Zhjun Zhang,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Zhimin Gao
gaozhimin@icbr.ac.cn

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

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

RECEIVED 17 August 2022
ACCEPTED 31 August 2022
PUBLISHED 21 September 2022

CITATION
Li Z, Xu X, Yang K, Zhu C, Liu Y and Gao Z (2022) Multifaceted analyses reveal carbohydrate metabolism mainly affecting the quality of postharvest bamboo shoots. *Front. Plant Sci.* 13:1021161. doi: 10.3389/fpls.2022.1021161

Multifaceted analyses reveal carbohydrate metabolism mainly affecting the quality of postharvest bamboo shoots

Zhen Li[1,2†], Xiurong Xu[1,2,3†], Kebin Yang[1,2], Chenglei Zhu[1,2], Yan Liu[1,2] and Zhimin Gao[1,2*]

[1]International Centre for Bamboo and Rattan, Institute of Gene Science and Industrialization for Bamboo and Rattan Resources, Beijing, China, [2]Key Laboratory of National Forestry and Grassland Administration/Beijing for Bamboo and Rattan Science and Technology, Beijing, China, [3]Zhejiang Academy of Forestry, Hangzhou, China

Bamboo shoot is one of nutritious vegetables in China. However, the edible quality of fresh bamboo shoots deteriorates easily after harvest. Here, morphological, physiological, transcriptomic and microRNA sequencing analyses were conducted to investigate the postharvest characteristics of moso bamboo (*Phyllostachys edulis*) shoots. Rapid decreases of soluble sugars, structural polysaccharides and hydrolyzed tannins, and increases of lignin and condensed tannins were observed in the postharvest bamboo shoots. Differentially expressed genes (DEGs) and miRNAs with opposite trends were mainly enriched in structural polysaccharide metabolism, starch and sucrose metabolism and glycolysis pathways, which were consistent with the changes of carbohydrates. A co-expression network of carbohydrate metabolism was constructed, which was verified by qPCR and yeast one-hybrid (Y1H) assay. Furthermore, the function of one hub glycosyltransferase gene was validated in *Arabidopsis*, which confirmed that it was involved in xylan biosynthesis. These results are of great significance for revealing the carbohydrate metabolism mechanisms of postharvest bamboo shoots and provide a potential candidate gene for molecular breeding related to xylan in the future.

KEYWORDS

***Phyllostachys edulis* shoot, postharvest physiology, carbohydrate metabolism, co-expression network, xylan biosynthesis**

Introduction

Bamboo shoots, a kind of immature and expanding culms generated from the nodes of bamboo pseudo-rhizomes, are usually considered as a highly desirable vegetable because of their sweet taste, crisp and crunchy texture and excellent nutrition (Singhal et al., 2013). Carbohydrates in bamboo shoots are abundant and available in multiple

forms, with the largest proportion of polysaccharides including starch as an energy source, and cellulose, and hemicellulose as dietary fibers. The dietary fiber of bamboo shoots accounts for 50% of carbohydrates (Wang et al., 2020), playing a critical role in hypolipidemic and anti-obesity activities as well as the taste (Tanabe et al., 2017). However, due to the perishability of bamboo shoots, the shelf life and market value of fresh bamboo shoots have been restricted. The soluble sugar content of bamboo shoots decreased significantly at the beginning of storage, e.g., the starch content of *Fargesia yunnanensis* shoots decreased significantly after 3 days of storage (Yu et al., 2021). Meanwhile, the changes in phenolic content of bamboo shoots led to browning and increased lignin content, along with the phenylalanine ammonia-lyase, and cinnamyl alcohol dehydrogenase activities increased rapidly (Li X. et al., 2019).

Therefore, several studies have focused on storage and food processing technology by both physical and chemical methods. In terms of physical methods, such as cooling, UV-C, gamma radiation and hypobaric atmosphere are often applied to extend the shelf life of bamboo shoots, which prolong the edible time of bamboo shoots by reducing the accumulation of lignin and chlorophyll (Wang et al., 2019; Wen et al., 2019). As for chemical methods, hormone treatment is commonly used to keep vegetable fresh. Exogenous melatonin and diphenyliodonium iodide treatment effectively delay the lignification of bamboo shoot and reduce the rate of yellowing during cold storage (Li C. et al., 2019; Li D. et al., 2019), while brassinolide treatment alleviates chilling injury of bamboo shoots by enhanced enzyme activities related to energy and proline metabolism (Liu et al., 2016). However, even low-temperature (LT) storage cannot stop the lignification process. Moreover, many techniques cannot be applied for ordinary households, especially bamboo shoot farmers living in remote mountainous areas who have difficulty using refrigerators. In view of the mechanisms underlying these physical and chemical methods, it suggests that the regulation of carbohydrate metabolism is important for affecting bamboo shoot edible quality and shelf life. Although similar biological processes occur in other harvested vegetables and fruit, and the effect of carbohydrate metabolism on quality is relatively clear, those in bamboo shoots are still unclear.

Structural polysaccharide metabolism, starch, and sucrose metabolism and glycolysis are the main carbohydrate metabolisms. Structural polysaccharides affected the juiciness and adhesion of cells, and promoted fruit soften or granulation. Propectin and cellulose showed significant increase in postharvest granulated orange, and the genes encoding enzymes responsible for pectin and cellulose, e.g., galacturonosyltransferase (GAUT) and cellulose synthase complex were up-regulated, while that of the gene encoding polygalacturonase responsible for pectin degradation showed significant decrease (Yao et al., 2020). In postharvest kiwifruits and persimmon fruit, the protopectin and cellulose content decreased while water-soluble pectin content increased progressively, accompanied by the increased activities of UDP-xylose 4-epimerase, UDP-Glc dehydrogenase, pectinesterase and cellulase (Tian et al., 2021). Genes encoding sucrose phosphate synthase (SPS), invertase (INV), and α-amylase (AMY) showed significantly differential expression in Powell (*Citrus sinensis*), fresh lotus seeds and kiwifruits during storage (Sun et al., 2021; Tian et al., 2021). Those genes encoding pyruvate kinase (PK) and phosphofructokinase related to glycolysis are highly variable during storage in many kinds of fruit and vegetables, such as sugarbeet roots, lotus seeds, and kiwifruits (Tian et al., 2021). Within sucrose and starch cycle, starch synthesis and degradation are simultaneously, during the preharvest period, taking place with net breakdown thereafter (Schouten et al., 2016). In addition, transcription factors (TFs) like bHLH, MYB, ERF, and LBD play important roles in regulating postharvest physiological changes (Wu et al., 2021).

Transcriptome sequencing is a powerful tool to rapidly obtain information on the expressed fraction of a genome, including the gene expression patterns, and the small ribonucleic acid (RNAs) of the targeted genes. The molecular mechanism of some physiological changes in postharvest bamboo shoots has been preliminarily analyzed, and the pathways of secondary cell walls and hormone biosynthesis are identified (Zhang et al., 2018). It was noticed that genes involved in lignin and hemicellulose biosynthesis were up-regulated and down-regulated, respectively (Li et al., 2018; Zhang et al., 2020). In addition, NAC, bHLH, bZIP, MYB, and WRKY were identified as critical TFs for shoot postharvest senescence, in which PheNAP2 and PheNAP3 were found to promote leaf senescence in *Arabidopsis* (Li X. et al., 2019). The expression of jasmonic acid (JA) biosynthesis related genes was highly consistent with that of lignin precursor biosynthesis genes under low temperature (LT), indicating a LT-lignification or LT-JA-lignification regulatory pathway existed in Lei bamboo shoots (Hou et al., 2022). However, the molecular mechanism affecting edible quality associated with carbohydrates in postharvest shoots is still unclear. In this study, the characteristics related to the main carbohydrate metabolism of postharvest moso bamboo shoots were revealed by integrated analyses of anatomical observation, chemical component and enzyme activity determination, transcriptomic and microRNA sequencing, which provided new insight into the modulation affecting the quality of postharvest moso bamboo shoots.

Materials and methods

Materials, treatment and ribonucleic acid extraction

The moso bamboo (*Phyllostachys edulis*) shoots of 15 cm ∼ 20 cm in length and 5 cm ∼ 8 cm in basal diameter were

harvested from a wild bamboo forest (28°45′58″N, 115°45′39″E, 399.0 m above sea level) in Nanchang, Jiangxi province, China, in December 2019. The shoots were stored under room temperature conditions (22 ± 2°C, RH = 75%) wrapped in paper bags. Middle part of the shoots in **Figure 1A** were collected at different storage time (S1, S2, S3, and S4 represented bamboo shoots stored for 0, 3, 6, and 12 d), one part of samples was immediately frozen in liquid nitrogen and then stored at −80°C, another part of samples was fixed with FAA and stored at 4°C until use, respectively. Three biological replicates were set for each time point. The total RNA and small RNA from 12 samples was isolated using the TRIzol and miRcute plant miRNA extraction and isolation reagent (item numbers DP504, TIANGEN Biotech Co., Ltd., Beijing), respectively.

Histological staining, determination of compound content and enzyme activity

The shoot samples were stained with 5% (v/v) of toluidine blue (TBO), as described in the reference (Yang et al., 2021). The content of soluble sugar, cellulose, hemicellulose, lignin, hydrolyzed tannins, and condensed tannins, and the enzyme activity of peroxidase (POD) and L-phenylalanine ammonia lyase (PAL) were determined with detection kits (item numbers G0501, G0519, G0527, G0708, G0119, G0120, G0114, and G0107) of Grace Biotechnology Co., Ltd., JiangSu. The content of xylan was determined by immunofluorescence with detection kits (item numbers S0169O) of KETE Biotechnology Co., Ltd., Jiangsu. Three replicates were performed for each determination.

Transcriptome and small ribonucleic acid sequencing

A total 12 transcriptome and 12 small RNA libraries were generated using NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) and NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, USA) following manufacturer's recommendations, respectively. Illumina platform was used to sequence 12 high quality RNA libraries at Biomarker Technologies Corporation (Beijing, China). Low quality reads, reads with adaptors and reads with unknown bases were removed from the data sets to obtain the clean reads, which were then mapped to the reference genome of moso bamboo, followed by unigene expression level calculation with fragments per kilobase of transcript per million fragments mapped (FPKM) in each sample.

Differential expression analysis was performed using DESeq2 (Love et al., 2014), and genes with Fold Change (FC) $\geq$ 2 and False Discovery Rate (FDR) $<$ 0.01 were assigned as differentially expressed (Tarazona et al., 2011). The putative functions of the differentially expressed genes (DEGs) and target genes were determined according to Non-redundant protein (NR) database. The cluster classification of DEGs based on expression of their transcriptome profiles were performed using OmicShare Tools[1] with default parameters. The Gene Ontology (GO) enrichment analysis was carried out using R-based GOseq package to identify the over-represented functional categories, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was implemented by KOBAS 3.0 (Xie et al., 2011). For co-expression network analysis, the weighted gene co-expression network analysis (WGCNA) package was used. Correlation analysis of gene expression was performed using programs in the BMKCloud.[2] Based on the results of WGCNA, the genes in the module with the highest carbohydrate correlation were selected to construct the co-expression network. DEGs related to structural polysaccharide metabolism, starch, and sucrose metabolism and glycolysis were screened from this module by gene function annotation. The upstream sequences of these DEGs were further analyzed to identify potential regulatory TFs in turquoise module combining with biological information in PlantTFDB databases.[3] The small RNA sequencing data from this study and psRNATarget database[4] were used to identify potential regulatory miRNAs targeting these DEGs.

For Small RNA data analysis, clean reads of small RNA were obtained by removing reads containing adapter and ploy-N besides low-quality reads using Trimmomatic v0.35. After screening against reads shorter than 18 nt and longer than 35 nt, rRNA, tRNA, snoRNA, and repeat sequences, the redundant reads were processed into unique sequences with associated read counts for miRNA prediction by comparing with genome and known miRNAs from miRBase using Bowtie alignment tool v1.1.2 with two mismatches. Randfold tools software was used for novel miRNA secondary structure prediction. Subsequently, the mRNA targets of the identified miRNAs were predicted using psRNATarget server with default parameters. The annotation of these target genes refers to the method of DEG annotation in 2.3. The miRNAs expression determination and differentially expressed miRNAs (DEMs) analysis refers to the method of transcriptome analysis.

1 https://www.omicshare.com/tools/Home/Soft/trend

2 www.biocloud.net

3 http://planttfdb.gao-lab.org/

4 https://www.zhaolab.org/psRNATarget/

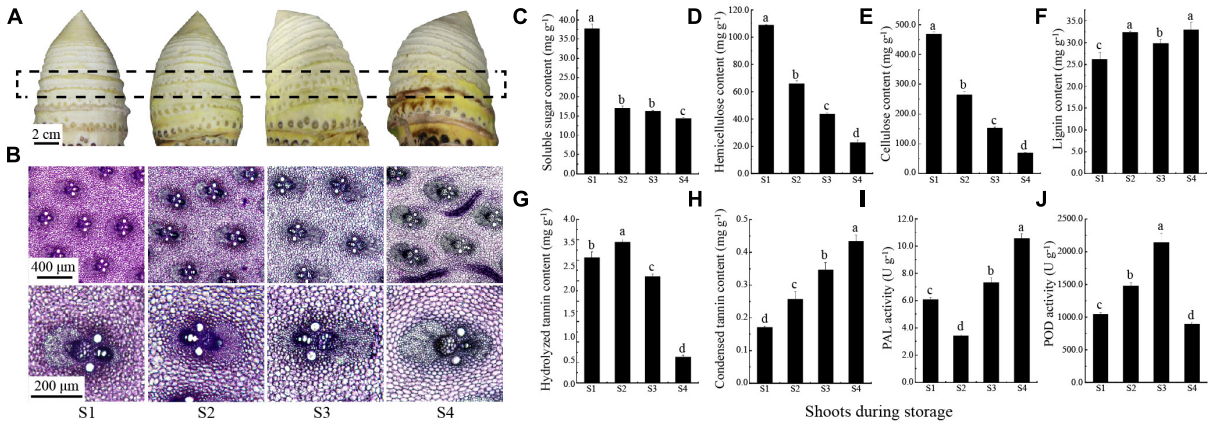


FIGURE 1
Changes of postharvest moso bamboo shoots. **(A)** Morphological changes of shoots after postharvest 0, 3, 6, and 12 d (S1: 0 d; S2: 3 d; S3: 6 d; S4: 12 d). **(B)** Staining intensity of vascular bundle transverse section of the boxed area in **(A)**. **(C–H)** The content of soluble sugar **(C)**, hemicellulose **(D)**, cellulose **(E)**, lignin **(F)**, hydrolyzed tannins **(G)** and condensed tannins **(H)**, the activities of PAL **(I)** and POD **(J)**. Lowercase letters a, b, c, and d indicate significant differences ($P < 0.05$).

Validation of differentially expressed genes and differentially expressed miRNAs

Primer Premier 5.0 software was used to design the specific primers based on the candidate genes and miRNA sequences (**Supplementary Table 1A**). The qPCR reactions were performed with the SYBR Green Master Mix in a qTOWER 2.2 system (Analytik Jena, Germany) with four technical replicates, using *PeTIP41* (Fan et al., 2013) and *U6* snRNA (Yang et al., 2021) as the endogenous controls for the genes and miRNAs, respectively. The qPCR conditions were as follows: 95°C for 6 min, followed by 40 cycles at 95°C for 15 s and 60°C for 15 s. The expression levels were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Yeast one-hybrid assay

The 5′ UTR and promoter sequences of PH02Gene04386 (a glycosyltransferase 43 gene) were segmented into four sequences with different lengths, including 5′ UTR sequence of 600 and 1,119 bp upstream of cDNA namely 5′ UTR-600 and 5′ UTR-1119, promoter sequence of 600 and 1,200 bp upstream of mRNA namely pro-600 and pro-1200, respectively. The coding sequence (CDS) of two *PeMYB*s (PH02Gene12213 and PH02Gene31429) were recombined into the pGADT7-Rec2 vector. Six SMRE element (SMRE2 ∼ SMRE7) regions and four 5′ UTR/promoter sequences of PH02Gene04386 were inserted into the pHIS2 vector, respectively. Two recombinant vectors were co-transformed into the Y187 strain. The transformants were screened on Leu-, Trp-, and His-deficient medium supplemented with 20 mM of 3-amino-1,2,4-triazole for 3 d, using cotransformants containing p53:His2 with pGADT7:53 and p53:His2 with pGADT7:PeMYBs as the positive control and negative control, respectively. The primers used in yeast one-hybrid (Y1H) assay were listed in **Supplementary Table 1B**.

Functional validation of PH02Gene04386

The CDS of PH02Gene04386 was isolated and cloned into the expression vector of pSuper1300. The transgenic plants overexpressing (OE) PH02Gene04386 were obtained through ectopic expression in *Arabidopsis* (Col-0) plants using the floral dip method, which were further verified by PCR and qPCR. Root length of transgenic and wild-type (WT) *Arabidopsis* seedlings was counted after 7 days growing on Murashige and Skoog medium with 50 mM xylose as the carbon source. The main inflorescence stem (1 ∼ 2 cm aboveground) of both transgenic and WT *Arabidopsis* plants of 6 weeks were used for histological staining with 5% (v/v) of TBO, xylan immunofluorescence localization with LM10 (Keppler and Showalter, 2010) and xylan content analysis, respectively. The fluorescence signals of FITC were recorded with a Zeiss LSM 980 confocal microscope. Images were taken with an Olympus CX31 light microscope (Tokyo, Japan).

Statistical analysis

A standard *t*-test is used to determine statistical significance with a 95% confidence interval. In all figures, data are represented as the mean ± standard error (SE).

Results

Changes of morphology, physiology, and enzyme activities in moso bamboo shoots

During storage at room temperature, the surface of bamboo shoots changed significantly, and gradually browned with time (**Figure 1A**). Meanwhile, the degree of staining intensity of lignin was deepening and the polysaccharides was lightening in vascular bundle (**Figure 1B**), which indicated a potential change in carbohydrate and phenols content. Since the carbohydrates and phenols are considered to be important factors affecting postharvest edible quality, the content of these substances such as soluble sugar, hemicellulose, cellulose, lignin, and procyanidins in postharvest moso bamboo shoots are measured. The results showed that the continuous decreases were observed in soluble sugar from 37.75 mg g^{-1} (S1) to 14.36 mg g^{-1} (S4), hemicellulose from 108.92 mg g^{-1} (S1) to 22.54 mg g^{-1} (S4) and cellulose from 468.68 mg g^{-1} (S1) to 69.99 mg g^{-1} (S4) (**Figures 1C–E**). In contrast, the lignin content showed an increase from 26.23 mg g^{-1} (S1) to 32.91 mg g^{-1} (S4) (**Figure 1F**). The content of hydrolyzed tannin increased slightly from 3.07 mg g^{-1} (S1) to 3.45 mg g^{-1} (S2) and then significantly decreased to 0.63 mg g^{-1} (S4) (**Figure 1G**). Meanwhile, the content of condensed tannin (procyanidin) increased continuously, and doubled that of S1 (0.17 mg g^{-1}) at the end (S4, 0.43 mg g^{-1}) (**Figure 1H**). The activity of PAL was first decreased from 6.08 U g^{-1} (S1) to 3.41 U g^{-1} (S2), and then increased significantly to 10.55 U g^{-1} (S4) (**Figure 1I**). The activity of POD peaked at 2140.90 U g^{-1} (S3), but decreased significantly to 897.27 U g^{-1} at the end (S4) (**Figure 1J**). These results were in agreement with our current understanding of taste and flavor decline in postharvest moso bamboo shoots.

Analysis of transcriptome sequencing and differentially expressed genes in moso bamboo shoots

To further understand the molecular mechanisms of above changes, 12 transcriptome libraries were constructed, with 325,040,670 clean reads generated. The clean data were mapped to the reference genome of moso bamboo with the mapping ratio varying from 96.16 to 96.66% (**Supplementary Table 2A**). Meanwhile, 7,458 unigenes were identified, among which 5,061 were functionally annotated (**Supplementary Table 2B**). A total 12,553 DEGs were identified by comparing the samples (**Supplementary Figures 1A,C**). Annotation analysis showed that there were 1,046 DEGs encoding TFs belonging to 44 families, in which MYB (112), bHLH (91), AP2/ERF (80), WRKY (79), and NAC (70) were the most over-represented TF families (**Supplementary Figure 1B**).

We performed expression analysis of the DEGs, and grouped genes according to the similar expression of their transcriptome profiles. A total of six clusters were identified (**Figures 2A–C**), and most of the DEGs belonged to two kinds of clusters: down-regulated (cluster I, 3,596 DEGs) and up-regulated (cluster II, 2,723 DEGs). In cluster I, GO and KEGG analyses showed that the DEGs were enriched in carbohydrate metabolism (**Figures 2D,E**). Different from cluster I, DEGs in cluster II were related to secondary metabolism (**Figures 2F,G**). The results of GO and KEGG analyses in cluster III showed that DEGs were enriched in a variety of pathways, including photosynthesis, carbohydrate and secondary metabolism (**Supplementary Figures 2A,B**). Considering the functional terms in combination with morphological characteristics, we proposed that the decrease of soluble sugar and structural polysaccharides along with the increase of lignin and procyanidins may be due to the expression changes of DEGs involved in carbohydrate metabolism and secondary metabolism. Among them, carbohydrate-related DEGs were the most abundant, which was consistent with the significant changes in carbohydrate content.

Expression patterns of carbohydrate metabolism related differentially expressed genes in moso bamboo shoots

To investigate the changes in carbohydrate metabolism, the expression patterns of DEGs in three key metabolic pathways were further analyzed. The result showed that DEGs were remarkably enriched in the pathway related to nucleotide sugar metabolism. Most of the DEGs were down-regulated, especially those involved in pectin and xylan biosynthesis, while the DEGs involved in arabinose biosynthesis were up-regulated during storage (**Figure 3**). Three DEGs encoding UDP-apiose/xylose synthase (AXS) and four DEGs encoding UDP-glucose 6-dehydrogenase (UGDH, EC:1.1.1.22) enriched in nucleotide sugar metabolism pathway were identified, with all three AXS DEGs and three of four UGDH DEGs down-regulated significantly. In addition, one DEG encoding UDP-arabinose 4-epimerase (UXE, EC:5.1.3.5) involved in arabinose biosynthesis was identified and its expression increased with storage time.

Further, the DEGs related to nucleotide sugar polymerization were identified, such as those encoding glycosyltransferase family 43 (GT43, EC:2.4.2.-) and glycosyltransferase family 47 (GT47, EC:2.4.2.-) associated with xylan biosynthesis. Most of nine GT43 DEGs were down-regulated significantly during storage, while the expression trends were significantly different among GT47 DEGs, in which the expression of PH02Gene05811 was significantly down-regulated and PH02Gene12063 was first up-regulated

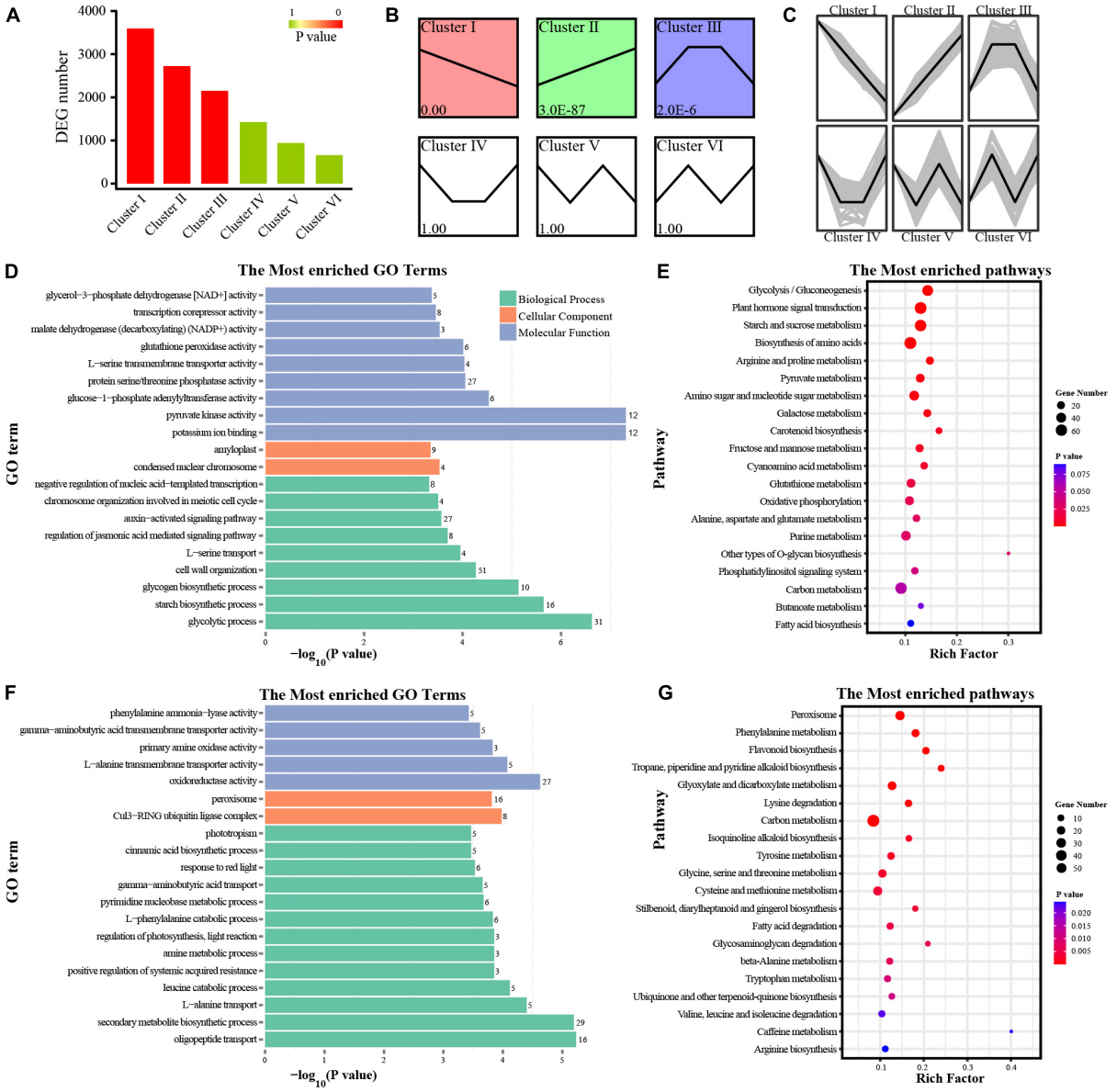


FIGURE 2
Function enrichment analysis of DEGs. **(A)** DEG number in different clusters. **(B)** Different clusters based on *P*-values. **(C)** DEGs with similar expression trends in a specific cluster. **(D–G)** GO and KEGG analyses of DEGs in cluster I **(D,E)**, cluster II **(F,G)**. Rich Factor represents the value of enrichment factor which is the quotient of foreground value (the number of DEGs), and the larger the value, the more significant enrichment; Coloring indicates *P*-value with higher in red and lower in blue, and the lower *P*-value, the more significantly enriched. Point size indicates the number of DEGs.

and then down-regulated. In addition, the DEGs involved in pectin synthesis were also identified. For example, DEGs encoding GAUT (EC:2.4.1.43) changed significantly during storage, with most of them showing a significant decrease at S2 immediately.

In starch and sucrose metabolic pathways, biosynthesis-related DEGs were down-regulated significantly at S2, e.g., three DEGs encoding starch synthase (glgA, EC:2.4.1.21) and four DEGs encoding SPS (EC:2.4.1.14), while hydrolysis-related DEGs were relatively up-regulated at S2 and S3 (**Supplementary Figure 3**), e.g., eight DEGs encoding amylase (AMY, EC:3.2.1.1, BMY, EC:3.2.1.2) and six DEGs encoding fructosidase (INV, EC:3.2.1.26). As for Glycolysis, the DEGs encoding enzymes involved in irreversible reactions were down-regulated rapidly at S2, e.g., hexokinase (HK, EC:2.7.1.1) and PK (EC:2.7.1.40) DEGs. However, there were some exceptions, e.g., three DEGs encoding fructose-bisphosphate aldolase

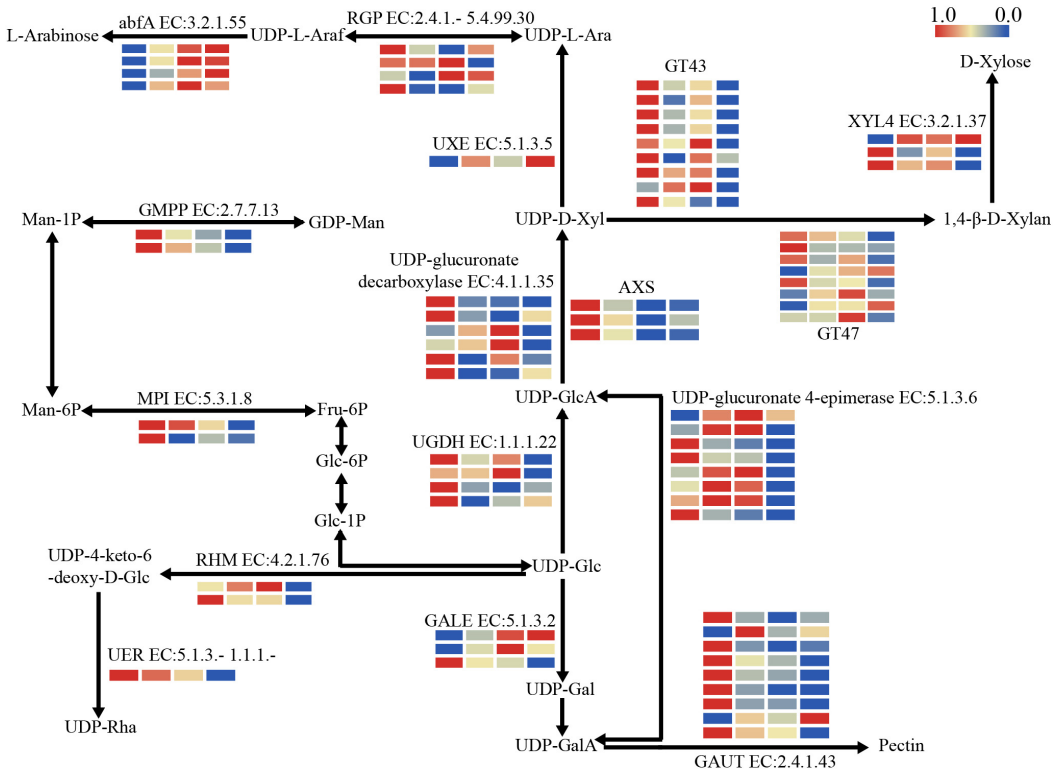


FIGURE 3
The expression patterns of DEGs involved in structural polysaccharide metabolism. abfA, alpha-L-arabinofuranosidase; AXS, UDP-apiose/xylose synthase; GALE, UDP-glucose 4-epimerase; GAUT, α-1,4-galacturonosyltransferase; GT43, glycosyltransferase 43; GT47, glycosyltransferase 47; RGP, eversibly glycosylated polypeptide; RHM, UDP-glucose 4,6-dehydratase; UER, 3,5-epimerase/4-reductase; UGDH, UDP-glucose 6-dehydrogenase; UXE, UDP-arabinose 4-epimerase; XYL4, xylan 1,4-beta-xylosidase. The color bar indicates $\log_2$-based fragments per kilobase per million (FPKM) with higher in red and lower in blue.

(ALDO, EC:4.1.2.13), which were first up-regulated and reached a peak at S3, and then down-regulated (**Supplementary Figure 4**).

Expression profiles and targets of small ribonucleic acids in moso bamboo shoots

A total 363.6 million of 418.9 million reads were generated after quality control (**Supplementary Table 3A**). The length distribution of the unique small RNA reads indicated that 24 nt were the most abundant (36.6%) (**Supplementary Table 3B**). Further, 47 unique conserved miRNAs and 415 novel miRNAs were identified, with lengths varied from 19 to 25 nt (**Supplementary Figures 5A,B**), which agreed with many other RNA-Seq experiments (Chen et al., 2020; Li et al., 2021). Of these miRNAs, 451 pre-miRNAs corresponding to 462 mature miRNAs were identified, which were grouped into 93 families (**Supplementary Figure 5C** and **Supplementary Table 3C**). The majority of mature miRNAs were characterized by a 5′-uridine residue and the proportion of A and U bases exceed 50%, which was consistent with the characteristics of plant miRNAs attributed to DCL1 cleavage and AGO1 association (Rajagopalan et al., 2006).

The majority of the identified miRNAs were expressed in more than one sample. In total, 182 miRNAs showed differential expression patterns across different combinations, including 26 known miRNAs and 156 novel miRNAs. There were 34 miRNAs with significant expression changes in moso bamboo shoots at S2, S3, and S4 compared with those at S1 (**Figures 4A,B**). We merged the DEMs with similar expression profiles to perform cluster analysis. A total of 84 DEMs were clustered into four major clusters with *P*-value less than 0.05 (**Figure 4C**). The DEMs in cluster I were up-regulated and those in cluster III were down-regulated, while those in cluster II were up-regulated firstly and then down-regulated during storage (**Figure 4D**). The opposite expression trends between DEMs and DEGs suggested that they might have a potential negative regulatory relationship.

In order to estimate the biological function of these miRNAs, their target genes were identified. A total of 13,748 miRNA-target pairs were identified by TargetFinder and psRNATarget, consisting of 424 miRNAs and 6,024 genes. There were 998

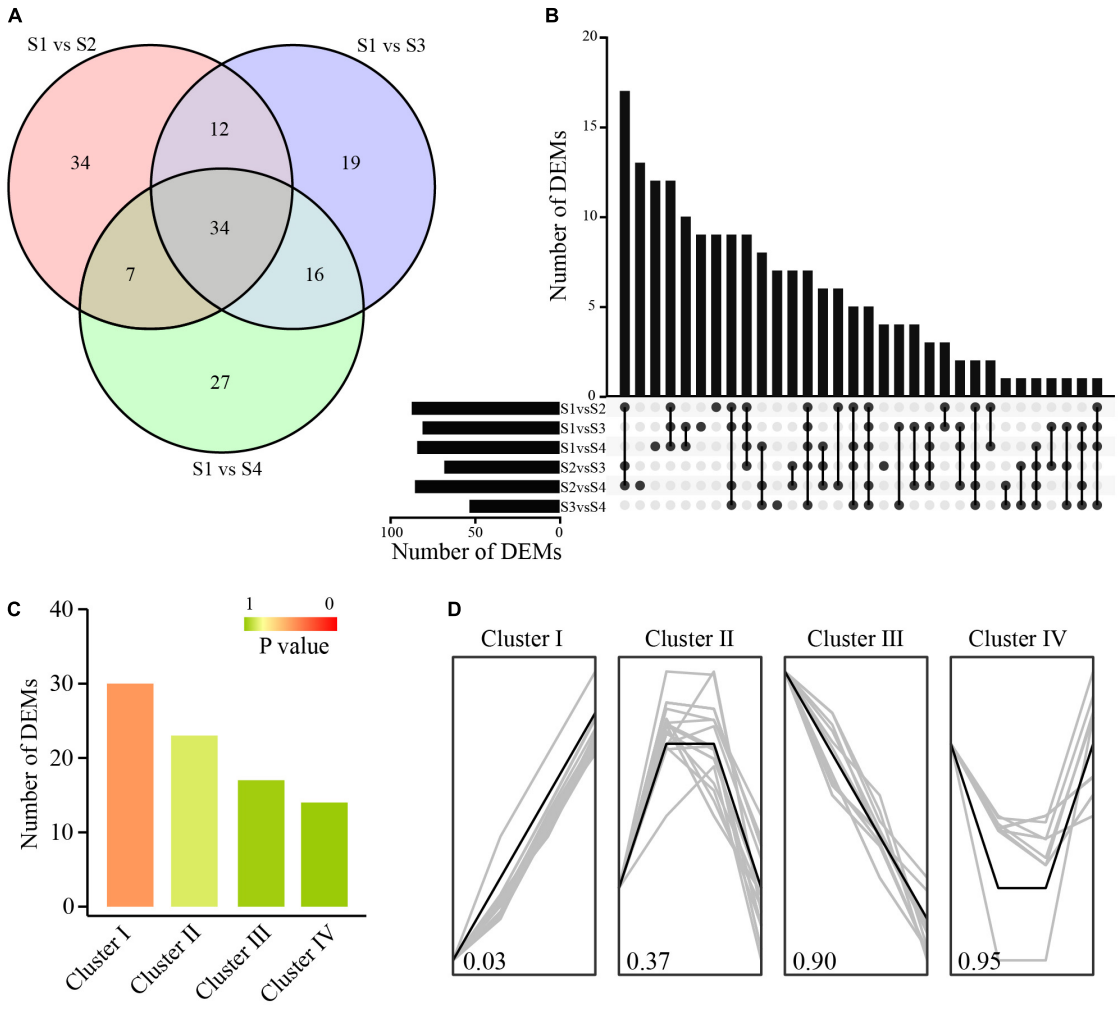


FIGURE 4
Identification and cluster analysis of DEMs. **(A)** Venn diagram of DEMs. **(B)** Number of DEMs between different comparison groups. **(C)** DEM number in different clusters. **(D)** DEMs with similar expression trends in a specific cluster.

target genes annotated to 125 KEGG pathways. Carbohydrate metabolism and terpenoid and polyketide metabolism were dominant pathways (**Supplementary Figure 6**), which indicated that transcriptional regulation of carbohydrate metabolism might have important effects on the physiological and biochemical changes in postharvest moso bamboo shoots.

Co-expression network of differentially expressed genes and differentially expressed miRNAs involved in carbohydrate metabolism of moso bamboo shoots

The co-expression network was comprehensively investigated using a WGCNA to predict the role of genes involved in carbohydrate metabolism. A total 4,048 genes with eigengene connectivity (kEM) value more than 0.7 were clustered into four major distinct modules (**Supplementary Figure 7A**). The turquoise module had a high correlation with sugar (0.99), cellulose (0.96), hemicellulose (0.96), and xylan (0.82), indicating that the genes in this module may be involved in carbohydrate metabolism of postharvest moso bamboo shoots. The correlation between the module and the compounds and expression trends of DEGs were shown in **Supplementary Figures 7B,C**. The expression of genes and their miRNA pairs in the turquoise module was further analyzed (**Figure 5A**). Combining the analysis of biological process, motif information, WGCNA, and coherent miRNA-target pairs, a network containing 15 TFs, 78 genes and 30 miRNAs was established, which supported that these TFs and miRNAs were significantly coupled with multiple key enzyme genes related to structural polysaccharides metabolism,

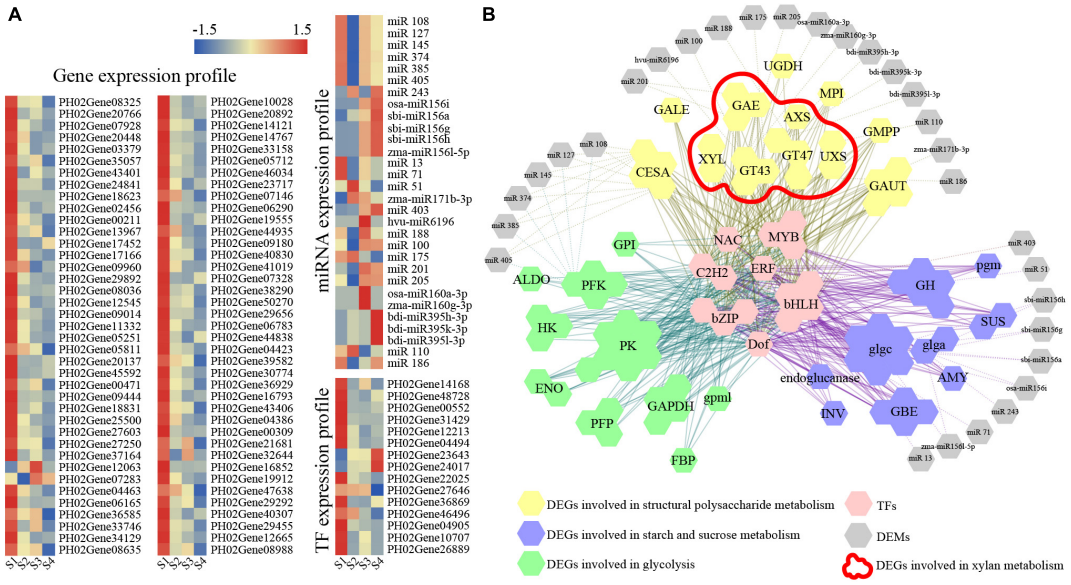


FIGURE 5
Analysis of DEGs and DEMs involved in carbohydrate metabolism of postharvest moso bamboo shoots. **(A)** Expression patterns of DEGs and DEMs. **(B)** Visualization of co-expression relationships between DEGs and DEMs. The color bar indicates $\log_2$-based fragments per kilobase per million (FPKM) with higher in red and lower in blue.

starch and sucrose metabolism and glycolysis (Figure 6B). The gene families involved in xylan metabolism were the most numerous, including GT43, GT47, AXS, XYL, and UXS families.

Validation of the regulatory relationships between genetic elements involved in carbohydrate metabolism

To validate the regulatory network, qPCR and Y1H assays were conducted. Opposite expression trends of miRNA and mRNA in each pair were observed, which was similar to the results of the high-throughput sequencing. Most miRNAs were up-regulated significantly, while their target genes were down-regulated significantly at S2 or S3 (Figure 6A). Additionally, the qPCR results of key TFs, such as two MYBs of PH02Gene12213 and PH02Gene31429, showed similar expression patterns to their target genes, with trends decreasing continuously along with storage time. These results indicated that the enzyme genes might be regulated positively by the TFs and negatively by the miRNAs.

To further validate the regulatory relationship between TFs and genes, the enriched xylan biosynthesis-related gene PH02Gene04386, which was predicted to be a target of MYBs (PH02Gene12213 and PH02Gene31429), was selected for validation of Y1H assay (Figure 6B). The results showed that the MYBs could bind to the 5′ UTR, promoter sequences and SMRE elements of PH02Gene04386. The yeast transformants harboring pro-600 grew best, followed by the transformants harboring pro-1200 and triple SMRE elements. In contrast, the negative control could not grow on the nutritional screening medium. These results indicated that PH02Gene12213 and PH02Gene31429 could bind to the promoter of PH02Gene04386 and might further regulate its expression.

PH02Gene04386 was a hub gene in the regulatory network and homologous gene of IRX9, which contributed to the xylan biosynthesis in *Arabidopsis*. The function of PH02Gene04386 was investigated in *Arabidopsis*, and three OE lines (L1, L3, and L9) were used for further analysis (Supplementary Figure 8). The results showed that the primary root length of OE plants was significantly greater than that of WT on 50 mM xylose medium ($P < 0.05$) (Figure 7A). The staining intensity was higher and the vascular bundle cell wall was more thickness in OE flower stems than those of WT (Figures 7B,C). To verify the effect of PH02Gene04386 on the xylan content and distribution in stems of *Arabidopsis* inflorescence, immunofluorescence localization and xylan content analysis were performed using OE and WT plants. The results showed that the fluorescence intensity was higher and the distribution range of xylan antibody (LM10) was broader in OE plants than in WT (Figure 7D). The chemical assay also supported the results that the xylan content was higher in OE plants than in WT (Figure 7E), indicating that overexpression of PH02Gene04386 could promote xylan biosynthesis in *Arabidopsis*. These results supported that PH02Gene04386 was

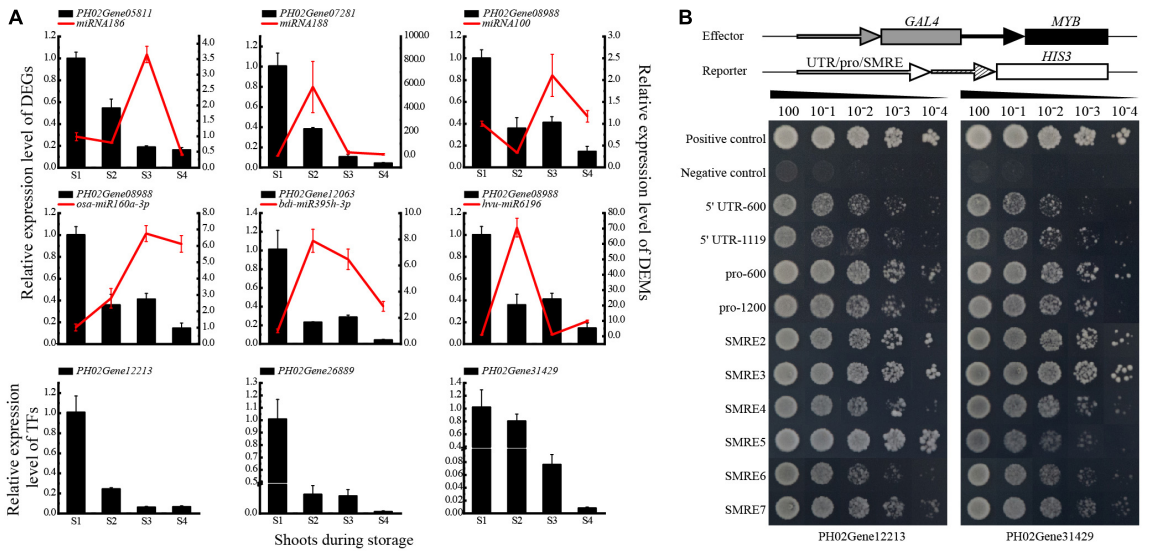


FIGURE 6
Validation of the genetic elements in the network of carbohydrate metabolism. **(A)** qPCR analysis of DEGs and DEMs. **(B)** Regulatory relationship analysis between MYBs (PH02Gene12213 and PH02Gene31429) and PH02Gene04386 by yeast one-hybrid assay. The red solid lines represent the miRNAs, the black blocks represent the structure genes targeted by miRNAs and MYBs. S1, S2, S3, and S4 represent bamboo shoots stored for 0, 3, 6, and 12 d.

the hub gene in the regulatory network involved in xylan biosynthesis.

Discussion

Physiological changes affected quality of postharvest moso bamboo shoots

The edible quality decline rapidly during storage, limiting the spread for fresh bamboo shoots in the market. The shoots of moso bamboo and *P. prominens* showed significant lignification during storage, with lower accumulation rate of lignin and cellulose in moso bamboo shoots (Zhang et al., 2020). PAL and POD catalyze the oxidation of polyphenols, which is presumed to be the cause of severe lignification and browning in *P. prominens* shoots (Nguyen et al., 2003; Roura et al., 2008). In present study, the degree of browning of moso bamboo shoots is lower than the results in the above study (**Figure 1A**), probably due to the retention of culm sheaths during storage, thus indicating that storage with culm sheaths is beneficial for preservation of moso bamboo shoots.

The consumption of carbohydrate was an important factor to reduce the edible quality of bamboo shoots. The soluble sugar and hemicellulose contents of the moso bamboo shoots in this study decreased significantly at S2 and continued to decrease during storage (**Figures 1C,D**), which was similar to the variation of non-structural sugar and hemicellulose in Lei bamboo and green bamboo shoots (Yu et al., 2021). However, the variation of cellulose content was different (**Figure 1E**; Li et al., 2018; Zhang et al., 2020). It might be explained by the different storage methods and bamboo shoot treatments, by which refrigeration was generally used in the former studies for the shoots with culm sheath removal and bagging in slices, while the shoots with culm sheaths at room temperature used in this study. These results are value for reasonable storage and transportation of bamboo shoots.

Expression changes of genes affected carbohydrate metabolism

The physiological changes of bamboo shoots during storage depend on the spatial and temporal expression of the relevant genes. In this study, we tried to interpret the molecular mechanism of the physiological changes in moso bamboo shoots based on the expression changes of DEGs. Carbohydrate-related DEGs were the most enriched (**Figure 2D**), which were closely related to the significant changes of carbohydrate content in moso bamboo shoots. Therefore, it could infer that the carbohydrate-related DEGs had important effect on the edible quality of moso bamboo shoots. Exogenous fructose treatment on Lei bamboo shoots could effectively reduce the lignification and prolong the shelf life by regulating the enzymes involved in starch and sucrose metabolism (Zhou et al., 2020). A large number of DEGs enriched in carbohydrate metabolism such as starch, cellulose and pectin were found in blueberry and African Pride (*Annona cherimola* × *A. squamosa*) during storage

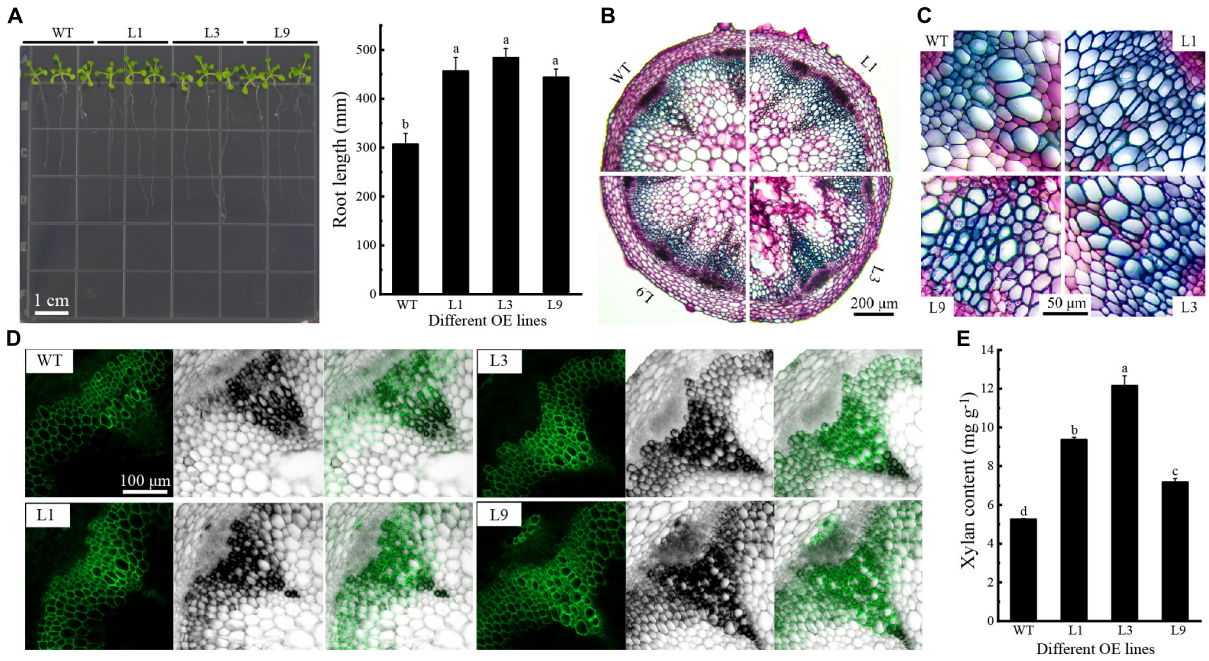


FIGURE 7
The function analysis of PH02Gene04386 overexpressed in *Arabidopsis*. **(A)** Root length of *Arabidopsis* growing on xylose medium. **(B)** Histochemical staining of cross section of inflorescence stems. **(C)** Vascular bundle in cross section of inflorescence stems in **(B)**. **(D)** Xylan immunofluorescence localization in inflorescence stems. **(E)** Xylan contents in inflorescence stems. Lowercase letters a, b, c, and d indicate significant differences ($P < 0.05$).

(Chen et al., 2019; Cárcamo de la Concepción et al., 2021). In the similar way, the changing pattern of the synthesis-related DEGs in non-structural sugar metabolism was characterized by a decreasing expression and those hydrolysis-related DEGs had an increasing expression along with the physiological changes in this study.

In addition, a large number of DEGs involved in structural polysaccharide metabolism were also identified, among which DEGs involved in the biosynthesis of mannose, xylan and pectin were significantly down-regulated (**Figure 3**), while those involved in the biosynthesis of arabinoxylans were either up-regulated or first up-regulated and then down-regulated. Studies have shown that the arabinose side chain of xylan binding to the ferulic acid of lignin could increase the stability of the cell wall, and the added xylan acts as a scaffold for lignin dehydrogenation polymer deposition in polysaccharides matrix (Li et al., 2015). This might explain the up-regulation of arabinose biosynthesis related DEGs in this study.

Carbohydrate metabolism of moso bamboo shoots involved a complex regulatory network

TFs and miRNAs affect carbohydrate metabolism through transcriptional regulation. GhMYB212 regulated the length of fibers and the accumulation of sucrose simultaneously by controlling the expression of a sucrose transporter gene *GhSWEET12* (Sun et al., 2019). Two MYBs of PH02Gene12213 and PH02Gene31429 were suggested to be involved in the biosynthesis of lignin, xylan and cellulose, and lignin deposition, respectively (Yang et al., 2019), both of which had potential regulatory relationships with xylan biosynthesis related DEGs like *GT43* and *GT47* by bioinformatics and Y1H assay (**Figures 5B**, **6B**). A number of miRNAs involved in carbohydrate metabolism were also identified in this study, such as *miRNA167*s and *miRNA160*s (**Figure 5B**), which had been confirmed to have regulatory roles in starch and sucrose metabolism and structural polysaccharide metabolism in cotton and Acacia (*Acacia auriculiformis* and *A. mangium*) (Wong et al., 2011; Wang et al., 2017). In addition, there were several novel miRNAs in the regulatory network, such as *miRNA100* and *miRNA175*, which provided important information for further investigating the function of miRNAs in polysaccharide metabolism.

Structural polysaccharides are important components of plant cell walls, which are synthesized by glycosyltransferases like GT43. In this study, overexpression of PH02Gene04386 encoding glycosyltransferase 43 promoted the elongation of OE *Arabidopsis* seedling roots on the medium with UDP xylose as a carbon source (**Figure 7A**), suggesting PH02Gene04386 could

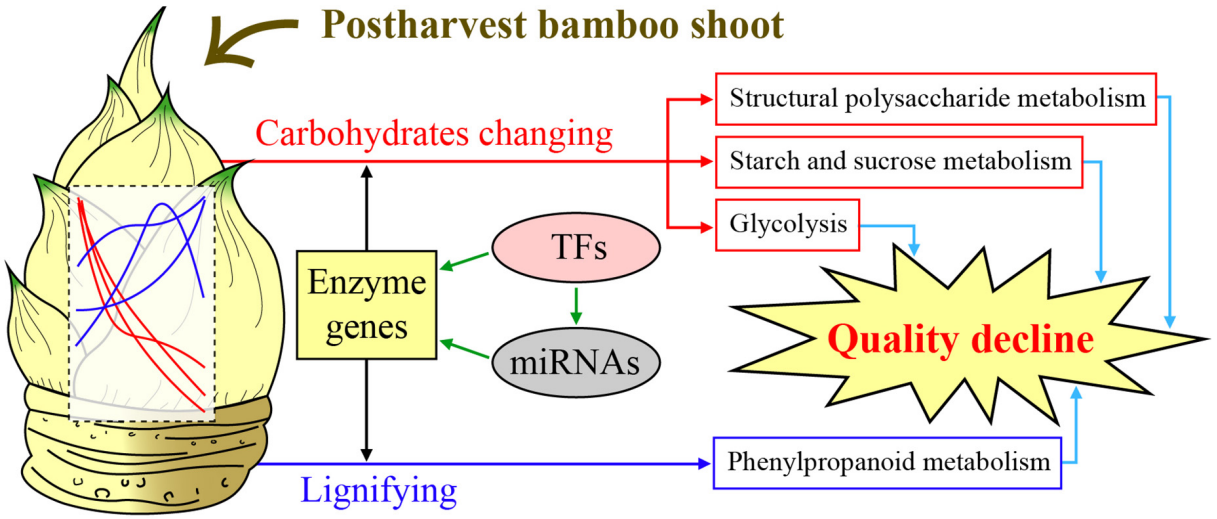


FIGURE 8
A model to elucidate the regulatory mechanism involved in carbohydrate metabolism of moso bamboo shoots during storage.

increase the efficiency of xylose utilization. The change of xylan content and its distribution in PH02Gene04386 OE *Arabidopsis* in this study (**Figures 7D,E**) supported the important role of GT43 family members involved in xylan biosynthesis. Similar results were also found in *GhGT43A1* and *GhGT43C1* OE cotton plants (Li et al., 2014), *IRX9* and *IRX14* in *Arabidopsis*, rice and cotton (Wu et al., 2010), and *GT43A* and *GT43B* in polar (Ratke et al., 2018). A model to elucidate the regulatory mechanism involved in carbohydrate metabolism of moso bamboo shoots during storage was constructed (**Figure 8**). Overall, these results will not only be useful for understanding the physiological changes and molecular regulatory mechanisms of moso bamboo shoots during storage, but also provide an important potential candidate gene for the breeding related to xylan.

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/, PRJNA840860 and https://www.ncbi.nlm.nih.gov/, PRJNA840882.

Author contributions

ZL designed and performed the experiments and wrote the manuscript. XX constructed the conceptions and worked for the bamboo sample collection. KY and CZ assisted with Y1H assay. YL assisted in completing the histochemical staining. ZG checked the data analysis and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (No. 2021YFD2200502) and the National Natural Science Foundation of China (No. 31971736).

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 ZZ declared a shared affiliation with the author(s) XX to the handling editor at the time of review.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1021161/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Gene expression analysis in moso bamboo during storage. **(A)** Correlation analysis between different samples based on gene expression. **(B)** Number of DEGs encoding transcript factors. **(C)** Number of DEGs between different comparison groups. S1, S2, S3, and S4 represent bamboo shoots stored for 0, 3, 6, and 12 d.

SUPPLEMENTARY FIGURE 2
GO **(A)** and KEGG **(B)** analyses of DEGs in cluster III of Rich Factor represents the value of enrichment factor which is the quotient of foreground value (the number of DEGs), and the larger the value, the more significant enrichment; Coloring indicates *P*-value with higher in red and lower in blue, and the lower *P*-value, the more significantly enriched. Point size indicates the number of DEGs.

SUPPLEMENTARY FIGURE 3
The expression patterns of DEGs involved in starch and sucrose metabolism. AMY, α-amylase; BMY, β-amylases; GBE, glucan branching enzyme; GH, glycosyl hydrolase; INV, fructosidase; glgA, starch synthase; glgC, glucose-1-phosphate adenylyltransferase; MGAM, maltase-glucoamylase; pgm, phosphoglucomutase; SPS, sucrose phosphate synthase; SUS, sucrose synthase; UGP, glucose-1-phosphate uridylyltransferase. The color bar indicates $\log_2$-based fragments per kilobase per million (FPKM) with higher in red and lower in blue.

SUPPLEMENTARY FIGURE 4
The expression patterns of DEGs involved in glycolysis. ALDO, fructose-bisphosphate aldolase; ENO, enolase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; gpml. phosphoglycerate mutase; HK, hexokinase; PFK, phosphofructokinase; PGK, phosphoglycerate kinase; PK, pyruvate kinase. The color bar indicates $\log_2$-based fragments per kilobase per million (FPKM) with higher in red and lower in blue.

SUPPLEMENTARY FIGURE 5
Length and family classification of miRNA. **(A)** Length of known miRNA. **(B)** Length of novel miRNA. **(C)** Percentage of miRNA members in each family.

SUPPLEMENTARY FIGURE 6
Function analysis of miRNA-target pairs.

SUPPLEMENTARY FIGURE 7
WGCNA modules based on DEG data. **(A)** Cluster dendrogram. **(B)** Module–trait relationships. Each column corresponds to a flavor compound. The color of each cell at the row-column intersection indicates the correlation coefficient between the module and the substance. **(C)** Expression heat map and expression level of DEGs in the module.

SUPPLEMENTARY FIGURE 8
Validation of PH02Gene04386 overexpressed plants. **(A)** Overexpression vector of PH02Gene04386. **(B)** Screening of transgenic plants by PCR. **(C)** Detection of PH02Gene04386 expression in transgenic plants by qPCR. WT: Wild type *Arabidopsis*. L1, L3, L6, L7, L9: Transgenic lines. Lowercase letters a and b indicate significant differences ($P < 0.05$).

References

Cárcamo de la Concepción, M., Sargent, D. J., Šurbanovski, N., Colgan, R. J., and Moretto, M. (2021). De novo sequencing and analysis of the transcriptome of two highbush blueberry (*Vaccinium corymbosum* L.) cultivars 'Bluecrop' and 'Legacy' at harvest and following post-harvest storage. *PLoS One* 16:e0255139. doi: 10.1371/journal.pone.0255139

Chen, C., Xie, F., Hua, Q., Zur, N. T., Zhang, L., Zhang, Z., et al. (2020). Integrated sRNAome and RNA-Seq analysis reveals miRNA effects on betalain biosynthesis in pitaya. *BMC Plant Biol.* 20:437. doi: 10.1186/s12870-020-02622-x

Chen, J., Duan, Y., Hu, Y., Li, W., Sun, D., Hu, H., et al. (2019). Transcriptome analysis of atemoya pericarp elucidates the role of polysaccharide metabolism in fruit ripening and cracking after harvest. *BMC Plant Biol.* 19:219. doi: 10.1186/s12870-019-1756-4

Fan, C., Ma, J., Guo, Q., Li, X., Wang, H., and Lu, M. (2013). Selection of reference genes for quantitative real-time PCR in bamboo (*Phyllostachys edulis*). *PLoS One* 8:e56573. doi: 10.1371/journal.pone.0056573

Hou, D., Lu, H., Zhao, Z., Pei, J., Yang, H., Wu, A., et al. (2022). Integrative transcriptomic and metabolomic data provide insights into gene networks associated with lignification in postharvest Lei bamboo shoots under low temperature. *Food Chem.* 368:130822. doi: 10.1016/j.foodchem.2021.130822

Keppler, B. D., and Showalter, A. M. (2010). IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in *Arabidopsis*. *Mol. Plant* 3, 834–841. doi: 10.1093/mp/ssq028

Li, C., Suo, J., Xuan, L., Ding, M., Zhang, H., Song, L., et al. (2019). Bamboo shoot-lignification delay by melatonin during low temperature storage. *Postharvest Biol. Technol.* 156:110933. doi: 10.1016/j.postharvbio.2019.110933

Li, D., Limwachiranon, J., Li, L., Zhang, L., Xu, Y., Fu, M., et al. (2019). Hydrogen peroxide accelerated the lignification process of bamboo shoots by activating the phenylpropanoid pathway and programmed cell death in postharvest storage. *Postharvest Biol. Technol.* 153, 79–86. doi: 10.1016/j.postharvbio.2019.03.012

Li, X., Xie, L., Zheng, H., Cai, M., Cheng, Z., Bai, Y., et al. (2019). Transcriptome profiling of postharvest shoots identifies PheNAP2- and PheNAP3-promoted shoot senescence. *Tree Physiol.* 39, 2027–2044. doi: 10.1093/treephys/tpz100

Li, C., Xuan, L., He, Y., Wang, J., Zhang, H., Ying, Y., et al. (2018). Molecular mechanism of xylogenesis in moso bamboo (*Phyllostachys edulis*) shoots during cold storage. *Polymers* 11:38. doi: 10.3390/polym11010038

Li, H., Lin, Q., Yan, M., Wang, M., Wang, P., Zhao, H., et al. (2021). Relationship between secondary metabolism and mirna for important flavor compounds in different tissues of tea plant (*Camellia sinensis*) as revealed by genome-wide mirna analysis. *J. Agric. Food Chem.* 69, 2001–2012. doi: 10.1021/acs.jafc.0c07440

Li, L., Huang, J., Qin, L., Huang, Y., Zeng, W., Rao, Y., et al. (2014). Two cotton fiber-associated glycosyltransferases, GhGT43A1 and GhGT43C1, function in hemicellulose glucuronoxylan biosynthesis during plant development. *Physiol. Plant.* 152, 367–379. doi: 10.1111/ppl.12190

Li, Q., Koda, K., Yoshinaga, A., Takabe, K., Shimomura, M., Hirai, Y., et al. (2015). Dehydrogenative polymerization of coniferyl alcohol in artificial polysaccharides matrices: Effects of xylan on the polymerization. *J. Agric. Food Chem.* 63, 4613–4620. doi: 10.1021/acs.jafc.5b01070

Liu, Z., Li, L., Luo, Z., Zeng, F., Jiang, L., and Tang, K. (2016). Effect of brassinolide on energy status and proline metabolism in postharvest bamboo shoot during chilling stress. *Postharvest Biol. Technol.* 111, 240–246. doi: 10.1016/j.postharvbio.2015.09.016

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

Nguyen, T., Ketsa, S., and Doorn, W. (2003). Relationship between browning and the activities of polyphenoloxidase and phenylalanine ammonia lyase in banana peel during low temperature storage. *Postharvest Biol. Technol.* 30, 187–193. doi: 10.1016/S0925-5214(03)00103-0

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Gene Dev.* 20, 3407–3425. doi: 10.1101/gad.1476406

Ratke, C., Terebieniec, B. K., Winestrand, S., Derba-Maceluch, M., Grahn, T., Schiffthaler, B., et al. (2018). Downregulating aspen xylan biosynthetic GT43 genes in developing wood stimulates growth via reprograming of the transcriptome. *New Phytol.* 219, 230–245. doi: 10.1111/nph.15160

Roura, S. I., Pereyra, L., and Del Valle, C. E. (2008). Phenylalanine ammonia lyase activity in fresh cut lettuce subjected to the combined action of heat mild

shocks and chemical additives. *LWT-Food Sci. Technol.* 41, 919–924. doi: 10.1016/j.lwt.2007.06.005

Schouten, R. E., Woltering, E. J., and Tijskens, L. M. M. (2016). Sugar and acid interconversion in tomato fruits based on biopsy sampling of locule gel and pericarp tissue. *Postharvest Biol. Technol.* 111, 83–92. doi: 10.1016/j.postharvbio.2015.07.032

Singhal, P., Bal, L. M., Satya, S., Sudhakar, P., and Naik, S. N. (2013). Bamboo shoots: A novel source of nutrition and medicine. *Crit. Rev. Food Sci.* 53, 517–534. doi: 10.1080/10408398.2010.531488

Sun, H., Liu, Y., Ma, J., Wang, Y., Song, H., Li, J., et al. (2021). Transcriptome analysis provides strategies for postharvest lotus seeds preservation. *Postharvest Biol. Technol.* 179:111583. doi: 10.1016/j.postharvbio.2021.111583

Sun, W., Gao, Z., Wang, J., Huang, Y., Chen, Y., Li, J., et al. (2019). Cotton fiber elongation requires the transcription factor GhMYB212 to regulate sucrose transportation into expanding fibers. *New Phytol.* 222, 864–881. doi: 10.1111/nph.15620

Tanabe, C., Furuta, K., Maeda, M., and Kimura, Y. (2017). Structural feature of N-glycans of bamboo shoot glycoproteins: Useful source of plant antigenic N-glycans. *Biosci. Biotechnol. Biochem.* 81, 1405–1408. doi: 10.1080/09168451.2017.1320519

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. *Genome Res.* 21, 2213–2223. doi: 10.1101/gr.124321.111

Tian, X., Zhu, L., Yang, N., Song, J., Zhao, H., Zhang, J., et al. (2021). Proteomics and metabolomics reveal the regulatory pathways of ripening and quality in post-harvest Kiwifruits. *J. Agric. Food Chem.* 69, 824–835. doi: 10.1021/acs.jafc.0c05492

Wang, J., Jiang, J., Wang, J., Wang, Z., Yang, X., and Jia, L. (2019). The influence of gamma irradiation on the storage quality of bamboo shoots. *Radiat. Phys. Chem.* 159, 124–130. doi: 10.1016/j.radphyschem.2019.02.021

Wang, M., Sun, R., Li, C., Wang, Q., and Zhang, B. (2017). MicroRNA expression profiles during cotton (*Gossypium hirsutum* L) fiber early development. *Sci. Rep.* 7:44454. doi: 10.1038/srep44454

Wang, Y., Chen, J., Wang, D., Ye, F., He, Y., Hu, Z., et al. (2020). A systematic review on the composition, storage, processing of bamboo shoots: Focusing the nutritional and functional benefits. *J. Funct. Foods* 71:104015. doi: 10.1016/j.jff.2020.104015

Wen, B., Cheng, Z., Hu, Y., Boon-Ek, Y., Wongs-Aree, C., and Supapanich, S. (2019). Ultraviolet-C treatment maintains physicochemical quality of water bamboo (*Zizania latifolia*) shoots during postharvest storage. *Postharvest Biol. Technol.* 152, 65–72. doi: 10.1016/j.postharvbio.2019.02.017

Wong, M. M., Cannon, C. H., and Wickneswari, R. (2011). Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via de novo transcriptome sequencing. *BMC Genom.* 12:342. doi: 10.1186/1471-2164-12-342

Wu, A. M., Hörnblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., et al. (2010). Analysis of the Arabidopsis *IRX9*/*IRX9-L* and *IRX14*/*IRX14-L* pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiol.* 153, 542–554. doi: 10.1104/pp.110.154971

Wu, M. X., Zou, Y., Yu, Y. H., Chen, B. X., Zheng, Q. W., Ye, Z. W., et al. (2021). Comparative transcriptome and proteome provide new insights into the regulatory mechanisms of the postharvest deterioration of *Pleurotus tuoliensis* fruitbodies during storage. *Food Res. Int.* 147:110540. doi: 10.1016/j.foodres.2021.110540

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. *Nucl. Acids Res.* 39:W316–W322. doi: 10.1093/nar/gkr483

Yang, K., Li, L., Lou, Y., Zhu, C., Li, X., and Gao, Z. (2021). A regulatory network driving shoot lignification in rapidly growing bamboo. *Plant Physiol.* 187, 900–916. doi: 10.1093/plphys/kiab289

Yang, K., Li, Y., Wang, S., Xu, X., Sun, H., Zhao, H., et al. (2019). Genome-wide identification and expression analysis of the MYB transcription factor in moso bamboo (*Phyllostachys edulis*). *PeerJ* 6:e6242. doi: 10.7717/peerj.6242

Yao, S., Wang, Z., Cao, Q., Xie, J., Wang, X., Zhang, R., et al. (2020). Molecular basis of postharvest granulation in orange fruit revealed by metabolite, transcriptome and methylome profiling. *Postharvest Biol. Technol.* 166:111205. doi: 10.1016/j.postharvbio.2020.111205

Yu, L., Pei, J., Zhao, Y., and Wang, S. (2021). Physiological changes of bamboo (*Fargesia yunnanensis*) shoots during storage and the related cold storage mechanisms. *Front. Plant Sci.* 12:731977. doi: 10.3389/fpls.2021.731977

Zhang, H., Ying, Y. Q., Wang, J., Zhao, X. H., Zeng, W., Beahan, C., et al. (2018). Transcriptome analysis provides insights into xylogenesis formation in Moso bamboo (*Phyllostachys edulis*) shoot. *Sci. Rep.* 8:3951. doi: 10.1038/s41598-018-21766-3

Zhang, Z., Li, C., Zhang, H., Ying, Y., Hu, Y., and Song, L. (2020). Comparative analysis of the lignification process of two bamboo shoots stored at room temperature. *Plants* 9:1399. doi: 10.3390/plants9101399

Zhou, D., Wang, K. T., Kuang, W. L., Lei, C. Y., Qiu, L. L., Li, C. H., et al. (2020). Effects of fructose treatment on quality and lignification of bamboo shoots (*Phyllostachys praecox*) during cold storage and its regulation mode involved [in Chinese]. *Food Ferment. Ind.* 46, 175–183. doi: 10.13995/j.cnki.11-1802/ts.023916

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fpls.2022.1041068

OPEN ACCESS

EDITED BY
Lin Chen,
Institute of Animal Sciences (CAAS), China

REVIEWED BY
Qinggang Zhu,
Northwest A&F University, China
Chen Shen,
Yangzhou University, China
Xiaojing Wang,
Guizhou University, China

*CORRESPONDENCE
Ronggao Gong
gongronggao@sicau.edu.cn

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

RECEIVED 10 September 2022
ACCEPTED 23 September 2022
PUBLISHED 18 October 2022

CITATION
Chen C, Chen H, Chen Y, Yang W, Li M, Sun B, Song H, Tang W, Zhang Y and Gong R (2022) Joint metabolome and transcriptome analysis of the effects of exogenous GA_3 on endogenous hormones in sweet cherry and mining of potential regulatory genes.
Front. Plant Sci. 13:1041068.
doi: 10.3389/fpls.2022.1041068

Joint metabolome and transcriptome analysis of the effects of exogenous GA_3 on endogenous hormones in sweet cherry and mining of potential regulatory genes

Chaoqun Chen, Hongxu Chen, Yuanfei Chen, Wenlong Yang, Mengyao Li, Bo Sun, Haiyan Song, Wenjing Tang, Yao Zhang and Ronggao Gong*

College of Horticulture, Sichuan Agricultural University, Chengdu, China

Gibberellin (GA) is an important phytohormone that can participate in various developmental processes of plants. The study found that application of GA_3 can induce parthenocarpy fruit and improve fruit set. However, the use of GA_3 affects endogenous hormones in fruits, thereby affecting fruit quality. This study mainly investigates the effect of exogenous GA_3 on endogenous hormones in sweet cherries. The anabolic pathways of each hormone were analyzed by metabolome and transcriptome to identify key metabolites and genes that affect endogenous hormones in response to exogenous GA_3 application. Results showed that exogenous GA_3 led to a significant increase in the content of abscisic acid (ABA) and GA and affected jasmonic acid (JA) and auxin (IAA). At the same time, the key structural genes affecting the synthesis of various hormones were preliminarily determined. Combined with transcription factor family analysis, *WRKY* genes were found to be more sensitive to the use of exogenous GA_3, especially the genes belonging to Group III (*PaWRKY16*, *PaWRKY21*, *PaWRKY38*, *PaWRKY52*, and *PaWRKY53*). These transcription factors can combine with the promoters of *NCED*, *YUCCA*, and other genes to regulate the content of endogenous hormones. These findings lay the foundation for the preliminary determination of the mechanism of GA_3's effect on endogenous hormones in sweet cherry and the biological function of *WRKY* transcription factors.

KEYWORDS

sweet cherry, transcriptome, metabolome, endogenous hormones, WRKY, gibberellin

1 Introduction

Sweet cherries (*Prunus avium* L.), belonging to the Rosaceae family, originated in Europe and Western Asia (Papapetros et al., 2018). Sweet cherries are rich in nutrients, such as flavonoids, ascorbic acid, and anthocyanins. These substances have beneficial effects on health, including preventive and regulatory effects on several chronic diseases (diabetes, cancer, cardiovascular, and other inflammatory diseases) (Faienza et al., 2020). Versatile and economical, sweet cherries are among the most popular fruits and are widely cultivated worldwide. However, sweet cherriy is a typical self-incompatible species controlled by the multi-allele expressed by a single gametophyte (Kivistik et al., 2022), and artificial pollination is often used to improve the fruit set. The workload of manual pollination is large and consumes considerable manpower and material resources. Therefore, people gradually adjust the program for higher efficiency, such as exogenous spraying of GA_3 (Askarieh et al., 2021).

GA, an important plant hormone and signaling molecule, plays an important role in fruit reproductive development and stress response (Rachappanavar et al., 2022). Numerous studies have shown that GA affects fruit development and ripening, especially fruit-enhancing fruit sets. In pepper, exogenous use of GA_3 can significantly increase yield (Tiwari et al., 2012). Treatment with GA_3 during flowering stimulates cell division and ovary growth, thereby improving citrus fruit set, whereas paclobutrazol (GA_3 biosynthesis inhibitor) inhibits cell division and reduces fruit set (Mesejo et al., 2016). Similarly, the use of GA_3 before apple flowering can induce parthenocarpic fruit and increase the fruit setting rate (Watanabe et al., 2008). The accumulation of sucrose and organic acids in single-core pears induced by GA_{4+7} was lower than that in pollinated pears (Niu et al., 2015). Citrus carotenoids were significantly reduced after GA_3 treatment (Zhang et al., 2012). Similarly, in sweet cherries, changes in fruit bioactivity and soluble sugars were found after the use of GA_3 (Ozkan et al., 2016). Exogenous use of GA_3 can improve the fruit set of sweet cherries, but the fruit quality changes (Kuhn et al., 2020).

Studies have shown that changes in endogenous hormones are one of the main reasons for the changes in fruit quality after GA_3 is applied (Tijero et al., 2019). However, plant hormones do not play an independent role in the process of plant growth and development, but they interact with each other to form a complex multihormone regulatory network to jointly regulate the life activities of plants (Jaillais and Chory, 2010). Therefore, the relationship between hormones should be considered when studying the effect of exogenous GA_3 on endogenous hormones. Studies have shown that GA_3 and IAA induce cell wall expansion by activating the expression of *EXP* genes, the combination of which regulates stem formation (Kou et al., 2021). Meanwhile, GA and IAA regulated soybean lower ovule elongation under the interaction of low light and high-temperature stress (Bawa et al., 2020). Through the crosstalk between the ABA and GA signaling pathways, root growth and tillering can be maintained, and the plant structure can be regulated (Lin et al., 2020). GA, ABA, and IAA interact to regulate strawberry fruit development (Liao et al., 2018). ABA, JA, and SA together with ETH modulate some abiotic stress defense responses of trees exposed to sunlight during photooxidative and thermal stress (Torres et al., 2017). These findings further indicate a crosstalk mechanism between endogenous hormones to jointly regulate fruit growth, development, and quality.

At present, most studies on the effect of exogenous GA_3 on the endogenous hormones in sweet cherry fruit focus on content determination, whereas systematic studies on the anabolic pathways of these hormones are limited. Therefore, in this study, the change patterns of endogenous hormones (ABA, GA_3, GA4, IAA, and JA) were determined by spraying different concentrations of GA_3. At the same time, by combining metabolome and transcriptome, the anabolic pathways of these hormones were analyzed to identify the differential metabolites (DEM) and differential genes (DEG) that affect endogenous hormones in response to exogenous GA_3. In addition to this, we performed a comprehensive genome-wide analysis of the sweet cherry *WRKY* gene family. Gene identification, phylogenetic analysis, and analyses of gene structure, conserved motifs, promoter *cis*-elements, and protein-protein interactions were performed, respectively, to investigate the potential relevance of *WRKY* genes to the application of exogenous GA_3. The results of this study will provide a reference for subsequent studies on the application of exogenous GA_3 in sweet cherry production and lay a foundation for studying the biological function of sweet cherry *WRKY* transcription factors.

2 Materials and methods

2.1 Plant materials and processing

In this study, 'Hongdeng' sweet cherry was selected as the experimental material, and the experimental site was located in the sweet cherry experimental base of Hanyuan County, Ya'an City, Sichuan Province, China. A total of 24 sweet cherry fruit trees with good growing conditions and the same developmental period were selected for listing. One control (CK) and seven GA_3 treatment concentration gradients were set up in the experiment: 10, 20, 30, 40, 60, 80, and 100 mg/L (Named A, B, C, D, E, F, G, respectively.), with three trees in each treatment. The entire tree was sprayed with GA_3 meticulously at 9:00 a.m. in the early flowering period and one week after full bloom, and the control group was treated with clean water. Each sweet cherry fruit tree uses 5 liters of GA_3 solution or water each time. Sampling was started after flowering, and then every three days until the fruit matured. A total of 12 periods of samples were collected (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12). After

harvesting, the fruits were brought back to the laboratory immediately, and then frozen in liquid nitrogen and stored in a −80°C refrigerator for subsequent experiments. Three biological replicates were prepared for each sample.

2.2 Hormone extraction and purification

Endogenous GA_3, GA_4, IAA, ABA and JA levels were determined using an indirect ELISA method. A 0.5 g sample was homogenized in liquid nitrogen and extracted in cold 80% (v/v) methanol containing 1 mM 2-tert-butylated hydroxytoluene as an antioxidant. The extracts were incubated at 4°C for 1 hour and centrifuged at 3500r/min for 8 minutes at the same temperature. The supernatant was then filtered through a Chromoseq C18 column (C18 Sep-Pak Cartridge, Waters, Millford, Massachusetts, USA). The resulting eluate was concentrated to dryness in vacuo and dissolved in 1 mL of phosphate buffered saline (PBS) containing 0.1% (v/v) Tween-20 and 0.1% (w/v) gelatin (pH 7.5) for ELISA analyze.

2.3 Metabolite profiling and data analysis

The freeze-dried sample was pulverized to a powder, and 100 mg was extracted with 600 μL of 2-chlorophenylalanine (4 ppm) in methanol overnight at 4°C. The supernatant was then collected by centrifugation at 12,000 rpm for 10 minutes. For each experimental sample, an equal volume of samples was obtained and mixed as a quality control (QC) sample, which was inserted in the front, middle, and back of the sample to test the repeatability of the experiment. Subsequently, these extracts were absorbed, filtered, and analyzed by a UHPLC-MS/MS system. ACQUITY UPLC® HSS T3 1.8 μm (2.1 × 150 mm) columns were used in this study. Mobile phase A is positive and negative ion 0.1% formic acid-water solution, and mobile phase B is formic acid acetonitrile. Chromatographic gradient elution program: 0–1 min, 98% A, 2% B; 1–9 min, 74% A, 26% B; 9–12 min, 26% A, 74% B; 12–13.5 min, 2% A, 98% B; 13.5–14 min, 50% A, 50% B; 14–20 min, 98% A, 2% B. Raw data were processed using Compound Discoverer 3.1 (CD3.1). Functional and taxonomic annotations were performed on the metabolites to investigate the functional properties and taxonomy of the identified metabolites. The data were logarithmically transformed and centrally formatted using MetaX software (http://metax.genomics.cn/). Differential metabolites were screened by three parameters, variable importance in the projection (VIP), fold change (FC), and P-value. The thresholds were set as VIP > 1.0, FC > 1.2 or FC < 0.833, and P value < 0.05. Six independent replications were included for each sample. Finally, OmicShare tools (https://www.omicshare.com/tools/) were used to perform cluster heatmap, correlation, and metabolic pathway analysis of screened metabolites.

2.4 RNA extraction, library construction, RNA sequencing, and data analysis

Total RNA was extracted with a total RNA kit (TIANGEN Biotech, Beijing, China). Using the polyA structure at the end of mRNA, the sample mRNA was separated from the total RNA by Oligo (dT) magnetic beads, and the obtained mRNA was randomly interrupted with divalent cations in NEB Fragmentation Buffer reagent. Using the fragmented mRNA as a template, the first strand of cDNA was synthesized with random oligonucleotide primers. Then, the second strand of cDNA was synthesized by using dNTPs as raw material and DNA polymerase I, and the double-stranded cDNA fragment was purified and recovered. The purified double-stranded cDNA is end-repaired and A is added to the end, and then the sequencing adapter is ligated to the double-stranded cDNA. Fragment selection was performed on the size of the cDNA using AMPure XP beads, and the 200 bp sequence was enriched. The enriched sequences were amplified by PCR, the PCR products were purified using AMPure XP beads, and the library was further constructed. Qubit2.0 Fluorometer was used for preliminary quantification, and Agilent 2100 Bioanalyzer was used for quality inspection of the constructed library. After pooling, as required, Illumina sequencing (TSINGKE, Beijing, China) was performed to generate 150 bp paired-end reads. Gene expression levels were analyzed by the fragments per kilobase per million reads (FPKM) method. DESeq2 v1.22.1 was used for differential expression analysis between sample groups, the original readcount was normalized, the significant P-value was corrected using the Benjamini and Hochberg methods, and finally, the corrected P-value (p-adjust), which is the false discovery rate value (FDR). |log2 (Fold Change)| > 2 and p-adjust ≤ 0.05 were used as the screening criteria for the significance of differentially expressed genes. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed using the clusterProfiler R package to clarify the signaling pathways involved in differential genes. The PlantTFDB database (http://planttfdb.gao-lab.org) was used to screen and classify possible transcription factors. At the same time, the log2(FPKM) values of differential genes were used to draw a clustering heatmap. The clustering heatmap passed the TBtools software (http://www.tbtools.org) for drawing.

2.5 Retrieval and identification of members of the sweet cherry *WRKY* transcription factor family

The *WRKY* family module sequence (PF03106) was downloaded from the Pfam database (http://pfam.xfam.org/), and then the sweet cherry protein sequence was downloaded from NCBI (https://www.ncbi.nlm.nih.gov). The HMMER

software was used to retrieve WRKY protein sequences from the sweet cherry genome sequence. Candidate proteins were further submitted to NCBI-CDD and Pfam for WRKY domain confirmation. The ExPASy website (http://web.expasy.org/protparam/) was used to analyze the physicochemical properties of the confirmed WRKY protein sequence, such as protein molecular weight and isoelectric point, and WOLF PSORT (http://www.genscript.com/wolf -psort.html) was use for subcellular localization analysis. The *WRKY* gene sequence of *Arabidopsisthalian* was downloaded from the TAIR (https://www.arabidopsis.org/) website.

2.6 Phylogenetic analysis of *WRKY* family in different species

Clustal X (v.2.1) software was used to perform multiple sequence alignment of all *WRKY* protein sequences of sweet cherry and *Arabidopsis thaliana*, and MEGA 6.06 software was used to perform phylogenetic analysis on the results of the multiple sequence alignment using the neighbor-joining method. The parameter bootstrap repeated 1000 times was verified, and a phylogenetic tree was constructed. The phylogenetic tree was modified in Evolview (http://www.evolgenius.info/evolview/#/login).

2.7 Sequence structure and conserved motif analysis of *WRKY* family members in sweet cherry

The GSDS server (http://gsds.cbi.pku.edu.cn/) was accessed, and the structural pattern diagram of the introns and exons of the *WRKY* family genes was designed by comparing the coding sequence with the gene sequence information. The conserved motifs of the members of the sweet cherry *WRKY* family gene were analyzed using MEME (http://meme-suite.org/). The parameter size of the conserved motif is set to be 10-100 amino acids, and the maximum number of domains to be exported is 10. Finally, the visualization of the results is realized by using TBtools software. The analysis results of these exon–intron structures and conserved motifs were arranged in the order shown on the phylogenetic tree.

2.8 Analysis of *cis*-acting elements of sweet cherry *WRKY* gene family

The Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) online software was used, considering the 2000 bp upstream of the start codon of the sweet cherry *PaWRKY* gene family as the sequence to predict and analyze the *cis*-acting elements of the gene family. Finally, the TBtools software was used to map.

2.9 Prediction of protein–protein interaction networks

All PaWRKY protein sequences were submitted to the STRING (https://cn.string-db.org) website, and *Arabidopsis thaliana* was selected as the reference organism. After blast analysis, the highest scoring *Arabidopsis* homolog (Bitscore) was used to construct the network. Genes that do not interact with any other genes are removed.

2.10 Correlation analysis of metabolite and transcript profiles

All the obtained DEMs and DEGs were mapped to the KEGG pathway database to obtain their common pathway enrichment information. The top 10 significantly enriched metabolic pathways in the three comparison groups were histogram plotted. The obtained DEM and DEG were analyzed based on the Pearson correlation coefficient, and the Cytoscape v3.9.1 software was used to make a correlation network diagram.

2.11 qRT-PCR analysis

Total RNA from the four samples was extracted with an RNA extraction kit (TSINGKE, Beijing, China). Subsequently, cDNA was synthesized using the Goldenstar RT6 cDNA Synthesis Kit Ver.2 kit (Beijing TsingKe Biotech Co., Ltd.). qRT-PCR was performed using the CFX96TM real-time system (Bio-Rad, California, USA) and 2 × TSINGKE® Master qPCR Mix(SYBR Green I)(TSINGKE, Beijing, China) reagents. The amplification program was as follows: pre-denaturation at 95°C for 30 s, denaturation at 95°C for 0.05 s, annealing at 59°C for 30 s, and the number of amplification cycles was 39. Gene expression was normalized with ACTIN as an internal control. Gene expression was calculated using 2−ΔΔCt, and primers were designed using Primer Premier 6.0 software. Table S1 lists the primers for qRT-PCR.

3 Results

3.1 Changes in endogenous hormones after exogenous GA_3 treatment

The endogenous hormones in sweet cherry fruit changed after the use of exogenous GA_3. The specific situation is shown in

Figure 1. Among them, the GA_3 content mainly showed a trend of initially decreasing and then increasing. The content of endogenous GA_3 was significantly increased after exogenous GA_3 treatment, and the change was most severe in the 5h to 6th period. In addition, the peak of GA_3 in CK appeared in the 6th to 7th period, whereas the peak in each treatment group was advanced to the 4th to 6th period. GA_4 generally showed an upward trend and reached its peak at maturity. After GA_3 treatment, the content of endogenous GA_4 was reduced in all periods except 1st, 6th, and 7th. ABA content, one of the most abundant hormones in sweet cherries, increased rapidly from stage 6, indicating that it may play a role mainly in the later stages of fruit growth. After treatment, the endogenous ABA content increased significantly, and the content increased the most in the 8th to 9th period, among which the E treatment group showed the greatest change, and the ABA content increased by approximately 80 ng/g FW. JA showed a trend of initially decreasing sharply to the 6th period and then increasing slightly. The lowest JA content in the seventh period was 20.71 ng/g FW. After GA_3 treatment, the content of endogenous JA was down-regulated, and a slight difference was observed between the treatment groups. The overall IAA showed a trend of initially decreasing and then increasing, and reached a peak value at the mature stage, which was 74.15 ng/g FW. After treatment, except for E treatment, the other treatments reduced the content of endogenous IAA, and the effect was most severe in the 6th to 10th period.

3.2 Metabolome analysis

These results showed that after exogenous use of GA_3, the effects of endogenous hormones were more severe in the three periods of the sixth, ninth, and 12th, and treatment group E had the most significant effect on endogenous GA_3, ABA, and IAA in the three periods. Therefore, CK6, CK9, CK12, E6, E9, and E12 were selected for metabolome and transcriptome analysis.

The PCA results of the metabolome profiles are shown in Figure 2A, the first two principal components could separate 36 samples, accounting for 17.23% and 12.18% of the total variability. On the PCA analysis chart, each group showed a separation trend, and each replicate was clustered, indicating that the data reproducibility was good during the experiment. Evident differences were observed among the six sweet cherry samples; CK6, E6, and CK9 were distributed in the positive end of PC1, whereas CK12, E9, and E12 were distributed in the negative end of PC1. In addition, in PC2, the CK group was distributed on the positive end, and the processing group E was distributed on the negative end. A total of 3011 metabolites were identified in the metabolome, including 2013 positive and 998 negative ions (Figure 2B). Through KEGG functional annotation, the identified metabolites were divided into seven categories, of which the metabolism group had the most metabolites, reaching 2486, accounting for 79.81%.

Thresholds were set to VIP > 1.0, FC > 1.2 or FC < 0.833, and P-value < 0.05 to screen for DEM. A total of 2256 DEMs were identified, and their expression patterns are shown in Figure 2C. Significant differences were found among the groups, and the higher expression of many DEMs in the treatment groups may be the main substance causing the differences in endogenous hormones in fruits. Each CK was compared with group E at different times, as shown in Figure 2D. Among the combinations of different treatments, the combination with the most differential metabolites was E6 vs CK6 (1088 in total, 580 up-regulated, and 508 down-regulated). In summary, fruit metabolites were changed

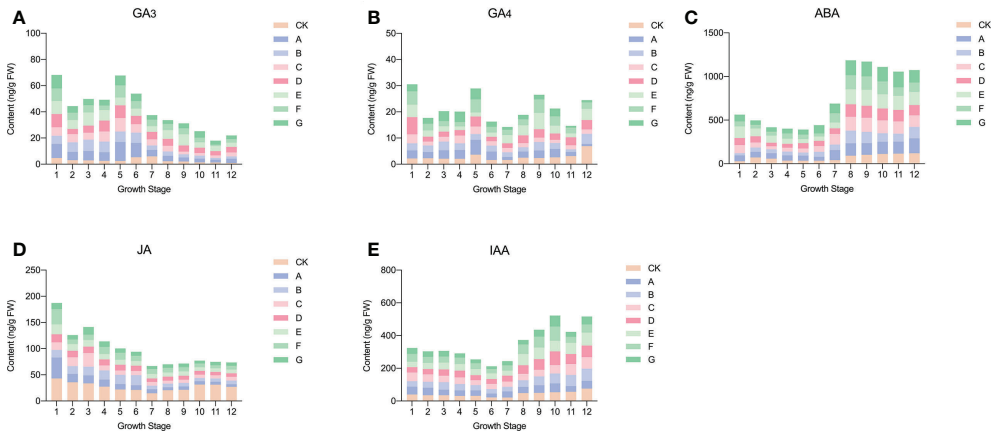


FIGURE 1
Changes of endogenous hormones in fruits after exogenous GA_3 treatment. **(A)** Changes in endogenous GA_3 after exogenous GA_3 treatment. **(B)** Changes in endogenous GA_4 after exogenous GA_3 treatment. **(C)** Changes in endogenous ABA after exogenous GA_3 treatment. **(D)** Changes in endogenous JA after exogenous GA_3 treatment. **(E)** Changes in endogenous IAA after exogenous GA_3 treatment.

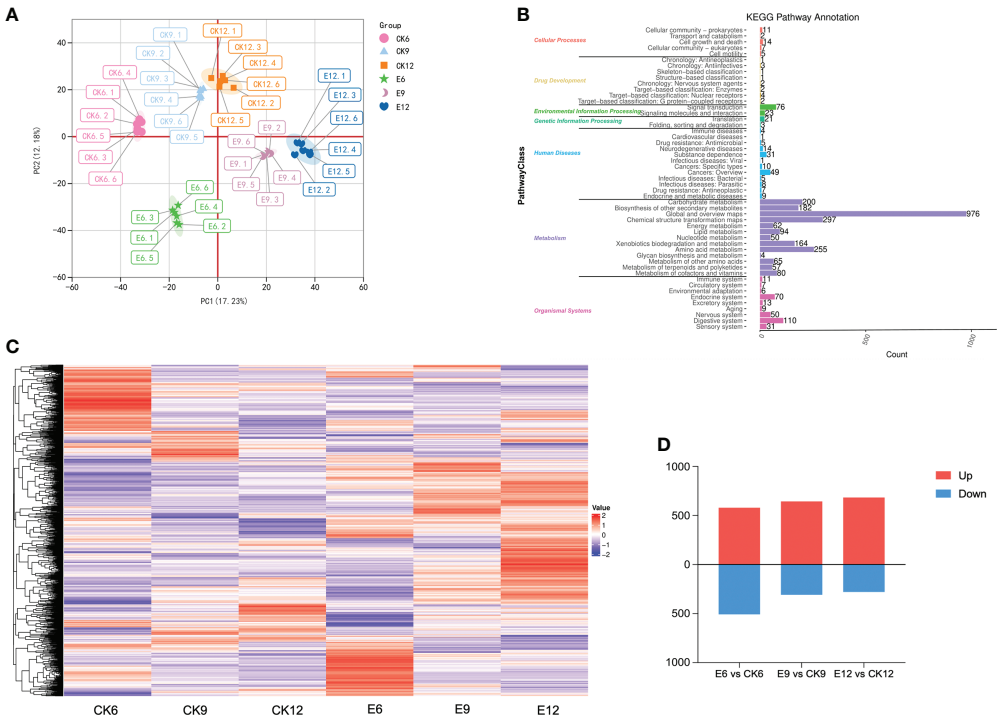


FIGURE 2
Metabolome analysis of fruits in three periods after gibberellin treatment. **(A)** PCA analysis of metabolites in different treatment groups. **(B)** KEGG Pathway classification annotation of metabolites. **(C)** DEMs expression heat map, color indicates the relative level content of each DEM, from low (purple) to high (red). **(D)** The number of DEMs in the treatment group and CK in the three periods, the number of up-regulated and down-regulated DEMs are represented by the bars above and below the x-axis, respectively.

in multiple periods after the exogenous use of GA_3 changed, and most expression levels showed an upward-regulated trend.

3.3 Transcriptome analysis

The samples of each treatment group were analyzed by RNA-seq technology. A total of 441,209,814 raw data were generated in three periods. After filtering out linker sequences, uncertain reads, and low-quality reads, 798,325,520 high-quality clean reads were obtained, and an average of 93.25% of the clean reads were mapped to the sweet cherry genome (Detailed results are shown in Table S2). The expression of transcript samples was analyzed by PCA, and the results are shown in Figure 3A. The figure shows that each sample can be clearly distinguished on the score map, and the resets are closely focused, indicating that the fruit transcripts are different after using exogenous GA_3. Similar to metabolome, CK6 and E6 are at the minus end of PC1, and the remainder is at the plus end of PC2. Interestingly, the score map shows that the discrimination between CK12 and E9 on PC1 is weak, indicating that the two samples are similar. Moreover, the sample gene expression correlation between the replicates of each sample was the highest, indicating that the samples had good repeatability, and CK12 and E9 had a high degree of correlation (Figure 3B).

Differential gene screening was performed with a threshold of P-value < 0.05 and |log2FoldChange|>2, and a total of 10,154 DEGs were identified. The results of DEGs compared between different developmental stages are shown in Figure 3C. Among the three groups, E12 vs CK12 had the most DEGs, reaching 3905. Different from the metabolome results, E12 vs CK12 had more differential genes than E9 vs CK9, and the number of up-regulated DEGs was less than that of down-regulated DEGs. The results of analyzing common or unique DEGs between the three comparison groups are shown in Figure 3D. The genes expressed in each sample were the least, only 236, and the genes expressed only in E12 vs CK12 were the greatest (2606).

Furthermore, a hierarchical clustering heatmap was drawn for the three developmental periods, with good repeatability of each treatment and large transcriptional differences between groups (Figure 3E). The expression patterns of all DEGs were divided into 10 groups, with more highly expressed genes in CK6, E6, and E12 (Figure 3F). Cluster2 contained the most DEGs, reaching 1593, mainly expressed in CK6 and E6. The expression levels of most DEGs were up-regulated in E12 vs CK12, mainly in Cluster8 and Cluster9.

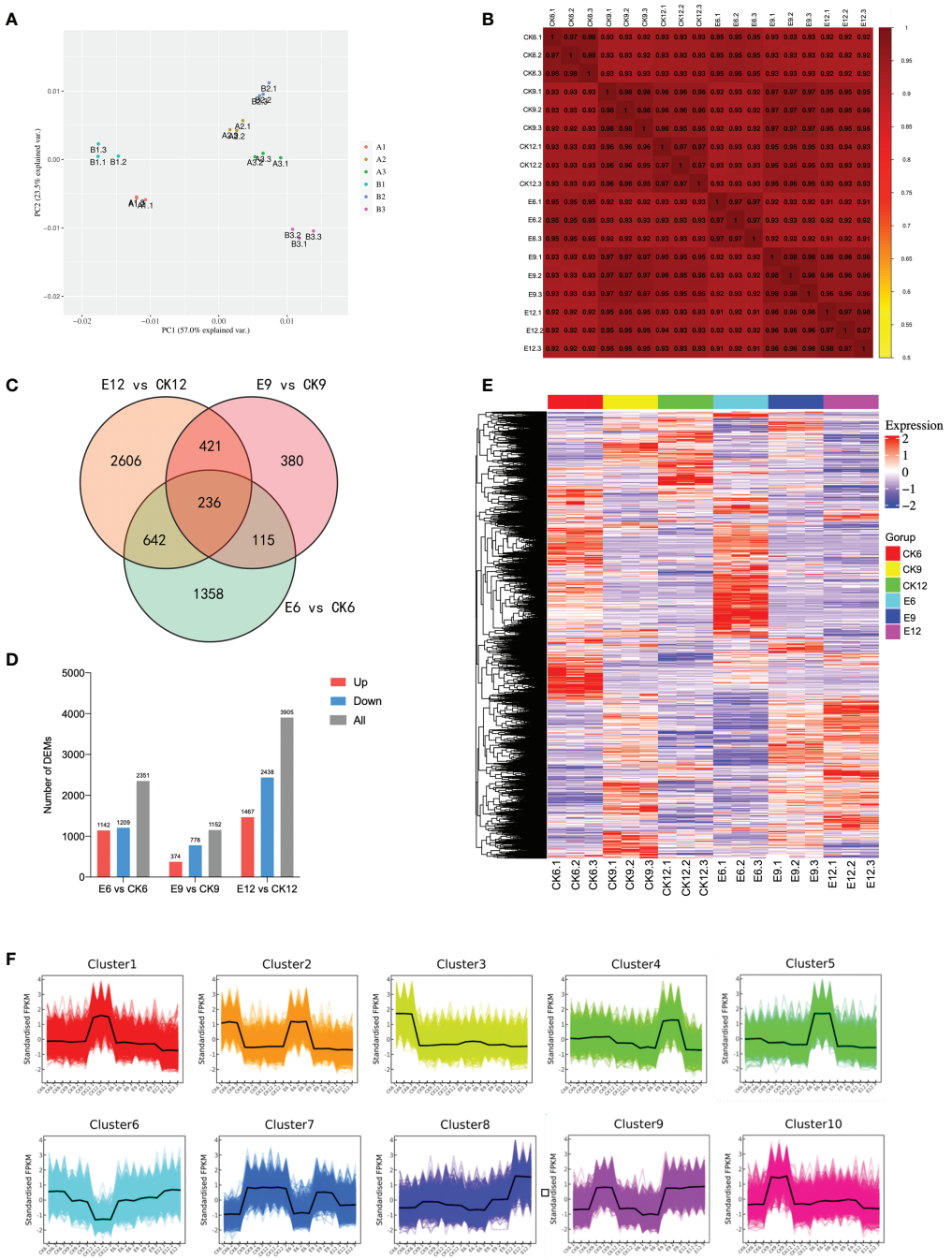


FIGURE 3
Transcriptome analysis of fruits in three periods after gibberellin treatment. **(A)** PCA analysis of gene expression under different treatments. **(B)** Spearman correlation coefficient of gene expression under different treatments. **(C)** Venn plot of commonly and exclusively expressed DEGs in the comparison of control and treatment groups at each developmental stage. **(D)** The number of DEGs compared at any two different developmental stages, the number of up-regulated and down-regulated genes is represented by the bars above and below the X-axis, respectively. **(E)** Hierarchical clustering of DEGs of all samples. **(F)** K-means clustering of DEGs expression trends, the expression profiles of genes in each cluster are represented by different colors, and the average expression profiles of all genes in each sample are represented by a black line.

3.4 Combined metabolome and transcriptome analysis

3.4.1 KEGG enrichment analysis

A KEGG pathway analysis was performed to further determine the main biochemical pathways and signal transduction pathways jointly participated by DEGs and DEMs, and the results are shown in Figure 4. The figure shows the top 10 significantly enriched metabolic pathways for each comparison group, with a total of 21 different pathways. Among them, the pathways that were significantly enriched in the top 10 in the three comparison groups were sesquiterpenoid and triterpenoid biosynthesis and plant hormone signal transduction. In the comparison group of E12 vs CK12, plant hormone signal transduction had the smallest p-value and was the most significant pathway. Meanwhile, we found that pathways, such as flavonoid biosynthesis and phenylalanine metabolism, were significantly enriched in the E6 vs CK6 combination, suggesting the probable differences in fruit coloration after GA_3 treatment. In addition, the enrichment of fructose and mannose metabolism pathways was higher in E12 vs CK12, indicating that the use of exogenous GA_S may affect the sugar content of sweet cherries.

3.4.2 Analysis of hormone anabolic pathways

Therefore, we carried out a detailed analysis of the anabolism of the five hormones, and the results are shown in Figure 5 and Table S3. In the ABA synthesis pathway, ABA is the only DEM, which gradually increases with fruit ripening. After the administration of exogenous GA_3, the content of ABA was significantly changed, especially in the latter two periods. The ABA content of E9 was 3.79 times higher than that of CK9, whereas that of E12 was 7.42 times higher than that of CK12. The expression of two *NCED* genes and one *ABA2* gene increased after the use of GA_3, which promoted the synthesis of xanthoxin and abscisic aldehyde and established sufficient precursor substances for the accumulation of ABA. At the same time, the expression of two *CYP707A* genes was inhibited after treatment, resulting in the massive accumulation of ABA. Interestingly, we found that the expressions of *lcyE* and *LUT1* genes were significantly up-regulated in E9 and E12 after GA_3 treatment, indicating that the use of exogenous GA_3 may affect the content of δ-Carotene and Lutein in fruits.

The metabolic pathway of GA indicates that exogenous GA_3 has a greater impact on the content of endogenous GA_3 than GA_4. Similar to the previous results (Figure 1A), exogenous GA_3 significantly increased the content of endogenous GA_3 in the three periods. At maturity, the expression of endogenous GA_3 in the treatment group increased by 39717583.24 compared with CK. At the same time, we found that the content of GA_7 changed significantly, thereby providing a sufficient material basis for the increase in GA_3. In this pathway, the expression of *CYP701* and *GA2ox* genes was significantly up-regulated and down-regulated, respectively, after treatment, promoting the massive synthesis of GA_7 and GA_3.

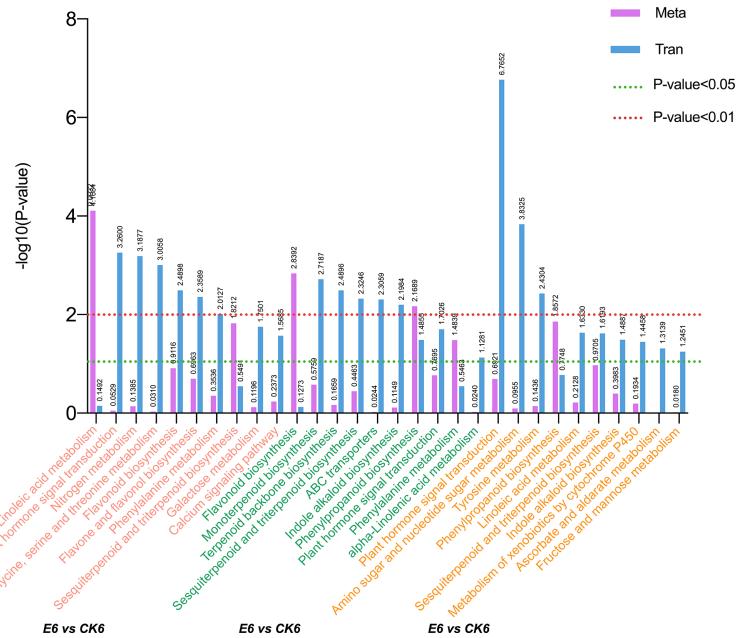


FIGURE 4
Annotation analysis of KEGG pathway in comparison groups of each period after gibberellin treatment. Select DEGs with P-value ≤ 0.5 for KEGG enrichment analysis.

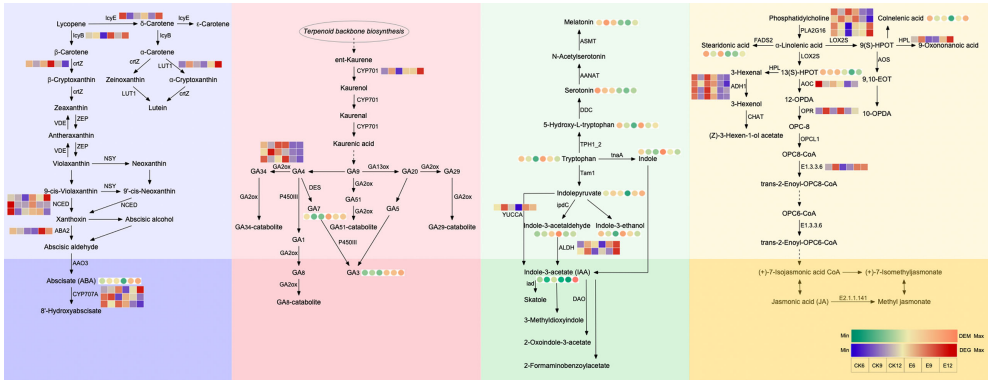


FIGURE 5
Analysis of each hormone anabolic pathway. The circle heatmap represents the expression of DEMs, and the square heatmap represents the expression of DEGs.

The DEMs in the IAA synthesis pathway are more abundant. The use of exogenous GA_3 decreased the content of Melatonin and Serotonin, but increased the contents of 5-Hydroxy-L-tryptophan, tryptophan, and indole, especially in E6. The accumulation of these substances lays a sufficient foundation for the synthesis of IAA. At the same time, the *ALDH* gene regulating IAA synthesis was highly expressed after treatment, especially the expression of E12 increased by 2.61 times compared with CK12. *YUCCA* gene expression was also up-regulated.

DEMs in the JA synthesis pathway are mainly concentrated in the anterior part. Exogenous GA_3 decreased the contents of 13 (S)-HPOT and stearidonic acid, and the expression of *AOC* and *OPR* genes was inhibited after treatment. Therefore, the repressed expression of these genes may be the main reason for the reduction of endogenous JA content.

In summary, the expression of *NCED*, *ABA2*, *CYP701*, *ALDH*, and other genes was affected after GA_3 treatment; thus, the contents of endogenous ABA, GA_3, and IAA changed.

3.5 Transcription factor family analysis

In the transcription factor family analysis, 4986 DEGs were identified as transcription factors, belonging to 55 transcription factor families. The transcription factor family mainly includes bHLH, MYB, NAC, and WRKY. The specific transcription factor family is shown in Figure 6A. Among them, transcription factor families, such as bHLH and MYB, accounted for the highest proportion, that is, 11.53% and 11.52%, respectively. Moreover, the proportion of transcription factor families, such as NAC (8.92%), ERF (6.38%), WRKY (5.17%), and B3 (4.19%) was relatively high.

Therefore, the expression patterns of these transcription factors were further analyzed, as shown in Figure 6B. Most transcription factors were expressed in each treatment, and their expression was affected by exogenous GA_3. Among them, the expression of transcription factors, such as MYB and bZIP, showed a downward trend after treatment. However, transcription factors, such as MYB and bHLH, were closely related to the synthesis of anthocyanin in fruit, indicating that the use of exogenous GA_3 may affect fruit color. Interestingly, the expression of most WRKY transcription factors was increased after exogenous GA_3 treatment, and more genes were up-regulated in E6.

3.5.1 Identification of sweet cherry *PaWRKYs*

The results obtained show that the expression of genes related to the synthesis of hormones is up-regulated after exogenous use of GA_3, thereby promoting the accumulation of these endogenous hormones. Therefore, according to the expression patterns of these genes, further screening was performed in cluster4, cluster5, cluster6, cluster8, and cluster9 to further identify the transcription factors that regulate these endogenous hormone synthesis genes. Among them, the genes of the *WRKY* family had the largest number of differential genes among the five groups, indicating that *WRKY* may be the main internal regulator in response to changes in exogenous GA_3.

According to the HMMER search results, we finally obtained 58 sequences with typical *WRKY* domains, namely, *PaWRKY1–PaWRKY58*, using the Pfam tool to identify the domains. The physicochemical properties (gene name, gene ID, amino acid size, molecular weight, theoretical isoelectric point, and subcellular localization) were analyzed according to the sequence, and the results are shown in Table 1. The results showed that the amino acid size, molecular weight, and isoelectric point of these *PaWRKY* genes exhibited great differences. The protein encoded by the *PaWRKY26* gene has the shortest amino acid length, containing only 97 amino acids, and its protein molecular mass is 11708.2 Da. The longest amino acid is PaWRKY8 protein, which contains 740 amino acids, its protein molecular mass is 79829.63 Da, and its theoretical

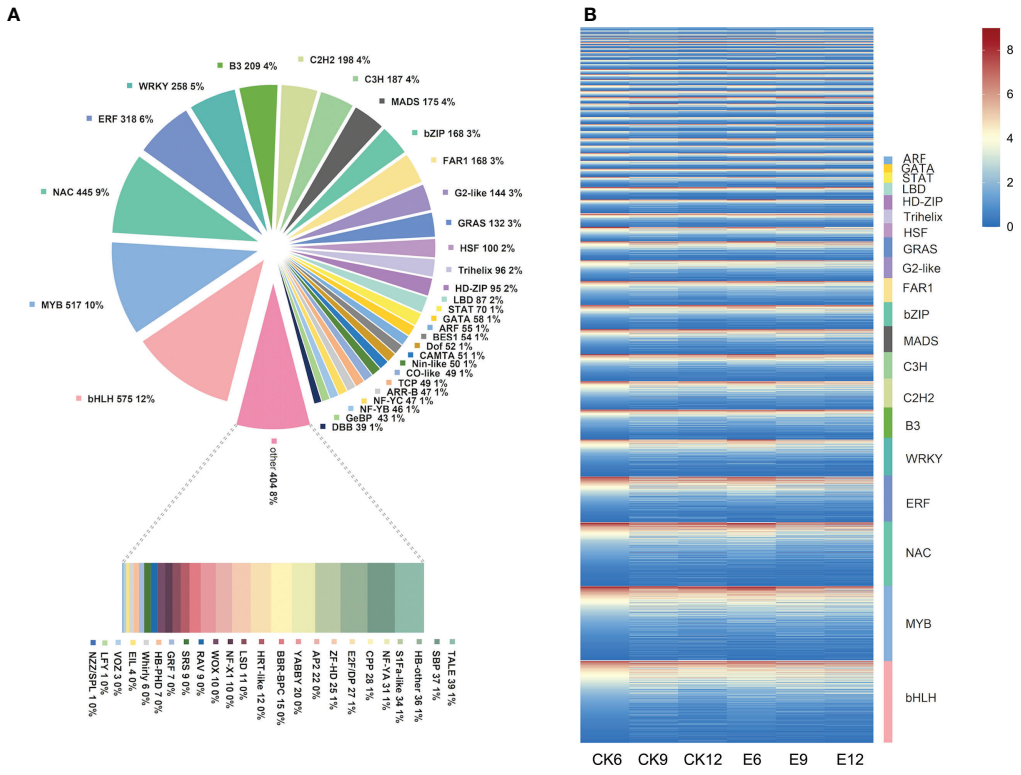


FIGURE 6
Transcription factor analysis. **(A)** Statistical summary of transcription factor families, the number and proportion of each transcription factor family are displayed in a pie chart. **(B)** Heat map of transcription factor expression patterns. The FPKM values of transcription factors were transformed according to log2, and heatmaps were constructed using Tbtools software.

isoelectric point is predicted to be 4.92–9.92. According to the subcellular localization analysis, most genes were mainly located in the nucleus, similar to the WRKY transcription factor genes reported in other species.

3.5.2 Analysis of phylogenetic relationship, gene structure, conserved motifs, and conserved elements in promoter regions

This study selected *Arabidopsis thaliana* and sweet cherry to construct a phylogenetic tree to analyze the evolutionary relationship of the *PaWRKY* gene family (Figure 7A). A total of 58 PaWRKY proteins and 70 ATWRKY proteins showed evident clustering in the phylogenetic tree. According to gene clustering, the *PaWRKY* gene family can be divided into three subfamilies, Group I, Group II, and Group III. The Group II subfamily can also be subdivided into five subfamilies: Group II-a, Group II-b, Group II-c, Group II-d, and Group II-e. Among the seven subgroups, Group II-a subgroup contains the least number of *WRKY* genes and only three in sweet cherry and *Arabidopsis*. The largest *WRKY* gene family is Group II-c, which contains 17 *PaWRKY* genes and 17 *ATWRKY* genes. The second groups containing more *WRKY* genes were Group I (10 *PaWRKYs* and 14 *ATWRKYs*) and Group III (9 *PaWRKYs* and 13 *ATWRKYs*). The number of *WRKY* transcription factor family members contained in each combination is shown in Figure 7B.

According to the gene structure analysis (Figure 7C), all 58 members of the *PaWRKY* gene family have coding regions (CDS) and untranslated regions (UTR), of which the longer intron sequence is *PaWRKY6*. At the same time, *PaWRKYs* with higher sequence similarity are similar in structure; for example, the coding regions of *PaWRKYs* belonging to Group III are similar in length and structure. Among them, the gene structure of *PaWRKYs* in Group III is more conserved than that of other subfamilies.

In addition, to better understand the conserved structure in the PaWRKY protein sequence, MEME software was used to analyze the PaWRKY protein structure, and 10 different conserved motifs were identified (Figure 7D). Most *PaWRKY* members located in the same subfamily have similar conserved motifs, which are consistent with the grouping of the family phylogenetic tree, showing a certain arrangement. All *PaWRKY* genes contain the most basic family motif 1, motif 2, and motif 3. In addition, gene members of the Group I subfamily contain motif

10.3389/fpls.2022.1041068

TABLE 1 Basic information of members of *PaWRKYs* gene family.

Gene name	Gene ID	Amino acids	Molecular weight/Da	Theoretical pI	Group	WoLF PSORT
PaWRKY1	LOC110767577	437	47414.97	4.92	II-e	nucl
PaWRKY2	LOC110749399	280	32165.39	5.08	II-d	nucl
PaWRKY3	LOC110762470	314	35398.78	5.11	II-e	nucl
PaWRKY4	LOC110751810	357	39763.26	5.15	III	nucl
PaWRKY5	LOC110764860	558	61645.64	5.2	II-b	nucl
PaWRKY6	LOC110760656	356	39964.36	5.24	III	nucl
PaWRKY7	LOC110769738	269	29407.61	5.25	II-e	nucl
PaWRKY8	LOC110765346	162	18707.37	5.34	II-c	nucl
PaWRKY9	LOC110771999	334	37718.47	5.47	III	nucl
PaWRKY10	LOC110768952	323	36275.34	5.47	III	nucl
PaWRKY11	LOC110756877	506	55912.98	5.49	I	nucl
PaWRKY12	LOC110767143	350	38306.45	5.61	III	nucl
PaWRKY13	LOC110771993	339	38129.3	5.64	III	nucl
PaWRKY14	LOC110767511	740	79829.63	5.7	I	nucl
PaWRKY15	LOC110771978	357	40735.01	5.74	III	nucl
PaWRKY16	LOC110756098	282	30821.59	5.8	II-e	nucl
PaWRKY17	LOC110768951	371	40804.43	5.81	III	nucl
PaWRKY18	LOC110751361	330	36066.52	5.81	II-c	nucl
PaWRKY19	LOC110749956	486	52935.3	5.84	I	nucl
PaWRKY20	LOC110752494	389	42694.75	5.89	II-c	nucl
PaWRKY21	LOC110760420	738	80782.4	5.93	I	nucl
PaWRKY22	LOC110752839	546	59639.64	5.95	II-b	nucl
PaWRKY23	LOC110764924	523	56802.51	6.04	II-e	nucl
PaWRKY24	LOC110758282	587	63999.36	6.07	I	nucl
PaWRKY25	LOC110764283	650	70731.02	6.16	II-b	nucl
PaWRKY26	LOC110771979	336	37449.64	6.21	III	nucl
PaWRKY27	LOC110751302	633	69129.29	6.38	II-b	nucl
PaWRKY28	LOC110752464	506	55452.73	6.52	II-b	nucl
PaWRKY29	LOC110745365	196	22119.27	6.52	II-c	nucl
PaWRKY30	LOC110745366	196	22119.27	6.52	II-c	nucl
PaWRKY31	LOC110763060	612	66386.49	6.66	II-b	nucl
PaWRKY32	LOC110766559	334	37034.8	6.68	II-c	nucl
PaWRKY33	LOC110755676	321	35641.8	6.75	II-c	nucl
PaWRKY34	LOC110753889	590	64696.37	6.77	I	nucl
PaWRKY35	LOC110767257	536	59527.31	6.82	I	nucl
PaWRKY36	LOC110754876	364	41145.25	6.87	II-c	nucl
PaWRKY37	LOC110751328	516	56174.74	7.28	I	nucl
PaWRKY38	LOC110750648	244	27715.36	7.29	II-c	nucl
PaWRKY39	LOC110750649	244	27715.36	7.29	II-c	nucl
PaWRKY40	LOC110752057	683	73473.62	7.61	II-b	nucl
PaWRKY41	LOC110762547	327	36551.11	7.62	II-a	nucl
PaWRKY42	LOC110755743	490	52847.36	7.68	II-b	nucl
PaWRKY43	LOC110751738	357	39319.46	7.68	II-e	nucl
PaWRKY44	LOC110755553	530	58044.2	7.74	I	nucl
PaWRKY45	LOC110762653	285	31618.43	8.54	II-a	nucl
PaWRKY46	LOC110763398	320	35314.43	8.77	II-a	nucl
PaWRKY47	LOC110767234	475	51742.18	8.91	I	nucl
PaWRKY48	LOC110764429	223	25562.39	8.97	II-c	nucl

(Continued)

TABLE 1 Continued

Gene name	Gene ID	Amino acids	Molecular weight/Da	Theoretical pI	Group	WoLF PSORT
PaWRKY49	LOC110772429	239	27237.52	9.03	II-c	nucl
PaWRKY50	LOC110756733	210	24000.97	9.08	II-c	pero
PaWRKY51	LOC110751840	221	24558.69	9.24	II-c	nucl
PaWRKY52	LOC110760359	170	19398.71	9.37	II-c	nucl
PaWRKY53	LOC110748572	97	11708.2	9.41	II-c	nucl
PaWRKY54	LOC110774544	342	37313.06	9.47	II-d	nucl
PaWRKY55	LOC110769338	184	20809.31	9.56	II-c	nucl
PaWRKY56	LOC110763432	355	40071.31	9.59	II-d	nucl
PaWRKY57	LOC110746394	326	35616.29	9.6	II-d	nucl
PaWRKY58	LOC110758321	286	31070.13	9.92	II-d	nucl

nucl means nucleus; pero means peroxisome.

4, Group III specifically motif 10, Group II-b specifically motif 8, and Groups II-a and II-b specifically motifs 6 and 7, respectively. This finding suggests that the reason why *WRKY* family members of different subfamilies are involved in coordinating specific processes of fruit growth and development may be that they have specific conserved structures.

To elucidate the possible regulatory mechanism of the conserved elements in the promoter region of the *PaWRKY* genes, *cis*-element analysis was performed on the 2000 bp upstream of the *PaWRKYs* gene family in sweet cherry. Nine *cis*-acting elements related to hormones and abiotic stresses were screened among the numerous response elements (Figure 7E). In terms of the number of elements, the light-responsive elements are the most, and they are distributed in the promoter regions of each *PaWRKY* gene. Every gene contains at least one hormone action element. Among them, ABA-acting and MeJA-acting elements are more abundant than other hormones.

3.5.3 Expression pattern analysis of *PaWRKYs*

The expression patterns of *PaWRKY* genes were analyzed according to the FPKM values determined by transcriptome analysis (Figure 8A and Table S4). Among them, except for *PaWRKY11*, *PaWRKY24*, *PaWRKY26*, *PaWRKY34*, *PaWRKY35*, *PaWRKY47*, and *PaWRKY51*, the expression of the remaining genes was detected in at least one group. The expression patterns of *PaWRKY* genes had high similarity in different subgroups. During sweet cherry development, the expression of most *PaWRKY* genes increased with fruit growth,

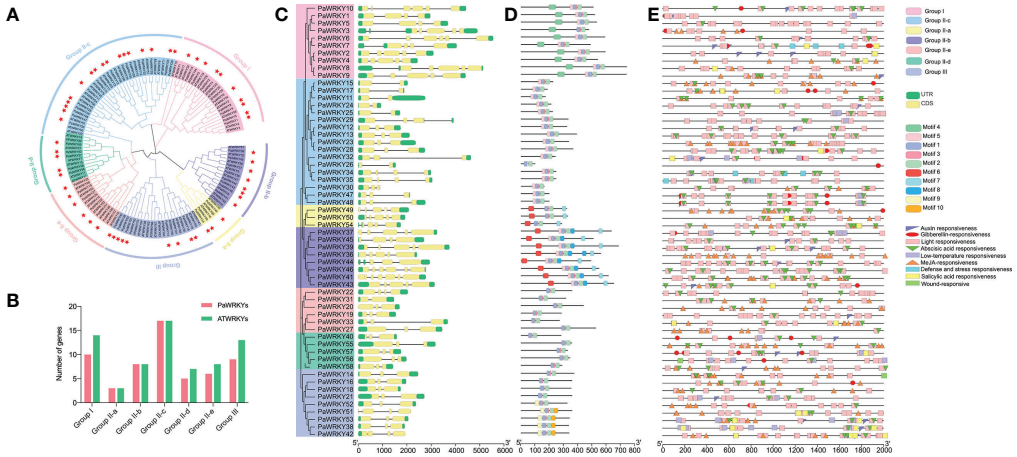


FIGURE 7
WRKY gene analysis of sweet cherry. **(A)** The phylogenetic relationship of the WRKY gene in sweet cherry. After aligning the sequences with Clustal W, a phylogenetic tree was constructed according to the neighbor-joining method. **(B)** The number statistics of *PaWRKYs* and *AtWRKYs* in each subfamily. **(C)** WRKY of sweet cherry Gene families were constructed based on the alignment of full-length amino acid sequences to construct a neighbor-joining tree and analyze gene structure. **(D)** Conserved motif analysis of the *WRKY* gene family in sweet cherry. **(E)** Analysis of functional elements in the promoter region of the *WRKY* gene family in sweet cherry.

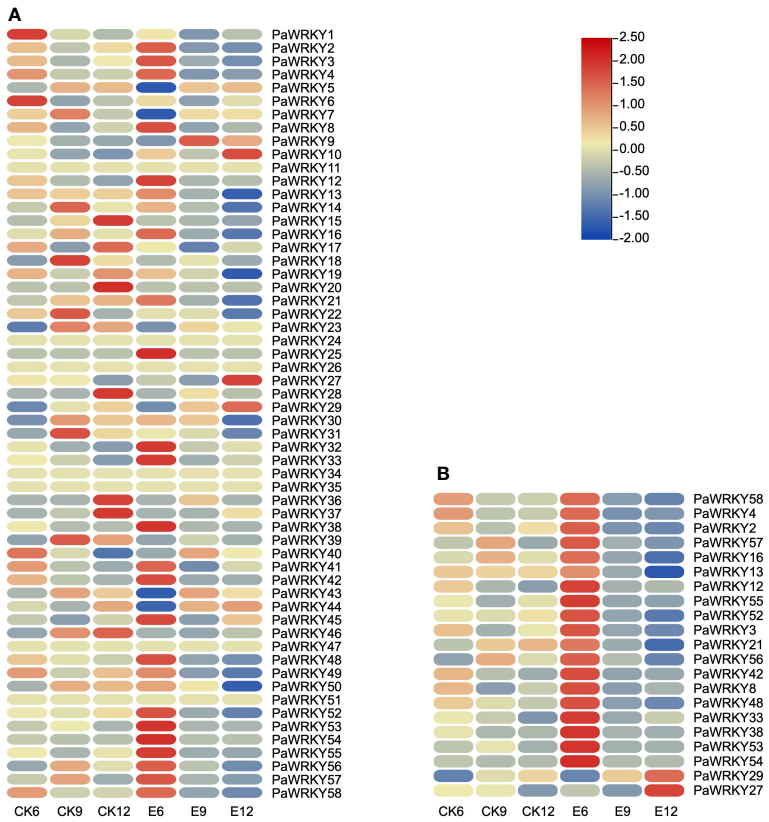


FIGURE 8
Expression profile of *PaWRKYs* genes. **(A)** The expression profile of *PaWRKYs* gene in three periods after treatment, and the FPKM values of transcription factors were transformed according to log2. **(B)** The relative expression levels of differentially expressed *PaWRKYs* family members after gibberellin treatment. The FPKM values of transcription factors were transformed and plotted according to log2.

such as *PaWRKY5*, *PaWRKY15*, and *PaWRKY37*. In CK, the expression of *PaWRKYs* was generally low, and after treatment, the expression of most genes was up-regulated.

The up-regulated expression of these genes may be positive regulators of endogenous hormone synthesis. The differentially expressed *PaWRKY* genes were screened in the transcriptome, and the results are shown in Figure 8B. The figure shows that exogenous use of GA_3 generally up-regulated the expression level of the 6th stage and inhibited the expression of the 9th and 12th stages. On the contrary, we found that the high expression of *PaWRKY27* and *PaWRKY29* was at E12, consistent with the changing pattern of *NCED*, *CYP701*, and other genes.

3.5.4 Protein interaction analysis of *PaWRKYs*

The cognate *WRKYs* of *Arabidopsis thaliana* were used to predict the protein interaction network of PaWRKYs (Figure 9A, see the Tables S5 and S6 for detailed annotation information in the figure). The results showed that most PaWRKY proteins interact with multiple proteins, of which 15 proteins can interact with more than four other PaWRKY proteins. For example, PaWRKY2 and PaWRKY4 are expected to interact directly with 13 WRKY proteins, namely, PaWRKY58 (AT4G31550.1 ortholog), PaWRKY18 (AT4G23810.1 ortholog), PaWRKY16 (AT4G23810.1 ortholog), PaWRKY21 (AT2G46400.1 ortholog), and others.

Figure 9A shows that differential genes have more complex relationships in protein interaction prediction. Therefore, we selected PaWRKYs with complex interactions in Figure 9A to further predict protein interaction networks (Figure 9B). These differential PaWRKYs (PaWRKY2, PaWRKY4, PaWRKY12, PaWRKY13, PaWRKY16, PaWRKY21, PaWRKY29, PaWRKY38, PaWRKY48, PaWRKY52, PaWRKY53, PaWRKY54, and PaWRKY58) can directly interact with various proteins, such as MKS1, SIB1, and ACS6. Among them, PaWRKY2 and PaWRKY4 (AT4G31550.1 ortholog) interact directly with 13 proteins, indicating that they may have important regulatory roles. In addition, PaWRKY21, PaWRKY16, PaWRKY53, and PaWRKY58 interact with stress-related proteins, such as MKS1 and MPK4, in response to growth and developmental changes. Overall, the predicted network provides an important reference for functional studies of PaWRKY proteins.

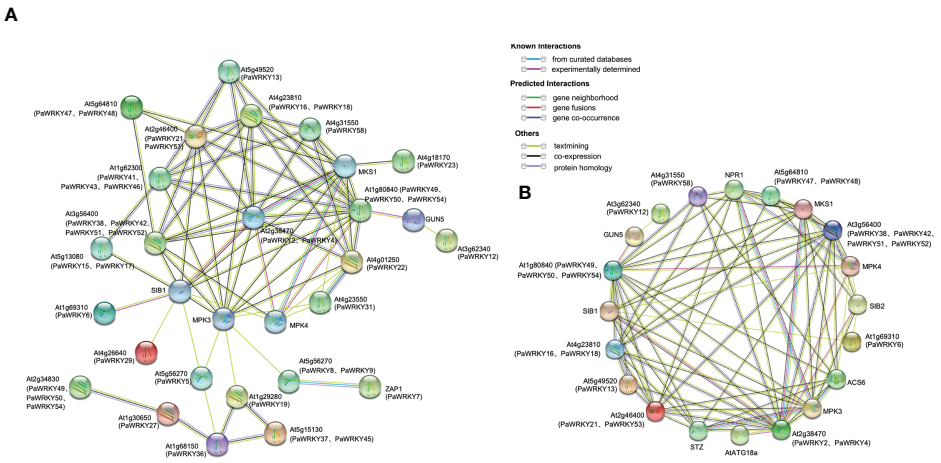


FIGURE 9
Protein interaction network. **(A)** Protein interaction network based on *Arabidopsis* homologous PaWRKYs. Abbreviated names are genes reported in *Arabidopsis*. The PaWRKYs proteins homologous to *Arabidopsis* are shown in parentheses. **(B)** PaWRKYs protein interaction network prediction based on differential expression of homology in *Arabidopsis*. Abbreviated names are genes reported in *Arabidopsis*. The PaWRKYs proteins homologous to *Arabidopsis* are shown in parentheses.

3.6 Correlation network analysis

Based on these studies, the *PaWRKY* genes may play an important role in the growth of sweet cherries. Therefore, the correlation analysis of transcription factors, hormone synthesis-related genes, and related metabolites was performed using Cytoscape software, and the results are shown in Figure 10. The figure shows that these *PaWRKYs* are closely related to the structural genes involved in the synthesis of various hormones. Among them, *PaWRKY29* has a significant positive correlation with ABA and Indolepyruvate and has a strong correlation with *NCED*. It is a potential positive regulator of ABA and IAA synthesis. Similarly, *PaWRKY27* may have the same effect. In addition, *PaWRKY38* was positively correlated with *crtZ*, and the expression of *crtZ* was inhibited by exogenous GA_3, suggesting that *PaWRKY38* may inhibit ABA synthesis. Moreover, *PaWRKY38* was significantly negatively correlated with *GA2ox2*, which inhibited the decomposition of GA_4 and promoted the accumulation of GA_3. Interestingly, the differential genes belonging to Group III (*PaWRKY16*, *PaWRKY21*, *PaWRKY38*, *PaWRKY52*, and *PaWRKY53*) were significantly negatively correlated with indolepyruvate, and their correlations ranged from −0.87 to −0.75. These results suggest that genes within Group III are more sensitive to the use of exogenous GA_3.

3.7 qRT-PCR validation

Key genes sensitive to exogenous GA_3 were selected from each pathway for qRT-PCR analysis to verify the validity of the transcripts, and the results are shown in Figure 11. The use of exogenous GA_3 increased the expression of *NCED1* and *ABA2*, thereby increasing the synthesis of ABA. In addition, two *CYP707A* genes were inhibited, and the decomposition of ABA was inhibited. The significant decrease in the expression of *GA2ox* after treatment may be the main reason for the accumulation of GA_3. The direct regulatory genes *YUCCA* and *ALDH* in the IAA synthesis pathway were highly expressed at E12, which in turn promoted the increase in IAA. JA synthesis-related genes were less sensitive to exogenous GA_3, but were generally inhibited and showed low expression. At the same time, some *WRKY* genes of *PaWRKY27*, *PaWRKY29*, and Group III were selected for qRT-PCR verification. Overall, the qRT-PCR results for most structural genes and transcription factors were consistent with the transcriptome data, indicating a high level of confidence in the transcript data.

4 Discussion

Self-incompatibility is a serious problem in sweet cherry production, affecting fruit yield and restricting the development of the planting industry (Wu et al., 2013). Artificial pollination can solve this production problem, but the accompanying labor costs increase the cost of sweet cherry production. The study found that exogenous use of GA_3 can improve fruit set, reduce labor costs, and effectively solve the cost problem of sweet cherry production (Devasirvatham and Tan, 2022). Aliyu et al. (2011) found that GA_3 was the most effective phytohormone to improve fruit yield in cashew nuts treated with various

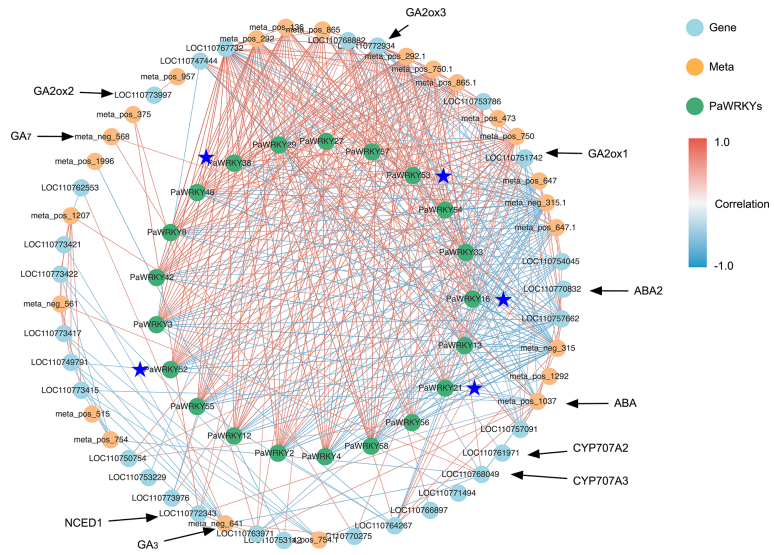


FIGURE 10
Correlation network analysis. The blue points represent genes, the yellow points represent metabolites, and the green points represent *PaWRKYs*; the red line represents positive correlation, the blue line represents negative correlation, and the darker the color, the higher the correlation; the genes marked with asterisk belong to *PaWRKYs* of Group III.

phytohormones at the flowering stage. In addition, similar conclusions are found on sweet cherries. Exogenous use of GA_3 can significantly promote parthenocarpic fruit and increase fruit yield (Wen et al., 2019). These results showed the importance of GA_3 to the growth of sweet cherry fruit.

In plants, GA_3 is a plant hormone with various biological functions. It can not only stimulate plant growth and development, but also mediate various stress responses (Hou et al., 2013). However, the fruit quality also changed after the exogenous use of GA_3. When pineapples were treated with GA_3 at the flowering stage, the higher concentration indicates higher volume of pulp cells, and the increase in fruit weight becomes more significant (Li et al., 2011). In citrus production, the method of spraying GA_3 is often used to reduce the phenomenon of peeling and puffing of citrus, but it brings the problem of the delayed coloring of citrus (Ma et al., 2021). Liu et al. (2022) treated apples with exogenous GA to control cell growth at flowering and young fruit stages, thereby reducing

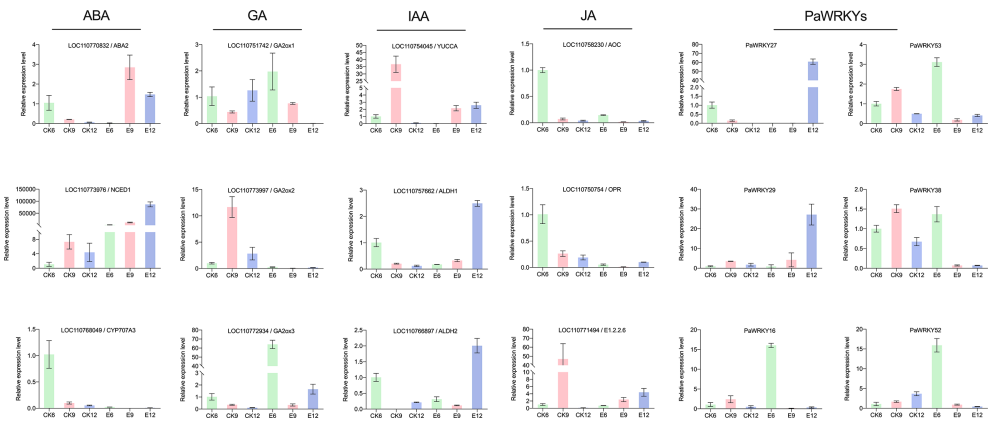


FIGURE 11
qRT-PCR analysis of genes and transcription factors related to hormone synthesis in sweet cherry after exogenous GA_3 treatment.

asymmetric fruit. Chen et al. (2020) found that by regulating the level of GA_3, the shape and ripening of tomato fruit can be affected. The study found that exogenous use of GA_3 often affects fruit quality by affecting the content of endogenous hormones. For example, exogenous use of GA_3 increased endogenous GA_3 content in citrus and inhibited the rate of fruit browning (Cai et al., 2021). Liu et al. (2011) found that exogenous GA_3 could regulate the growth of tiller buds by changing the endogenous ABA, IAA, and ZR contents of rice plants. Therefore, understanding the changing pattern of endogenous hormones after exogenous use of GA_3 is greatly important to explore the mechanism of fruit quality changes after GA3 treatment. In this study, after exogenous use of GA_3, the content of endogenous ABA, GA_3, GA_4, IAA, and JA changed with the concentration of use, and the hormone with the most drastic change was ABA. The contents of ABA and GA_3 increased significantly after exogenous GA_3 treatment, whereas the contents of GA_4 and JA were inhibited. Only when the treatment concentration was 60 mg/L, did the content of IAA increase. These results indicated that the effect of exogenous GA_3 on fruit yield and quality may be through changing the content of endogenous hormones. This is a systematic study on the changing pattern of endogenous hormones after administering exogenous GA_3. This provides a theoretical reference for the application of GA_3 in the production of sweet cherries and some ideas for further research on the effect of GA_3 on the quality of stone fruit.

We further used the combined transcriptome and metabolome analysis to determine the relevant metabolites and genes in response to exogenous GA_3 and identified a total of 2256 DEMs and 10154 DEGs. A large number of metabolites and genes jointly respond to the use of exogenous GA3 to regulate the content of endogenous hormones, thereby affecting the growth and development of sweet cherry fruit and changing fruit quality (Information on the differential metabolites related to the synthesis and metabolism of fruit sugars, acids and flavonoids is shown in Table S7. In this study, the expression levels of *NCED* and *ABA2* were significantly up-regulated, thereby directly affecting the accumulation of endogenous ABA. Similar findings were also found in grapes, where *NCED* and ABA3 affect fruit quality by participating in ABA biosynthesis (Li et al., 2021). At the same time, the study found that exogenous use of allantoin (Moriyama et al., 2020), acetic acid (Sun et al., 2022), and synthetic strigolactone analogue (Ferrero et al., 2018) improved fruit quality by affecting the expression of the *NCED* gene. The finding indicates that *NCED* is highly sensitive to external conditions and is the main regulator of ABA synthesis. In the GA pathway, the expression of the three *GA2ox* genes that regulate GA_4 breakdown was significantly inhibited during maturation, and the expression of the *CYP701* gene was significantly upregulated after treatment, indicating that *GA2ox* and *CYP701* cooperate to promote the massive accumulation of GA_7 and GA_3. In the study of alkaline stress, the authors found that alkaline stress also changed the content of endogenous GA_3 by changing the expression of *GA2ox* (Ma et al., 2022). Bermejo et al. (2018) also made similar conclusions in the study of strawberry, the expression of *GA2ox* was significantly changed after exogenous use of IAA. Similarly, exogenous use of GA_3 and paclobutrazol modulates anthocyanin accumulation in Arabidopsis by affecting *GA2ox* (Zhang et al., 2017). In tomatoes, specific overexpression of SlGA2ox1 reduced endogenous GA concentrations in the fruit (Chen et al., 2016). In the study of IAA synthesis, we found that after exogenous use of GA_3, a large number of DEMs appeared, whereas DEGs were less, indicating the importance of *YUCCA* and *ALDH* genes. Similar findings were also found in loquat (Jiang et al., 2016). After exogenous use of GA_3, the expression of the *YUCCA* gene was up-regulated, increasing the IAA content. Compared with the four other hormones, the content of JA was significantly inhibited and down-regulated. The use of exogenous GA_3 mainly acts on the genes in front of the JA synthesis pathway, such as *OPR*, *AOC*, and *E1.3.3.6*. In Arabidopsis, *AOC* and *OPR* are also key regulatory genes for JA synthesis (Leon-Reyes et al., 2010).

Transcription factor family molecules indicated that *WRKY* transcription factors were greatly affected by exogenous GA_3, and the difference ratio was higher. *WRKY* transcription factors play crucial roles in regulating various plant growth and developmental processes. However, the *WRKY* gene family of the sweet cherry has not been widely studied, and its role remains to be explored. Therefore, we systematically analyzed the *WRKY* gene family of sweet cherry. The 58 *PaWRKY* gene families were studied in detail by analyzing the phylogeny, gene structure, promoter region, and sequence characteristics. However, the number of *WRKY* genes in sweet cherries was less than that in apple (127) (Meng et al., 2016), kiwifruit (97) (Jing and Liu, 2018), and pear (103) (Huang et al., 2015), which may be caused by the differences between plant genomes. We classified the 58 PaWRKY proteins into seven subfamilies on the basis of phylogeny. Similar findings were also found in the study of strawberry FaWRKY protein, and the 47 *FaWRKY* genes were divided into seven subfamilies (Chen and Liu, 2019). Gene structure and conserved motif analysis of *PaWRKYs* showed that genes belonging to the same subfamily had similar exon and intron organization and similar conserved motifs. These results suggest that *PaWRKYs* of the same subfamily are closely related in evolution. Among the seven subgroups, Group II contained the largest number of subgroups and genes, indicating its variability. According to previous research results, Group III gene members have the highest activity and play an important role in plant evolution (Wu et al., 2005). In this study, although Group III genes had fewer members, most genes showed differences after the use of GA_3, suggesting that it may be a potential regulator of fruit growth. In addition, the protein function prediction indicated that *PaWRKY16*, *PaWRKY21*, and *PaWRKY53* could directly interact with multiple proteins to regulate fruit growth.

WRKY gene also plays an important role in regulating the content of plant endogenous hormones. Wang et al. (2016) isolated a new *WRKY* gene *CsWRKY2* from Camellia, which can participate in the signaling pathway of ABA synthesis, regulate ABA synthesis, and further improve plant defense against cold and drought stress. *CaWRKY40*, found in pepper, can be induced by JA-mediated signaling mechanisms, coordinating call responses to heat stress (Dang et al., 2013). In tobacco, overexpression of *NtWRKY50* resulted in altered JA levels and increased plant resistance (Liu et al., 2017). *WRKY* gene can cooperate with structural genes related to endogenous hormone synthesis to influence endogenous hormones. For example, Yan et al. (2014) found that the cotton transcription factor *GhWRKY17* can reduce ABA levels by inhibiting the expression of the *NCED* gene, thereby regulating the sensitivity to drought. Similarly, *GhWRKY1* was found to interact with the "W-box" *cis*-elements of the promoters of *AtNCED2*, *AtNCED5*, *AtNCED6*, and *AtNCED9* in Arabidopsis to promote ABA biosynthesis (Hu et al., 2021). In Pyrus betulaefolia, *PbrWRKY53* can bind to the W-box element in the promoter region of *PbrNCED1* to promote the synthesis of vitamin C and ABA, thereby improving drought tolerance (Liu et al., 2019). In addition, *WRKY* was found to be involved in the IAA signaling process in plants (Jin et al., 2018). The correlation between the *WRKY* gene and IAA metabolites or genes in this study also indicates that the *PaWRKY* gene in sweet cherries may be involved in the IAA signaling process

In summary, the *WRKY* gene can participate in the signal response process of various plant hormones, such as ABA and JA, and is widely involved in the growth and development of plants. In this study, potential regulatory genes in response to exogenous GA_3 changes were initially screened by a combination of transcription and metabolism methods, which laid the foundation for sweet cherry fruit production. However, the specific regulatory mechanism of *WRKY* transcription factor and endogenous hormone synthesis related structural genes is still unclear and requires further analysis.

Data availability statement

The datasets presented in this study can be found in online repositories and Supplementary Material. The metabolome and transcriptome proposed in the study are deposited in the National Genomics Data Center database. You can query the metabolome data by visiting the link (https://ngdc.cncb.ac.cn/omix/release/OMIX001762) (BioProject: PRJCA010046; Accession number: OMI001762); You can query transcriptome data by visiting the link (https://ngdc.cncb.ac.cn/gsa/browse/CRA007287) (BioProject: PRJCA010046; accession number: CRA007287).

Author contributions

RG and CC: conceptualization. CC and HC: data curation. YC and WY: formal analysis. CC, WT, HS, and YZ: investigation. CC and HC: software. CC: writing–original draft. ML, BS, RG, and HS: writing–editing. RG: supervision. All authors read and approved the final manuscript.

Funding

This research was funded by the Sichuan Science and Technology Plan Project (Key R&D Project) (2021YFN0081, 2021YFN0082). The funders had no rolein the design of the study in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in thedecision to publish the results.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1041068/full#supplementary-material

References

Aliyu, O. M., Adeigbe, O. O., and Awopetu, J. A. (2011). Foliar application of the exogenous plant hormones at pre-blooming stage improves flowering and fruiting in cashew (*Anacardium occidentale* l.). *J. Crop Sci. Biotechnol.* 14, 143–150. doi: 10.1007/s12892-010-0070-3

Askarieh, A., Suleiman, S., and Tawakalna, M. (2021). Sweet cherry (*Prunus avium* l.) fruit drop reduction by plant growth regulators (Naphthalene acetic acid NAA and gibberellic acid GA3). *Am. J. Plant Sci.* 12, 1338–1346. doi: 10.4236/ajps.2021.129094

Bawa, G., Feng, L., Chen, G., Chen, H., Hu, Y., Pu, T., et al. (2020). Gibberellins and auxin regulate soybean hypocotyl elongation under low light and high-temperature interaction. *Physiologia plantarum.* 170, 345–356. doi: 10.1111/ppl.13158

Bermejo, A., Granero, B., Mesejo, C., Reig, C., Tejedo, V., Agustí, M., et al. (2018). Auxin and gibberellin interact in citrus fruit set. *J. Plant Growth Regul.* 37, 491–501. doi: 10.1007/s00344-017-9748-9

Cai, N., Chen, C., Wan, C., and Chen, J. (2021). Effects of pre-harvest gibberellic acid spray on endogenous hormones and fruit quality of kumquat (*Citrus japonica*) fruits. *New Z. J. Crop Hortic. Science.* 49, 211–224. doi: 10.1080/01140671.2020.1806084

Chen, P., and Liu, Q. (2019). Genome-wide characterization of the *WRKY* gene family in cultivated strawberry (*Fragaria× ananassa* duch.) and the importance of several group III members in continuous cropping. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-44479-7

Chen, S., Wang, X. J., Tan, G. F., Zhou, W. Q., and Wang, G. L. (2020). Gibberellin and the plant growth retardant paclobutrazol altered fruit shape and ripening in tomato. *Protoplasma.* 257, 853–861. doi: 10.1007/s00709-019-01471-2

Chen, S., Wang, X. J., Zhang, L. Y., Lin, S. S., Liu, D. C., Wang, Q. Z., et al. (2016). Identification and characterization of tomato gibberellin 2-oxidases (GA2oxs) and effects of fruit-specific *SlGA2ox1* overexpression on fruit and seed growth and development. *Horticulture Res.* 3, 16059. doi: 10.1038/hortres.2016.59

Dang, F. F., Wang, Y. N., Yu, L., Eulgen, T., Lai, Y., Liu, Z. Q., et al. (2013). *CaWRKY40*, a WRKY protein of pepper, plays an important role in the regulation of tolerance to heat stress and resistance to ralstonia solanacearum infection. *Plant Cell Environment.* 36, 757–774. doi: 10.1111/pce.12011

Devasirvatham, V., and Tan, D. K. Y. (2022). Key determinants of the physiological and fruit quality traits in sweet cherries and their importance in a breeding programme. *Horticulturae.* 8, 694. doi: 10.3390/horticulturae8080694

Faienza, M. F., Corbo, F., Carocci, A., Catalano, A., Clodoveo, M. L., Grano, M., et al. (2020). Novel insights in health-promoting properties of sweet cherries. *J. Funct. foods.* 69, 103945. doi: 10.1016/j.jff.2020.103945

Ferrero, M., Pagliarani, C., Novák, O., Ferrandino, A., Cardinale, F., Visentin, I., et al. (2018). Exogenous strigolactone interacts with abscisic acid-mediated accumulation of anthocyanins in grapevine berries. *J. Exp. botany.* 69, 2391–2401. doi: 10.1093/jxb/ery033

Hou, X., Ding, L., and Yu, H. (2013). Crosstalk between GA and JA signaling mediates plant growth and defense. *Plant Cell Rep.* 32, 1067–1074. doi: 10.1007/s00299-013-1423-4

Huang, X., Li, K., Xu, X., Yao, Z., Jin, C., and Zhang, S. (2015). Genome-wide analysis of *WRKY* transcription factors in white pear (*Pyrus bretschneideri*) reveals evolution and patterns under drought stress. *BMC Genomics* 16, 1–14. doi: 10.1186/s12864-015-2233-6

Hu, Q., Ao, C., Wang, X., Wu, Y., and Du, X. (2021). *GhWRKY1-like*, a *WRKY* transcription factor, mediates drought tolerance in *Arabidopsis via* modulating ABA biosynthesis. *BMC Plant Biol.* 21, 458. doi: 10.1186/s12870-021-03238-5

Jaillais, Y., and Chory, J. (2010). Unraveling the paradoxes of plant hormone signaling integration. *Nat. Struct. Mol. Biol.* 17, 642–645. doi: 10.1038/nsmb0610-642

Jiang, S., Luo, J., Xu, F., and Zhang, X. (2016). Transcriptome analysis reveals candidate genes involved in gibberellin-induced fruit setting in triploid loquat (*Eriobotrya japonica*). *Front. Plant science.* 7. doi: 10.3389/fpls.2016.01924

Jing, Z., and Liu, Z. (2018). Genome-wide identification of *WRKY* transcription factors in kiwifruit (*Actinidia* spp.) and analysis of *WRKY* expression in responses to biotic and abiotic stresses. *Genes Genomics* 40, 429–446. doi: 10.1007/s13258-017-0645-1

Jin, W., Zhou, Q., Wei, Y., Yang, J., Hao, F., Cheng, Z., et al. (2018). *NtWRKY-R1*, a novel transcription factor, integrates IAA and JA signal pathway under topping damage stress in *Nicotiana* tabacum. *Front. Plant science.* 8. doi: 10.3389/fpls.2017.02263

Kivistik, A., Jakobson, L., Kahu, K., and Laanemets, K. (2022). Wild and rare self-incompatibility allele S17 found in 24 sweet cherry (Prunus avium l.) cultivars. *Plant Mol. Biol. Reporter.* 40, 376–388. doi: 10.1007/s11105-021-01327-1

Kou, E., Huang, X., Zhu, Y., Su, W., Liu, H., Sun, G., et al. (2021). Crosstalk between auxin and gibberellin during stalk elongation in flowering Chinese cabbage. *Sci. Rep.* 11, 1–9. doi: 10.1038/S41598-021-83519-Z

Kuhn, N., Ponce, C., Arellano, M., Time, A., Sagredo, B., Donoso, J. M., et al. (2020). Gibberellic acid modifies the transcript abundance of ABA pathway orthologs and modulates sweet cherry (*Prunus avium*) fruit ripening in early- and mid-season varieties. *Plants.* 9, 1796. doi: 10.3390/plants9121796

Leon-Reyes, A., van der Does, D., De, Lange, E. S., et al. (2010). Salicylate-mediated suppression of jasmonate-responsive gene expression in *Arabidopsis* is targeted downstream of the jasmonate biosynthesis pathway. *Planta.* 232, 1423–1432. doi: 10.1007/s00425-010-1265-z

Liao, X., Li, M., Liu, B., Yan, M., Yu, X., Zi, H., et al. (2018). Interlinked regulatory loops of ABA catabolism and biosynthesis coordinate fruit growth and ripening in woodland strawberry. *Proc. Natl. Acad. Sci.* 115, E11542–E11550. doi: 10.1073/pnas.1812575115

Lin, Q., Zhang, Z., Wu, F., Feng, M., Sun, Y., Chen, W., et al. (2020). The APC/CTE E3 ubiquitin ligase complex mediates the antagonistic regulation of root growth and tillering by ABA and GA. *Plant Cell.* 32, 1973–1987. doi: 10.1105/tpc.20.00101

Li, D., Pang, Y., Li, H. , Wang, R., and Ma, C. (2021). Comparative analysis of the gene expression profile under two cultivation methods reveals the critical role of ABA in grape quality promotion. *Scientia Horticulturae.* 281, 109924. doi: 10.1016/j.scienta.2021.109924

Liu, Y., Ding, Y. F., Wang, Q. S., Guo, D., Wang, R., Ma, C., et al. (2011). Effect of plant growth regulators on growth of rice tiller bud and changes of endogenous hormones. *Acta Agronomica Sinica.* 37, 670–676. doi: 10.1016/S1875-2780(11)60019-9

Liu, Q., Liu, Y., Tang, Y., Chen, J., and Ding, W. (2017). Overexpression of *NtWRKY50* increases resistance to ralstonia solanacearum and alters salicylic acid and jasmonic acid production in tobacco. *Front. Plant science.* 8. doi: 10.3389/fpls.2017.01710

Liu, C., Xiao, P., Jiang, F., Wang, S., Liu, Z., Song, G., et al. (2022). Exogenous gibberellin treatment improves fruit quality in self-pollinated apple. *Plant Physiol. Biochem.* 174, 11–21. doi: 10.1016/j.plaphy.2022.01.029

Liu, Y., Yang, T., Lin, Z., Gu, B., Xing, C., Zhao, L., et al. (2019). A *WRKY* transcription factor *PbrWRKY53* from *Pyrus betulaefolia* is involved in drought tolerance and AsA accumulation. *Plant Biotechnol. J.* 17, 1770–1787. doi: 10.1111/pbi.13099

Li, Y. H., Wu, Y. J., Wu, B., Zou, M. H., Zhang, Z., and Sun, G. M. (2011). Exogenous gibberellic acid increases the fruit weight of 'Comte de paris' pineapple by enlarging flesh cells without negative effects on fruit quality. *Acta Physiol. Plant* 33, 1715–1722. doi: 10.1007/s11738-010-0708-2

Ma, Q., Yuan, Y., Wu, E., Wang, H., Dang, K., Feng, Y., et al. (2022). Endogenous bioactive gibberellin/abscisic acids and enzyme activity synergistically promote the phytoremediation of alkaline soil by broomcorn millet (Panicum miliaceum l.). *J. Environ. Management.* 305, 114362. doi: 10.1016/j.jenvman.2021.114362

Ma, G., Zhang, L., Kudaka, R., Inaba, H., Furuya, T., Kitamura, M., et al. (2021). Exogenous application of ABA and NAA alleviates the delayed coloring caused by puffing inhibitor in citrus fruit. *Cells.* 10, 308. doi: 10.3390/cells10020308

Meng, D., Li, Y., Bai, Y., Li, M., and Cheng, L. (2016). Genome-wide identification and characterization of *WRKY* transcriptional factor family in apple and analysis of their responses to waterlogging and drought stress. *Plant Physiol. Biochem.* 103, 71–83. doi: 10.1016/j.plaphy.2016.02.006

Mesejo, C., Yuste, R., Reig, C., Martínez-Fuentes, A., Iglesias, D. J., Muñoz-Fambuena, N., et al. (2016). Gibberellin reactivates and maintains ovary-wall cell division causing fruit set in parthenocarpic citrus species. *Plant Science.* 247, 13–24. doi: 10.1016/j.plantsci.2016.02.018

Moriyama, A., Nojiri, M., Watanabe, G., Enoki, S., and Suzuki, S. (2020). Exogenous allantoin improves anthocyanin accumulation in grape berry skin at early stage of ripening. *J. Plant Physiol.* 253, 153253. doi: 10.1016/j.jplph.2020.153253

Niu, Q., Wang, T., Li, J., Yang, Q., Qian, M., and Teng, Y. (2015). Effects of exogenous application of GA_{4+7} and n-(2-chloro-4-pyridyl)-N′-phenylurea on induced parthenocarpy and fruit quality in *Pyrus pyrifolia* 'Cuiguan'. *Plant Growth Regul.* 76, 251–258. doi: 10.1007/s10725-014-9995-8

Ozkan, Y., Ucar, M., Yildiz, K., and Ozturk, B. (2016). Pre-harvest gibberellic acid (GA3) treatments play an important role on bioactive compounds and fruit quality of sweet cherry cultivars. scientia. *Horticulturae.* 211, 358–362. doi: 10.1016/j.scienta.2016.09.019

Papapetros, S., Louppis, A., Kosma, I., Kontakos, S., Badeka, A., and Kontominas, M. G. (2018). Characterization and differentiation of botanical and geographical origin of selected popular sweet cherry cultivars grown in Greece. *J. Food Composition Analysis.* 72, 48–56. doi: 10.1016/j.jfca.2018.06.006

Rachappanavar, V., Padiyal, A., Sharma, J. K., and Gupta, S. K. (2022). Plant hormone-mediated stress regulation responses in fruit crops-a review. scientia. *Horticulturae.* 304, 111302. doi: 10.1016/j.scienta.2022.111302

Sun, T., Zhang, J., Zhang, Q., Li, X., Li, M., Yang, Y., et al. (2022). Exogenous application of acetic acid enhances drought tolerance by influencing the MAPK signaling pathway induced by ABA and JA in apple plants. *Tree Physiol.* 42, 1827–40. doi: 10.1093/treephys/tpac034

Tijero, V., Teribia, N., and Munné-Bosch, S. (2019). Hormonal profiling reveals a hormonal cross-talk during fruit decay in sweet cherries. *J. Plant Growth Regulation.* 38, 431–437. doi: 10.1007/s00344-018-9852-5

Tiwari, A., Offringa, R., and Heuvelink, E. (2012). Auxin-induced fruit set in *Capsicum annuum* l. requires downstream gibberellin biosynthesis. *J. Plant Growth Regulation.* 31, 570–578. doi: 10.1007/s00344-012-9267-7

Torres, C. A., Sepúlveda, G., and Kahlaoui, B. (2017). Phytohormone interaction modulating fruit responses to photooxidative and heat stress on apple (Malus domestica borkh.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02129

Wang, Y., Shu, Z., Wang, W., Jiang, X., Li, D., Pan, J., et al. (2016). *CsWRKY2*, a novel *WRKY* gene from *Camellia sinensis*, is involved in cold and drought stress responses. *Biol. Plant* 60, 443–451. doi: 10.1007/s10535-016-0618-2

Watanabe, M., Segawa, H., Murakami, M., Sagawa, S., and Komori, S.. (2008). Effects of plant growth regulators on fruit set and fruit shape of parthenocarpic apple fruits. *J. Japanese Soc. Hortic. Science.* 77, 350–357. doi: 10.2503/jjshs1.77.350

Wen, B., Song, W., Sun, M., Chen, M., Mu, Q., Zhang, X., et al. (2019). Identification and characterization of cherry (*Cerasus pseudocerasus* g. don) genes responding to parthenocarpy induced by GA3 through transcriptome analysis. *BMC Genet.* 20, 65. doi: 10.1186/s12863-019-0746-8

Wu, J., Gu, C., Khan, M. A., Wu, J., Gao, Y., Wang, C., et al. (2013). Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of rosaceae. *Crit. Rev. Plant Sci.* 32, 53–68. doi: 10.1080/07352689.2012.715986

Wu, K. L., Guo, Z. J., Wang, H. H., Li, J., et al. (2005). The *WRKY* family of transcription factors in rice and arabidopsis and their origins. *DNA Res.* 12, 9–26. doi: 10.1093/dnares/12.1.9

Yan, H., Jia, H., Chen, X., Hao, L., An, H., and Guo, X. (2014). The cotton *WRKY* transcription factor *GhWRKY17* functions in drought and salt stress in transgenic *Nicotiana benthamiana* through ABA signaling and the modulation of reactive oxygen species production. *Plant Cell Physiol.* 55, 2060–2076. doi: 10.1093/pcp/pcu133

Zhang, Y., Liu, Z., Liu, J., Lin, S., Wang, J., Lin, W., et al. (2017). GA-DELLA pathway is involved in regulation of nitrogen deficiency-induced anthocyanin accumulation. *Plant Cell Rep.* 36 (4), 557–569. doi: 10.1007/s00299-017-2102-7

Zhang, L., Ma, G., Kato, M., Yamawaki, K., Takagi, T., Kiriiwa, Y., et al. (2012). Regulation of carotenoid accumulation and the expression of carotenoid metabolic genes in citrus juice sacs *in vitro*. *J. Exp. botany.* 63, 871–886. doi: 10.1093/jxb/err318

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 November 2022
DOI 10.3389/fpls.2022.1061196

OPEN ACCESS

EDITED BY
Guo-Fei Tan,
Guizhou Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Shengwu Hu,
Northwest A&F University, China
Xiaoming Song,
North China University of Science and Technology, China
Li Cai,
Huazhong Agricultural University, China

*CORRESPONDENCE
Xiaohui Cheng
chengxiaohui@caas.cn
Meili Xie
xiemeili0101@163.com

†These authors have contributed equally to this work

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

RECEIVED 04 October 2022
ACCEPTED 20 October 2022
PUBLISHED 02 November 2022

CITATION
Zhao C, Yang L, Tang M, Liu L, Huang J, Tong C, Xiang Y, Liu S, Cheng X and Xie M (2022) Genome-wide association study reveals a *GLYCOGEN SYNTHASE KINASE 3* gene regulating plant height in *Brassica napus*.
Front. Plant Sci. 13:1061196.
doi: 10.3389/fpls.2022.1061196

Genome-wide association study reveals a *GLYCOGEN SYNTHASE KINASE 3* gene regulating plant height in *Brassica napus*

Chuanji Zhao[1†], Li Yang[1,2†], Minqiang Tang[3], Lijiang Liu[1], Junyan Huang[1], Chaobo Tong[1], Yang Xiang[4], Shengyi Liu[1], Xiaohui Cheng[1*] and Meili Xie[1*]

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, The Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, Hubei, China, [2]Biosystematics Group, Wageningen University and Research, Wageningen, Netherlands, [3]Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Ministry of Education), School of Forestry, Hainan University, Haikou, China, [4]Guizhou Rapeseed Institute, Guizhou Academy of Agricultural Sciences, Guiyang, Guizhou, China

Rapeseed (*Brassica napus*) is an allotetraploid crop that is the main source of edible oils and feed proteins in the world. The ideal plant architecture breeding is a major objective of rapeseed breeding and determining the appropriate plant height is a key element of the ideal plant architecture. Therefore, this study aims to improve the understanding of the genetic controls underlying plant height. The plant heights of 230 rapeseed accessions collected worldwide were investigated in field experiments over two consecutive years in Wuhan, China. Whole-genome resequencing of these accessions yielded a total of 1,707,194 informative single nucleotide polymorphisms (SNPs) that were used for genome-wide association analysis (GWAS). GWAS and haplotype analysis showed that *BnaA01g09530D*, which encodes BRASSINOSTEROID-INSENSITIVE 2 and belongs to the *GLYCOGEN SYNTHASE KINASE 3* (*GSK3*) family, was significantly associated with plant height in *B. napus*. Moreover, a total of 31 *BnGSK3s* with complete domains were identified from *B. napus* genome and clustered into four groups according to phylogenetic analysis, gene structure, and motif distribution. The expression patterns showed that *BnGSK3s* exhibited significant differences in 13 developmental tissues in *B. napus*, suggesting that *BnGSK3s* may be involved in tissue-specific development. Sixteen *BnGSK3* genes were highly expressed the in shoot apical meristem, which may be related to plant height or architecture development. These results are important for providing new haplotypes of plant height in *B. napus* and for extending valuable genetic information for rapeseed genetic improvement of plant architecture.

KEYWORDS

plant height, genome-wide association study (GWAS), rapeseed (*B. napus* L.), RNA sequencing (RNA-Seq), *GSK3* gene family

Introduction

Rapeseed (*Brassica napus* L., 2n = 38, AACC) is the main source of edible oils and feed proteins worldwide. However, the rapeseed industry is currently confronted with multiple bottlenecks, i.e. low yield, low planting density, low mechanization degree, large amount of fertilization, and high labor costs, which seriously impacts the sustainable development of the rapeseed industry. Shaping the ideal plant architecture of rapeseed is helpful to break through these bottlenecks, but the lack of a clear genetic basis and constituent elements has hindered the development of this research. Plant height is one of the most important determinants of ideal plant architecture. Since lodging is a common phenomenon and yield loss caused by lodging is severe (16.2%) in rapeseed production (Islam and Evans, 1994). Therefore, moderate dwarfing of crop plant height increased the harvest index.

Plant height is an agronomic trait with complex genetic basis. It is easily affected by environment and usually regulated by both major and minor genes. In recent years, with the rise of the green revolution in wheat, breeders have identified a large number of quantitative trait loci (QTLs) controlling wheat plant height on 21 chromosomes using different populations and markers (Chu et al., 2008; Buerstmayr et al., 2011; Guo et al., 2018). The green revolution in rice began with the application of a semi-dwarf gene *sd1* (Monna et al., 2002; Sasaki et al., 2002; Spielmeyer et al., 2002). The discovery and utilization of dwarf mutants and corresponding genes have greatly promoted the development of new rice varieties. The cloned dwarf genes in rice are mainly involved in the biosynthesis and signal pathways of plant hormones (e.g., gibberellin, brassinolide, and strigolactone). Some of these genes contain special domains, including *sd1* (Ye et al., 2015), *D1* (Ferrero-Serrano et al., 2018; Sun et al., 2018), *GID1* and *GID2* (Hirano et al., 2010), *OsDWARF4* (Fang et al., 2016), and *OsTB1* (Fang et al., 2020). Currently, the only known gene responsible for ideal plant architecture gene in rice is *IPA1*, which encodes the squamosa-like promoter-binding protein OsSPL14. Mutations in *OsSPL14* reduced tillering, increased grain number per ear and 1000-grain weight, thickened stem, and enhanced lodging resistance, thereby increasing the yield (Jiao et al., 2010; Miura et al., 2010).

In *B. napus*, the identification of QTLs highly related with plant height is an important task in genetic maps and genome-wide association analysis (GWAS). Fourteen QTLs for plant height were identified in different linkage groups using a recombinant inbred line (Cai et al., 2014). A major plant height QTL on chromosome A10, was identified by whole-genome resequencing (WGS) based genetic mapping (Dong et al., 2021). Using the Illumina Brassica 60 K Bead Chip Array and a diversity of 520 accessions, a total of 68 plant height-related loci were obtained by GWAS under six environments. Most of the genes in these loci were involved in gibberellin synthesis and signal pathway (Sun et al., 2016). In recent years, progress has been made in the exploitation of dwarf genetic resources and genes in *B. napus*. Most dwarf mutants belong to gibberellin, auxin, and brassinolide-insensitive mutants. In two *B. napus* dwarf mutants of approximately 70 cm height, their candidate genes were mapped on chromosomes A06 and C07, both of which encode DELLA proteins, a negative regulator of the gibberellin signal transduction pathway, and have missense mutations in the VHYNP domain (Liu et al., 2010; Zhao et al., 2017). Mutations at different sites of *BnaC05g29300D*, encoding an auxin signaling transport repressor, resulted in rapeseed plant heights of only 25 cm (Zhao et al., 2019; Zheng et al., 2019). Mutation of *BnaA3.IAA7*, which encodes an auxin-inducible protein, disrupted the conserved degradation motif GWPPV and reduced the affinity between BnaA3.IAA7 and the transport inhibitor in an auxin dose-dependent manner, thus inhibiting BnaA3.IAA7 degradation and auxin signaling in *B. napus* dwarf mutant *sca* (Li et al., 2019a). The dwarf locus *BnDWARF2* was mapped to a 34.62 kb interval, in which *BnaC04g41660D* encoding a GLYCOGEN SYNTHASE KINASE 3 (GSK3-like) in the brassinosteroid signaling, was the causal gene controlling plant height in oilseed rape (Yang et al., 2021). In addition, other genes unrelated to plant hormones may also be involved in the regulation of plant height in *B. napus*; for example, the Octicosapeptide/Phox/Bem1p family protein encoding gene *BnaC09g20450D* contains a single nucleotide polymorphism (SNP) that co-segregates with the dwarf phenotype in *df59* mutant (Wang et al., 2020a).

Although many plant height QTLs and dwarf genes have been identified, they have not been fully utilized in breeding, and cultivars with dwarf or semi-dwarf phenotypes are still the major objective in rapeseed breeding. This study aims to better understand the genetic control of plant height and to unearth more valuable information from the genome of polyploid rapeseed based on GWAS for plant height in 230 core rapeseed accessions around the world. We identified *BnaA01g09530D*, a *BnGSK3* gene involved in the cross-talk between auxin and brassinosteroid signaling pathways, was significantly associated with plant height. We also analyzed the expression pattern in various tissues, overall distribution in the rapeseed genome, and phylogenetic analysis of the *BnGSK3s* family.

Results

Phenotype variation of plant height in 230 *B. napus* accessions

Extensive phenotypic variations of plant height were observed in 230 inbred accessions over two consecutive years (Table 1). The plant height ranged from 149.23–230.59 cm in 2017–2018 and from 115.12–189.28 cm in 2018–2019, suggesting that the environment factors had a great impact on

TABLE 1 Phenotypic variations of plant height in rapeseed natural population.

Environment	Min	Max	Mean	SE	SD	Var	Kurtosis	Skewness	CV (%)
2017–2018	149.23	230.59	190.88	0.93	14.08	198.33	0.051	0.055	7.38
2018–2019	115.12	189.28	150.07	0.9	13.66	186.67	-0.039	0.209	9.1
BLUP	132.75	206.85	170.25	0.86	13.09	171.28	0.138	0.098	7.69

Min, minimum value; Max, maximum value; Mean, mean value; SE, standard error; SD, standard deviation; Var, variance; CV, coefficient of variation; BLUP, best linear unbiased prediction.

plant height (Supplementary Figure 1A and Table 1). The plant heights in 2017–2018, 2018–2019, and the BLUP of the 230 rapeseed accessions displayed normal distributions (Supplementary Figure 1A). The coefficient of variation in 2018–2019 was 9.10%, which was higher than that in 2017–2018 (7.38%) and BLUP (7.69%) (Table 1). Nevertheless, no significant difference was observed between the phenotype of 2017-2018 and 2018-2019, as shown by the correlation analysis ($R^2 > 0.70$) (Supplementary Figure 1B). These analyses revealed that the phenotype of 230 rapeseed accessions were reliable and feasible for association analysis.

Genomic variation of rapeseed resequencing population

A total of 230 rapeseed accessions, consisting of 25 spring-, 33 winter-, and 172 semi-winter ecotypes, were employed for WGS (Supplementary Table 1). Approximately 1,097.37 Gb data were generated, with an average size of 4.77 Gb and an average depth of 6.46 × depth per accession (Supplementary Table 1). The average coverage of *B. napus* reference genome was 82.19% (Supplementary Table 1). A total of 1,707,194 informative SNPs were acquired with an average of 94,844 SNPs on each chromosome (Table 2). The density of SNPs on different chromosomes ranged from 1.01 to 4.86 SNP/kb, with chromosome C09 having the lowest density and A10 the highest (Table 2). These results suggested the reliability of SNP information and could be used for further analyses.

Identification of *BnGSK3* significantly associated with plant height

According to the Q+K model, the associated population could be divided into nine subgroups (Supplementary Figure 2A, B). More than 90% of the relative kinship coefficients among these accessions were found to be lower than 0.1, suggesting that most accessions in this population lacked or had weak genetic relatedness (Supplementary Figure 2C). The average linkage disequilibrium (LD) decay of the A and C sub-genomes were 4.1 and 120.3 kb, respectively. It was 33.4 kb for the whole genome (A + C) when r^2 decayed to its half (Figure 1D).

To dissect the genetic control of plant height in *B. napus*, we performed GWAS in two consecutive years. A significant locus on chromosome A01 was simultaneously identified using GLM, MLM, and BLINK models (Figures 1A–C, and Supplementary Table 3). Quantile-quantile plots showed obvious deviations between the observed and expected values, indicating the selected models were correct and suitable for GWAS (Supplementary Figure 3). Within the significance interval, 806 SNPs were repeatedly identified in different environments and models (GLM and MLM) (Supplementary Table 3). According to the MLM model in BLUP and LD decay of A sub-genome (Figure 1), three genes (*BnaA01g09530D*, *BnaA01g09540D*, and *BnaA01g09550D*) near the significant SNPs were strongly associated (Figure 2A). Based on the annotation of the *B. napus* reference genome, *BnaA01g09530D*, encoding BRASSINOSTEROID-INSENSITIVE 2 (BIN2) and involving in the brassinosteroid signaling pathway, may be a candidate gene controlling plant height in *B. napus*. In addition, the position of co-identified SNP by BLINK model in different environments was 4,772,232 (Supplementary Table 3), which was far away from the co-identified significant SNPs in GLM and MLM, due to the different algorithm principle of BLINK. In the application of BLINK, a bin, containing all the linked SNPs in a region, is taken as a unit, rather than a single SNP as a unit like GLM and MLM (Huang et al., 2019), suggesting that *BnaA01g09530D* was also identified in the BLINK models. A total of ten SNPs variations were observed in the sequence of *BnaA01.BIN2*. Haplotype analysis of these ten SNPs revealed favorable allelic variation (Hap_II), conferring a significant reduction in plant height (Figure 2B).

The expression pattern of *BnaA01.BIN2* showed that it was highly expressed in leaves, buds, and roots, followed by SAM, suggesting that it plays an important role in plant development. Subcellular localization, as indicated by green fluorescent protein (GFP), showed that BnaA01.BIN2 was localized in nucleus and cytoplasm (Figure 2D).

In silico analysis of *BnGSK3s* in rapeseed genome

Candidate gene *BnaA01.BIN2* belongs to the *glycogen synthase kinase 3* (GSK3) gene family. Using protein sequences of AtGSK3s as the query of BLAST, a total of 38 *BnGSK3s* were identified in

TABLE 2 Statistics of SNP number and density on each chromosome.

Chromosome	Length	SNPs	SNP/kb
chrA01	23,267,856	80,834	3.47
chrA02	24,793,737	75,701	3.05
chrA03	29,767,490	122,543	4.12
chrA04	19,151,660	83,705	4.37
chrA05	23,067,598	104,826	4.54
chrA06	24,396,386	117,969	4.84
chrA07	24,006,521	114,454	4.77
chrA08	18,961,941	68,741	3.63
chrA09	33,865,340	126,210	3.73
chrA10	17,398,227	84,541	4.86
chrC01	38,829,317	109,044	2.81
chrC02	46,221,804	82,943	1.79
chrC03	60,573,394	129,746	2.14
chrC04	48,930,237	120,199	2.46
chrC05	43,185,227	53,918	1.25
chrC06	37,225,952	74,092	1.99
chrC07	44,770,477	78,036	1.74
chrC08	38,477,087	79,692	2.07
chrC09	48,508,220	49,213	1.01

"*Darmor-bzh*" rapeseed genome, and 31 *BnGSK3s* with Pkinase domain were finally extracted (Table 3). Of these *BnGSK3s*, 16% (5) resulted from dispersed duplications and 84% (26) originated from whole-genome duplication (WGD) or segmental duplication (Table 3). Most *AtGSK3s* have several syntenic genes in *B. napus*, among which *BnBIN2* contains six homologous genes and is the largest member of *BnGSK3s* (Supplementary Table 2). However, there were no homologous genes for *AtBIL2* and *AtSK42* in *B. napus* (Supplementary Table 2). We identified 15 *BrGSK3s* and 16 *BoGSK3s* according to the *Brassica* Database (BRAD) (http://brassicadb.cn/) and no homologous genes of *AtBIL2* and *AtSK42* were identified in the reference genomes of *B. rapa* (Brara_Chiifu_V3.5) and *B. oleracea* (Braol_JZS_V2.0) (Supplementary Table 2). These results suggested that the *GSK3s* family is highly conserved in *Brassicaceae*, whereas the loss of *BIL2s* and *SK42s* may occur prior to *Brassicaceae* speciation.

The 31 *BnGSK3s* were unevenly distributed in 13 chromosomes and four random chromosomes, 16 and 15 *BnGSK3s* were located on the A and C sub-genomes, respectively (Figure 3 and Table 3).

Phylogenetic, syntenic relationship, and conservation analysis of *BnGSK3s*

To explore the phylogenetic relationship of *GSK3s* family, we constructed a phylogenetic tree using GSK3s protein sequences from *Arabidopsis* and *B. napus*. The 10 *AtGSK3s* and 31 *BnGSK3s* were divided into four groups: Group I (SK11, SK12, and SK13), Group II (BIN2, BIL1, and BIL2), Group III (SK31 and SK32), and Group IV (SK41 and SK42) (Figure 4A and Supplementary Table 2). Group I had 22 *GSK3s*, including 11 *BnGSK3s*, 5 *BrGSK3s*, and 6 *BoGSK3s*, accounting for the largest group. Group IV was the smallest, with only eight *GSK3s* (Figure 4A and Supplementary Table 2). This suggests that Group I of *GSK3s* was more expanded compared to that of Group IV. Within each group, *BnGSK3s* belonging to the A and C sub-genomes in *B. napus*, along with the *AtGSK3s* in *Arabidopsis*, clustered into a small clade (Figure 4A), suggesting that the phylogenetic relationship of GSK3 was consistent with the evolution of rapeseed. The syntenic analysis between *AtGSK3s* and *BnGSK3s* showed that most of *AtGSK3s* have over two syntenic genes in B. napus (Figure 4B), which is consistent with phylogenetic relationship of *GSK3s*.

To explore the conservation of *BnGSK3s*, gene structure and protein motifs were analyzed (Figure 5). In general, gene structures of the 31 *BnGSK3s* differed obviously between different groups. Among them, the syntenic genes showed relatively similar gene structures (Figures 5A, C). The gene structures of approximately 74% of *BnGSK3s* (23) exhibited 5' and 3' untranslated regions (UTR) (Figure 5C). Six *BnGSK3s* (*BnA05g31460D*, *BnA03g05700D*, *BnA09g38810D*, *BnA09g51790D*, *BnCnng52760D*, *and BnAnng35300D*) only had 5-' or 3-UTR (Figure 5C). In addition, all *BnGSK3s* contained exons and introns (Figure 5C). As for motif analysis, except *BnaA05g11700D* possessed seven conserved motifs, the remaining *BnGSK3s* had ten conserved motifs (Figure 5B). These results suggested that the core sequences of the *BnGSK3s* were conserved.

Expression patterns of *BnGSK3s*

Based on published transcriptome data (Li et al., 2019b; Dong et al., 2021), the expression patterns of the 31 *BnGSK3s* in 13 tissues of ZS11 were analyzed, which showed that the *BnGSK3s* were expressed in different tissues (Figure 6A). However, a set of homologous genes, including *BnSK31*, *BnBIL1*, and *BnBIN2*, showed similar expression patterns, suggesting a potential redundancy of function (Figure 6A). Different expression patterns were observed within the same group, suggesting functional divergence in *BnGSK3s*. For example, *BnBIL1* was highly expressed in SAM, whereas *BnBIN2* was highly expressed in roots (Figure 6A). In addition, *BnSK13s* were prone to express in pistils and buds, suggesting that these genes may be involved in flower development (Figure 6A). Thirteen *BnGSK3s* were highly expressed in SAM, indicating that *BnGSK3s* have a certain effect on the development of plant architecture. In addition, we selected eight *BnGSK3s* from different groups to perform qRT-PCR in six tissues, which suggested that the expression pattern was consistent with the RNA-seq data (Figure 6B).

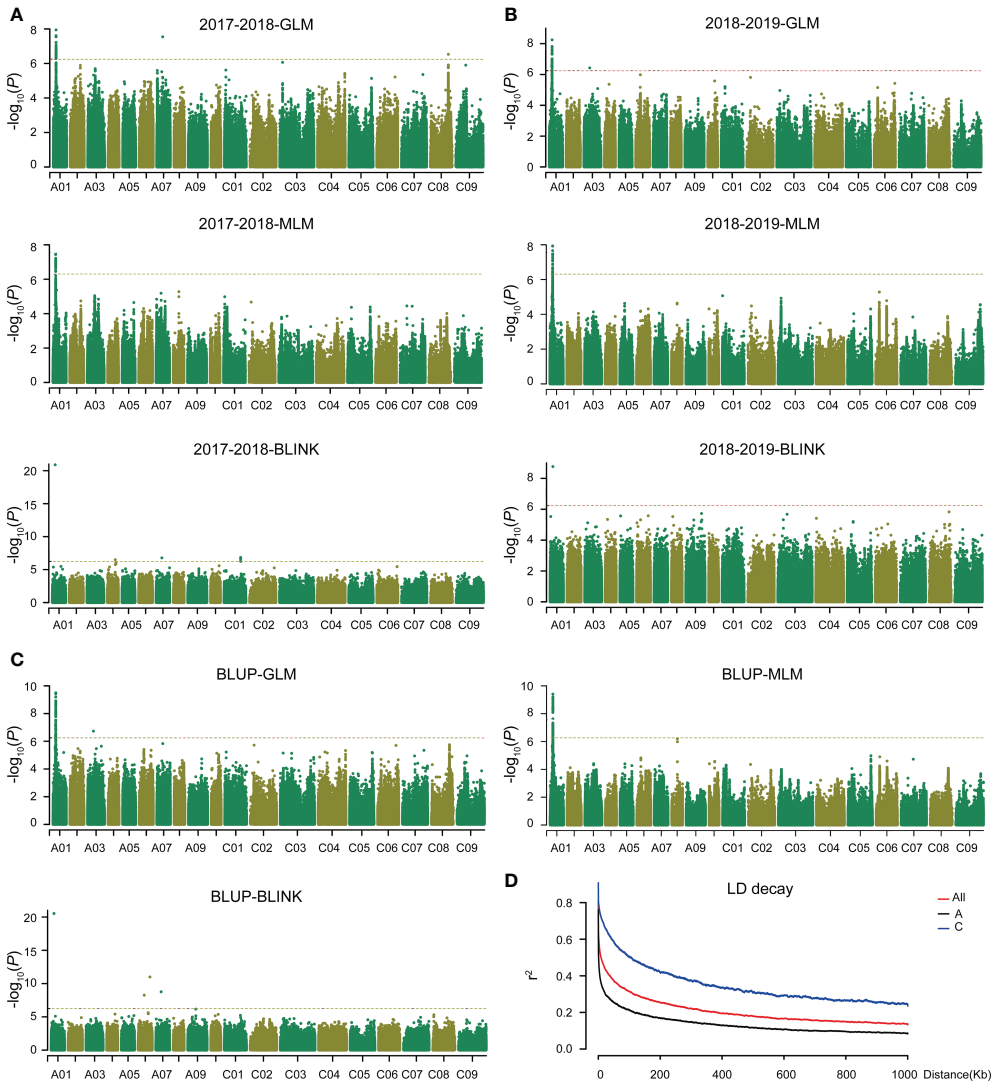


FIGURE 1
GWAS of plant height in *Brassica napus* and LD decay analysis. The threshold value is -$\log_{10}$(1/SNPs number). **(A)** GWAS of plant height in 2017–2018 based on GLM, MLM, and BILNK models. **(B)** GWAS of plant height in 2018–2019 based on GLM, MLM, and BILNK models. **(C)** GWAS of plant height for BLUP based on GLM, MLM, and BLINK models. **(D)** Linkage disequilibrium (LD) decay of A and C sub-genomes and whole genome.

Discussion

Dilemma of plant architecture breeding and the lack of genetic basis for plant height in rapeseed

Since the "Green Revolution" in the 1960s, researchers have carried out extensive research to come up with ideal plant architecture models for many crops (Teichmann and Muhr, 2015; Liu et al., 2020; Pearce, 2021). Several novel genes controlling aboveground plant architecture have been identified and their regulatory mechanisms have been expounded, laying the foundation for breeding new high-yielding varieties of rice (Monna et al., 2002; Sasaki et al., 2002; Wang and Li, 2008; Wang et al., 2020b), wheat (Chai et al., 2022; Xiong et al., 2022), maize (Phillips et al., 2011; Li et al., 2020a), and soybean (Guo et al., 2020; Chen et al., 2021b). Many researchers have proposed models of ideal rapeseed plant architecture (Liu et al., 2022; Zheng et al., 2022). However, these are only concepts and cannot solve actual problems in production. There are several difficulties in studying the ideal plant architecture of rapeseed: 1) the lack of materials with good plant architecture materials; 2) the uncertainty of proper index used for the research of rapeseed plant architecture; 3) severe

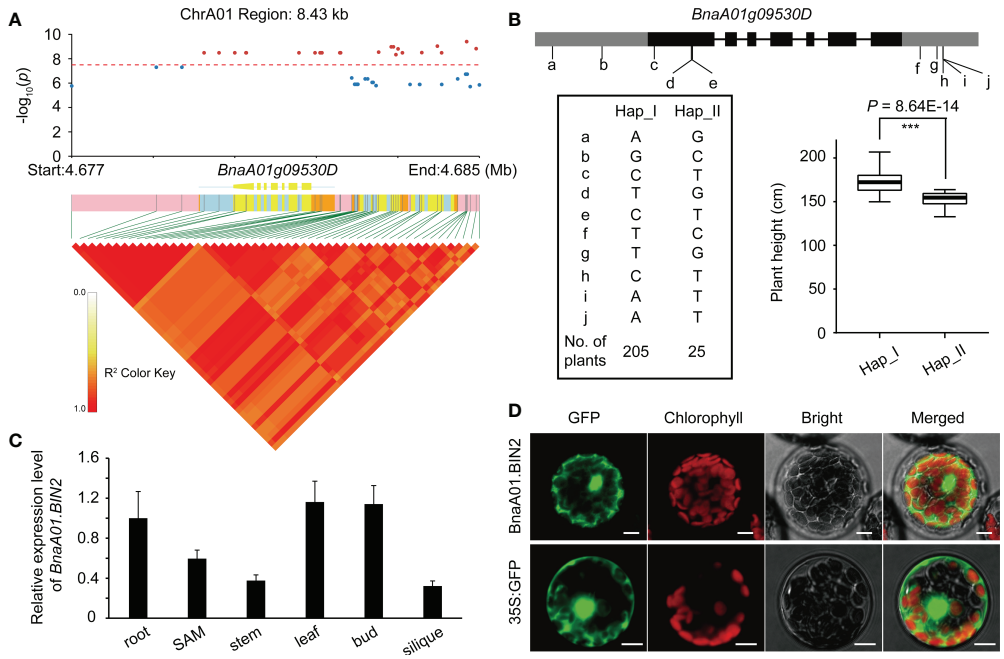


FIGURE 2
Integrated characteristic of *BnA01.BIN2*. **(A)** LD block analysis of significant SNPs in BLUP based on MLM models. The threshold value is -$\log_{10}$ (0.05/SNPs number). **(B)** Gene structure of *BnaA01.BIN2* and haplotype analysis. **(C)** Subcellular localization of *BnaA01.BIN2* in *Arabidopsis* protoplasts. Green fluorescence, GFP; red fluorescence, chloroplast autofluorescence; Merged, merged images of all channels. Bar = 10 μm. **(D)** Expression pattern of *BnaA01.BIN2*.

environmental impact on plant architecture-related traits; 4) the lack of clear genetic basis. For many crops, such as rice and wheat, plant height has been used as a breakthrough point to study plant architecture (Peng et al., 1999; Hedden, 2003). Therefore, plant height is essential in shaping the ideal plant architecture of crops. Although research progress has been made in the study of plant height traits of rapeseed (Liu et al., 2010; Wang et al., 2016a; Wang et al., 2016b), the genetic basis of rapeseed plant height remains unclear.

Currently, a single genetic resource cannot effectively improve the present plant architecture of rapeseed. GWAS is often used as an effective method to unravel the genetic architecture of complex agronomic traits in crops. Combined with association analysis and linkage analysis, 61 SNPs significantly associated with low zinc tolerance and 15 QTLs were identified in maize. Expression and haplotype analyses were used to mine the favorable allele conferring low zinc tolerance (Xu et al., 2022). Similar study could be found in Guo et al. (2021), in which 63 loci related to stem strength and yield were identified and favorable alleles for both high stem strength and high yield were discovered using 524 rice germplasm resources and 193 recombinant inbred lines (Guo et al., 2021). Based on GWAS and a transcriptome-wide association study, 15 stable QTLs and 1,854 candidate genes were detected in *B. napus*, which were significantly associated with seed glucosinolate content. Haplotype analysis showed that seed low glucosinolate was mainly resulted by the co-action of multiple favorable alleles (Tan et al., 2022). In this study, GWAS was performed on plant height of 230 *B. napus* accessions using three models (GLM, MLM, and BLINK). An unreported gene, *BnaA01.BIN2*, was simultaneously identified by all three models (Figures 1, 2), which increased the confidence of the results. However, no other loci or reported genes were co-identified, probably due to population structure constraints. These results provide insights for subsequent adjustment of population structure to more effectively detect available loci, genes, or favorable alleles.

BnBIN2, a core member of *BnGSK3s*, is involved in plant development and stress response

GSK3 is a group of highly conserved cytoplasmic serine/threonine protein kinases that are widely present in animal and plant cells. These proteins perform their functions mainly by phosphorylating key substrate proteins of different signaling pathways. GSK3 is regulated by a variety of post-translational modification mechanisms. *BnaA01.BIN2*, identified by GWAS in this study (Figures 1, 2), encodes BRASSINOSTEROID-INSENSITIVE 2 (BIN2) and, belongs to the *BnGSK3* family.

TABLE 3 The information of *BnGSK3s* family in rapeseed.

Gene ID	Chromosome	AAs	pI	MW (kDa)	Duplication type
BnaAnng02930D	Ann_random	405	6.38	46.08	WGD or Segmental
BnaA09g38810D	A09	438	7.61	49.67	WGD or Segmental
BnaAnng35300D	Ann_random	375	8.85	42.45	Dispersed
BnaA09g04100D	A09	407	8.7	46.21	WGD or Segmental
BnaA05g31460D	A05	515	8.97	58.72	WGD or Segmental
BnaC07g50210D	C07_random	375	8.74	42.43	WGD or Segmental
BnaA08g26100D	A08	422	8.37	47.67	WGD or Segmental
BnaCnng48480D	Cnn_random	422	8.37	47.66	Dispersed
BnaA05g11700D	A05	225	7.57	25.35	WGD or Segmental
BnaC03g62810D	C03	381	8.58	43.08	WGD or Segmental
BnaCnng52760D	Cnn_random	375	8.85	42.42	Dispersed
BnaCnng13510D	Cnn_random	433	6.87	49.3	WGD or Segmental
BnaAnng31110D	Ann_random	341	8.72	38.99	Dispersed
BnaC03g34380D	C03	412	8.56	46.8	WGD or Segmental
BnaA09g51790D	A09_random	479	7.92	53.5	WGD or Segmental
BnaA07g18960D	A07	433	7.2	49.35	WGD or Segmental
BnaC05g07320D	C05	418	8.39	47.43	WGD or Segmental
BnaC04g41660D	C04	411	8.74	46.28	WGD or Segmental
BnaC01g11150D	C01	382	8.44	43.11	WGD or Segmental
BnaA03g29180D	A03	412	8.56	46.8	WGD or Segmental
BnaA03g27010D	A03	469	6.71	52.69	WGD or Segmental
BnaC05g46010D	C05	411	8.52	46.65	WGD or Segmental
BnaCnng02170D	Cnn_random	472	8.2	52.69	WGD or Segmental
BnaA03g05700D	A03	569	8.89	62.99	Dispersed
BnaA01g09530D	A01	375	8.7	42.41	WGD or Segmental
BnaC07g28590D	C07	403	8.7	45.85	WGD or Segmental
BnaC03g06580D	C03	410	8.65	46.03	WGD or Segmental
BnaC09g03480D	C09	407	8.7	46.18	WGD or Segmental
BnaA06g28290D	A06	404	8.59	45.83	WGD or Segmental
BnaC03g31970D	C03	467	6.89	52.36	WGD or Segmental
BnaA06g05770D	A06	418	8.39	47.43	WGD or Segmental

AAs, amino acids; pI, isoelectric point; MW, molecular weight; WGD, whole-genome duplication; random, contigs unassembled on chromosomes.

AtBIN2 plays a role in the crosstalk between auxin and brassinosteroid signaling pathways (https://www.arabidopsis.org/index.jsp). In *B. napus*, *BnaC04.BIL1*, which has been isolated from the dwarf mutant *Bndwarf2* (Yang et al., 2021), encodes BIN2-LIKE 1, and is also a member of *GSK3s*.

As a core member of *GSK3s*, *BIN2* is a constitutively active kinase in plants, whose activity is affected by various regulatory mechanisms, including nucleocytoplasmic distribution, protein-protein interaction strength, phosphorylation and dephosphorylation, acetylation, and ubiquitination (Mao and Li, 2020). The direct function of BIN2 is to participate in the signal transduction pathway of brassinolide, which plays an important role in plant development (Anne et al., 2015). BIN2 directly controls the transcriptional regulatory complex composed of WEREWOLF (WER), transcription factor GLABRA3 (GL3), and WD40 repeat protein TRANSPARENT TESTA GLABRA1 (TTG1), It can phosphorylates GL3 and TTG1 in the WER-GL3-TTG1 complex to inhibit their transcriptional activity, thereby regulating root hair development (Cheng et al., 2014). BIN2 participates in photomorphogenesis by interacting with HY5, an important transcription factor for photomorphogenesis (Li et al., 2020b). In addition, BIN2 is involved in osmotic stress and adverse effects, and can promote lateral root development by phosphorylating auxin-responsive factor ARF7 (Cho et al., 2014). BIN2 is also involved in abscisic acid signal transduction to regulate the osmotic stress response (Wang et al., 2018) and enhances plant drought tolerance by phosphorylating RSPONSIVE OT DESICCATION 26, NAC family transcription factor (Jiang et al., 2019).

GSK3 is involved in the regulation of plant growth and development. However, only one gene has been reported to be

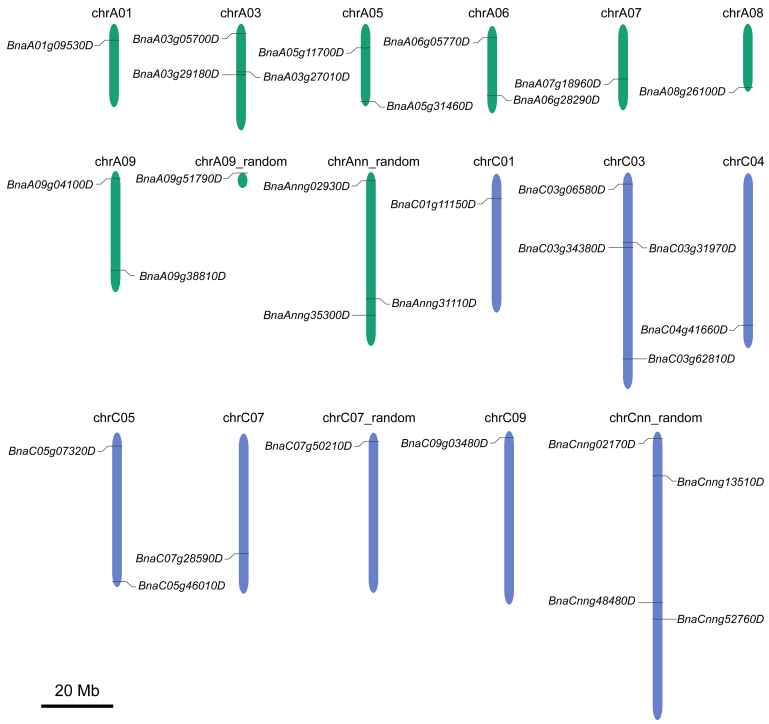


FIGURE 3
Chromosomal distribution of *BnGSK3s* in the genome of *Brassica Napus*. Green and blue chromosomes represent A and C sub-genome of *B. napus*, respectively.

related with plant height in rapeseed (Yang et al., 2021). In this study, to determine the relationship between *BnGSK3s* and plant height in allotetraploid rapeseed, we investigated the *BnGSK3s* family, which consists of 16 homologs in A sub-genome and 15 in C sub-genome (Figure 3 and Table 3). Based on the transcriptome data of 13 tissues in ZS11, the expression pattern of *BnGSK3s* were found to show obvious expression preference difference in rapeseed (Figure 6), in which 16 genes were highly expressed in SAM and three were highly expressed in the pistil (Figure 6A). Moreover, we identified favorable allelic variations in *BnaA01.BIN2* among 230 *B. napus* accessions, whereas we failed to detect any SNP variation in the corresponding syntenic gene *BnaC01.BIN2* (*BnaC01g11150D*). This could be caused by the limited numbers of accessions used for GWAS in this study. As such, more rapeseed genetic resources should be collected to dissect more favorable allelic variations in *BnGSK3s* for plant height and plant architecture.

Conclusions

In this study, GWAS was performed on plant heights of a bio-panel of 230 rapeseed accessions in two consecutive years based on three models. The results showed that *BnaA01.BIN2* belonging to *BnGSK3s* family, was significantly associated with plant height in *B. napus*. A total of 31 *BnGSK3s* were identified and clustered into four groups. Expression pattern analysis suggests that *BnGSK3s* may be involved in tissue-specific development. Sixteen *BnGSK3* genes were highly expressed in SAM, which may be related to plant height development. These findings are important for the genetic improvement of plant height and architecture in rapeseed.

Materials and methods

Plant materials, growth conditions, and phenotypic analysis

A total of 230 rapeseed cultivars or inbred lines (Supplementary Table 1) were collected worldwide, representing the genetic diversity of *B. napus* for GWAS of plant height. Field trials were conducted by a randomized design with three replications. For each accession, 45 individuals were grown in a 2.0 × 1.0 m^2 plot with three rows in each environment (2017–2018, 2018–2019, winter-spring growing season) in the Yangluo experimental field of Oil Crops Research Institute of the Chinese Academy of Agricultural

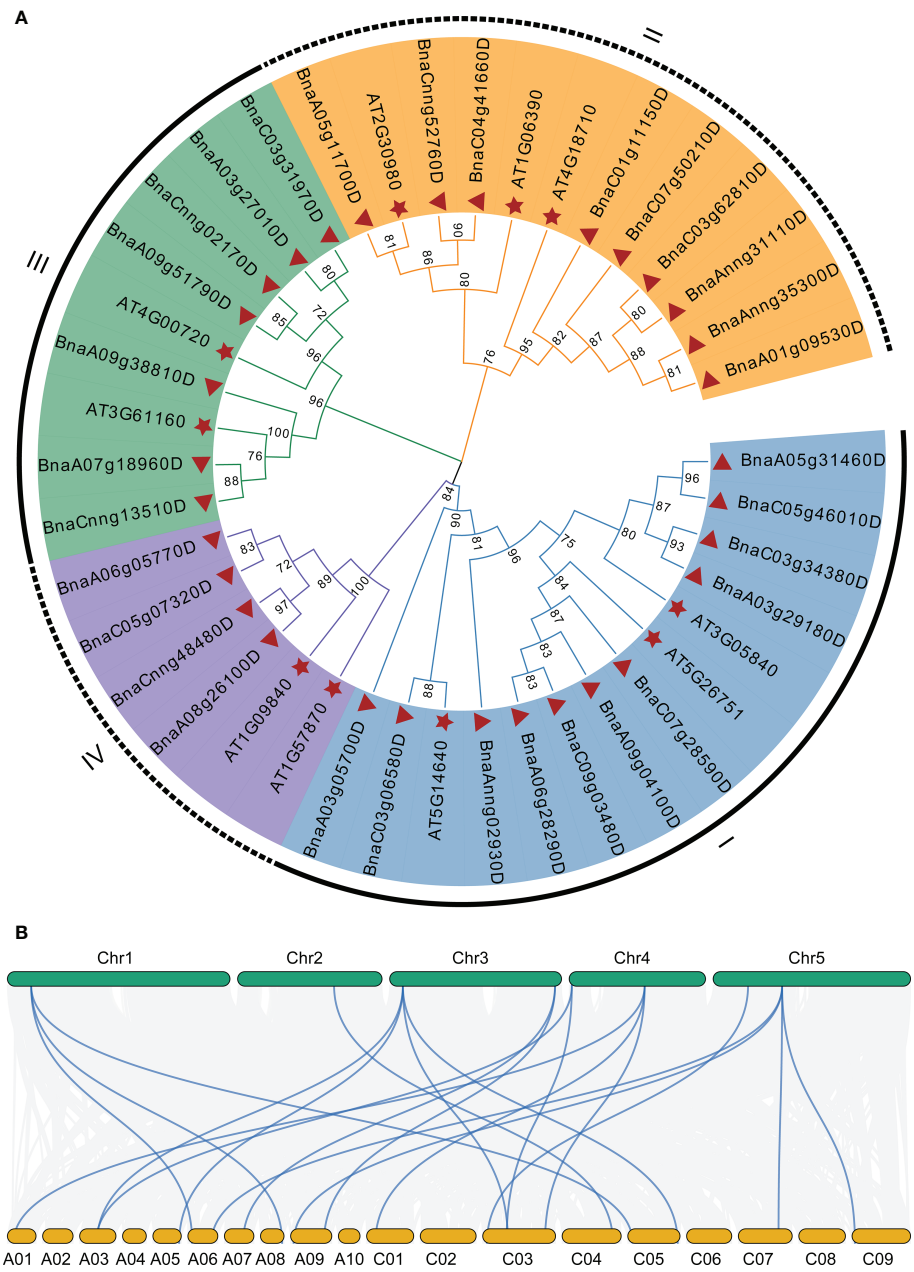


FIGURE 4
Phylogenetic and syntenic analysis of *AtGSK3s* and *BnGSK3s*. **(A)** Phylogenetic analysis. **(B)** syntenic analysis. The green and yellow blocks represent *Arabidopsis* and *B. napus* chromosome, respectively.

Sciences, Wuhan, China. The R script lme4 (CRAN-Package lme4 (r-project.org)) and lsmeans were used to calculate the best linear unbiased prediction (BLUP) of each inbred line in the natural population (Zhao et al., 2022).

At the mature stage, 10 plants with good growth and development were randomly selected from each plot for phenotype investigation. The length from the cotyledon node to the apical position of the whole plant was measured and recorded as plant height. The statistics of the phenotype variation and frequency distribution were calculated using SPSS 22 (IBM SPSS, Armonk, NY, United States) (Zhao et al., 2022). The Pearson's product-moment correlation analysis of plant height between 2017-2018 and 2018-2019 was carried out by R Package.

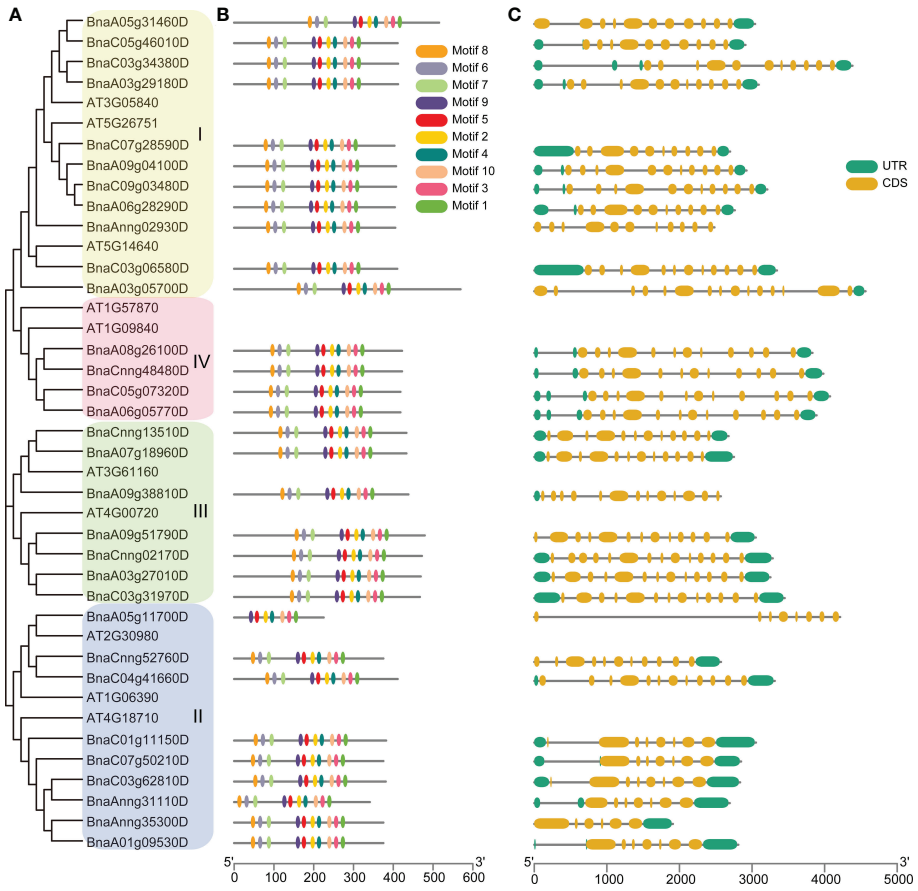


FIGURE 5
Gene structure and conserved motif analyses of *AtGSK3s* and *BnGSK3s*. **(A)** Phylogenetic tree of *AtGSK3s* and *BnGSK3s*. **(B)** Conserved motifs of *AtGSK3s* and *BnGSK3s*. **(C)** Gene structures of *AtGSK3s* and *BnGSK3s*.

Whole-genome sequencing, variant identification and annotation

Total genomic DNA from fresh young leaf tissue of each inbred line (230 accessions) was extracted using a Hi-DNAsecure Plant Kit (TIANGEN, Beijing). DNA libraries were constructed with high-quality genomic DNA and whole-genome resequencing (WGS) was performed using the Illumina NovaSeq 6000 system. Clean data (clean reads) were obtained by filtering the raw data. All clean reads were mapped to the *B. napus* reference genome (*Darmor-bzh* V5, https://www.genoscope.cns.fr/brassicanapus/data/) using the Burrows-Wheeler Aligner software (Li and Durbin, 2009; Chalhoub et al., 2014). SAMTools (parameter: -q 30; http://samtools.sourceforge.net/) and Sentieon Genomics (parameter: –algo Dedup –rmdup) software were used to filter alignment duplications (Li et al., 2009; Freed et al., 2017). GATK (version 4.1.4.0) and vcftools (version 4.2) were used for SNP identification and filtration (parameters: MQ < 50.0 || QD < 2.0, -min-alleles 2 -max-alleles 2 -maf 0.05 -max-missing 0.9, and -cluster-size 3 -cluster-window size 10) (McKenna et al., 2010; Danecek et al., 2011). At last, a total of 1,707,194 informative SNPs were acquired, and the original SNPs were obtained from published data of our lab (Tang, 2019; Ding et al., 2020).

Association study of plant height

To analyze the natural population structure and linkage disequilibrium (LD) decay, ADMIXTURE (Version 1.3.0) (Alexander et al., 2009), Q+K model, and PopLDdecay (Zhang et al., 2018) were performed according to detailed descriptions from previous studies (Zhao et al., 2022). Three software and models were used, including the general linear model (GLM) in trait analysis by association, evolution, and linkage (TASSEL, Version 5.0) (http://www.maizegenetics.net/tassel); mixed linear model (MLM) in Efficient Mixed-Model Association eXpedited (EMMAX); and Bayesian information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) (Huang

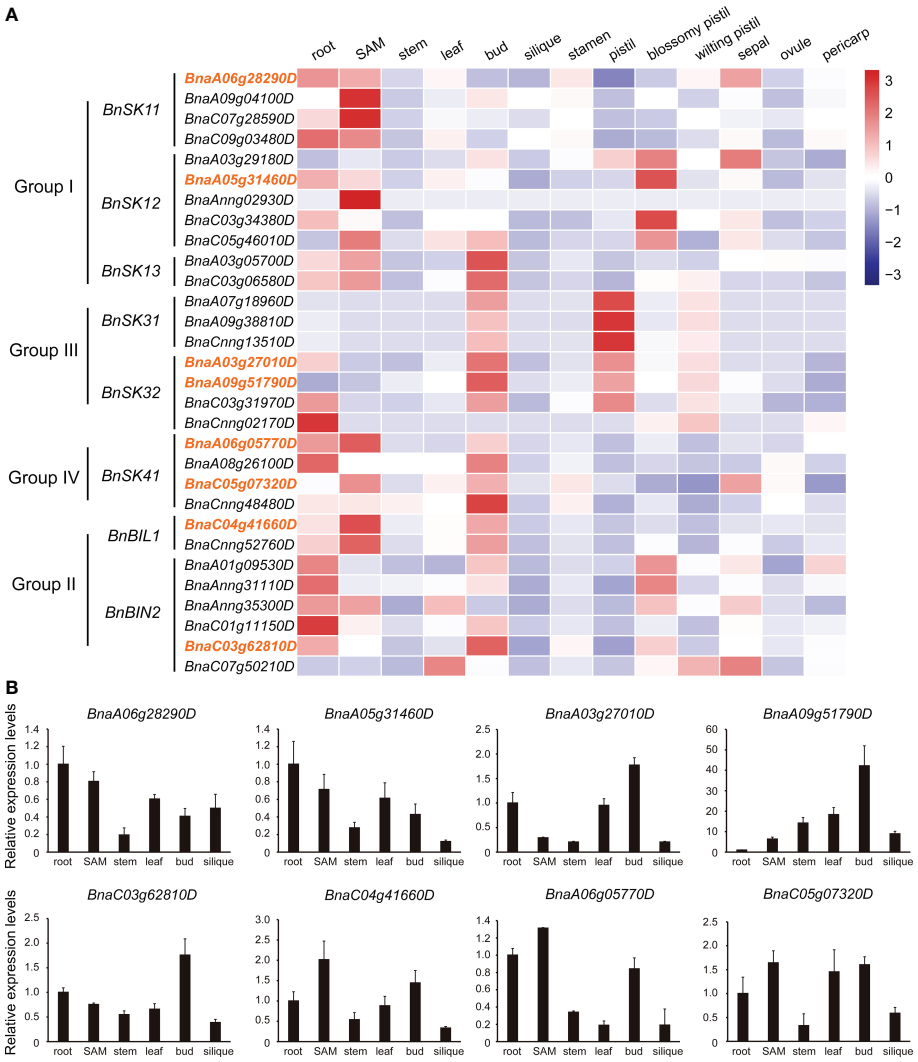


FIGURE 6
Expression pattern of *BnGSK3s*. **(A)** Expression pattern of the 31 *BnGSK3s* in 13 tissues of *B. napus* cv. ZS11 based on RNA-seq. Bar represents the normalized transformed counts of FPKM. The expression of eight genes marked in orange were verified by qRT-PCR.**(B)** Expression patterns of eight *BnGSK3s* in six tissues of *B. napus* cv. ZS11 based on qRT-PCR.

et al., 2019). TASSEL was used to calculate the kinship of 230 *B. napus* accessions (Yu et al., 2006). The LD block was displayed using LDBlockShow software (Dong et al., 2020).

Subcellular localization

Complete coding sequence of *BnaA01.GSK3* was amplified from the *B. napus* cv. Zhongshuang11 (ZS11). The purified DNA fragment was fused with green fluorescent protein (GFP) in the backbone vector pBWA(V)HS-gfp, resulting in the plasmid *35S: BnaA01.BIN2-GFP via* the ClonExpressMultiS One Step Cloning Kit C113-01 (Vazyme). The *35S:GFP* plasmid was used as the mock control. These plasmids were transiently transformed into *Arabidopsis* protoplast cells using the Agrobacterium-mediated method. The subcellular localization of BnaA01.BIN2 was determined by observing GFP using a Nikon C2-ER confocal microscope (Nikon, Japan) (Zhao et al., 2021). The primers used for amplification of *BnaA01.BIN2* are listed in Supplementary Table 4.

Identification and distribution, structure and conserved domain analysis of *BnGSK3s* family

The amino acid sequences of *AtGSK3s* family were obtained from the database "The Arabidopsis Information Resource

(TAIR; https://www.arabidopsis.org/)," which were used to build a Hidden Markov Model, and HMMER3.0 was used to search the annotation and genome information of *B. napus* "*Darmor-bzh*" in the *Brassicaceae* Database (BRAD) (Mistry et al., 2013; Chalhoub et al., 2014; Chen et al., 2021a). The isoelectric point (pI) and molecular weight (MW) of BnGSK3s proteins were predicted using ProtParam online software (https://web.expasy.org/protparam/).

The National Center for Biotechnology Information (NCBI) Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the SMART database (http://smart.embl.de/) were performed to verify the candidate *BnGSK3s* (Lu et al., 2020; Letunic et al., 2021). Chromosomal locations of the candidate *BnGSK3s* were visualized *via* MapGene2Chromosome V2 (MG2C, http://mg2c.iask.in/mg2c_v2.0/).

The sequences of *BnaGSK3s* were downloaded from BRAD and the gene structures were displayed by Tbtools. The conserved motifs were analyzed by Multiple Expectation Maximization for Motif Elicitation (MEME, http://meme-suite.org) (Bailey et al., 2015; Chen et al., 2020).

Phylogenetic and syntenic analysis of BnGSK3s

The alignment of the amino acid sequences of AtGSK3s and BnGSK3s was performed by ClustalW (Larkin et al., 2007). Phylogenetic tree was constructed and visualized using the neighbor-joining (NJ) method in MEGA11 software with 1,000 bootstrap replications (Tamura et al., 2021), and visualized by Evolview7 software (He et al., 2016). The syntenic analysis of *GSK3s* between *AtGSK3s* and *BnGSK3s* were obtained from the BRAD database.

RNA-seq, synthesis of cDNA, and quantitative real-time PCR analysis

The RNA-seq data generated from 13 tissues of ZS11, including roots, SAM, stems, leaves, buds, siliques, stamens, pistils, blossomy petals, wilting petals, sepals, ovules, and pericarps, was previously published in our lab (Sequence Read Archive accession: PRJNA474576 in NCBI and CNP0001630 in China National GeneBank DataBase) were used for the expression pattern analysis of the *BnGSK3s* family (Li et al., 2019b; Dong et al., 2021). FastPure Plant Total RNA Isolation Kit RC401 (Vazyme) was used to extracted total RNA from three biological replicates of different tissues of ZS11, including the roots, SAM, stems, leaves, buds, and siliques using the. First-strand cDNA was generated using a HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper) R312 (Vazyme). Quantitative real-time PCR (qRT-PCR) was performed according to a previously described protocol, and the *BnActin* gene was used as an internal control to quantify the relative expression levels of target genes (Zhao et al., 2021). Gene-specific primers for *BnGSK3s* used for qRT-PCR were obtained from the qPrimerDB qPCR Primer Database (Lu et al., 2018) and the corresponding sequences were listed in Supplementary Table 4. The heatmap is illustrated using OmicShare Tools (https://www.omicshare.com/tools/).

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://bnaomics.ocri-genomics.net/tools/jb-dev/?data=data%2FBna_darmor_v4.1.

Author contributions

CZ, MX, and XC designed this study and provided the funding. SL supervised the study. CZ and MX performed experiments and wrote the manuscript. LY, XC, MT, YX, and LL provided the plant materials and collected the data. LY assisted in data analysis. LY and MX revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32101813, 32070217), Central Public-interest Scientific Institution Basal Research Fund (CAAS-OCRI-XKPY-202104), China Agriculture Research System of MOF and MARA (CARS-12), and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2013-OCRI). Precursor projects of Guizhou province for biological breeding supporting by science and technology in 2022 (Fine identification and evaluation of crop germplasm resources). LY was supported by China Scholarship Council (201903250085).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1061196/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Phenotype distribution and correlation of plant height in two consecutive years in *B. napus*. (A) Phenotype distribution of plant height in 2017-2018, 2018-2019, and BLUP. (B) Correlation of plant height between 2017-2018 and 2018-2019.

SUPPLEMENTARY FIGURE 2
Population structure of 230 *B. napus* accessions. (A) Cross–validation error under different K values. (B) Model-based population structure under K = 9. The y axis represents clusters memberships and the x axis represents the 230 *B. napus* accessions. (C) Relative kinship of 230 rapeseed accessions.

SUPPLEMENTARY FIGURE 3
Quantile-quantile plots of GWAS in two consecutive years under three models.

References

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

Anne, P., Azzopardi, M., Gissot, L., Beaubiat, S., Hématy, K., and Palauqui, J. C. (2015). OCTOPUS negatively regulates BIN2 to control phloem differentiation in *Arabidopsis thaliana*. *Curr. Biol.* 25, 2584–2590. doi: 10.1016/j.cub.2015.08.033

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Buerstmayr, M., Lemmens, M., Steiner, B., and Buerstmayr, H. (2011). Advanced backcross QTL mapping of resistance to fusarium head blight and plant morphological traits in a triticum macha x t. aestivum population. *Theor. Appl. Genet.* 123, 293–306. doi: 10.1007/s00122-011-1584-x

Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., et al. (2014). Association mapping of six yield-related traits in rapeseed (*Brassica napus* l.). *Theor. Appl. Genet.* 127, 85–96. doi: 10.1007/s00122-013-2203-9

Chai, L., Xin, M., Dong, C., Chen, Z., Zhai, H., Zhuang, J., et al. (2022). A natural variation in ribonuclease h-like gene underlies *Rht8* to confer "Green revolution" trait in wheat. *Mol. Plant* 15, 377–380. doi: 10.1016/j.molp.2022.01.013

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

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

Cheng, Y., Zhu, W., Chen, Y., Ito, S., Asami, T., and Wang, X. (2014). Brassinosteroids control root epidermal cell fate *via* direct regulation of a MYB-bHLH-WD40 complex by GSK3-like kinases. *Elife* 3, e02525. doi: 10.7554/eLife.02525.022

Chen, H., Wang, T., He, X., Cai, X., Lin, R., Liang, J., et al. (2021a). BRAD V3.0: an upgraded brassicaceae database. *Nucleic Acids Res.* 50, D1432–D1441. doi: 10.1093/nar/gkab1057

Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., et al. (2021b). Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnol. J.* 19, 702–716. doi: 10.1111/pbi.13496

Cho, H., Ryu, H., Rho, S., Hill, K., Smith, S., Audenaert, D., et al. (2014). A secreted peptide acts on BIN2-mediated phosphorylation of ARFs to potentiate auxin response during lateral root development. *Nat. Cell Biol.* 16, 66–76. doi: 10.1038/ncb2893

Chu, C. G., Xu, S. S., Friesen, T. L., and Faris, J. D. (2008). Whole genome mapping in a wheat doubled haploid population using SSRs and TRAPs and the identification of QTL for agronomic traits. *Mol. Breed.* 22, 251–266. doi: 10.1007/s11032-008-9171-9

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

Ding, L. N., Li, M., Guo, X. J., Tang, M. Q., Cao, J., Wang, Z., et al. (2020). *Arabidopsis* GDSL1 overexpression enhances rapeseed sclerotinia sclerotiorum resistance and the functional identification of its homolog in *Brassica napus*. *Plant Biotechnol. J.* 18, 1255–1270. doi: 10.1111/pbi.13289

Dong, Z., Alam, M. K., Xie, M., Yang, L., Liu, J., Helal, M. M. U., et al. (2021). Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome resequencing in *Brassica napus*. *G3-GENES Genom. Genet.* 11, jkab118. doi: 10.1093/g3journal/jkab118

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., and Yang, T.-L. (2020). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22, bbaa227. doi: 10.1093/bib/bbaa227

Fang, Z., Ji, Y., Hu, J., Guo, R., Sun, S., and Wang, X. (2020). Strigolactones and brassinosteroids antagonistically regulate the stability of the D53-OsBZR1 complex to determine FC1 expression in rice tillering. *Mol. Plant* 13, 586–597. doi: 10.1016/j.molp.2019.12.005

Fang, N., Xu, R., Huang, L., Zhang, B., Duan, P., Li, N., et al. (2016). SMALL GRAIN 11 controls grain size, grain number and grain yield in rice. *Rice* 9, 64. doi: 10.1186/s12284-016-0136-z

Ferrero-Serrano, Á., Su, Z., and Assmann, S. M. (2018). Illuminating the role of the gα heterotrimeric G protein subunit, RGA1, in regulating photoprotection and photoavoidance in rice. *Plant Cell Environ.* 41, 451–468. doi: 10.1111/pce.13113

Freed, D., Aldana, R., Weber, J. A., and Edwards, J. S. (2017). The sentieon genomics tools - a fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv* 115717. doi: 10.1101/115717

Guo, W., Chen, L., Herrera-Estrella, L., Cao, D., and Tran, L.-S. P. (2020). Altering plant architecture to improve performance and resistance. *Trends Plant Sci.* 25, 1154–1170. doi: 10.1016/j.tplants.2020.05.009

Guo, Z., Liu, X., Zhang, B., Yuan, X., Xing, Y., Liu, H., et al. (2021). Genetic analyses of lodging resistance and yield provide insights into post-Green-Revolution breeding in rice. *Plant Biotechnol. J.* 19, 814–829. doi: 10.1111/pbi.13509

Guo, J., Shi, W., Zhang, Z., Cheng, J., Sun, D., Yu, J., et al. (2018). Association of yield-related traits in founder genotypes and derivatives of common wheat (*Triticum aestivum* l.). *BMC Plant Biol.* 18, 38. doi: 10.1186/s12870-018-1234-4

Hedden, P. (2003). The genes of the green revolution. *Trends Genet.* 19, 5–9. doi: 10.1016/S0168-9525(02)00009-4

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* 44, W236–W241. doi: 10.1093/nar/gkw370

Hirano, K., Asano, K., Tsuji, H., Kawamura, M., Mori, H., Kitano, H., et al. (2010). Characterization of the molecular mechanism underlying gibberellin perception complex formation in rice. *Plant Cell* 22, 2680–2696. doi: 10.1105/tpc.110.075549

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8, giy154. doi: 10.1093/gigascience/giy154

Islam, N., and Evans, E. J. (1994). Influence of lodging and nitrogen rate on the yield and yield attributes of oilseed rape (*Brassica napus* l.). *Theor. Appl. Genet.* 88, 530–534. doi: 10.1007/BF01240914

Jiang, H., Tang, B., Xie, Z., Nolan, T., Ye, H., Song, G. Y., et al. (2019). GSK3-like kinase BIN2 phosphorylates RD26 to potentiate drought signaling in *Arabidopsis*. *Plant J.* 100, 923–937. doi: 10.1111/tpj.14484

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic. Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937

Li, Y., Dong, C., Hu, M., Bai, Z., Tong, C., Zuo, R., et al. (2019b). Identification of flower-specific promoters through comparative transcriptome analysis in *Brassica napus*. *Int. J. Mol. Sci.* 20, 5949. doi: 10.3390/ijms20235949

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, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H., Li, J., Song, J., Zhao, B., Guo, C., Wang, B., et al. (2019a). An auxin signaling gene BnaA3.IAA7 contributes to improved plant architecture and yield heterosis in rapeseed. *New Phytol.* 222, 837–851. doi: 10.1111/nph.15632

Li, J., Terzaghi, W., Gong, Y., Li, C., Ling, J. J., Fan, Y., et al. (2020b). Modulation of BIN2 kinase activity by HY5 controls hypocotyl elongation in the light. *Nat. Commun.* 11, 1592. doi: 10.1038/s41467-020-15394-7

Liu, S., Raman, H., Xiang, Y., Zhao, C., Huang, J., and Zhang, Y. (2022). *De novo* design of future rapeseed crops: Challenges and opportunities. *Crop J.* 10, 587–596. doi: 10.1016/j.cj.2022.05.003

Liu, C., Wang, J., Huang, T., Wang, F., Yuan, F., Cheng, X., et al. (2010). A missense mutation in the VHYNP motif of a DELLA protein causes a semi-dwarf mutant phenotype in *Brassica napus*. *Theor. Appl. Genet.* 121, 249–258. doi: 10.1007/s00122-010-1306-9

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

Li, H., Wang, L., Liu, M., Dong, Z., Li, Q., Fei, S., et al. (2020a). Maize plant architecture is regulated by the ethylene biosynthetic gene ZmACS7. *Plant Physiol.* 183, 1184–1199. doi: 10.1104/pp.19.01421

Lu, K., Li, T., He, J., Chang, W., Zhang, R., Liu, M., et al. (2018). qPrimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. *Nucleic Acids Res.* 46, D1229–D1236. doi: 10.1093/nar/gkx725

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Mao, J., and Li, J. (2020). Regulation of three key kinases of brassinosteroid signaling pathway. *Int. J. Mol. Sci.* 21, 4340. doi: 10.3390/ijms21124340

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41, e121. doi: 10.1093/nar/gkt263

Miura, K., Ikeda, M., Matsubara, A., Song, X. J., Ito, M., Asano, K., et al. (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* 42, 545–549. doi: 10.1038/ng.592

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., et al. (2002). Positional cloning of rice semidwarfing gene, sd-1: Rice "Green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res.* 9, 11. doi: 10.1093/dnares/9.1.11

Pearce, S. (2021). Towards the replacement of wheat 'Green revolution' genes. *J. Exp. Bot.* 72, 157–160. doi: 10.1093/jxb/eraa494

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400, 256. doi: 10.1038/22307

Phillips, K. A., Skirpan, A. L., Liu, X., Christensen, A., Slewinski, T. L., Hudson, C., et al. (2011). Vanishing tassel2 encodes a grass-specific tryptophan aminotransferase required for vegetative and reproductive development in maize. *Plant Cell* 23, 550–566. doi: 10.1105/tpc.110.075267

Sasaki, A., Ashikari, M., Ueguchitanaka, M., Itoh, H., Nishimura, A., Swapan, D., et al. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416, 701–702. doi: 10.1038/416701a

Spielmeyer, W., Ellis, M. H., and Chandler, P. M. (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *P. Natl. Acad. Sci. U.S.A.* 99, 9043–9048. doi: 10.1073/pnas.132266399

Sun, S., Wang, L., Mao, H., Shao, L., Li, X., Xiao, J., et al. (2018). A G-protein pathway determines grain size in rice. *Nat. Commun.* 9, 851. doi: 10.1038/s41467-018-03141-y

Sun, C., Wang, B., Yan, L., Hu, K., Liu, S., Zhou, Y., et al. (2016). Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus* l.). *Front. Plant Sci.* 7, 1102. doi: 10.3389/fpls.2016.01102

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tang, M. Q. (2019). *Population genome variations and subgenome asymmetry in brassica napus l* (Wuhan, China: Huazhong Agricultural University), D50–D59.

Tan, Z., Xie, Z., Dai, L., Zhang, Y., Zhao, H., Tang, S., et al. (2022). Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*. *Plant Biotechnol. J.* 20, 211–225. doi: 10.1111/pbi.13707

Teichmann, T., and Muhr, M. (2015). Shaping plant architecture. *Front. Plant Sci.* 6, 233. doi: 10.3389/fpls.2015.00233

Wang, Y., Chen, W., Chu, P., Wan, S., Yang, M., Wang, M., et al. (2016a). Mapping a major QTL responsible for dwarf architecture in *Brassica napus* using a single-nucleotide polymorphism marker approach. *BMC Plant Biol.* 16, 178. doi: 10.1186/s12870-016-0865-6

Wang, Y., He, J., Yang, L., Wang, Y., Chen, W., Wan, S., et al. (2016b). Fine mapping of a major locus controlling plant height using a high-density single-nucleotide polymorphism map in brassica napus. *Theor. Appl. Genet.* 129, 1479–1491. doi: 10.1007/s00122-016-2718-y

Wang, Y., and Li, J. (2008). Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* 59, 253–279. doi: 10.1146/annurev.arplant.59.032607.092902

Wang, Y., Shang, L., Yu, H., Zeng, L., Hu, J., Ni, S., et al. (2020b). A strigolactone biosynthesis gene contributed to the green revolution in rice. *Mol. Plant* 13, 923–932. doi: 10.1016/j.molp.2020.03.009

Wang, H., Tang, J., Liu, J., Hu, J., Liu, J., Chen, Y., et al. (2018). Abscisic acid signaling inhibits brassinosteroid signaling through dampening the dephosphorylation of BIN2 by ABI1 and ABI2. *Mol. Plant* 11, 315–325. doi: 10.1016/j.molp.2017.12.013

Wang, X., Zheng, M., Liu, H., Zhang, L., Chen, F., Zhang, W., et al. (2020a). Fine-mapping and transcriptome analysis of a candidate gene controlling plant height in *Brassica napus* l. *Biotechnol. Biofuels.* 13, 42. doi: 10.1186/s13068-020-01687-y

Xiong, H., Zhou, C., Fu, M., Guo, H., Xie, Y., Zhao, L., et al. (2022). Cloning and functional characterization of *Rht8*, a "Green revolution" replacement gene in wheat. *Mol. Plant* 15, 373–376. doi: 10.1016/j.molp.2022.01.014

Xu, J., Ni, Z., Chen, F., Fu, X., and Yu, F. (2022). Integrated linkage mapping and genome-wide association study to dissect the genetic basis of zinc deficiency tolerance in maize at seedling stage. *Crop J.* doi: 10.1016/j.cj.2022.05.004

Yang, M., He, J., Wan, S., Li, W., Chen, W., Wang, Y., et al. (2021). Fine mapping of the *BnaC04.BIL1* gene controlling plant height in *Brassica napus* L. *BMC Plant Biol.* 21, 359. doi: 10.1186/s12870-021-03137-9

Ye, H., Feng, J., Zhang, L., Zhang, J., Mispan, M. S., Cao, Z., et al. (2015). Map-based cloning of seed Dormancy1-2 identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. *Plant Physiol.* 169, 2152–2165. doi: 10.1104/pp.15.01202

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

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2018). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhao, B., Li, H., Li, J., Wang, B., Dai, C., Wang, J., et al. (2017). Brassica napus DS-3, encoding a DELLA protein, negatively regulates stem elongation through gibberellin signaling pathway. *Theor. Appl. Genet.* 130, 727–741. doi: 10.1007/s00122-016-2846-4

Zhao, C., Safdar, L. B., Xie, M., Shi, M., Dong, Z., Yang, L., et al. (2021). Mutation of the PHYTOENE DESATURASE 3 gene causes yellowish-white petals in *Brassica napus*. *Crop J.* 9, 1124–1134. doi: 10.1016/j.cj.2020.10.012

Zhao, B., Wang, B., Li, Z., Guo, T., Zhao, J., Guan, Z., et al. (2019). Identification and characterization of a new dwarf locus DS-4 encoding an Aux/IAA7 protein in brassica napus. *Theor. Appl. Genet.* 132, 1435–1449. doi: 10.1007/s00122-019-03290-8

Zhao, C., Xie, M., Liang, L., Yang, L., Han, H., Qin, X., et al. (2022). Genome-wide association analysis combined with quantitative trait loci

mapping and dynamic transcriptome unveil the genetic control of seed oil content in *Brassica napus* l. *Front. Plant Sci.* 13, 929197. doi: 10.3389/fpls.2022.929197

Zheng, M., Hu, M., Yang, H., Tang, M., Zhang, L., Liu, H., et al. (2019). Three BnaIAA7 homologs are involved in auxin/brassinosteroid-mediated plant morphogenesis in rapeseed (*Brassica napus* l.). *Plant Cell Rep.* 38, 883–897. doi: 10.1007/s00299-019-02410-4

Zheng, M., Terzaghi, W., Wang, H., and Hua, W. (2022). Integrated strategies for increasing rapeseed yield. *Trends Plant Sci.* 27, 742–745. doi: 10.1016/j.tplants.2022.03.008

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fpls.2022.1080999

OPEN ACCESS

EDITED BY
Lin Chen,
Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, China

REVIEWED BY
Xuehai Zhang,
Henan Agricultural University, China
Jinwu Deng,
Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, China
Guo Jian,
Yangzhou University, China

*CORRESPONDENCE
Xiaohui Cheng
chengxiaohui@caas.cn
Junyan Huang
huangjy@oilcrops.cn

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

RECEIVED 26 October 2022
ACCEPTED 01 December 2022
PUBLISHED 15 December 2022

CITATION
Dong Z, Tang M, Cui X, Zhao C, Tong C, Liu Y, Xiang Y, Li Z, Huang J, Cheng X and Liu S (2022) Integrating GWAS, linkage mapping and gene expression analyses reveal the genetic control of first branch height in *Brassica napus* L.
Front. Plant Sci. 13:1080999.
doi: 10.3389/fpls.2022.1080999

Integrating GWAS, linkage mapping and gene expression analyses reveal the genetic control of first branch height in *Brassica napus* L

Zhixue Dong [1,2], Minqiang Tang [2,3], Xiaobo Cui [2], Chuanji Zhao [2], Chaobo Tong [2], Yueying Liu [2], Yang Xiang [4], Zaiyun Li [1], Junyan Huang [2]*, Xiaohui Cheng [2]* and Shengyi Liu [2]

[1]National Key Lab of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]Key Laboratory of Biology and Genetic Improvement of Oil Crops, the Ministry of Agriculture and Rural Affairs of the People's Republic of China (PRC), Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China, [3]Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, School of Forestry, Hainan University, Haikou, China, [4]Guizhou Rapeseed Institute, Guizhou Academy of Agricultural Science, Guiyang, China

Rapeseed (*Brassica napus* L.) is a crucial oil crop cultivated worldwide. First branch height, an essential component of rapeseed plant architecture, has an important effect on yield and mechanized harvesting; however, the underlying genetic mechanism remains unclear. In this study, based on the 60K single nucleotide polymorphism array and a recombinant inbred lines population derived from M083 and 888-5, a total of 19 QTLs were detected in five environments, distributed on linkage groups A02, A09, A10, C06, and C07, which explained phenotypic variation ranging from 4.87 to 29.87%. Furthermore, 26 significant SNPs were discovered on Chr.A02 by genome-wide association study in a diversity panel of 324 re-sequencing accessions. The major QTL of the first branch height trait was co-located on Chr.A02 by integrating linkage mapping and association mapping. Eleven candidate genes were screened *via* allelic variation analysis, inter-subgenomic synteny analysis, and differential expression of genes in parental shoot apical meristem tissues. Among these genes, *BnaA02g13010D*, which encodes a TCP transcription factor, was confirmed as the target gene according to gene function annotation, haplotype analysis, and full-length gene sequencing, which revealed that TATA insertion/deletion in the promoter region was closely linked to significantly phenotypic differences *BnaA02.TCP1*^{M083} overexpression resulted in decreased branch height and increased branch number in *Arabidopsis*. These results provide a genetic basis for first branch height and the ideal architecture of *B. napus*.

KEYWORDS

Brassica napus*, first branch height, RIL, GWAS, *TCP1

Introduction

Rapeseed (*Brassica napus* L.) is an important source of edible oil, biodiesel and protein feed for livestock in the world (Chalhoub et al., 2014). It is the third largest oil crop with an annual seed yield of 72 million tons that takes up ~13% of global oilseed production after oil palm and soybean (FAOSTAT, https://www.fao.org/faostat/). Global rapeseed production has increased significantly since 1961, but there has been no change in seed yield, at least in the last five years (Liu et al., 2022). With the growing human population and changing environment, the energy demand for rapeseed and crop production has increased. However, owing to the increasing urbanization, dwindling resources and competition with staple crops for land, the development of the rapeseed industry has plateaued in recent years (Liu et al., 2022). Therefore, there is an urgent to study and cultivate high-yield rapeseed varieties using genomics and biotechnology to meet global demands.

Crop yield mainly comprises the number of plants per unit area, number of grains per plant, and thousand-seed weight. Plant architecture, defined as the three-dimensional organization of the plant body, plays an extremely important role in crop growth, yield, and harvest index (Reinhardt and Kuhlemeier, 2002; Sarlikioti et al., 2011; Gou et al., 2017; Wang et al., 2017; Sun et al., 2019; Tian et al., 2019; Yang et al., 2019; Zheng et al., 2020; Guo et al., 2020). Crop breeding research has focused on shoot architecture, including the component traits of plant height, branch or tiller number and angle, leaf shape, size and angle, and inflorescence morphology (Wang and Li, 2008; Wang et al., 2018). Among these traits, plant height is the most prominent determinant of plant architecture and is often selected in domesticated crops. In the 1960s, the central theme of the "Green Revolution" was reducing plant height by introducing the semi-dwarf gene *sd1* into rice and the dwarf genes *Rht-B1b*/*Rht-D1b* into wheat, which greatly improved lodging resistance and the harvest index (Khush, 2001; Liu et al., 2018). The upright and sparsely branched phenotype of maize (*Zea mays*) is suitable for high-density planting, which greatly improved the harvest index (Tian et al., 2019). Currently, one of the major challenges in modern agriculture is the development of an ideotype that produces high yields in various environments.

Plant architecture is controlled by a complex regulatory network that includes phytohormones, miRNAs, and several transcription factors (Lo et al., 2017; Tang and Chu, 2017; Liu et al., 2019; Schneider et al., 2019; Chen et al., 2020; Guo et al., 2020). The molecular mechanisms regulating plant aerial architecture have been identified in several crops, including rice (*OsIPA1*), maize (*ZmUPA2*, *Zmtin1*), wheat (*TaSPL8*), cucumber (*CsBRC1*), soybean (*GmWRI1b*), and rapeseed (*BnaMAX1*) (Liu et al., 2019; Shen et al., 2019; Sun et al., 2019; Tian et al., 2019; Zhang et al., 2019c; Zheng et al., 2020; Guo et al., 2020), which improved plant performance and yield. Simultaneously, most identified genes that shape plant architecture are related to plant height, which participates in biosynthesis or signaling pathways of phytohormones, including auxin, gibberellins, brassinosteroids and strigolactones (Wang et al., 2018). Genetic studies in tomato, *Arabidopsis*, rice, and maize have shown that many genes are directly involved in the initiation and outgrowth of axillary buds, which significantly affects crop yield by influencing branch or tiller numbers (Wang and Li, 2006; Wang et al., 2016; Wang and Jiao, 2018).

Dicot plant architecture is typical of rapeseed plants, including the shoot apical meristem (SAM), which establishes the shoot as the primary growth axis of the plant by continuously initiating phytomers, and the root apical meristem, which establishes the primary root that can branch to form secondary or higher-order lateral roots (Teichmann and Muhr, 2015). Recently, an ideotype-like rapeseed plant stature mutant with compact inflorescence and high pod density was isolated, which could be utilized to shorten the flowering period and increase production and was highly desired for mechanized harvest (Fan et al., 2021; Kuai et al., 2022). In general, rapeseed with semi-dwarf plant type tends to show many branches and high primary branch height, which is beneficial for improving production by increasing pod number per plant and enhancing lodging resistance. Therefore, first branch height (FBH), the distance between the growing point of the first effective branch and the cotyledon node, is also a non-negligible factor shaping rapeseed plant architecture and influencing rapeseed yield (Zheng et al., 2019). Nevertheless, there are no reports on FBH gene localization and cloning in rapeseed.

Quantitative trait loci (QTL) mapping combined with genome-wide association study (GWAS) is an effective approach for identifying loci controlling complex traits (Han et al., 2018; Wang et al., 2020b). Zhao et al. (2016) detected 27 QTLs related to FBH trait distributed on 11 chromosomes with a doubled haploid population derived from KenC-8 and N53-2. Cai et al. (2016) also used a doubled haploid population derived from Huashuang 5 and J7005 to detect16 QTLs that contribute to FBH trait in rapeseed. Luo et al. (2017) genotyped and reanalyzed the previous phenotypes of a doubled haploid population (Tapidor and Ningyou7), and detected 35 QTLs related to FBH, which were distributed on A01, A02, A03, A05, A06, A09, A10, C05, C06, C07, C08, and C09. The contribution rate of these QTLs ranged from 3.19 to 22.91%. Shen et al. (2018) detected five plant height-related QTLs and five FBH-related QTLs on chromosomes A02 and A07 using a doubled haploid population derived from Westar and Y689. All the above reports identified QTLs for rapeseed FBH, but all QTLs were screened using a single method. QTL mapping and GWAS can accurately and rapidly locate and screen for candidate genes. Recently, several studies detected and

screened candidate genes for agronomic traits in rapeseed (Sun et al., 2016; Wang et al., 2020b; Zhao et al., 2022). However, this approach was not used to rapeseed FBH.

In this study, a recombinant inbred lines (RILs) population derived from M083 and 888-5 containing 210 lines was used for linkage mapping to detect QTLs for FBH in different environments with a 60 K Brassica Infinium single nucleotide polymorphism (SNP) array. We further identified several candidate genes for FBH with a GWAS of 324 natural accessions that our lab collected and re-sequenced previously based on QTL mapping. The results obtained here lay the foundation for further study of the genetic architecture and molecular mechanisms of FBH in rapeseed, and facilitate molecular breeding to improve plant architecture for easy mechanical harvest and increase rapeseed yield.

Materials and methods

Plant materials

An association panel composed of 324 diverse rapeseed inbred lines (25 winter, 259 semi-winter, and 40 spring types) collected worldwide was used for GWAS. These lines were grown in Yangluo, Wuhan, China in 2016, 2017, 2018, and 2019. Each variety was planted in a plot with three rows (33-cm line width and 20-cm plant distance), with 12 plants per row. The lines were randomly designed with three replicates.

A RIL population containing 210 lines derived from the crossing of 888-5 and M083 was used in a previous study. The parent line 888-5 had a lower FBH than the M083 inbred line (Liu et al., 2005). The genetic linkage map contained 9,278 SNPs covering 4,071.08 cM, which mapped to 2,771 loci (bins) in a previous study (Zhang et al., 2019b). The population was planted in the standard breeding trial fields of Xiantao and Wuhan, Hubei Province, and Yangzhou, Jiangsu Province, China in 2016 and 2017. The planting method was the same as that of association population.

Phenotyping and data analysis

At maturity, the association population, RIL population, and the parental lines (888-5 and M083) were assessed for FBH (from the base of the stem to the first effective branch) in each environment. The phenotypic values of the lines in each environment were calculated as the average of ten plants in triplicate. The two parents were compared using two-sample *t*-test. SPSS 21 (Armonk, NY, United States) was used to calculate the frequency distribution. Analysis of variance was performed using the GLM procedure of SAS. The best linear unbiased prediction (BLUP) of each inbred line in the association population and the effects of genotype, environment, and genotype × environment interactions in linkage and association mapping population were obtained by using the "*lmer*" function in the "*lme4*" package of R version 3.6.3 (http://www.R-project.org). The broad-sense heritability (H^2) of FBH was calculated as follows:

$$H^2 = G/(G + \frac{G \times E}{n} + \frac{e}{nr})$$

where G is the genetic variance, G×E is the variance of the genotype-environment interaction, e is the residual error variance, n is the number of environments, and r is the number of replicates within the environment.

QTL identification and analysis

The QTL for FBH was detected using composite interval mapping(CIM) (Zeng, 1994) in Windows QTL Cartographer 2.5 (Wang et al., 2005). The QTL evidence was checked at 1cM intervals with a 10 cM window size using the likelihood ratio test (LRT), after which the LRT values were converted to a logarithm of odds (LOD) scale (LOD¼; 0.217_LRT). The threshold LOD value was estimated using 1,000 permutations at a significance level of $P< 0.05$. The confidence interval of the QTL position was determined by the 2-LOD interval method. The contribution rate (R^2) and additive effect of a presumptive QTL were simultaneously calculated using Windows QTL Cartographer 2.5. The QTL nomenclature described by McCouch et al. (1997) was adopted in this study, each designation was named begins with '*q*' followed by the abbreviation of the trait name, the linkage group, and the serial number of the QTL on the linkage group, such as *qFBH*02-1. The QTLs position was drawn using MapChart software (Voorrips, 2002).

GWAS for FBH

GWAS for FBH was performed using a mixed linear model (MLM) to detect associations between phenotypes and genotypes in TASSEL v5.0 (http://www.maizegenetics.net/tassel) (Zhang et al., 2010). The population structure and genetic relationships between accessions were also calculated using TASSEL v5.0 in previous study (Zhang et al., 2019b). The *P* value was calculated for each SNP as $P = 1/N$ according to the adjusted Bonferroni method, where N is the number of SNPs calculated by the simple program in R software (Li et al., 2012; Li et al., 2021), and defined as the genome-wide control threshold. The quantile-quantile (Q-Q) plot shown with the expected *P*-value and Manhattan plot were displayed using the R package "*qqman*" (Turner, 2014). Linkage disequilibrium (LD) of the whole genome was analyzed using PopLDdecay software (Zhang et al., 2019a).

The LD heat map of 0.5 LD sequence regions adjacent to the significantly associated SNPs were generated using the "*LD heatmap*" package in R.

RNA extraction and gene expression quantification

Parental SAM samples were collected at the initial developmental stage of the shoot bud. Total high-quality RNA was extracted using the FastPure Plant Kotal RNA Isolation Kit (Vazyme, Nanjing, China). Then, cDNA was synthesized using HiScript III 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China). Quantitative real-time PCR (qRT-PCR) was performed using SYBR Green Real-time PCR (Takara, Kusatsu, Japan) in a CFX Connect Real-time PCR system (BioRad, United States), and *BnaActin* was used as the reference gene. The $2^{-\Delta\Delta CT}$ method was used to calculate the relative gene expression levels (Livak and Schmittgen, 2001). Each qRT-PCR experiment contained three technical replicates. The specific primers used for qRT-PCR are listed in Supplementary Table 1. Genes were considered as differentially expressed between parents when the $\log_2$fold change was ≥ 1 or ≤ -1, and they can be selected as candidate genes.

Gene analysis in the common interval

Allele variation and sub-genome synteny analysis were used to exclude genes when predicting candidate genes for FBH (Zhang et al., 2019b). Based on the *B.napus* cv. Darmor-*bzh* reference genome (Chalhoub et al., 2014), we matched the genetic markers for linkage mapping to the physical positions, and compared the linkage and association mapping results. Then, parallel region was defined as co-located interval, which is the major QTL for trait. Parental allele variation analysis was performed using homozygous polymorphic SNP and InDel derived from deeply re-sequenced parent data. The genes within the major QTL were analyzed using sub-genome synteny, and synteny genes existed in another sub-genome of *B.napus* were excluded. Candidate genes were identified combined with gene expression levels derived from qRT-PCR data.

Haplotype analysis of candidate genes

The 324 accessions of the association population were used to analyze the haplotypes of candidate genes in the genomic regions. The SNPs within the 1.5 kb promoter region of the candidate genes and the nonsynonymous SNPs in the coding sequence were used for haplotype analysis.

Analyses of parental *BnaA02.TCP* sequence

The total parent DNA was extracted from the leaf samples using the Hi-DNA secure Plant Kit (TianGen, Beijing, China). The sequence of the genomic regions ~1.5 kb upstream of the *BnaA02.TCP1* coding sequence was cloned using the primers listed in Supplementary Table 2. The cloned products were sent to QingKe Biotechnology Co., LTD for sequencing, and the genomic sequences were compared using MEGA-X (Tamura et al., 2011).

Construction of the *35S:: BnaA02.TCP1*^{M083}molecular cassette and plant transformation

The *BnaA02.TCP1*^{M083} coding sequence was isolated using the primers listed in Supplementary Table 2 to generate the *35S:: BnaA02.TCP1*^{M083} construct. The PCR fragment was cloned into the pBI121 vector and was driven by the 35S promoter. *Agrobacterium tumefaciens* strain GV3101 carrying the sequenced construct was confirmed *via* colony PCR and transformed into col-0 plants using the floral dipping method in *Arabidopsis*. Surviving seedlings with kanamycin were positive transgenic plants and were validated using qRT-PCR experiments, *AtActin* was used as the reference gene and three repeats of samples were collected for RNA extraction. Six transgenic lines were obtained, three of which were randomly chosen for phenotypic observation.

Results

Phenotypic variation in RIL and association populations

The FBH values of 888-5 and M083 was 14.9 ± 3.8 cm and 70.5 ± 10.2 cm, respectively, which showed highly significant difference in the field trial (Figures 1A, B). The phenotypic data of the FBH of RIL lines were investigated in five environments (Supplementary Table 3), and all displayed normal frequency distributions with continuous variation (Figure 1C). Significant variation among the FBH genotypes was observed: for instance, the minimum FBH was 6.15 cm, while the maximum was 121.4 cm, and the coefficient of variation ranged from 19.84 to 40.94% (Table 1), which showed great distinction. In the association population, the phenotypic frequency distributions of FBH in the four environments and BLUP all presented approximately continuous and normal distributions (Figure 1D; Supplementary Table 5), indicating that the group was suitable for association analysis. Significant variation was also observed in the

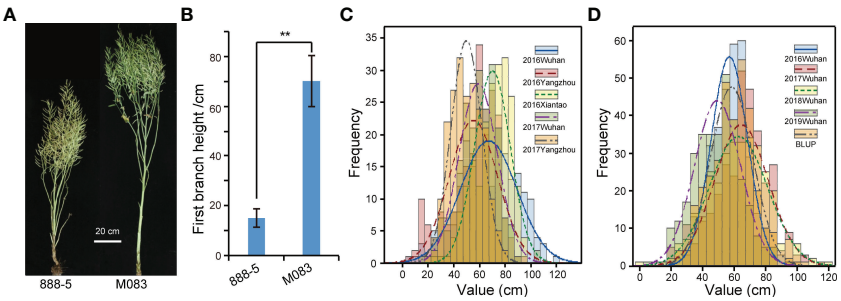


FIGURE 1
Parent phenotype and phenotypic frequency distribution of recombinant inbred lines (RILs) and natural accessions. **(A)** The mature stage phenotype of 888-5 and M083. Scale bar = 20 cm. **(B)** Comparison of first branch height (FBH) between 888-5 and M083. Data are presented as means ± SD (n > 20). The double asterisks represent a significant difference determined *via* Student's *t*-test at $P < 0.01$. **(C)** FBH phenotypic frequency distribution in RILs under five environments. **(D)** FBH phenotypic frequency distribution in 324 collected accessions under four consecutive years and best linear unbiased prediction (BLUP).

association population, for example, FBH ranged from 23.41 to 99.11 cm, with an average of 58.75 cm in BLUP; the coefficient of variation was 23.07% (Table 1). The coefficient of variation was greater than 19.84% in both RIL and association populations (Table 1), which indicated dispersion measures were high among accessions or lines. Broad-sense heritability (H^2) of FBH in linkage and association populations were 88.96% and 92.61% with significant variation ($P < 0.01$), respectively (Supplementary Table 4).

FBH association analysis

Based on the whole genome sequencing (WGS) data, a total of 2,465,230 high-quality SNPs (minor allele frequency, MAF > 0.05, missing rate< 0.5) were filtered in previous studies and the LD distances of the A and C sub-genomes were 21 and 128 kb ($r^2 > 0.2$), respectively. The quantile-quantile plots for GWAS based on the four environments and BLUP values of FBH implied that the associations were well controlled for population structure, and the Manhattan plots revealed only one significant peak located on ChrA02 in the four environments and BLUP (Figures 2). At a significance level of $P < 1/2{,}465{,}230$, 26 significant SNPs were mapped on A02 chromosome in BLUP within the interval of 6.92-7.11 Mb between SNPs p6926697 and p7113686. (Figure 2A; Supplementary Table 6). These results suggest the presence of an important genetic locus in A02 that regulates FBH.

QTL linkage mapping of FBH

Based on the modified *B.napus* 60 K SNP array and a high density genetic map developed in a previous study (Zhang

TABLE 1 Phenotypic variation in the linkage and association populations.

Population	Env	Min (cm)	Max (cm)	Mean (cm)	SE	SD	Var	CV (%)	Kurtosis	Skewness
RIL	2016Wuhan	7.33	121.40	67.54	1.52	21.95	481.99	32.51	-0.225	-0.310
	2016Yangzhou	6.15	93.07	52.81	1.49	21.62	467.47	40.94	-0.637	-0.801
	2016Xiantao	27.38	95.87	70.41	0.96	13.97	195.13	19.84	-0.617	-0.093
	2017Wuhan	14.00	105.50	59.34	1.03	14.95	223.47	25.19	-0.267	0.449
	2017Yangzhou	19.41	78.50	50.13	0.83	12.08	145.83	24.09	0.099	-0.617
GWAS	2016Wuhan	24.96	91.19	57.35	0.64	11.59	134.23	20.20	-0.160	-0.007
	2017Wuhan	8.62	116.30	64.73	0.97	17.47	305.07	26.98	-0.471	0.703
	2018Wuhan	1.95	118.79	62.65	1.07	19.21	369.15	30.67	-0.192	0.467
	2019Wuhan	5.60	94.60	49.72	0.81	14.65	214.61	29.47	0.028	-0.133
	BLUP	23.41	99.11	58.75	0.75	13.55	183.65	23.07	-0.137	-0.089

Env, environment; Min, minimum value; Max, maximum value; Mean, mean value; SE, standard error; SD, standard deviation; Var, variance; CV, coefficient of variation.

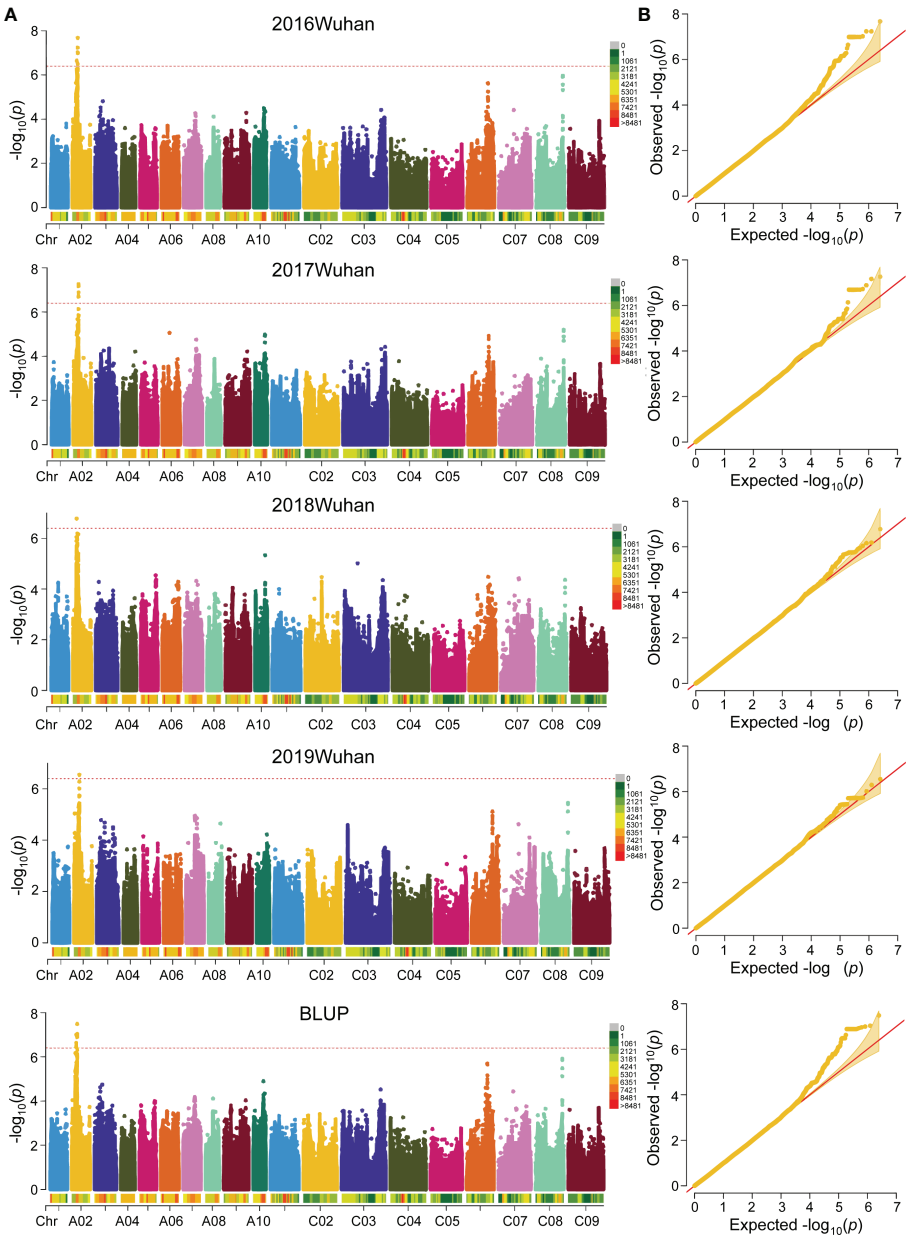


FIGURE 2
Genome-wide association study (GWAS) of FBH trait in 324 accessions. **(A)** Manhattan plot of FBH identified by GWAS in four consecutive years and BLUP based on a mixed linear model (MLM) of Efficient Mixed-Model Association eXpedited (EMMAX). The red dotted lines and color blocks represent threshold lines and SNP number in 1 Mb window size, respectively. **(B)** Quantile-Quantile plot of FBH in four consecutive years and BLUP.

et al., 2019b), a total of 19 QTLs were detected for FBH in single environment using WinQTL Cartographer 2.5 (Figure3A; Table 2). These QTLs were distributed on linkage groups A02, A09, A10, C06, and C07, with phenotypic variation explained (PVE) rates ranging from 4.87 to 29.87% (Table 2). Among these QTLs, six were located in an approximate region on linkage group A02, and *qFBH*02-1, *qFBH*02-2, and *qFBH*02-3 were repeatedly detected in more than one environment, with the PVE ranging from 15.77 to 29.87% (Figures 3B, C; Table 2). In addition, *qFBH*02-5 demonstrated a PVE as high as 18.96%, but it was only detected in E5. Simultaneously, four QTLs were detected in group C07 under all environments, except E1, and explained 5.61 ~ 9.02% of phenotypic variation (Table 2). Among the 19 QTLs, *qFBH*09-1, *qFBH*09-2 and *qFBH*09-3

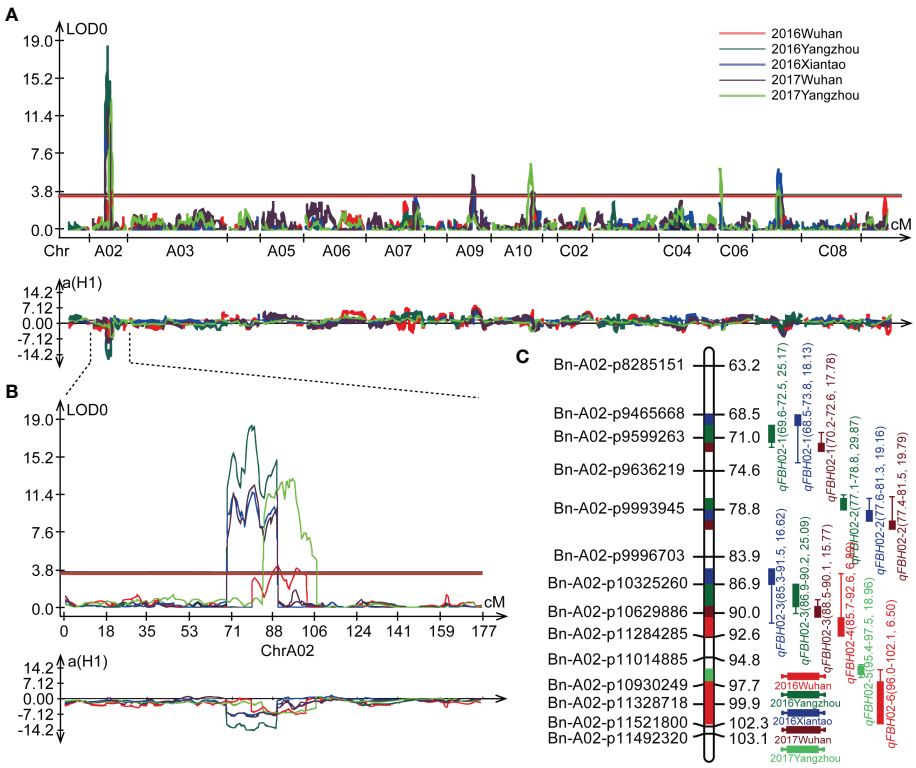


FIGURE 3
Quantitative trait loci (QTL) identification for FBH in five environments. **(A)** QTL distribution on the whole genome; diverse colors represent the QTL identified in different environments. **(B)** Localization of the identified QTLs on linkage group A02 in different environments. **(C)** Genetic markers and QTL localizations on linkage group A02.

low-FBH alleles came from 888-5, and the remaining high-FBH alleles QTLs were provided by M083.

Overlaps between linkage and association mapping

Integrated linkage and association mapping results, showed that the 26 significant SNP loci in GWAS and *qFBH*02-3 obtained *via* linkage mapping had overlapping regions (Figure 4A), which indicated that the overlapped region could be considered the major QTL for FBH trait. LD decay analysis of 324 rapeseed germplasm resources showed that the decay distances ($r^2 > 0.2$) of sub-genomes A and C were approximately 20 kb and 60 kb, respectively (Figure 4B). The attenuation distance of sub-genome A was significantly smaller than that of sub-genome C, indicating that sub-genome A had a higher degree of chromosome variation and genetic diversity, whereas sub-genome C was more stable. Simultaneously, LD heatmap analysis 60 Kb upstream and downstream of the significant SNPs revealed that three tightly linked blocks within an interval of 6.89 - 7.12 Mb on chromosome A02, named *qA02.FBH* (Figure 4C).

Candidate gene predictions

The *qA02.FBH* region covers 230 kb and contains 31 annotated genes according to the reference genome annotation (Darmor-*bzh*), three of which (*BnaA02g12830D*, *BnaA02g12900D*, and *BnaA02g12930D*) with several hundreds of nucleotides had no orthologous genes in *Arabidopsis*. Four genes, *BnaA02g12770D*, *BnaA02g12780D*, *BnaA02g12910D*, and *BnaA02g13070D*, encoded unknown proteins. The remaining genes encoded functional proteins, including four pairs of genes (*BnaA02g12870D* and *BnaA02g12880D*, *BnaA02g12940D* and *BnaA02g12950D*, *BnaA02g12980D* and *BnaA02g12990D*, *BnaA02g13030D* and *BnaA02g13040D*) had the same functions as orthologous gene in *Arabidopsis* (Supplementary Table 7). At the DNA level, the SNP and InDel variations of the *qA02.FBH* between the two parents were analyzed using deep re-sequencing data (Figure 5A). In total, 1,398 variations were identified, most of

10.3389/fpls.2022.1080999

TABLE 2 Putative QTLs for rapeseed FBH in RIL population under different environments.

QTL	Peak position (cM)	PVE (%)	Additive effect	LOD	Interval (cM)	Env
*qFBH*02-1	71.1	25.17	-12.94	14.67	69.6-72.5	E2
	71.1	18.13	-7.13	14.32	68.5-73.8	E3
	71.1	17.78	-7.57	14.45	70.2-72.6	E4
*qFBH*02-2	78.8	29.87	-14.16	8.79	78.2-78.8	E2
	78.8	19.16	-7.32	8.88	77.6-81.3	E3
	78.8	19.79	-8.10	8.94	77.4-81.5	E4
*qFBH*02-3	89.5	16.62	-7.14	11.43	85.3-91.5	E3
	89.5	25.09	-13.17	13.98	86.9-90.2	E2
	89.5	15.77	-7.33	11.29	88.5-90.1	E4
*qFBH*02-4	90.0	6.89	-6.96	4.69	85.7-92.6	E1
*qFBH*02-5	95.9	18.96	-6.32	5.38	95.4-97.5	E5
*qFBH*02-6	97.7	6.50	-6.83	5.12	96.0-102.1	E1
*qFBH*09-1	128.8	6.93	5.82	3.74	122.4-130.4	E1
	128.8	4.87	4.82	3.81	128.3-130.4	E2
	128.8	7.69	4.25	4.21	123.1-131.6	E4
*qFBH*09-2	134.2	5.07	4.90	2.76	132.8-140.8	E2
*qFBH*09-3	137.6	5.71	3.43	4.57	134.8-144.8	E3
*qFBH*10-1	205.5	8.65	-3.73	4.61	194.6-209.5	E5
*qFBH*10-2	211.7	8.64	-3.65	5.25	209.5-213.7	E5
*qFBH*10-3	212.7	6.09	-3.75	2.55	206.6-214.8	E4
*qFBH*10-4	218.5	5.98	-3.71	4.45	214.8-220.9	E4
*qFBH*16-1	2.0	8.20	-3.48	3.67	0.0-4.3	E5
*qFBH*16-2	8.7	6.56	-3.14	3.72	7.5-10.0	E5
*qFBH*17-1	127.0	6.61	-4.09	4.05	122.9-130.7	E4
	127.0	5.61	-2.97	4.12	122.9-129.1	E5
*qFBH*17-2	133.8	9.02	-4.45	3.73	130.3-135.4	E3
	134.2	5.86	-3.04	3.63	129.1-134.3	E5
*qFBH*17-3	145.6	6.84	-5.92	4.06	413.5-148.4	E2
	145.6	8.46	-4.32	3.89	145.2-149.2	E3
*qFBH*17-4	153.8	5.70	-5.45	3.67	151.2-157.1	E2

PVE, phenotypic variation explained; LOD, logarithm of odds; Env, environment; E1, 2016Wuhan; E2, 2016Yangzhou; E3, 2016Xiantao; E4, 2017Wuhan; E5, 2017Yangzhou.

which were located in the intergenic and intragenic region of the candidate genes, and 114 SNPs and 10 InDels resulting in missense mutations and frameshift or codon changes in the coding region, respectively (Supplementary Tables 8, 9).

Synteny analysis revealed that, the syntenic block of *qA02.FBH* ranged from 12,397,768 to 13,133,808 bp on chromosome C02 (Figure 5B). There was no syntenic genes for *BnaA02g12830D*, *BnaA02g12840D*, *BnaA02g12890D*, *BnaA02g12910D*, *BnaA02g12920D*, and *BnaA02g12940D* in the syntenic block (Table 3). In contrast, there were no QTLs or significant SNPs identified in the syntenic block based on the association and linkage mapping results (Figures 2A, 3A), suggesting that these six genes should be considered as candidate genes. Simultaneously, seven genes, *BnaA02g12840D*, *BnaA02g12850D*, *BnaA02g12860D*, *BnaA02g12890D*, *BnaA02g13010D*, *BnaA02g13020D*, and *BnaA02g13060D*, were

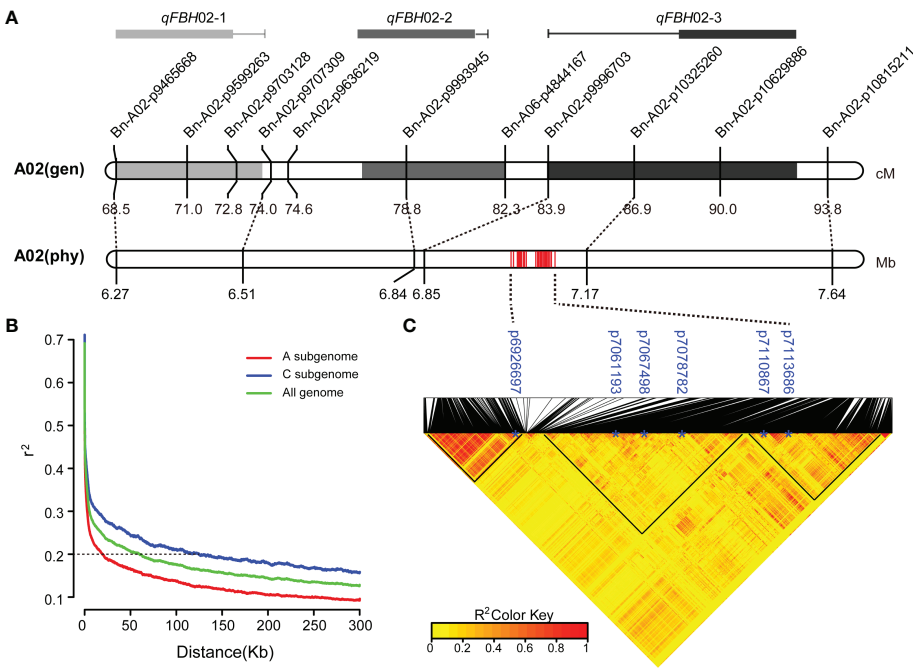


FIGURE 4
Co-localization of *qA02.FBH*. **(A)** The genetic location corresponds to the physical location. **(B)** Linkage disequilibrium (LD) decay of the **(A, C)** subgenomes and the whole genome. **(C)** LD heatmap surrounding the peak on A02. The black triangles represent closely linked loci.

differentially expressed between SAM tissues of the two parents, as demonstrated using RT-qPCR (Figure 5C; Table 3). The variations and haplotype results of above eleven genes revealed that *BnaA02g13010D* had TATA variation in the promoter region, resulting in a significant difference in BLUP (Figure 5D; Table 3). A BLAST search revealed that *BnaA02g13010D* was annotated as a transcription factor with a TCP (TB1, CYC, PCF) domain (*BnaA02.TCP1*) gene in BRAD (http://brassicadb.cn/), and its homologous genes were associated with plant architecture in *Arabidopsis* (Guo et al., 2010; Gao et al., 2015) and maize (Dixon et al., 2020).

BnaA02.TCP altered plant architecture in transgenic *Arabidopsis*

Sequencing results showed that *BnaA02.TCP1*^{M083} had a TATA deletion in the promoter region and 14 missense mutations in the coding region compared with *BnaA02.TCP1*$^{888\text{-}5}$ (Figure 6A; Supplementary Figure 1).*BnaA02.TCP1* expression was higher in 888-5 SAM than in M083 SAM (Figure 5C). Consequently, an overexpression (OE) vector containing *BnaA02.TCP1*^{M083} was transferred into wild-type *Arabidopsis* (col-0). Transgenic lines were validated by qRT-PCR and three highly expressed lines were randomly selected for phenotyping (Supplementary Figure 2). At the mature stage, *BnaA02.TCP1*^{M083}-OE transgenic plants exhibited higher FBH and more branches than the wild type plants (Figures 6B-D). Phylogenetic tree analysis of *BnaA02.TCP1* showed that *AtTCP1* was most similar in sequence, suggesting that they are homologous and perform the same function. These findings implied that *BnaA02.TCP1* may respond to rapeseed plant architecture development, as in *Arabidopsis*.

Discussion

Comparisons with previous study of FBH

Improvements in traits of agronomic importance are the primary objective of crop improvement programs. Most agronomic traits are controlled by complex or quantitative loci (Abe et al., 2012). With the increase in rapeseed yield demand, labor cost, and decrease of farmer profits, many challenges require a quantum leap in seed yield, reducing yield loss caused by biotic and abiotic stresses, and applying gene editing and genome breeding biotechnology tools (Liu et al., 2022). As a component of plant architecture, FBH is not as important as other traits, such as plant height and flowering time, but it is also a concern for breeders. Appropriate branch height can significantly increase rapeseed yield during practical production (Cai et al., 2016), so the underlying molecular mechanisms must be elucidated. However, only a few studies

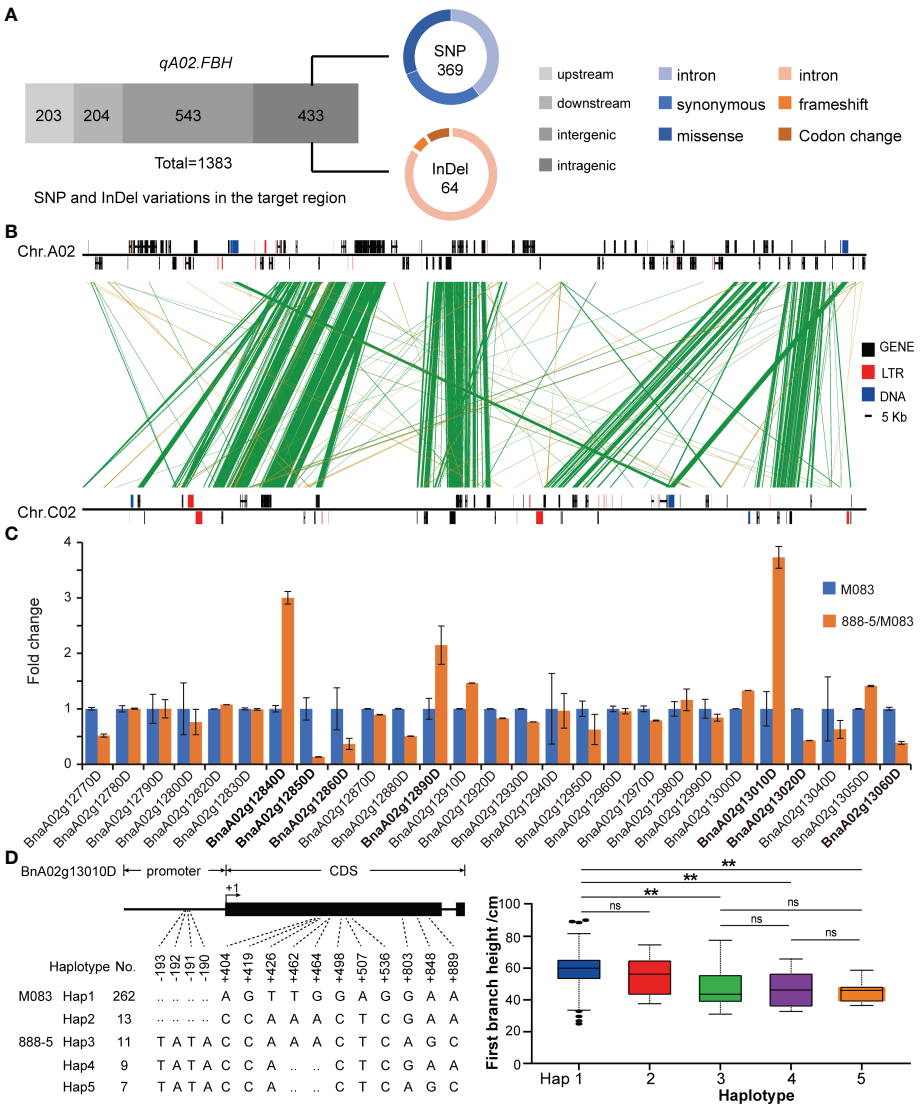


FIGURE 5
Candidate gene prediction *via* manifold analysis. **(A)** The identified variations in *qA02.FBH*. **(B)** Synteny analysis for *qA02.FBH* between subgenomes. **(C)** Expression levels of annotated genes in *qA02.FBH*. **(D)** Haplotype analysis of *BnaA02g13010D* based on association population in BLUP. Data are presented as means ± SD. The double asterisks represent a significant difference determined *via* Student's *t*-test at $P < 0.01$. ns represents no significance.

have reported on FBH in rapeseed, and studied on cloning genes related to FBH still lacking.

Chen et al. (2007) identified 10 QTLs distributed on chromosomes A02, A07, A08, C04, C06, and C07, associated with the height of the lowest primary effective branch using two rapeseed populations. Another study detected four peak SNPs located on chromosomes A02, A07, A08, and A09 using GWAS with 333 collected rapeseed accessions (Zheng et al., 2017). Candidate genes for FBH were identified, but no functional genes discovered or demonstrated. In this study, we used an F10 RIL population derived from M083 × 888-5 cross in QTL analysis of FBH. Nineteen QTLs were identified on chromosomes A02, A09, A10, C06, and C07 (Figure 3A; Table 2). The QTLs located on chromosome A10 were different from QTLs associated with FBH in previous studies, while the QTLs on chromosome A02 were consistent with previous study, which suggests that our results were reliable, novel, and worthy of further investigation. Using the same RIL population, Zhang et al. (2019b) performed QTL mapping for flowering time and *Sclerotinia* resistance traits. Nineteen QTLs for flowering time and twenty-five QTLs for *Sclerotinia* resistance were identified. Interestingly, *qFBH*02-1 and

TABLE 3 Candidate genes for FBH in the *qA02.FBH* interval.

B.napus gene ID	*Ath* homolog	Gene annotation	Variations	Syntenic gene	log_2fold change
BnaA02g12830D	#N/A	#N/A	√	√	-0.014
BnaA02g12840D	*AT5G38430.1*	Ribulose bisphosphate carboxylase (small chain) family protein	√	√	1.586
BnaA02g12850D	*AT1G67100.1*	LOB domain-containing protein 40	×	×	-2.913
BnaA02g12860D	*AT1G67110.1*	cytochrome P450, family 735, subfamily A, polypeptide 2	√	×	-1.444
BnaA02g12890D	*AT5G40480.1*	embryo defective 3012 (EMB3012), embryo development ending in seed dormancy	√	√	1.103
BnaA02g12910D	*AT1G67170.1*	unknown protein	√	√	0.550
BnaA02g12920D	*AT1G67180.1*	zinc finger (C3HC4-type RING finger) family protein/BRCT domain-containing protein	√	√	-0.271
BnaA02g12940D	*AT2G31470.1*	F-box and associated interaction domains-containing protein	√	√	0.312
BnaA02g13010D	*AT1G67260.2*	TCP family transcription factor	√	×	1.900
BnaA02g13020D	*AT1G67780.1*	Zinc-finger domain of monoamine-oxidase A repressor R1 protein	√	×	-1.221
BnaA02g13060D	*AT1G67320.2*	DNA primase, large subunit family, DNA replication, synthesis of RNA primer	√	×	-1.381

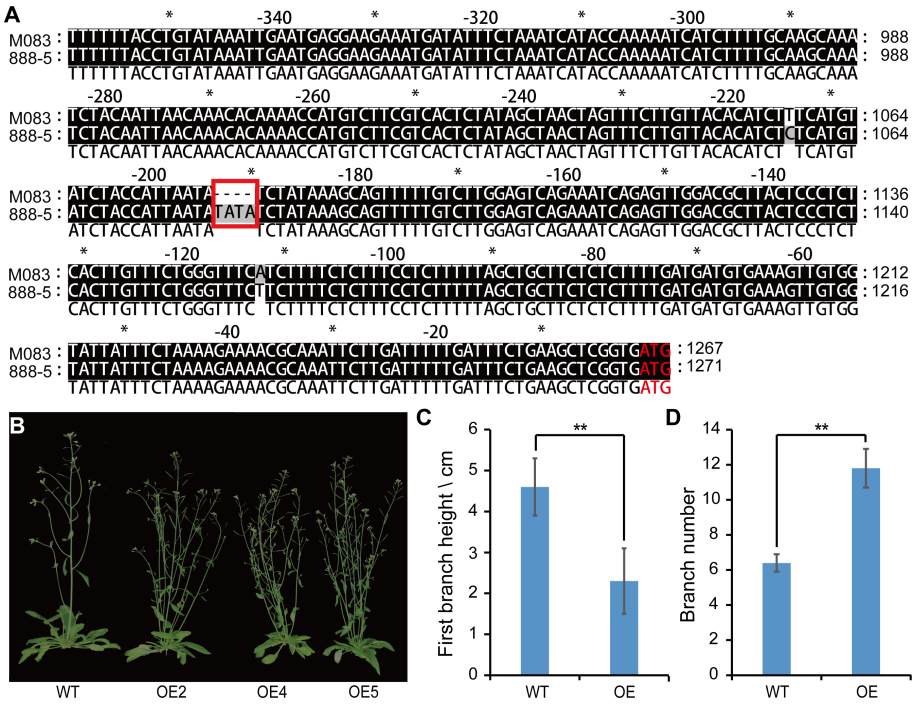


FIGURE 6
Sequence analysis and phenotypic comparison of wild type (WT) plants and the *BnaA02.TCP1*^{M083}-OE lines in *Arabidopsis*. **(A)** Alignment of *BnaA02.TCP1* promotor sequences between 888-5 and M083. The red rectangle represents the nucleotide change between 888-5 and M083. **(B)** Phenotype comparison of WT and *BnaA02.TCP1*^{M083} overexpression (OE) lines in *Arabidopsis* at the grain filling stage. **(C, D)** Digital comparison of **(C)** FBH and **(D)** branch number between WT and OE lines in *Arabidopsis*. Data are presented as means ± SD (n = 10). The double asterisks represent a significant difference determined using Student's *t*-test at $P < 0.01$.

*qFBH*02-3 identified in the current study co-localized with QTLs controlling flowering time. *qFBH*02-3 and *qFBH*02-4 were consistently correlated with the QTLs of *Sclerotinia* resistance, indicating that *qFBH*02-3 might have pleiotropic effects on controlling FBH, flowering time, and *Sclerotinia* resistance.

Association and linkage mapping combination for rapid candidate interval identification

QTL identification, followed by mapping and cloning of candidate genes/QTLs are central to agronomic traits analysis. Linkage mapping, a traditional QTL mapping method, was used to accurately and slowly detect QTL regions with genetic makers. GWAS has been widely used to detect variations associated with complex agronomic traits in natural populations with the development of whole-genome sequencing over last 10 years, which may circumvent the cumbersome procedures for fine mapping (Jaganathan et al., 2020; Wang et al., 2020a). The complex genetic background of allotriploid rapeseed and the agronomic traits usually affected by the environment complicate integral and accurate detection using a single method. Therefore, integrating association and linkage mapping would be quick and accurate. Previous studies have used this approach to map QTL in crops (Sun et al., 2016; Han et al., 2018; Wang et al., 2020b; Zhang et al., 2020). In this study, only a single locus contained 26 SNPs associated with the FBH trait was identified in the natural population with GWAS (Figure 2A), which is very different from linkage mapping results (Table 2). The reasons may be the association population size was small, SNPs with low MAF were filleted out (MAF< 0.05), and trait was influenced by environment (Qiu et al., 2015). Therefore, major QTL could be detected but minor QTLs can't in small panel size (Si et al., 2016). Meanwhile, these significant SNPs were co-located with *qFBH*2-3, detected using linkage mapping in multiple environments and exhibited high phenotypic variation (Figure 4). Overall, the findings of the present study suggest that the *qA02.FBH* interval would be the major QTL for FBH (Figure 4C).

BnaA02.TCP possibly affected rapeseed plant architecture by mediating strigolactone signaling

Hormones are crucial regulators of plant growth and development. Several studies have demonstrated that strigolactone is an important factor that regulates axillary bud growth and a determinant of plant architecture (Doebley et al., 1995; Martín-Trillo and Cubas, 2010; Dixon et al., 2020; Wang et al., 2020c). In plants, it remains unclear which and how genes involved in the strigolactone pathway could be utilized in breeding, but one possibility is that strigolactones regulate transcription of the TCP-domain transcription factor in shoot branching (Dun et al., 2012). In this study, *BnaA02.TCP1* was differentially expressed at the transcriptional level in parental SAM tissues owing to TATA variation in the promoter region (Figures 5D, 6A). *BnaA02.TCP1*^{M083}-OE plants showed lower FBH and more branches compared with wild type in *Arabidopsis* (Figures 6B-D). Although *BnaA02.TCP1* over-expression and loss-of-function in rapeseed and the proteins interactions remain elusive in this study, the above findings are still credible, because rapeseed originates from *Arabidopsis* (Chalhoub et al., 2014). Therefore, we concluded that BnaA02.TCP1 closely related *Arabidopsis* TCP proteins, may be involved in strigolactone pathway activity to regulate the growth of rapeseed axillary buds.

Conclusion

In this context, the genetic mechanism underlying FBH was analyzed through multiple strategies (integrating GWAS, linkage mapping, and gene expression analysis) in *B. napus*. In summary, 26 SNPs and 19 QTLs associated with FBH were identified in 324 accessions and RIL population derived from M083 and 888-5, respectively. Meta-analysis revealed that one unique QTL for FBH was co-located on chromosome A02. An integrated analysis of allelic variations, syntenic genes, and DEGs identified 11 candidate genes for FBH. Among 11 genes, *BnaA02g13010D* was confirmed as the target gene according to haplotype analysis and full-length gene sequencing, which revealed that TATA insertion/deletion in the promoter region was closely linked to significantly phenotypic differences. *BnaA02.TCP1*^{M083} overexpression resulted in decreased branch height and increased branch number in *Arabidopsis*. Therefore, our data suggested that *BnaA02.TCP1* might play important role as regulator of FBH and provided a genetic basis for shaping ideal architecture in *B. napus*.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NGDC (https://ngdc.cncb.ac.cn/gsa), CRA008925.

Author contributions

ZD, ZL, YX, JH, XCh, and SL designed the research. ZD, CZ, XCu, YL, MT, CT, and JH performed the experiments. ZD and JH analyzed the data. YL provided the materials and phenotypic data. ZD, XCh, and JH wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Key Research and Development Program of China (2021YFD1600500), Young Top-notch Talent Cultivation Program of Hubei Province for CT, National Natural Science Foundation of China (31770250), and Precursor projects of Guizhou province for biological breeding supporting by science and technology in 2022 (Fine identification and evaluation of crop germplasm resources).

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 JD declared a shared affiliation with the authors ZD, MT, XC, CZ, CT, YL, JH, XCh, SL to the handling editor at the time of the 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.2022.1080999/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Alignment of the deduced amino acid sequence between 888-5 and M083. The red rectangle represents the TCP domain.

SUPPLEMENTARY FIGURE 2
Expression levels of *BnaA02.TCP1* in wild-type (WT) and transgenic *Arabidopsis* plants (OE2, OE4, and OE5).

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095

Cai, G., Yang, Q., Chen, H., Yang, Q., Zhang, C., Fan, C., et al. (2016). Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Sci. Re.p.* 6, 21625. doi: 10.1038/srep21625

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chen, Y., Dan, Z., Gao, F., Chen, P., Fan, F., and Li, S. (2020). Rice *GROWTH-REGULATING FACTOR7* modulates plant architecture through regulating GA and indol-3-acetic acid metabolism. *Plant Physiol.* 184, 393–406. doi: 10.1104/pp.20.00302

Chen, W., Zhang, Y., Liu, X., Chen, B., Tu, J., and Fu, T. (2007). Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F_2 populations. *Theor. Appl. Genet.* 115, 849–858. doi: 10.1007/s00122-007-0613-2

Dixon, L. E., Pasquariello, M., and Boden, S. A. (2020). *TEOSINTE BRANCHED1* regulates height and stem internode length in bread wheat. *J. Exp. Bot.* 71, 4742–4750. doi: 10.1093/jxb/eraa252

Doebley, J., Stec, A., and Gustus, C. (1995). *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333–346. doi: 10.1093/genetics/141.1.333

Dun, E., De Saint Germain, A., Rameau, C., and Beveridge, C. A. (2012). Antagonistic action of strigolactone and cytokinin in bud outgrowth control. *Plant Physiol.* 158, 487–498. doi: 10.1104/pp.111.186783

Fan, S., Zhang, L., Tang, M., Cai, Y., Liu, J., Liu, H., et al. (2021). CRISPR/Cas9-targeted mutagenesis of the *BnaA03.BP* gene confers semi-dwarf and compact architecture to rapeseed (*Brassica napus* L.). *Plant Biotechnol. J.* 19, 2383–2385. doi: 10.1111/pbi.13703

Gao, Y., Zhang, D., and Li, J. (2015). *TCP1* modulates *DWF4* expression *via* directly interacting with the GGNCCC motifs in the promoter region of *DWF4* in *Arabidopsis thaliana*. *J. Genet. Genomics* 42, 383–392. doi: 10.1016/j.jgg.2015.04.009

Gou, J., Fu, C., Liu, S., Tang, C., Debnath, S., Flanagan, A., et al. (2017). The miR156-SPL4 module predominantly regulates aerial axillary bud formation and controls shoot architecture. *New Phytol.* 216, 829–840. doi: 10.1111/nph.14758

Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., et al. (2020). Overexpression of *GmWRI1b* in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. *Plant Biotechnol. J.* 18, 1639–1641. doi: 10.1111/pbi.13324

Guo, Z., Fujioka, S., Blancaflor, E. B., Miao, S., Gou, X., and Li, J. (2010). TCP1 modulates brassinosteroid biosynthesis by regulating the expression of the key biosynthetic gene *DWARF4* in *Arabidopsis thaliana*. *Plant Cell* 22, 1161–1173. doi: 10.1105/tpc.109.069203

Han, K., Lee, H. Y., Ro, N. Y., Hur, O. S., Lee, J. H., Kwon, J. K., et al. (2018). QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnol. J.* 16, 1546–1558. doi: 10.1111/pbi.12894

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet.* 133, 1791–1810. doi: 10.1007/s00122-020-03560-w

Kuai, J., Li, X., Ji, J., Li, Z., Xie, Y., Wang, B., et al (2020). Response of leaf carbon metabolism and dry matter accumulation to density and row spacing in two rapeseed (*Brassica napus* L.) genotypes with differing plant architectures. *Crop J.* 10, 680–691. doi: 10.1016/j.cj.2021.10.006

Khush, G. S. (2001). Green revolution: The way forward. *Nat. Rev. Genet.* 2, 815–822. doi: 10.1038/35093585

Liu, K., Cao, J., Yu, K., Liu, X., Gao, Y., Chen, Q., et al. (2019). Wheat *TaSPL8* modulates leaf angle through auxin and brassinosteroid signaling. *Plant Physiol.* 181, 179–194. doi: 10.1104/pp.19.00248

Liu, S., Raman, H., Xiang, Y., Zhao, C., Huang, J., and Zhang, Y. (2022). *De novo* design of future rapeseed crops: Challenges and opportunities. *Crop J.* 10, 587–596. doi: 10.1016/j.cj.2022.05.003

Liu, S., Wang, H., Zhang, J., Fitt, B. D., Xu, Z., Evans, N., et al. (2005). *In vitro* mutation and selection of doubled-haploid *Brassica napus* lines with improved resistance to *Sclerotinia sclerotiorum*. *Plant Cell Rep.* 24, 133–144. doi: 10.1007/s00299-005-0925-0

Liu, C., Zheng, S., Gui, J., Fu, C., Yu, H., Song, D., et al. (2018). Shortened basal internodes encodes a gibberellin 2-oxidase and contributes to lodging resistance in rice. *Mol. Plant* 11, 288–299. doi: 10.1016/j.molp.2017.12.004

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

Li, M., Yeung, J., Cherny, S., and Sham, P. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial

genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. doi: 10.1007/s00439-011-1118-2

Li, W., Yu, Y., Wang, L., Luo, Y., Peng, Y., Xu, Y., et al. (2021). The genetic architecture of the dynamic changes in grain moisture in maize. *Plant Biotechnol. J.* 19, 1195–1205. doi: 10.1111/pbi.13541

Lo, S., Ho, T., Liu, Y., Jiang, M., Hsieh, K., Chen, K., et al. (2017). Ectopic expression of specific GA2 oxidase mutants promotes yield and stress tolerance in rice. *Plant Biotechnol. J.* 15, 850–864. doi: 10.1111/pbi.12681

Luo, Z., Wang, M., Long, Y., Huang, Y., Shi, L., Zhang, C., et al. (2017). Incorporating pleiotropic quantitative trait loci in dissection of complex traits: Seed yield in rapeseed as an example. *Theor. Appl. Genet.* 130, 1569–1585. doi: 10.1007/s00122-017-2911-7

Martín-Trillo, M., and Cubas, P. (2010). *TCP* genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003

McCouch, S., Cho, Y., Yano, M., Paule, E., Blinstrue, M., and Kinosita, T. (1997). Report on QTL nomenclature. *Rice Genet. Newsl.* 14, 11–13.

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., et al. (2015). Genome-wide association study of grain appearance and milling quality in a worldwide collection of *Indica* rice germplasm. *PloS One* 10, e0145577. doi: 10.1371/journal.pone.0145577

Reinhardt, D., and Kuhlemeier, C. (2002). Plant architecture. *EMBO Reps.* 3, 846–851. doi: 10.1093/embo-reports/kvf177

Sarlikioti, V., De Visser, P. H., Buck-Sorlin, G. H., and Marcelis, L. F. (2011). How plant architecture affects light absorption and photosynthesis in tomato: towards an ideotype for plant architecture using a functional-structural plant model. *Ann. Bots.* 108, 1065–1073. doi: 10.1093/aob/mcr221

Schneider, A., Godin, C., Boudon, F., Demotes-Mainard, S., Sakr, S., and Bertheloot, J. (2019). Light regulation of axillary bud outgrowth slong plant sxes: sn overview of the roles of dugars and hormones. *Front. Plant Sci.* 10, 1296. doi: 10.3389/fpls.2019.01296

Shen, Y., Xiang, Y., Xu, E., Ge, X., and Li, Z. (2018). Major co-localized QTL for plant height, branch initiation height, stem diameter, and flowering time in an alien introgression derived *Brassica napus* DH population. *Fron. Plant Sci.* 9, 390. doi: 10.3389/fpls.2018.00390

Shen, J., Zhang, Y., Ge, D., Wang, Z., Song, W., Gu, R., et al. (2019). *CsBRC1* inhibits axillary bud outgrowth by directly repressing the auxin efflux carrier CsPIN3 in cucumber. *Proc. Natl. Acad. Sci. U. S. A.* 116, 17105–17114. doi: 10.1073/pnas.1907968116

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Sun, F., Liu, J., Hua, W., Sun, X., Wang, X., and Wang, H. (2016). Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*. *Plant Sci.* 252, 388–399. doi: 10.1016/j.plantsci.2016.09.001

Sun, Z., Su, C., Yun, J., Jiang, Q., Wang, L., Wang, Y., et al. (2019). Genetic improvement of the shoot architecture and yield in soya bean plants *via* the manipulation of *GmmiR156b*. *Plant Biotechnol. J.* 17, 50–62. doi: 10.1111/pbi.12946

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tang, J., and Chu, C. (2017). MicroRNAs in crop improvement: fine-tuners for complex traits. *Nat. Plants* 3, 17077. doi: 10.1038/nplants.2017.77

Teichmann, T., and Muhr, M. (2015). Shaping plant architecture. *Front. Plant Sci.* 6, 233. doi: 10.3389/fpls.2015.00233

Tian, J., Wang, C., Xia, J., Wu, L., Xu, G., Wu, W., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* 365, 658–664. doi: 10.1126/science.aax5482

Turner, S. (2014). qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* 005165. doi: 10.1101/005165

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, S., Basten, C., and Zeng, Z. (2005). *Windows QTL cartographerver. 2.5* (Raleigh, NC: NC State Department of Statistics).

Wang, Q., Hasson, A., Rossmann, S., and Theres, K. (2016). Divide et impera: boundaries shape the plant body and initiate new meristems. *New Phytol.* 209, 485–498. doi: 10.1111/nph.13641

Wang, Y., and Jiao, Y. (2018). Axillary meristem initiation-a way to branch out. *Curr. Opin. Plant Biol.* 41, 61–66. doi: 10.1016/j.pbi.2017.09.001

Wang, Y., and Li, J. (2006). Genes controlling plant architecture. *Curr. Opin. Biotechnol.* 17, 123–129. doi: 10.1016/j.copbio.2006.02.004

Wang, Y., and Li, J. (2008). Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* 59, 253–279. doi: 10.1146/annurev.arplant.59.032607.092902

Wang, Y., Shang, L., Yu, H., Zeng, L., Hu, J., Ni, S., et al. (2020c). A strigolactone biosynthesis gene contributed to the green revolution in rice. *Mol. Plant* 13, 923–932. doi: 10.1016/j.molp.2020.03.009

Wang, B., Smith, S., and Li, J. (2018). Genetic regulation of shoot architecture. *Annu. Rev. Plant Biol.* 69, 437–468. doi: 10.1146/annurev-arplant-042817040422

Wang, Q., Tang, J., Han, B., and Huang, X. (2020a). Advances in genome-wide association studies of complex traits in rice. *Theor. Appl. Genet.* 133, 1415–1425. doi: 10.1007/s00122-019-03473-3

Wang, T., Wei, L., Wang, J., Xie, L., Li, Y. Y., Ran, S., et al. (2020b). Integrating GWAS, linkage mapping and gene expression analyses reveals the genetic control of growth period traits in rapeseed (*Brassica napus* L.). *Biotechnol. Biofuels* 13, 134. doi: 10.1186/s13068-020-01774-0

Wang, J., Yu, H., Xiong, G., Lu, Z., Jiao, Y., Meng, X., et al. (2017). Tissue-specific ubiquitination by *IPA1 INTERACTING PROTEIN1* modulates IPA1 protein levels to regulate plant srchitecture in rice. *Plant Cell* 29, 697–707. doi: 10.1105/tpc.16.00879

Yang, J., Wang, M., Li, W., He, X., Teng, W., Ma, W., et al. (2019). Reducing expression of a nitrate-responsive bZIP transcription factor increases grain yield and N use in wheat. *Plant Biotechnol. J.* 17, 1823–1833. doi: 10.1111/pbi.13103

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, C., Dong, S., Xu, J., He, W., and Yang, T. (2019a). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R., Tiwari, H., Gore, M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhang, X., Guan, Z., Li, Z., Liu, P., Ma, L., Zhang, Y., et al. (2020). A combination of linkage mapping and GWAS brings new elements on the genetic basis of yield-related traits in maize across multiple environments. *Theor. Appl. Genet.* 133, 2881–2895. doi: 10.1007/s00122-020-03639-4

Zhang, F., Huang, J., Tang, M., Cheng, X., Liu, Y., Tong, C., et al. (2019b). Syntenic quantitative trait loci and genomic divergence for sclerotinia resistance and flowering time in brassica napus. *J. Integr. Plant Biol.* 61, 75–88. doi: 10.1111/jipb.12754

Zhang, X., Lin, Z., Wang, J., Liu, H., Zhou, L., Zhong, S., et al. (2019c). The *tin1* gene retains the function of promoting tillering in maize. *Nat. Commun.* 10, 5608. doi: 10.1038/s41467-019-13425-6

Zhao, W., Wang, X., Wang, H., Tian, J., Li, B., Chen, L., et al. (2016). Genome-wide identification of QTL for seed yield and yield-related traits and construction of a high-density consensus map for QTL comparison in *Brassica napus*. *Front. Plant Sci.* 7, 17. doi: 10.3389/fpls.2016.00017

Zhao, C., Xie, M., Liang, L., Yang, L., Han, H., Qin, X., et al. (2022). Genome-wide association analysis combined with quantitative trait loci mapping and dynamic transcriptome unveil the genetic control of seed oil content in *Brassica napus* L. *Front. Plant Sci.* 13, 929197. doi: 10.3389/fpls.2022.929197

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., et al. (2017). Genome-wide association study reveals candidate genes for control of plant height, branch initiation height and branch number in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 8, 1246. doi: 10.3389/fpls.2017.01246

Zheng, B., Cui, C., Zhang, J., Li, H., Chai, L., and Jiang, J. (2019). Correlation analysis of yield per plant and agronomic traits in breeding lines in Brassica napus L.. *J. Plant Genet. Res.* 20, 113–121. doi: 10.13430/j.cnki.jpgr.20180705001

Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., et al. (2020). Knockout of two *BnaMAX1* homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnol. J.* 18, 644–654. doi: 10.1111/pbi.13228

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fpls.2022.1101199

OPEN ACCESS

EDITED BY
Guo-Fei Tan,
Guizhou Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Chen Shen,
Yangzhou University, China
Jian Zhang,
University of British Columbia, Okanagan Campus, Canada
Xiaobo Luo,
Guizhou Academy of Agricultural Sciences (CAAS), China

*CORRESPONDENCE
Yangxia Zheng
zhengyx13520@sicau.edu.cn

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

RECEIVED 17 November 2022
ACCEPTED 28 November 2022
PUBLISHED 15 December 2022

CITATION
Huang Y, Li Y, Liu Z, Chen W, Wang Y, Wang X, Liu Y and Zheng Y (2022) Combined analysis of the transcriptome and metabolome provides insights into the fleshy stem expansion mechanism in stem lettuce. *Front. Plant Sci.* 13:1101199. doi: 10.3389/fpls.2022.1101199

Combined analysis of the transcriptome and metabolome provides insights into the fleshy stem expansion mechanism in stem lettuce

Ying Huang[1], Yanwen Li[2], Zhenning Liu[1], Wanqin Chen[3], Yalin Wang[3], Xiaohua Wang[1], Yihua Liu[1] and Yangxia Zheng[2]*

[1]College of Agriculture and Forestry Sciences, Linyi University, Linyi, China, [2]College of Horticulture, Sichuan Agricultural University, Chengdu, China, [3]College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang, China

As a stem variety of lettuce, the fleshy stem is the main product organ of stem lettuce. The molecular mechanism of fleshy stem expansion in stem lettuce is a complex biological process. In the study, the material accumulation, gene expression, and morphogenesis during fleshy stem expansion process were analyzed by the comparative analysis of metabolome, transcriptome and the anatomical studies. The anatomical studies showed that the occurrence and activity of vascular cambium mainly led to the development of fleshy stems; and the volume of pith cells gradually increased and arranged tightly during the expansion process. A total of 822 differential metabolites and 9,383 differentially expressed genes (DEGs) were identified by the metabolomics and transcriptomics analyses, respectively. These changes significantly enriched in sugar synthesis, glycolysis, and plant hormone anabolism. The expression profiles of genes in the sugar metabolic pathway gradually increased in fleshy stem expansion process. But the sucrose content was the highest in the early stage of fleshy stem expansion, other sugars such as fructose and glucose content increased during fleshy stem expansion process. Plant hormones, including IAA, GA, CTK, and JA, depicted important roles at different stem expansion stages. A total of 1,805 DEGs were identified as transcription factors, such as MYB, bHLH, and bZIP, indicating that these transcription factor families might regulate the fleshy stems expansion in lettuce. In addition, the expression patterns identified by qRT-PCR were consistent with the expression abundance identified by the transcriptome data. The important genes and metabolites identified in the lettuce fleshy stem expansion process will provide important information for the further molecular mechanism study of lettuce fleshy stem growth and development.

KEYWORDS

stem lettuce, fleshy stem expansion, anatomical analysis, metabolic regulation, gene expression

Lettuce, an annual or biennial herb belonging to the *Lettuce* genus of the Compositae family, is widely cultivated throughout China. As an important vegetable and medicinal plant, lettuce is rich in vitamin C, anthocyanins, and flavonoids (Malarz et al., 2020; Medina-Lozano et al., 2021). Lettuce can be classified as a leaf or stem lettuce according to its edible organs. The stem diameter of stem lettuce is an important commercial trait and directly affects its quality and economic value. Therefore, it is necessary to explore the molecular mechanism of stem expansion process.

The formation and development of plant stems is a complex physiological process which can be regulated by nutrients, hormones, genes, and environmental factors. The type and ratio of sugar and acid also affects the quality and commercial value of stem lettuce. Soluble sugar and organic acids are important nutrients and flavor substances of fruits and vegetables, which also participate in the process of plant metabolism (Das, 2015; Das et al., 2019; Ahmed et al., 2020). During the fruit ripening period of 'Orin' apples, the content of four sugars (glucose, sucrose, fructose, and sorbitol) and the malic acid increased (Yang et al., 2021). As the important signaling substances, plant hormones could regulate plant development, physiology, growth, and reproduction (Durbak et al., 2012; Wu et al., 2022). Numerous studies have demonstrated that hormones such as auxin (IAA), gibberellin (GA), abscisic acid (ABA), cytokinin (CTK), and methyl jasmonate (MeJA) participated in the process of rhizome development (Saidi and Hajibarat, 2021; Chen et al., 2022). For example, CTK promoted the formation and development of potato tubers and significantly increased the tuberization rate (Pavlista, 2011). JA and MeJA promoted the expansion of apical cells and tubers by inhibiting stolon elongation (Sohn et al., 2011). In the process of carrot root development, GA-related genes were differentially expressed, which indicated that GA might play a crucial role in carrot elongation and expansion (Wang et al., 2015).

A variety of metabolites and genes have been confirmed to participate in the expansion process of plant organs (root, stem). During mustard stem development process, most differentially expressed genes (DEGs) were identified to involve in the synthesis, accumulation, and metabolism process of sugars, starch, and storage proteins (Li et al., 2020). A large number of genes have been identified to involve in the process of rhizome development. For instance, radish *CycD3* and *RsCLE22a* gene were confirmed to participate in the formation of fleshy roots (Lutova et al., 2008; Dong et al., 2022). Two *MADS-box* genes could regulate vascular cambium activity and secondary growth of poplar by regulating IAA homeostasis (Zheng et al., 2020). The initiation and activity of *Arabidopsis thaliana* cambium were controlled by the transcriptional regulator AHL15 (Rahimi et al., 2022). Tomato *SD1* gene played positively roles in tomato stem development by regulating the size and number of secondary phloem cells (Ye et al., 2020). The transcription factor (TF) gene *SlHB8* negatively regulated tomato stem development by inhibiting xylem width and xylem cell layers (Liu et al., 2021).

Previous studies on stem lettuce have mainly focused on the effect of light on the lettuce quality (Li et al., 2021; Qi et al., 2021). The molecular mechanism of fleshy stem expansion in stem lettuce is unclear. In the study, the anatomical structure changes during fleshy stem expansion of stem lettuce were analyzed. The combined analysis of metabolomo and transcriptome were also conducted to determine changes in the metabolites and genes at four fleshy stem expansion stages (S1: transverse diameter length of fleshy stem is 1 cm, S2: transverse diameter length of fleshy stem is 2 cm, S3: transverse diameter length of fleshy stem is 3 cm, S4: transverse diameter length of fleshy stem is 4 cm). The results will provide a theoretical basis for elucidating the expansion mechanism of fleshy stems in stem lettuce.

Materials and methods

Plant materials and treatment

In this study, the cultivated stem lettuce 'Yonganhong' was selected as the experimental material. The seeds were placed in the incubator for 12 h photoperiod at 22℃ and 18℃ (day vs. night) with a light intensity of 20,000 μmol/m^2/s (lux) to raise seedlings. When the seedlings grew to 2-3 true leaves, the healthy seedlings with the same growth were selected and transplanted into the nutrition bowl (23 cm × 18 cm). Then, the seedlings were transferred into the greenhouse to grow normally under natural light. The fleshy stem of stem lettuce began to expand as time goes on. To ensure the accuracy of experimental data, digital calipers were chosen to measure the fleshy stem diameter of the third internode. When the diameter of the swollen stem at this internode reached to 1cm (S1 stage), 2cm (S2 stage), 3cm (S3 stage), and 4cm (S4 stage), the samples were collected and frozen in liquid nitrogen for further analysis.

Anatomical structure of fleshy stem in stem lettuce

The fleshy stems of the stem lettuce 'Yonganhong' at four expansion stages (S1, S2, S3, and S4) were fixed in 50 mL 50% FAA fixative and placed at 4°C for 24 h. Then, the treated samples were embedded in paraffin and stained with safranin O-fast green reagent referring to previous methods (Li et al., 2020). The anatomical structure during the process of fleshy stem expansion was observed and recorded with CaseViewer (version 7.3).

Determination of plant hormones

Plant hormones, including IAA, ABA, JA, CTK, and GA, were extracted at four expansion stages (S1, S2, S3, and S4) of fleshy stem in stem lettuce. Each expansion stage contained three biological replicates. The endogenous hormones were quantified by Nanjing Ruiyuan Biotechnology Co., Ltd according to the previously described method (Liu et al., 2010).

Metabolite profiling and data analysis

The extraction of metabolites and metabolomics analysis was conducted by Biomarker Technologies Co., Ltd (Beijing, China). Six biological replicates at each expansion stages of fleshy stem (S1, S2, S3, and S4) were collected for the metabolomics analysis. The samples after liquid nitrogen grinding were added 500 μL extraction solution (methanol: acetonitrile = 1:1), followed by ultrasonic for 10 min and incubation at -20℃ for 1 h, then centrifuging to obtain the supernatants. The supernatants were dried in a vacuum concentrator, then added 160 μL extraction solution (acetonitrile: water = 1:1) for further analysis. Samples of equal volume were taken from each experimental sample and mixed to obtain quality control samples (QC), inserted before, during and after the samples to test the repeatability of the experiment. The liquid chromatography-mass spectrometry system in this study was composed of ultra-high performance liquid chromatography (Waters Acquity I-Class PLUS) and high resolution mass spectrometer (Waters Xevo G2-XS QTOF). The chromatographic and mass spectrometric conditions were as follows: Acquity UPLC HSS T3 column (1.8 μm, 2.1×100 mm, Waters) was used, the injection volume was 1 μl, and the flow rate was 400 μL/min. Mobile phase A was 0.1% formic acid aqueous solution, mobile phase B was 0.1% formic acid acetonitrile. The chromatographic gradient elution procedure consisted of 0~0.25 min, 98% A, 2% B; 0.25~10.00 min, 2% A, 98% B; 10.00~13.00 min, 2% A, 98% B; 13.00~13.10 min, 98% A, 2% B; 13.10~15.00 min, 98% A, 2% B. The mass spectrometer collected primary and secondary mass spectrum data in MSe mode controlled by acquisition software (MassLynx version 4.2, Waters). The data was analyzed by principal component analysis (PCA), partial least squares discrimination analysis (PLS-DA), and other multivariate statistical analysis. Then, the variable importance in the projection (VIP), fold change (FC), and p-value were used to screen metabolites using the following screening criteria: FC > 1.0, $p < 0.05$, and VIP > 1.0. HMDB (https://hmdb.ca/metabolites) and LIPIDMAPS database were used to conduct the annotation of the functions and classifications of the metabolites. The bioinformatics analysis of the metabolites was conducted using BMKCloud (www.biocloud.net).

Transcriptome data analysis

A total of 12 samples of four fleshy stem expansion stages (S1, S2, S3, and S4) with three biological replicates were subjected to RNA sequencing. The Agilent 2100 bioanalyzer was used to accurately detect the integrity and total amount of RNA. The cDNA libraries (S1, S2, S3, and S4) were sequenced on the Illumina HiSeq 4000 sequencing platform (Biomarker Technologies Co., Ltd., Beijing, China), as previously described (Zhang et al., 2017). FPKM (fragments per kilobase of transcript per million fragments mapped) was used as an index to measure the level of transcripts or gene expression. DESeq2 (version 1.20.0) was used to analyze the differential expression between sample groups. FC represented the ratio of the expression between two samples (stages). The false discovery rate (FDR) was obtained by correcting the p-value of the significance of the difference, and the screening criteria were FC ≥ 2 and FDR < 0.01. The bioinformatics analysis of the DEGs was conducted using BMKCloud (www.biocloud.net).

qRT-PCR analysis

The total RNA was extracted from the four fleshy stem expansion stages (S1, S2, S3, and S4) and subjected to qRT-PCR analysis using the Roche LightCycler 96 and the SYBR qPCR master mix. The amplification procedure was as follows: pre-denaturation at 95°C for 30 s, then 40 cycles of denaturation at 95°C for 10 s, and annealing at 60°C for 30 s. Gene expression was normalized using *LsTIP41* as an internal reference and quantified using $2^{-\Delta\Delta Ct}$ as described previously (Pfaffl, 2001; Borowski et al., 2014). The primers used for qRT-PCR were listed in Table S1.

Results

Anatomical analysis of lettuce fleshy stem

The anatomical structure analysis of lettuce fleshy stem displayed that the development of the fleshy stem was mainly caused by the occurrence and activity of vascular cambium (Figure 1). The lettuce fleshy stem comprised a cortex, vascular bundle, and pith from outside to inside; the vascular bundle contained phloem, cambium, and xylem (Figures 1E–H). Vascular bundles arranged in a ring on the inside of the cortex, and the pith rays were located between the vascular bundles, which connected the cortex and the medulla (Figure 1H). As displayed in Figures 1I, J, the cell volume increased significantly

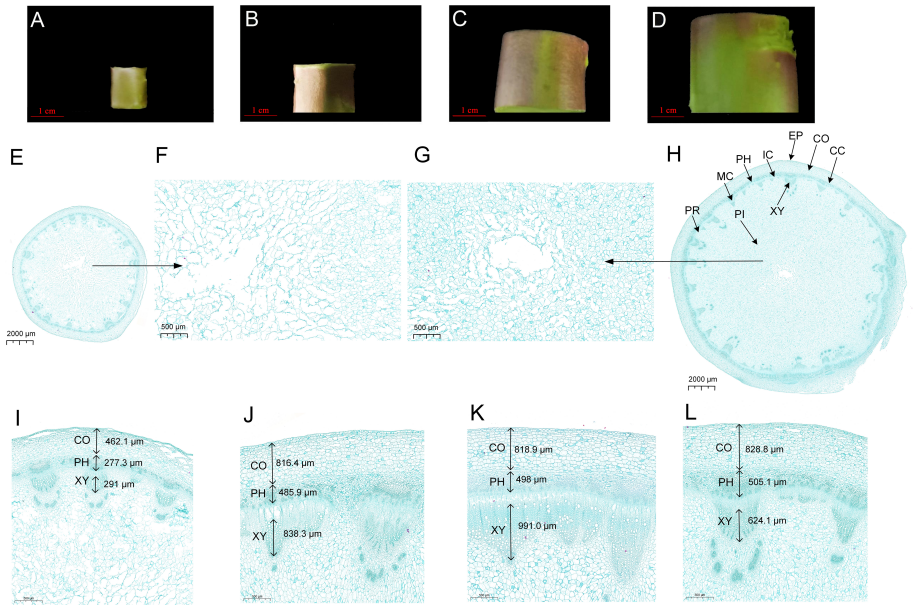


FIGURE 1
Anatomical analysis of lettuce fleshy stem at different expansion stages (S1, S2, S3, and S4). **(A–D)** The transverse diameters are 1 cm (S1 stage), 2 cm (S2 stage), 3 cm (S3 stage), and 4 cm (S4 stage), respectively, **(E)** Global cross-sections of S1 stage, **(F)** Local maps of medullary cells at S1 stage, **(G)** Local maps of medullary cells at S4 stage, **(H)** Global cross-sections at S4 stage, **(I–L)** Local maps of the epidermis to vascular bundles from S1 to S4 stages. EP, Epidermis; CO, Cortex; CC, Cork Cambium; PH, Phloem; PR, Pith Ray; IC, Interfascicular Cambium; MC, Midbundle Cambium; XY, Xylem; PI, Pith.

from S1 stage to S2 stage; and the cambium parenchyma cells divided tangentially, inducing a rapid increase in the diameter of the fleshy stem. Compared to S1 stage, the cortex, phloem, and xylem at S4 stage increased about two, two, and three times (Figure 1L). The transverse diameter of the xylem was larger than that of the phloem and increased more rapidly (Figures 1I–L). The pith cells at S1 stage were loosely arranged and numerous gaps existed between the pith cells. During the process of stem expansion, the number of pith cells increased rapidly and the pith cells were arranged very closely, which might be one of the main reasons to cause a thickening of the fleshy stem (Figures 1F, G).

Metabolomics analysis of lettuce fleshy stem

To investigate the metabolite changes during the fleshy stem expansion process, metabolomics analysis of S1, S2, S3, and S4 stages was conducted. PCA analysis showed that the first two principal components (PC1 and PC2) separated 24 samples, accounting for 49.02% and 17.43% of the total variability, respectively (Figure 2A). In addition, the samples on the PCA analysis map were clustered into four groups. The metabolite of S1 stage was mainly at the negative end of PC1; while, the metabolite of S2, S3, and S4 stages were mainly at the positive end of PC1. The metabolites of S2, S3, and S4 stages were similar, even with a large part of overlap. The results indicated that different metabolites existed in the expansion process of lettuce fleshy stem. A total of 1,251 metabolites were detected and clustered into two categories (cluster 1 and cluster 2) during the four expansion stages (S1, S2, S3, and S4 stages). The metabolite in cluster 1 highly expressed in S1 stage and the metabolite in cluster 2 mainly expressed in S2, S3, and S4 stages (Figure 2B).

The identified metabolites were annotated using the HMDB and LIPIDMAPS databases. As shown in Figure 2C, 457 metabolites were classified into twelve categories by the HMDB database. The metabolite belonged to lipids and lipid-like molecules were the most abundant (128 metabolites), followed by organic acids and derivatives (122 metabolites). In the LIPIDMAPS annotation analysis, the most metabolites belonged to Docosanoids (83 metabolites), followed by Hydrocarbons (82 metabolites) and Fatty amides (80 metabolites) (Figure 2D). The metabolites identified in S1, S2, S3, and S4 stages were compared between groups, and they were identified by VIP, FC, and *p*-value (FC > 1.0, p value < 0.05, and VIP > 1.0). The pairwise comparisons of the identified metabolites among the four stages are depicted in Figure 2E. The three comparison pairs of S1 vs. S2, S1 vs. S3, and S1 vs. S4

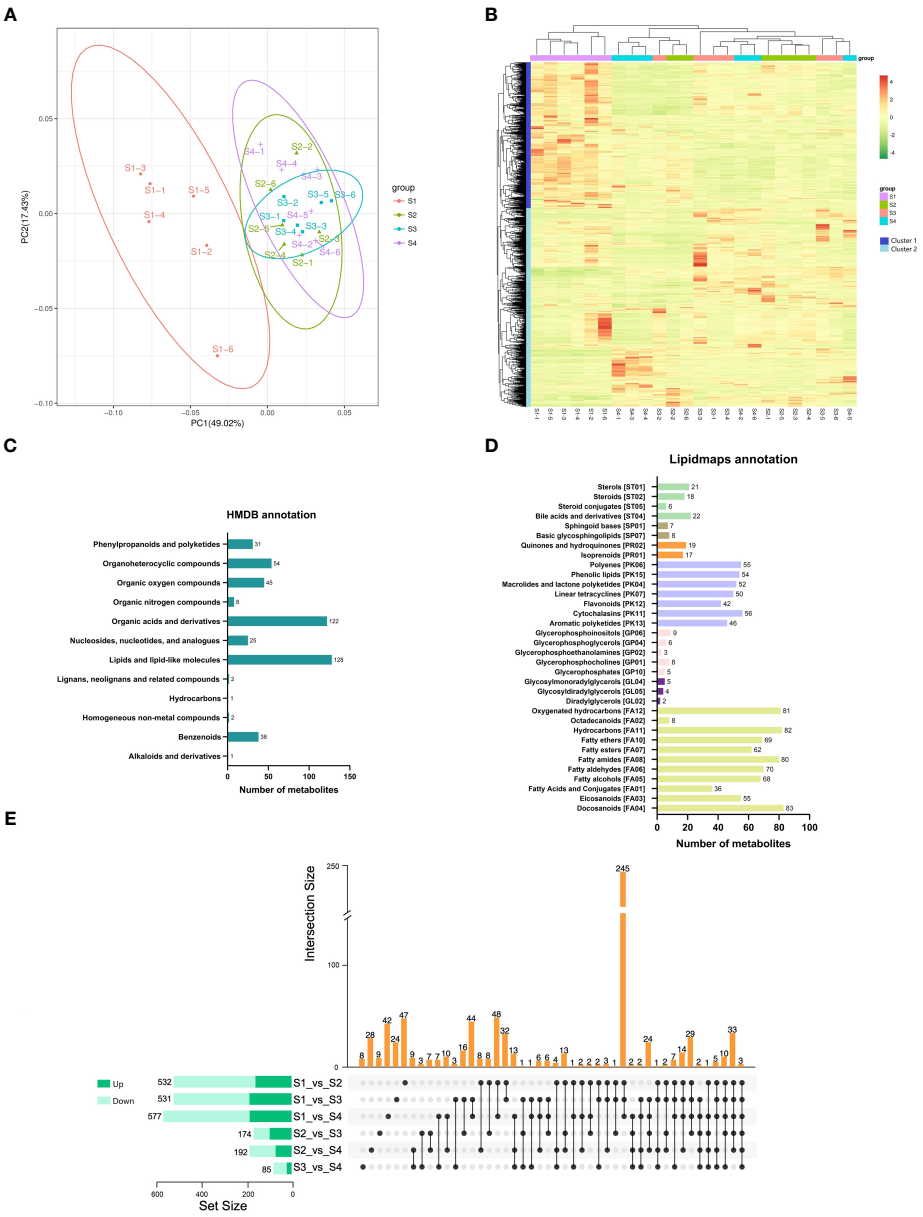


FIGURE 2
Metabolic analysis of lettuce fleshy stem during four expansion stages (S1, S2, S3, and S4). **(A)** Principal component analysis (PCA) of the identified metabolites in four expansion stages (S1, S2, S3, and S4). **(B)** Heatmaps of all metabolites during the four expansion stages (S1, S2, S3, and S4). **(C)** Taxonomic annotations of metabolites by HMDB database. **(D)** Taxonomic annotations of metabolites by LIPIDMAPS database. **(E)** Comparisons of up-regulated metabolites (dark green) and down-regulated metabolites (light green) in the paired comparison of each expansion period.

contained the most differentially expressed metabolites (DEMs) (532, 531, and 577 DEMs, respectively). The comparison pairs of S3 vs. S4 had the least metabolites (85 DEMs), indicating the similarity between these two stages. Only three DEMs were identified at all the four expansion stages (S1, S2, S3, and S4). A total of 245 DEMs were identified to exist in the three combinations of S1 vs. S2, S1 vs. S3, and S1 vs. S4, indicating that the development of S2, S3, and S4 stages was similar.

Transcriptome analysis of lettuce fleshy stem

To investigate the DEGs during different expansion stage, transcriptome analysis of S1, S2, S3, and S4 stages with three biological replicates was conducted. The cDNA libraries were constructed and sequenced using the Illumina HiSeq 4000 platform. A total of 596,863,836 raw reads were generated

from 12 samples, and 298,431,918 high-quality clean reads were identified after filtration. The average percentage of Q20 and Q30 bases was 97.74% and 93.46%, respectively. The percentage of GC was more than 43.65%. The clean reads were mapped to the lettuce genome, with an average efficiency of 94.33% (Table S2). A total of 42,916 transcripts were detected and the expression level of each gene was normalized by FPKM. The identified transcripts among four expansion stages (S1, S2, S3, and S4) were analyzed by PCA analysis. As shown in Figure 3A, the gene expression of S1 and S2 stages were located at the negative and positive ends of PC1, respectively. The gene expression of S3 and S4 stage were both located at the negative end of PC2, indicating the similarity between S3 and S4 stage. The correlation of gene expression levels was higher in the three biological replicates of the same stage. The gene expression level at S3 stage showed high correlation those at S2 and S4 stages. While, The gene expression level at S1 stage showed the lowest correlations with S2, S3, and S4 stages (Figure 3B).

The DEGs in the six comparison pairs (S1 vs. S2, S1 vs. S3, S1 vs. S4, S2 vs. S3, S2 vs. S4, S3 vs. S4) were analyzed (Figure 3C). Compared with the S1 stage, there were 5,028 (2,727 up-regulated DEGs and 2,301 down-regulated DEGs), 5,068 (2,478 up-regulated DEGs and 2,590 down-regulated DEGs), and 6,496 DEGs (3,446 up-regulated DEGs and 3,050 down-regulated DEGs) existed in S2, S3, and S4 stages, respectively. The results indicated that the development of fleshy stem at S2, S3, and S4 stages was significantly different with S1 stage. Only seven genes were co-expressed among the six comparison pairs (S1 vs. S2, S1 vs. S3, S1 vs. S4, S2 vs. S3, S2 vs. S4, S3 vs. S4). There were 2,359 co-expressed genes in the comparison pairs of S1 vs. S2, S1 vs.S3, and S1 vs. S4. Among the individual comparison pairs, a total of 935 DEGs were identified in the comparison pair of S1 vs. S4, indicating the largest difference existed in S1 and S4 stage. In the comparison of S3 vs. S4, 60 genes were identified, indicating the smallest difference existed in S3 and S4 stage. All the results showed that the development of fleshy stem in S2, S3, and S4 stages was significantly different from S1 stage.

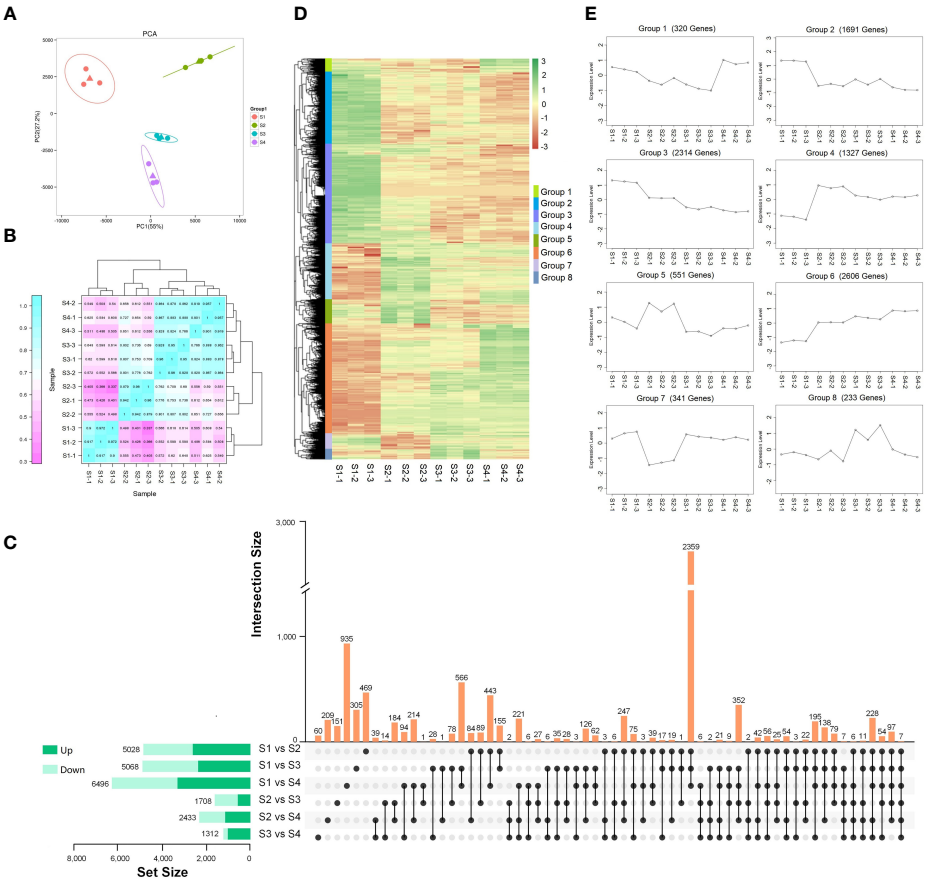


FIGURE 3
Transcriptome analysis of lettuce fleshy stem during four expansion stages (S1, S2, S3, and S4). **(A)** Principal component analysis (PCA) of gene expression at four expansion stages (S1, S2, S3, and S4). **(B)** Pearson correlation coefficient of gene expression in four expansion stages (S1, S2, S3, and S4). **(C)** Comparisons of up-regulated DEGs (dark green) and down-regulated DEGs (light green) in the paired comparison of each expansion period. **(D)** Hierarchical clustering heatmap of the identified DEGs. **(E)** K-means clustering of the DEGs expression.

A total of 9,383 DEGs were identified after merging the differential genes of all combinations, and a hierarchical clustering heatmap was drawn using the standardized FPKM Z-score value (Figures 3D, E). These DEGs were divided into eight groups (1-8). The 320 genes in group 1 showed highly expressed at S4 stage; 1,691 genes in group 2 were highly expressed at S1 stage. The 2,314 genes existed in group 3 showed gradually decreased expression during fleshy stem expansion process. The expression of genes in group 4 increased rapidly at S2 stage, then decreased slightly, but the overall expression was higher than those at S1 stage. 551 genes in group 5 showed higher expression at S2 stage. Group 6 contained the most number of genes (2,606), and the expression levels of these genes increased gradually with the expansion of the stem. The expression of 341 genes in group 7 was the lowest at S2 stage; and the expression of 233 genes in group 8 was the highest at S3 stage. In addition, most genes showed significant changes at S2 stage in these eight groups, indicating that S2 stage was very important for fleshy stem expansion.

Transcription factor (TF) family analysis

A total of 1,805 DEGs were identified as TFs by the further analysis, and these genes belonged to 46 TF families, including MYB, bHLH, AP2, and C2H2 (Figure 4A, Table S3). The expression levels of identified TF family gene were showed by heatmap (Figure 4B). The top 3 TF families with the highest number of genes were MYB (221), AP2 (214), and bHLH (140), with high overall expression in the four stages. The expression of some genes belonged to WRKY, MYB, and AP2/ERF TFs was higher at S1 stage and then decreased at S2 and S3 stages, indicating that these genes might play a regulatory role in the early stage of lettuce fleshy stem expansion. The expression of some TF genes such as bZIP and HSF was low at S1 stage but increased at S2 and S3 stages. PHD and NF-YC TF genes were not significantly expressed at S1 and S2 stages but up-regulated at S3 stage. These results suggested that the TFs played important regulatory roles during different expansion stages of lettuce fleshy stem.

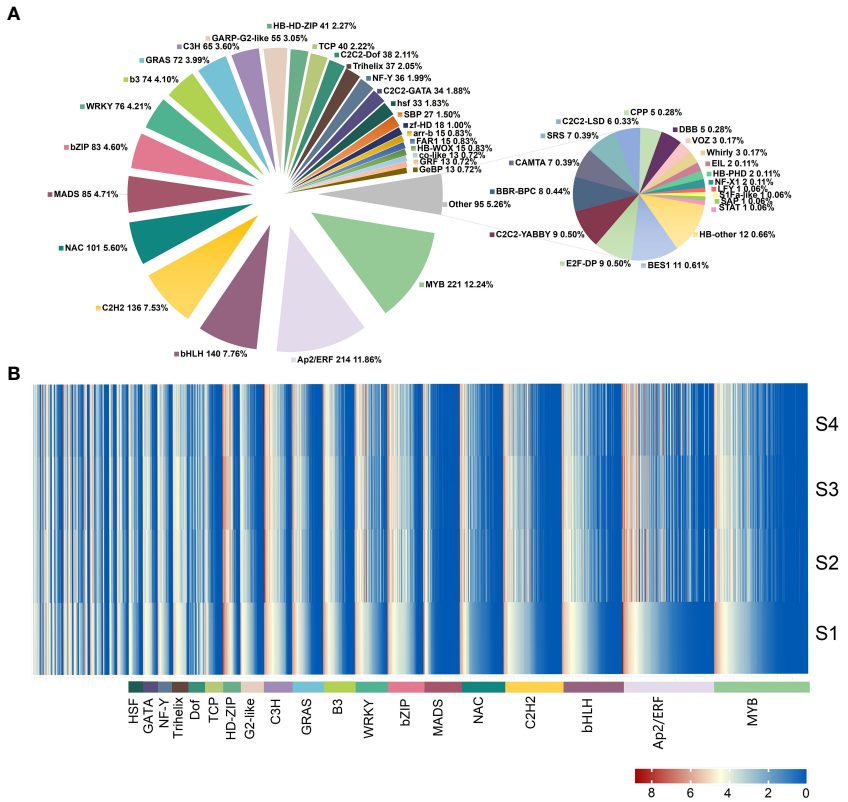


FIGURE 4
Transcription factor family analysis. **(A)** Summary of transcription factor families, the pie chart showed the number and proportion of each transcription factor family; **(B)** Expression pattern analysis of transcription factor family genes using heatmap. The data were calculated using log2 (FPKM+1).

Enrichment analyses of the transcriptome and metabolome

Significant enriched metabolic pathways related to DEGs and DEMs were identified by KEGG enrichment analysis (Figure 5A and Table S4). A total of 33 metabolic pathways were identified from the top 10 significantly enriched metabolic pathways for all six comparisons. The highly expressed metabolic pathways in the transcriptome and metabolome were 'Starch and sucrose metabolism', 'phenylpropanoid biosynthesis', and 'Plant hormone signal transduction'. The most significantly expressed pathways in transcriptome were 'Phenylpropanoid biosynthesis', 'Starch and sucrose metabolism', 'Zeatin biosynthesis', and 'Flavonoid biosynthesis'. The most significantly expressed pathways in the metabolome were 'Glycerophospholipid metabolism', 'Carbon metabolism', 'Pantothenate and CoA biosynthesis', and 'Glycine, serine, and threonine metabolism'. The pathways related to the growth and development of lettuce fleshy stem were mainly 'Starch and sucrose metabolism', 'Zeatin biosynthesis', and 'Plant hormone signal transduction' by the comprehensive analysis, which were related to the physiological processes of sugar and acid metabolism as well as hormone synthesis.

In order to further analyze the correlation between the various pathways, the association maps of the significantly enriched pathways in the six combinations were drawn (Figure 5B). A total of 119 different pathways were enriched. The pathways, which closely related to lettuce growth, were located in the middle of the figure. The glycolytic synthesis pathway had a central role in all pathways. The glycolytic synthesis pathway directly related to multiple significant enrichment pathways. First, glycolysis pathway directly connected starch and sucrose metabolism, citrate cycle (also known as TCA cycle), as well as various amino acid synthesis pathways. Glycolysis pathway also participated in the synthesis and metabolism of chlorophyll, endogenous hormones, alkaloids, and other substances, which provided a material basis for lettuce growth and development. Second, glycolysis pathway could regulate the synthesis of plant hormones in lettuce through hormone signal transduction pathways. The differential pathways of S1 vs. S2, S1 vs. S3, and S1 vs. S4 accounted for more than half of the total, indicating that the metabolic transcription differences in S2, S3, and S4 stages are important factors in lettuce stem expansion.

Analysis of the sugar acid synthesis metabolic pathway

The starch and sucrose metabolism pathways were identified through the transcriptomic and metabolomics pathway analysis. As shown in Figure 6 and Tables S5-S6, in the process of starch and sucrose metabolism, most enzyme genes (including 19 enzymes, 73 enzyme genes, and 24 metabolites) showed an upward trend during fleshy stem expansion stages. Compared with S1 stage, the expression of some genes, encoding sucrose

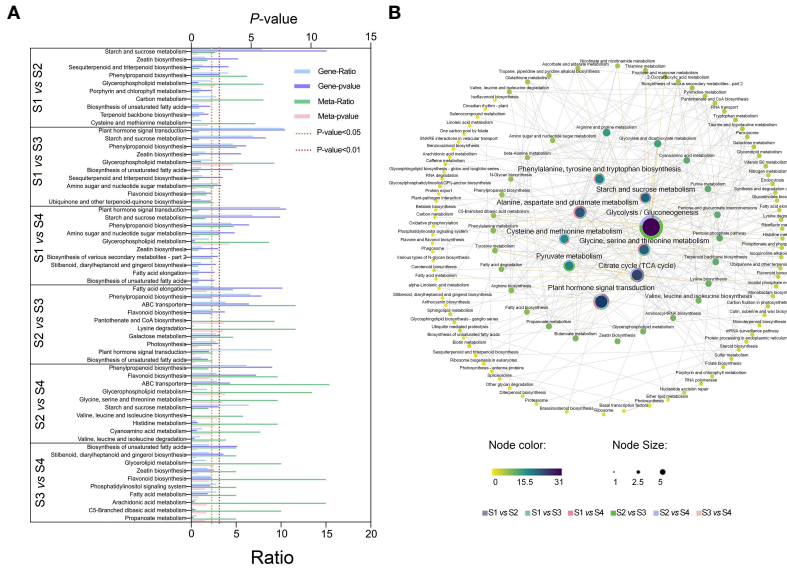


FIGURE 5
KEGG pathway annotated analysis in different lettuce stem expansion stages (S1, S2, S3, and S4 stages). **(A)** The top 10 pathways associated with transcriptome and metabolome. Ratio: the number of differential metabolites or differential genes enriched in this pathway/the number of metabolites or genes annotated in this pathway. **(B)** Correlation analysis of each pathway. Different dots represented different metabolic pathways; yellow to purple indicated the degree of enrichment in the whole expansion stage, and the central fan chart displayed the degree of enrichment of the pathway in each comparison group. The data were calculated by -log10 (*p*-value), and the higher the value, the greater the enrichment.

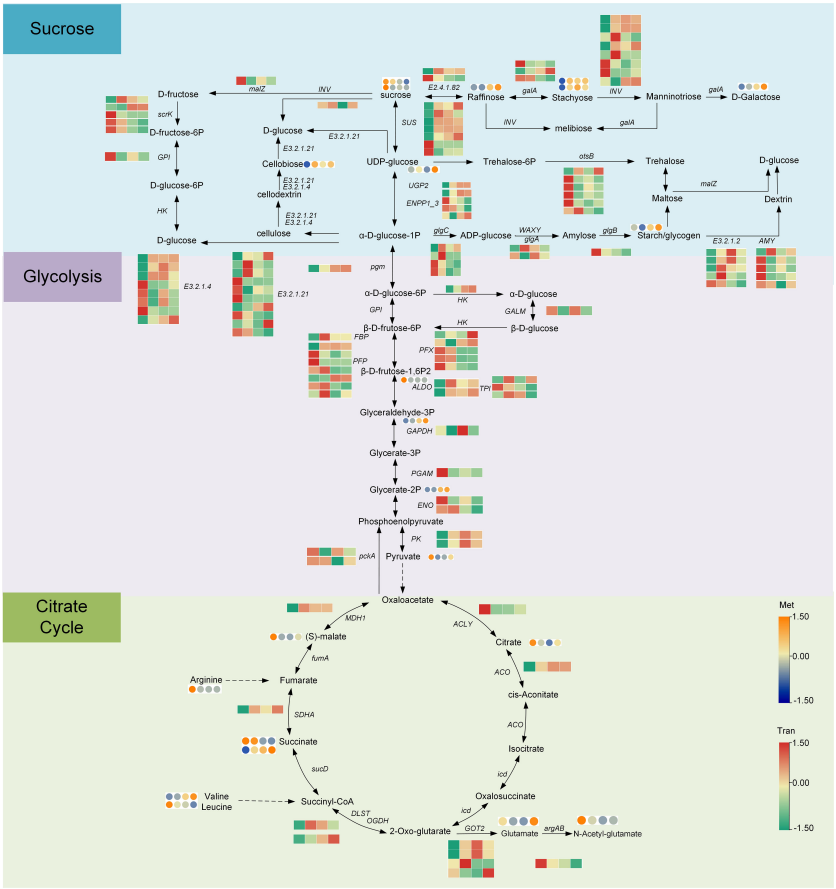


FIGURE 6
Analysis of genes and metabolites expressed in the sugar and acid biosynthesis pathways during lettuce fleshy stem expansion stages (S1, S2, S3, and S4). The red (up-regulated) and green (down-regulated) in the heatmap represent gene expression, and the yellow (up-regulated) and blue (down-regulated) in the heatmap represent the expression of metabolites. The expression level was calculated using log2(FPKM+1).

synthase (SUS), UTP–glucose-1-phosphate fructofuranosidase (UGP2), fructofuranosidase (INV), endoglucanase (E3.2.1.4), and raffinose synthase (E2.4.1.82), increased five-fold at S2, S3, and S4 stages. However, the expression patterns of some enzyme genes encoding fructokinase (scrK), gluc-1-phosphate adenylyltransferase (glgC), 1,4-alpha-glucan branching enzyme (glgB), and beta-amylase (E3.2.1.2) decreased in lettuce stem expansion process. The content of some metabolites, such as raffinose, stachyose, starch, and cellobiose, was low at S1 stage, but increased at S3 and S4 stages. The content of sucrose decreased in the fleshy stem expansion process. The content of sucrose decreased two-fold at S4 stage compared with S1 stage, indicating that sucrose began to accumulate at the beginning of the expansion for the later lettuce expansion.

The expression of most enzyme genes in the glycolysis pathway (including 9 enzymes, 30 enzyme genes, and 10 metabolites) increased gradually during expansion stages. Compared to S1 stage, the expression of genes, encoding some enzymes such as phosphoglucomutase (pgm), hexokinase (HK), class I (ALDO), and triosephosphate isomerase (TPI) increased ten-fold at S2, S3, and S4 stages. However, the expression of some genes, encoding 6-phosphofructokinase 1 (PFK), phosphoenolpyruvate carboxykinase ATP (pckA), and enolase (ENO), decreased at S2, S3, and S4 stages. The expression of most enzyme genes in the TCA cycle such as aconitate hydratase (ACO), flavoprotein subunit (SDHA), and malate dehydrogenase (MDH1) was the lowest at S1 stage but increased two-fold at S2, S3, and S4 stages. Overall, most metabolites and enzyme genes involved in sugar synthesis, glycolysis, and the citrate cycle were up-regulated during the process of fleshy stem expansion.

Analysis of the hormone synthesis metabolic pathway

Plant hormones play important roles in the process of plant organ expansion. The hormone pathways, including CTK, IAA, GA, JA, and ABA, were analyzed to explore its roles in lettuce

fleshy stem expansion stages (Figure 7, Tables S5-S6). In CTK, the overall expression level of genes encoding enzymes was high at S1 stage and decreased at S2, S3, and S4 stages. The expression of some genes encoding isopentenyl-diphosphate Delta-isomerase (IDI), adenylate dimethylallyltransferase (IPT), and cytokinin trans-hydroxylase (CYP735A) decreased two-fold at S2, S3, and S4 stages. Cytokinin content also showed a two-fold decrease at S2, S3, and S4 stages than at S1 stage. The expression profiles of genes such as *CYP701* and phytochrome-interacting factor 4 involved in GA pathway, were higher at S1 stage and lower at S2, S3, and S4 stages. However, the expression of some genes encoding GA20ox, GA receptor GID1, and DELLA protein, was high at S1 stage and decreased at S2 stage but showed an upward trend at S3 and S4 stages. The GA content initially decreased and then increased, which showed similar results to the gene expression trend.

In ABA pathway, the expression of genes before ABA synthesis was low at S1 stage, and then gradually increased with lettuce expansion process. For example, compared to S1 stage, the expression levels of *LUT5* and *NCED* increased by 2-15 times at S4 stage. The expression of genes such as the ABA receptor PYR/PYL family gene in ABA metabolism pathway highly expressed at S1 stage and decreased at S2, S3, and S4 stages. The ABA content was highest at S1 stage and reduced two-fold at S2, S3, and S4 stages, which was similar to the trend of gene expression. In the process of JA synthesis, most genes were highly expressed at S1 or S4 stage, such as *OPR* and *ACOX1*. In the process of JA metabolism, the expression of genes, such as *JAR1*, *JAZ*, and *MYC2*, was higher at S2, S3, and S4 stages. The JA content identified by metabolome was higher at S1 and S2 stages than at S3 and S4 stages, which might related to the high expression of JA metabolism genes at S4 stage. The overall trend of JA content at different expansion stages was similar to in result of metabolome.

The expression patterns of most genes in the auxin synthesis pathway, were highly expressed at S1 stage, such as aldehyde dehydrogenase *NAD+*(*ALDH*) and *YUCCA*. The expression profiles of most genes in the metabolic pathway of auxin were similar to the expression patterns of synthetic genes; and the highest expression levels of genes such as *AUX1*, *TIR1*, and

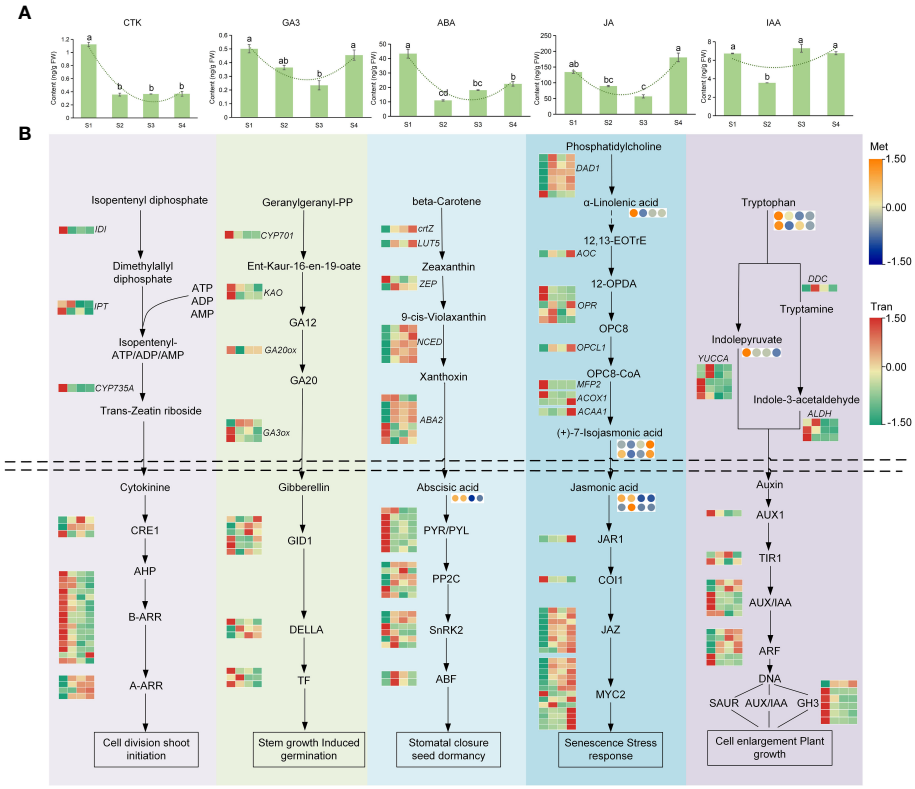


FIGURE 7
The anabolism of plant hormone during lettuce fleshy stem expansion stages (S1, S2, S3, and S4). **(A)** The contents of CTK, GA, ABA, JA, and IAA in the four expansion stages (S1, S2, S3, and S4). Lowercase letters (A–D) in each column denote significant differences. **(B)** Analysis of the synthetic metabolic pathway of CTK, GA, ABA, JA, and IAA in the four stages (S1, S2, S3, and S4). Above the dotted line is the synthetic pathway, and below the dotted line is the metabolic pathway. The red (up-regulated) and green (down-regulated) in the heatmap represented gene expression, and the yellow (up-regulated) and blue (down-regulated) in the heatmap represented the metabolite accumulation.

auxin-responsive protein IAA, were identified at S1 stage, which indicated that auxin had a fast metabolic transformation process in the early growth stage. In summary, during lettuce stem expansion process, the overall expression of CTK gradually decreased, whereas GA, ABA, IAA, and JA decreased first and then increased.

Correlation network analysis of anabolism pathways and gene verification

The sugar-acid synthesis correlation network analysis showed the correlation between sucrose and other metabolites was the complete opposite (Figure 8A). Sucrose showed a significant positive correlation with most genes in the pathway, while other metabolites depicted a significant negative correlation with these genes. To verify the validity of the transcripts, twelve genes in the sugar synthesis pathway were selected for qRT-PCR analysis (Figure 8B). The expression of *Lsat_1_v5_gn_3_67140*, *Lsat_1_v5_gn_2_125121*, *Lat_1_v5_gn_1_36821*, and *Lactuca_ sativa_newGene_5834* involved in the sugar synthesis process, showed an overall upward trend in both the four stages (S1, S2, S3, and S4). In the glycolysis pathway, *Lsat_1_v5_gn_4_13360*, *Lsat_1_v5_gn_8_88260*, and *Lsat_1_v5_gn_8_80661* showed an overall upward trend; and the overall expression of these genes at S2, S3, S4 stages increased about two-four times than at S1 stage.

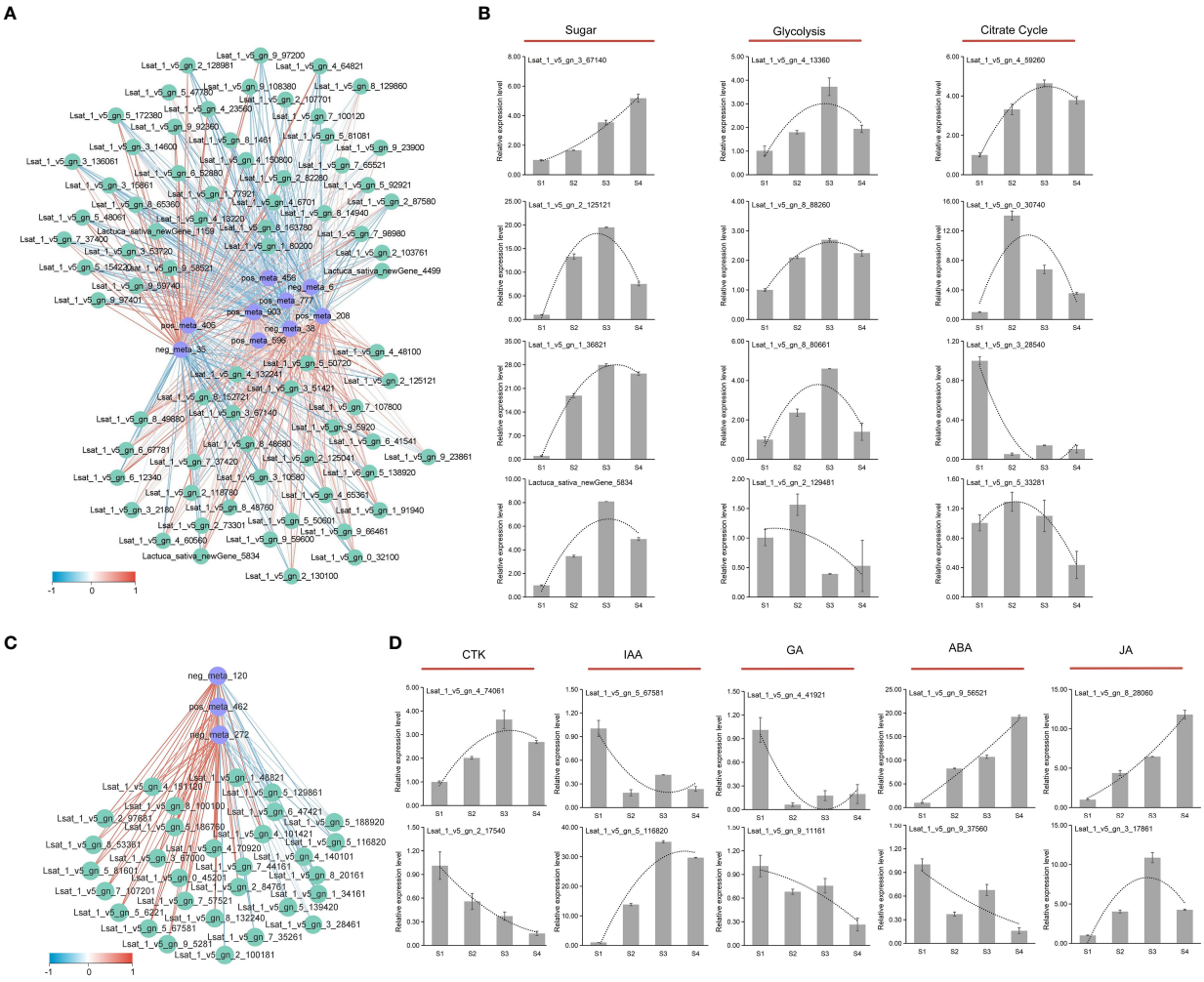


FIGURE 8
Correlation network analysis and gene verification of sugar, acid, and hormone metabolism. **(A)** Correlation network analysis diagram of DEMs (blue) and DEGs (green) related to glucose synthesis metabolism. The correlation was represented by blue and red lines from low to high. **(B)** qRT-PCR analysis of genes involved in sugar and acid biosynthesis. **(C)** The correlation network analysis diagram of DEMs (blue) and DEGs (green) related to auxin synthesis metabolism. The correlation was represented by blue and red lines from low to high. **(D)** qRT-PCR analysis of genes involved in hormone synthesis.

In the late stage of glycolysis, the expression level of *Lsat_1_v5_gn_2_129481* showed a downward trend, and the expression level at S4 stage decreased about twice compared with S1 stage. In Citrate Cycle, the expression levels of *Lsat_1_v5_gn_4_59260* and *Lsat_1_v5_gn_0_30740* both increased at S2, S3, and S4 stages compared with S1 stage, but *Lsat_ 1_v5_gn_3_28540* and *Lsat_1_v5_gn_5_33281* showed a downward trend. The qRT-PCR results of genes in the sugar and acid synthesis metabolic pathways were generally consistent with the trend in transcriptome data, indicating that the reliability of transcript data was high.

The strong positive correlation between metabolites and genes involved in IAA biosynthetic pathway was identified by the correlation network analysis (Figure 8C). Tryptophan and pyruvate had a high positive correlation with most genes, reaching 0.9. To verify the validity of the transcript, genes involved in hormone pathway (CTK, IAA, GA, JA, and ABA) were selected for qRT-PCR analysis (Figure 8D). In the CTK synthesis pathway, the expression of *Lsat_1_v5_gn_2_17540* depicted a downward trend during the fleshy stem expansion process. In the CTK metabolic pathway, the expression of *Lsat_1_v5_gn_4_74061* showed an upward trend during the fleshy stem expansion process. In the GA synthesis metabolic pathway, the expression levels of *Lsat_1_v5_gn_4_41921* and *Lsat_1_v5_gn_9_11161* decreased at S2, S3, and S4 stages than at S1 stage. In the ABA synthesis pathway, *Lsat_1_v5_gn_9_56521* depicted an upward trend expression at S2, S3, and S4 stages compared to S1 stage. In the ABA metabolic pathway, the expression of *Lsat_1_v5_gn_9_37560* showed a downward trend. In the JA synthesis pathway, the expression levels of *Lsat_1_v5_gn_8_28060* and *Lsat_1_v5_gn_3_17861* increased at S2, S3, and S4 stages than at S1 stage. Compared with the S1 stage, the expression level of *Lsat_1_v5_gn_5_116820*, which belonged to auxin metabolic pathway, increased about ten times or more at S2, S3, and S4 stages. The expression patterns of genes involved in hormone synthesis metabolic pathway identified by qRT-PCR were generally consistent with the trend in transcriptome data, indicating that the reliability of transcript data was high.

Discussion

'Omics' technologies, such as metabolomics, transcriptomics, and proteomics have been widely applied to identify a series of molecular mechanism involved in the process of plant' growth and development or the interaction between plants and external factor (Wang et al., 2021a; Li et al., 2022; Lin et al., 2022). As one of the most popular vegetables, domesticated lettuce contained some horticultural types such as butterhead, loose leaf, and stem lettuce. The molecular mechanism involved in leaf development in lettuce has been explored. Lettuce *LsKN1* has been identified to generate wavy leaves by regulating plant hormone signaling pathways (Jia et al., 2022; Wang et al., 2022). The genes regulating anthocyanin in lettuce has also been analyzed (Su et al., 2022). The domestication history and domestication-shaped genetic architecture of cultivated lettuce have been elucidated, which provides valuable genomic resources for cultivated lettuce breeding (Zhang et al., 2020; Wei et al., 2021). However, the metabolome and transcriptome combined analysis exposed to fleshy stem expansion is not common.

The growth and expansion of lettuce fleshy stem is a complex dynamic process coordinated by many regulatory elements, including the activity of cells, plant hormones, metabolites, and gene regulation. In the study, the changes in morphology, physiology, metabolites, and gene expression during lettuce fleshy stem expansion stages (S1, S2, S3, and S4) were identified. The occurrence and activity of the vascular cambium led to the development of the fleshy stem. Metabolome and transcriptome analyses revealed that the pathways of sugar synthesis, glycolysis, TCA cycle, and plant hormone anabolism played important roles during fleshy stem expansion. Plant hormone such as IAA, GA, and JA showed different promoting effects, while ABA exerted inhibitory effects. Sugar biosynthesis and metabolism gradually increased during lettuce stem expansion. TFs also played an important role in the fleshy stem expansion process. For example, MYB, bHLH, and bZIP changed significantly with lettuce expansion (Figure 9).

The distribution and allocation of photosynthate in source-sink units is of great significance to the formation of the vegetative organs (Tang et al., 2021). In cucumber, *CsAGA2* decomposed RFOs into hexose, and transported it to pulp cells through the apoplast, promoting the accumulation of sugar in fruit (Liu et al., 2022). The increase in the size and number of parenchyma cells in the cortex and medulla was the main reason for stem development (Kolomiets et al., 2001; Zhong et al., 2007). Hormone signal transduction, sugar starch metabolism, and phenylpropanoid synthesis metabolism were enriched in the main root enlargement of *Panax notoginseng* (Li et al., 2019). During the process of fleshy stem expansion, the photosynthate was rapidly distributed to the fleshy stem; cells divided inward to rapidly increase the internal diameter of the fleshy stem. At the same time, the pith cells were arranged tightly. Sugar synthesis, glycolysis, TCA cycle, and plant hormone anabolism were identified to play important roles in the fleshy stem expansion process in stem lettuce.

As an energy substance for plant growth and development, sugar involve in plant growth and development as a signaling substance (Peng et al., 2018). Sucrose played a key role in integrating developmental stages and environmental cues to regulate plant yield (Fernie et al., 2020; Liao et al., 2020; Wu et al., 2021; Ren et al., 2022). In this study, the sucrose content was highest in the early stage of stem expansion in lettuce, indicating that it stored a large amount of energy in the early stage for the later stem expansion. In the later stage of lettuce

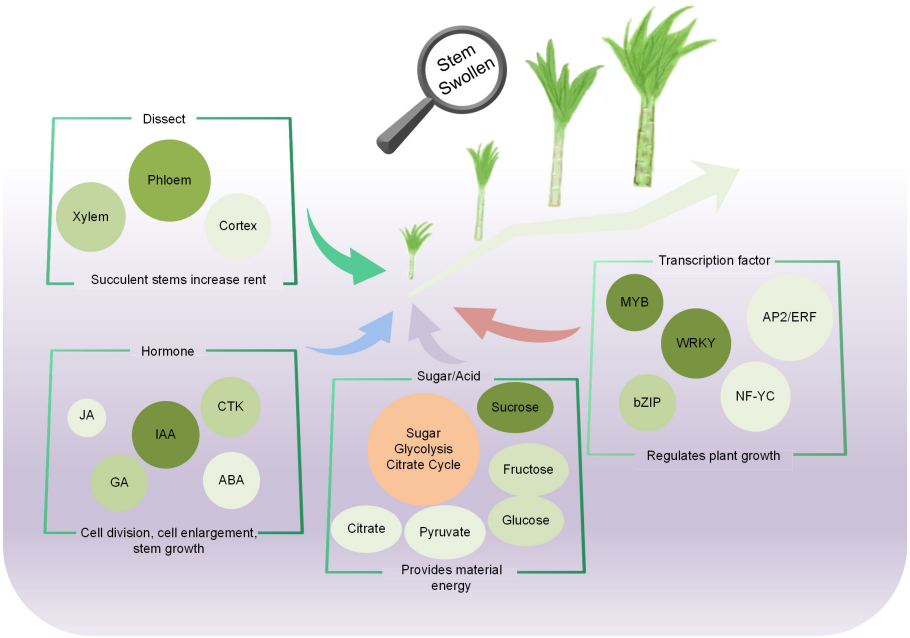


FIGURE 9
Regulation pattern of the fleshy stem expansion in stem lettuce.

expansion, sucrose was converted into glucose and fructose, which might also explain the increased glucose and fructose content and the decrease of sucrose in the later stage. Hormones have been identified to be essential endogenous substances for plant growth and development (Liang et al., 2014; Miyashima et al., 2019). GA was considered as a potentially important regulator of cell elongation and expansion in plants (Ayano et al., 2014). Some GA-related genes were differentially expressed in different stages of carrot root development (Xu et al., 2014; Wang et al., 2015). Auxin could stimulate GA biosynthesis, and exogenous use of 2,4-D could stimulate the expression of *GA20ox* and *GA3ox* (Cong et al., 2019). CTKs increased cell division in Arabidopsis, tomato, and tobacco and were also associated with cell proliferation in the early stages of tuber growth (Katsarou et al., 2016). ABA and JA could involve in many processes regulating plant growth and development. The β-glucosidase gene (*ClBG1*) was a key gene in regulating ABA content, reducing the seed size of watermelons (Wang et al., 2021b). At different stages of tulip bulb development, the lipoxygenase genes *TgLOX4* and *TgLOX5* involved in the biosynthesis of JA; and the silencing of *TgLOX4*, and *TgLOX5* genes inhibited the growth of tulip bulbs (Sun et al., 2022). In the study, the content of GA in lettuce fleshy stems was the highest in the early stage of stem expansion, and then decreased with expansion; and the related genes involved in GA pathway also demonstrated a similar trend. These results indicated that GA might play an important role in cell proliferation in the early growth and development of lettuce. The content of CTK and ABA was the highest in the early stage of lettuce expansion, gradually decreasing with expansion, indicating that CTK and ABA played an important role in the early stage of lettuce expansion. However, the content of JA was higher than that of other hormones, indicating that JA played a promoting role during lettuce stem expansion.

Plant growth and development are not only regulated by enzyme genes but also by transcription factors (Smet et al., 2019). The bHLH TF gene *TMO5* was expressed in the early provascular initial cells. The heterodimer TMO5/LHW formed by TMO5 and LHW could directly target and induce the expression of *LOG4* and *LOG3* to control cell proliferation (De Rybel et al., 2013; De Rybel et al., 2014). *IbNAC083* was the core initiation factor for storage root development (He et al., 2021). Several WRKYIII TF genes showed different expression patterns at stem expansion stages in lettuce (Du et al., 2022). Overall, transcription factors might participate in the key biological processes during fleshy stem expansion. Studying the molecular mechanism of specific TFs is of great significance for regulating the growth and expansion of lettuce stems. The function of transcription factors involved in controlling stem expansion should be analyzed in-depth to improve the cultivation of high-yield lettuce. In the study, a comprehensive analysis of global changes in the metabolites and gene expression was conducted during fleshy stem expansion in stem lettuce. The important genes and metabolites identified in the progress of

lettuce fleshy stem expansion will provide important information for the further analysis of the molecular regulatory mechanisms during the process of lettuce stem growth and development.

Data availability statement

The datasets presented in this study can be found in NCBI, the accession number is PRJNA844256.

Author contributions

YZ and YH contributed to conception and design of the study. YWL, ZL, WC, YW, XW, and YHL performed the experiments. YWL, ZL, WC, and YW organized the database. YH wrote the paper. YZ revised the paper. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32102405), the Natural Science Foundation of Shandong Province (ZR2020QC156, ZR2020QC063), Central Guidance on Local Science and Technology Development Fund of Shaanxi Province (2022ZY1-CGZY-07), the Innovation Team of Youth Technology Project of High School in Shandong Province (2021KJ055), and Key project of Science and Technology Research Plan of Education Department of Hubei Province (D20221206).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1101199/full#supplementary-material

References

Ahmed, M., Iqbal, A., Latif, A., Din, S., Sarwar, M., Wang, W., et al. (2020). Overexpression of a sucrose synthase gene indirectly improves cotton fiber quality through sucrose cleavage. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.476251

Ayano, M., Kani, T., Kojima, M., Sakakibara, H., Kitaoka, T., Kuroha, T., et al. (2014). Gibberellin biosynthesis and signal transduction is essential for internode elongation in deepwater rice. *Plant Cell Environ.* 37, 2313–2324. doi: 10.1111/pce.12377

Borowski, J. M., Galli, V., Messias, R., Perin, E., Buss, J., Silva, S., et al. (2014). Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. *Planta.* 239, 1187–1200. doi: 10.1007/s00425-014-2041-2

Chen, P., Yang, R., Bartels, D., Dong, T., and Duan, H. (2022). Roles of abscisic acid and gibberellins in stem/root tuber development. *Int. J. Mol. Sci.* 23, 4955. doi: 10.3390/ijms23094955

Cong, L., Yue, R., Wang, H., Liu, J., Zhai, R., Yang, J., et al. (2019). 2,4-d-induced parthenocarpy in pear is mediated by enhancement of GA(4) biosynthesis. *Physiol. Plant* 166, 812–820. doi: 10.1111/ppl.12835

Das, U. N. (2015). Sucrose, fructose, glucose, and their link to metabolic syndrome and cancer. *Nutrition.* 31, 249–257. doi: 10.1016/j.nut.2014.05.015

Das, P., Manna, I., Sil, P., Bandyopadhyay, M., and Biswas, A. (2019). Exogenous silicon alters organic acid production and enzymatic activity of TCA cycle in two NaCl stressed indica rice cultivars. *Plant Physiol. Biochem.* 136, 76–91. doi: 10.1016/j.plaphy.2018.12.026

De Rybel, B., Adibi, M., Breda, A. S., Wendrich, J., Smit, M., Novak, O., et al. (2014). Plant development integration of growth and patterning during vascular tissue formation in arabidopsis. *Science.* 345, 1255215. doi: 10.1126/science.1255215

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., Reuille, P., Boeren, S., et al. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in arabidopsis. *Dev. Cell.* 24, 426–437. doi: 10.1016/j.devcel.2012.12.013

Dong, J., Wang, Y., Xu, L., Li, B., Wang, K., Ying, J., et al. (2022). *RsCLE22a* regulates the taproot growth through auxin signaling-related pathway in radish (*Raphanus sativus* L.). *J. Exp. Bot.*, erac406. doi: 10.1093/jxb/erac406

Durbak, A., Yao, H., and McSteen, P. (2012). Hormone signaling in plant development. *Curr. Opin. Plant Biol.* 15, 92–96. doi: 10.1016/j.pbi.2011.12.004

Du, P., Wu, Q., Liu, Y., Cao, X., Yi, W., Jiao, T., et al. (2022). WRKY transcription factor family in lettuce plant (*Lactuca sativa*): Genome-wide characterization, chromosome location, phylogeny structures, and expression patterns. *PeerJ.* 10, e14136. doi: 10.7717/peerj.14136

Fernie, A. R., Bachem, C. W. B., Helariutta, Y., Neuhaus, H., Prat, S., Ruan, Y., et al. (2020). Synchronization of developmental, molecular and meta-bolic aspects of source-sink interactions. *Nat. Plants.* 6, 55–66. doi: 10.1038/s41477-020-0590-x

He, S., Wang, H., Hao, X., Wu, Y., Bian, X., Yin, M., et al. (2021). Dynamic network biomarker analysis discovers IbNAC083 in the initiation and regulation of sweet potato root tuberization. *Plant J.* 108, 793–813. doi: 10.1111/tpj.15478

Jia, Y., Yu, P., Shao, W., An, G., Chen, J., Yu, C., et al. (2022). Up-regulation of LsKN1 promotes cytokinin and suppresses gibberellin biosynthesis to generate wavy leaves in lettuce. *J. Exp. Bot.* 73, 6615–6629. doi: 10.1093/jxb/erac311

Katsarou, K., Wu, Y., Zhang, R., Bonar, N., Morris, J., Hedley, P., et al. (2016). Insight on genes affecting tuber development in potato upon potato spindle tuber viroid (PSTVd) infection. *PloS One* 11, e0150711. doi: 10.1371/journal.pone.0150711

Kolomiets, M. V., Hannapel, D. J., Chen, H., Tymeson, M., and Gladon, R. (2001). Lipoxygenase is involved in the control of potato tuber development. *Plant Cell.* 13, 613–626. doi: 10.1105/tpc.13.3.613

Liang, Y. C., Reid, M. S., and Jiang, C. Z. (2014). Controlling plant architecture by manipulation of gibberellic acid signalling in petunia. *Hortic. Res.* 1, 14061. doi: 10.1038/hortres.2014.61

Liao, S., Wang, L., Li, J., and Ruan, Y. (2020). Cell wall invertase is essential for ovule development through sugar signaling rather than provision of carbon nutrients. *Plant Physiol.* 183, 1126–1144. doi: 10.1104/pp.20.00400

Li, M., Li, J., Zhang, R., Lin, Y., Xiong, A., Tan, G., et al. (2022). Combined analysis of the metabolome and transcriptome to explore heat stress responses and adaptation mechanisms in celery (*Apium graveolens* L.). *Int. J. Mol. Sci.* 23, 3367. doi: 10.3390/ijms23063367

Lin, M., Zhou, Z., and Mei, Z. (2022). Integrative analysis of metabolome and transcriptome identififfies potential genes involved in the flavonoid biosynthesis in *Entada phaseoloides* stem. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.792674

Li, L., Tong, Y. X., Lu, J. L., Li, Y., Liu, X., and Cheng, R. (2021). Morphology, photosynthetic traits, and nutritional quality of lettuce plants as affected by green light substituting proportion of blue and red light. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.627311

Liu, H., Liu, X., Zhao, Y., Nie, J., Yao, X., Lv, L., et al. (2022). Alkaline α-galactosidase 2 (CsAGA2) plays a pivotal role in mediating source-sink communication in cucumber. *Plant Physiol.* 189, 1501–1518. doi: 10.1093/plphys/kiac152

Liu, Z., Wei, F., and Feng, Y. Q. (2010). Determination of cytokinins in plant samples by polymer monolith microextraction coupled with hydrophilic interaction chromatography-tandem mass spectrometry. *Anal. Methods* 2, 1676–1685. doi: 10.1039/c0ay00334d

Liu, X., Wu, C., Su, D., Yang, Y., Xian, Z., Yu, C., et al. (2021). The SlHB8 acts as a negative regulator in stem development and lignin biosynthesis. *Int. J. Mol. Sci.* 22, 13343. doi: 10.3390/ijms222413343

Li, M., Xie, F., Li, J., Sun, B., Luo, Y., Zhang, Y., et al. (2020). Tumorous stem development of *Brassica juncea*: a complex regulatory network of stem formation and identification of key genes in glucosinolate biosynthesis. *Plants.* 9, 1006. doi: 10.3390/plants9081006

Li, X. J., Yang, J. L., Hao, B., Lu, Y., Qian, Z., Li, Y., et al. (2019). Comparative transcriptome and metabolome analyses provide new insights into the molecular mechanisms underlying taproot thickening in *Panax notoginseng*. *BMC Plant Biol.* 19, 451. doi: 10.1186/s12870-019-2067-5

Lutova, L. A., Dolgikh, E. A., Dodueva, I. E., Osipova, M.A., and Ilina, E.L. (2008). Investigation of systemic control of plant cell division and differentiation in the model of tumor growth in radish. *Genetika.* 44, 1075–1083. doi: 10.1134/S1022795408080073

Malarz, J., Michalska, K., and Stojakowska, A. (2020). Stem lettuce and its metabolites: does the variety make any difference? *Foods.* 10, 59. doi: 10.3390/foods10010059

Medina-Lozano, I., Bertolín, J. R., and Díaz, A. (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin c and anthocyanin content. *Food Chem.* 359, 129864. doi: 10.1016/j.foodchem.2021.129864

Miyashima, S., Roszak, P., Sevilem, I., Toyokura, K., Blob, B., Heo, J., et al. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature.* 565, 490–494. doi: 10.1038/s41586-018-0839-y

Pavlista, A. D. (2011). Growth regulators increased yield of atlantic potato. *Am. J. Pot. Res.* 88, 479–484. doi: 10.1007/s12230-011-9214-3

Peng, Y., Chen, L., Li, S., Zhang, Y., Xu, R., Liu, Z., et al. (2018). BRI1 and BAK1 interact with G proteins and regulate sugar-responsive growth and development in arabidopsis. *Nat. Commun.* 9, 1522. doi: 10.1038/s41467-018-03884-8

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in realtime RT-PCR. *Nucleic Acids Res.* 29, e45. doi: 10.1093/nar/29.9.e45

Qi, J. S., Zhang, B., Ma, L. G., Ma, G., Qin, S., Li, C., et al. (2021). First report of *Pythium aphanidermatum* causing root rot of head lettuce in China. *Plant Dis.* 11, 3324–3770. doi: 10.1094/PDIS-09-20-1875-PDN

Rahimi, A., Karami, O., Lestari, A. D., Werk, T., Amakorova, P., Shi, D., et al. (2022). Control of cambium initiation and activity in arabidopsis by the transcriptional regulator AHL15. *Curr. Biol.* 32, 1764–1775. doi: 10.1016/j.cub.2022.02.060

Ren, Z. M., Zhang, D., Jiao, C., Li, D., Wu, Y., Wang, X., et al. (2022). Comparative transcriptome and metabolome analyses identified the mode of sucrose degradation as a metabolic marker for early vegetative propagation in bulbs of lycoris. *Plant J.* 112, 115–134. doi: 10.1111/tpj.15935

Saidi, A., and Hajibarat, Z. (2021). Phytohormones: plant switchers in developmental and growth stages in potato. *J. Genet. Eng. Biotechnol.* 19, 89. doi: 10.1186/s43141-021-00192-5

Smet, W., Sevilem, I., de Luis Balaguer, M. A., Wybouw, B., Mor, E., Miyashima, S., et al. (2019). DOF2.1 controls cytokinin-dependent vascular cell proliferation downstream of TMO5/LHW. *Curr. Biol.* 29, 520–529. doi: 10.1016/j.cub.2018.12.041

Sohn, H. B., Lee, H. Y., Seo, J. S., Jung, C., Jeon, J., Kim, J., et al. (2011). Overexpression of jasmonic acid carboxyl methyltransferase increases tuber yield and size in transgenic potato. *Plant Biotechnol. Rep.* 5, 27–34. doi: 10.1007/s11816-010-0153-0

Sun, Q., Zhang, B., Yang, C., Wang, W., Xiang, L., Wang, Y., et al. (2022). Jasmonic acid biosynthetic genes *TgLOX4* and *TgLOX5* are involved in daughter bulb development in tulip (*Tulipa gesneriana*). *Hortic. Res.* 9, uhac006. doi: 10.1093/hr/uhac006

Su, W., Tao, R., Liu, W., Yu, C., Yue, Z., He, S., et al. (2022). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* 18, 479–409. doi: 10.1111/pbi.13213

Tang, M., Zhao, W., Xing, M., Zhao, J., Jiang, Z., You, J., et al. (2021). Resource allocation strategies among vegetative growth, sexual reproduction, asexual reproduction and defense during growing season of *Aconitum kusnezoffii* reichb. *Plant J.* 105, 957–977. doi: 10.1111/tpj.15080

Wang, M., Lavelle, D., Yu, C., Zhang, W., Chen, J., Wang, X., et al. (2022). The upregulated *LsKN1* gene transforms pinnately to palmately lobed leaves through auxin, gibberellin, and leaf dorsiventrality pathways in lettuce. *Plant Biotechnol. J.* 20, 1756–1769. doi: 10.1111/pbi.13861

Wang, Y., Wang, J., Guo, S., Tian, S., Zhang, J., Ren, Y., et al. (2021b). CRISPR/Cas9-mediated mutagenesis of ClBG1 decreased seed size and promoted seed germination inwatermelon. *Hortic. Res.* 8, 70. doi: 10.1038/s41438-021-00506-1

Wang, G. L., Xiong, F., Que, F., Xu, Z., Wang, F., and Xiong, A. (2015). Morphological characteristics, anatomical structure, and gene expression: novel insights into gibberellin biosynthesis and perception during carrot growth and development. *Hortic. Res.* 2, 15028. doi: 10.1038/hortres.2015.28

Wang, Q., Xu, Y., Zhang, M., Zhu, F., Sun, M., Lian, X., et al. (2021a). Transcriptome and metabolome analysis of stress tolerance to aluminium in *Vitis quinquangularis*. *Planta.* 254, 105. doi: 10.1007/s00425-021-03759-1

Wei, T., Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., et al. (2021). Whole-genome resequencing of 445 lactuca accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* 53, 752–760. doi: 10.1038/s41588-021-00831-0

Wu, Y., Ren, Z., Gao, C., Sun, M., Li, S., Min, R., et al. (2021). Change insucrose cleavage pattern and rapid starch accumulation govern lilyshoot-to-bulblet transition *in vitro*. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.564713

Wu, P., Zhang, L., Zhang, K., Yin, Y., Liu, A., Zhu, Y., et al. (2022). The adaptive evolution of *Euryale ferox* to the aquatic environment through paleo-hexaploidization. *Plant J.* 110, 627–645. doi: 10.1111/tpj.15717

Xu, Z. S., Tan, H. W., Wang, F., Hou, X., and Xiong, A. (2014). CarrotDB: a genomic and transcriptomic database for carrot. *Database.* 2014, bau096. doi: 10.1093/database/bau096

Yang, S., Meng, Z., Li, Y., Chen, R, Yang, Y., and Zhao, Z. (2021). Evaluation of physiological characteristics, soluble sugars, organic acids and volatile compounds in 'Orin' apples (*Malus domestica*) at different ripening stages. *Molecules.* 26, 807. doi: 10.3390/molecules26040807

Ye, J., Tian, R., Meng, X., Tao, P., Li, C., Liu, G., et al. (2020). Tomato SD1, encoding a kinase interacting protein, is a major locus controlling stem development. *J. Exp. Bot.* 71, 3575–3587. doi: 10.1093/jxb/eraa144

Zhang, W., Alseekh, S., Zhu, X., Zhang, Q., Fernie, A., Kuang, H., et al. (2020). Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *Plant J.* 104, 613–630. doi: 10.1111/tpj.14950

Zhang, X., Zhang, L., Zhang, Q., Xu, J., Liu, W., and Dong, W. (2017). Comparative transcriptome profiling and morphology provide insights into endocarp cleaving of apricot cultivar (*Prunus armeniaca* L.). *BMC Plant Biol.* 17, 72. doi: 10.1186/s12870-017-1023-5

Zheng, S., He, J., Lin, Z., Zhu, Y., Sun, J., and Li, L. (2020). Two MADS-box genes regulate vascular cambium activity and secondary growth *via* modulating auxin homeostasis in populus. *Plant Commun.* 2, 100134. doi: 10.1016/j.xplc.2020.100134

Zhong, R., Richardson, E. A., and Ye, Z. H. (2007). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in arabidopsis. *Plant Cell.* 19, 2776–2792. doi: 10.1105/tpc.107.053678

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 December 2022
DOI 10.3389/fpls.2022.1090621

OPEN ACCESS

EDITED BY
Lin Chen,
Institute of Animal Sciences (CAAS), China

REVIEWED BY
Mingjun Li,
Northwest A&F University, China
Liangju Wang,
Nanjing Agricultural University, China

*CORRESPONDENCE
Liyi Zhang
zhangliyi@caas.cn
Peihua Cong
congph@163.com

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

RECEIVED 05 November 2022
ACCEPTED 09 December 2022
PUBLISHED 29 December 2022

CITATION
Liang Z, Liu K, Jiang C, Yang A, Yan J, Han X, Zhang C, Cong P and Zhang L (2022) Insertion of a TRIM-like sequence in *MdFLS2-1* promoter is associated with its allele-specific expression in response to *Alternaria alternata* in apple.
Front. Plant Sci. 13:1090621.
doi: 10.3389/fpls.2022.1090621

Insertion of a TRIM-like sequence in *MdFLS2-1* promoter is associated with its allele-specific expression in response to *Alternaria alternata* in apple

Zhaolin Liang[1,2], Kai Liu[1,2], Chunyang Jiang[1,2], An Yang[1,2], Jiadi Yan[1,2], Xiaolei Han[1,2], Caixia Zhang[1,2], Peihua Cong[1,2]* and Liyi Zhang[1,2]*

[1]Research Institute of Pomology, Chinese Academy of Agricultural Sciences, Xingcheng, China, [2]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Germplasm Resources Utilization), Research Institute of Pomology, Chinese Academy of Agricultural Sciences, Ministry of Agriculture, Xingcheng, China

Alternaria blotch disease, caused by *Alternaria alternata* apple pathotype (AAAP), is one of the major fungal diseases in apple. Early field observations revealed, the anther-derived homozygote Hanfu line (HFTH1) was highly susceptible to AAAP, whereas Hanfu (HF) exhibited resistance to AAAP. To understand the molecular mechanisms underlying the difference in sensitivity of HF and HFTH1 to AAAP, we performed allele-specific expression (ASE) analysis and comparative transcriptomic analysis before and after AAAP inoculation. We reported an important immune gene, namely, *MdFLS2*, which displayed strong ASE in HF with much lower expression levels of HFTH1-derived alleles. Transient overexpression of the dominant allele of *MdFLS2-1* from HF in GL-3 apple leaves could enhance resistance to AAAP and induce expression of genes related to salicylic acid pathway. In addition, *MdFLS2-1* was identified with an insertion of an 85-bp terminal-repeat retrotransposon in miniature (TRIM) element-like sequence in the upstream region of the nonreference allele. In contrast, only one terminal direct repeat (TDR) from TRIM-like sequence was present in the upstream region of the HFTH1-derived allele *MdFLS2-2*. Furthermore, the results of luciferase and β-glucuronidase reporter assays demonstrated that the intact TRIM-like sequence has enhancer activity. This suggested that insertion of the TRIM-like sequence regulates the expression level of the allele of *MdFLS2*, in turn, affecting the sensitivity of HF and HFTH1 to AAAP.

KEYWORDS

***alternaria alternata* apple pathotype, *malus domestica* borkh., allele-specific expression, MdFLS2, TRIM element**

1 Introduction

Apple (*Malus domestica* Borkh.) is one of the significant fruit species that is widely consumed globally. However, apple is susceptible to fungal diseases during its cultivation. Alternaria blotch disease, caused by *Alternaria alternata* apple pathotype (AAAP), is one of the most serious fungal diseases of apples (Filajdic and Sutton, 1991; Zhang et al., 2015). As a pathogenic variant of *Alternaria*, AAAP mainly infects apple leaves and causes early defoliation of the tree, severely damaging the yield and quality of apples (Moriya et al., 2019). The most effective measure for overcoming fungal diseases in apple is the selective breeding of resistant cultivars (Zhang et al., 2018). However, the mechanism of resistance to AAAP in apples is still not clearly studied (Hou et al., 2021). Therefore, it is urgent to investigate the molecular mechanism underlying Alternaria blotch disease resistance in apple.

To defend themselves against pathogens, plants have evolved with strong immune mechanisms (Ngou et al., 2022). Pattern recognition receptors (PRRs) located on plant cell membranes recognize conserved molecular pathogen-molecular patterns (PAMPs) secreted by pathogens, thereby activating pattern-triggered immunity (PTI) to limit pathogenicity (Zipfel, 2008). In contrast, nucleotide-binding leucine-rich repeat receptors located inside plant cells can activate effector-triggered immune (ETI) responses by directly or indirectly sensing specific effectors secreted by pathogens (Nomura et al., 2011). PTI is an important component of basal resistance in plants. Recent studies have reported that defense responses triggered by PTI are equally essential for ETI activation (Ngou et al., 2021; Yuan et al., 2021). As PRRs trigger PTI responses, they have been at the forefront of plant immune research, with the receptor kinase *FLS2* being the most extensively studied (Zipfel et al., 2004; Lu et al., 2011; Shi et al., 2016; Zhang et al., 2020b). FLS2 can form a complex with the co-receptor kinase BAK1 to recognize the conserved protein polypeptide flg22 in the flagellin of pathogenic bacteria to activate the immune response of plants (Sun et al., 2013). In addition, FLS2 may be involved in the recognition of other ligands, e.g., the sensitivity of *Arabidopsis AtFLS2* mutants to the protein polypeptide *Ax21* secreted by *Xanthomonas oryzae* pv. *oryzae* (Xoo) is significantly reduced by the action of FLS2 (Danna et al., 2011). PTI-mediated immune responses involve MAPK cascade responses and phytohormone signaling (Asai et al., 2002; Kong et al., 2016). Upon recognition of flg22 by FLS2, the MAPK cascade response in plants is activated to further phosphorylate WRKY transcription factors, thereby regulating the expression of immune-related genes (Sarowar et al., 2019). Previous studies in apple reported one *FLS2* ortholog, namely, *MdFLS2* (Chen-Hui et al., 2018). Overexpression of *MdFLS2* in *Arabidopsis* increased its resistance to *Botryosphaeria dothidea*, whereas the expression levels of genes related to salicylic acid pathway were significantly enhanced (Liu et al., 2018). However, it is unclear whether *MdFLS2* has a broad-spectrum resistance to various plant pathogens, similar to *AtFLS2*, and the mechanisms regulating its expression level should be further studied.

Apple is typically hetero-pollinated plant, and its highly heterozygous genotypes confer diversity in response to biotic (Reim et al., 2019) and abiotic (Ren et al., 2017) stresses. After 3 years of field observations, we observed that Hanfu (HF) exhibited resistance, whereas the anther-derived homozygote Hanfu line (HFTH1) was highly susceptible to AAAP (Figure 1A). The homozygosis of some important genes predisposes plants to exhibit extreme phenotypes (Wang et al., 2014; Zhang et al., 2020a). However, heterozygotes can overcome these adverse effects by selecting more favorable allele expression under specific conditions, i.e., allele-specific expression (ASE) (Shao et al., 2019). Recent studies have reported that ASE is involved in the regulation of many important traits in apples, such as fruit ripening (Sun et al., 2020) and flower color formation (Tian et al., 2022). To understand the molecular mechanisms underlying the difference in sensitivity of HF and HFTH1 to AAAP, we aimed to investigate the ASE patterns in HF under AAAP stress. Previously, we have obtained the HFTH1 haplotype reference genome (Zhang et al., 2019), providing a genomic resource for ASE analysis in HF. In this study, we integrated ASE analysis and comparative transcriptomic analysis of HF and HFTH1 before and after AAAP inoculation to identify key genes and regulatory patterns that influence differences in susceptibility to AAAP. These results can provide new insights and references for breeding with Alternaria blotch disease resistance in apple.

2 Materials and methods

2.1 Plant materials, AAAP culture, and leaf inoculation

To reduce the genetic difference between two apple species, we selected HF (exhibiting resistance to AAAP) and HFTH1 (an anther-derived homozygote Hanfu line susceptible to AAAP) as experimental materials. Both materials were grafted on GM256 rootstock in 2010 and grown in a research orchard at Research Institute of Pomology, Chinese Academy of Agricultural Sciences (CAAS). AAAP was obtained from Plant Protection Center of Institute of Pomology, CAAS. AAAP was cultured on potato dextrose agar (PDA; 200 g potato extract, 20 g dextrose, and 20 g agar in 1 L water) medium and incubated in dark until the mycelia spread to two-third of the plate. The mycelium was inoculated onto 20-day-old HF and HFTH1 leaves according to the experimental method described previously (Zhu et al., 2017). In each treatment, four cakes of mycelium were used to inoculate both sides of the midvein on the abaxial surface of the leaves.

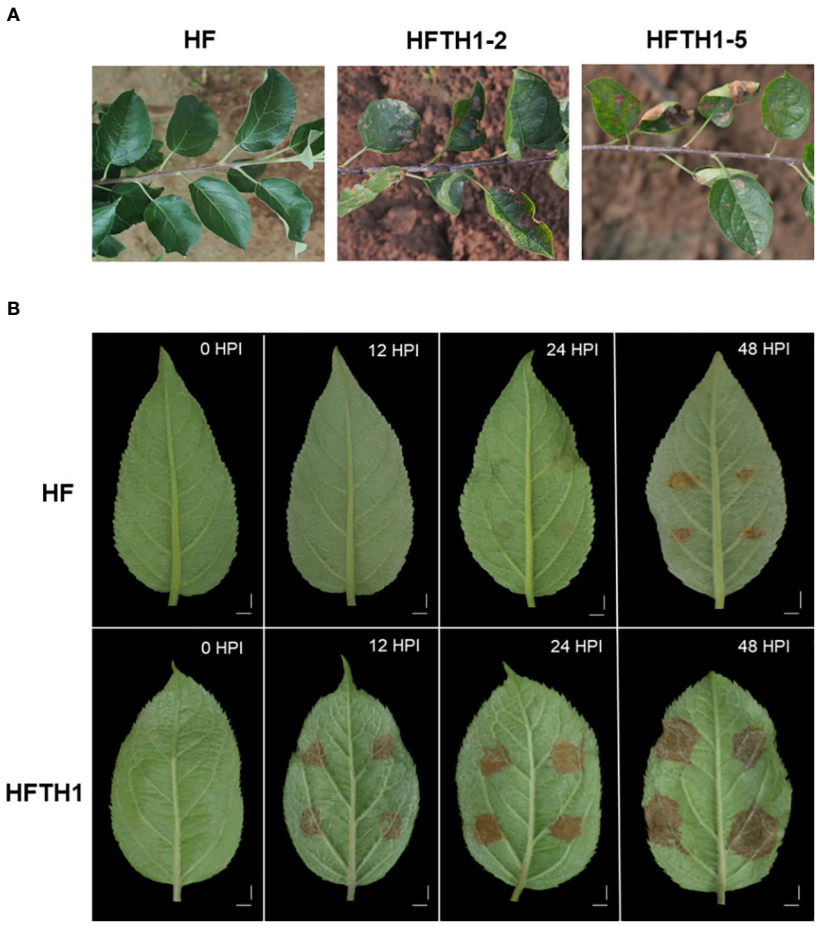


FIGURE 1
Different susceptibility of HF and HFTH1 to AAAP. **(A)** Investigation of the susceptibility of HF and HFTH1 to AAAP in the field experiments. **(B)** Changes in symptoms of HF and HFTH1 at 12, 24, and 48 h after inoculation with AAAP *in vitro*. Mock inoculation of HF and HFTH1 leaves without AAAP was used as the control (0HPI). Scale bar = 1 cm.

Six treatment groups were set up at 12-h intervals for both plant materials, and the control group was mock-inoculated using PDA medium cakes without mycelium. After inoculation, the apple leaves were incubated in a sterilized culture chamber at 25°C under a 14 h light/10 h dark conditions with moisture control. Further, 48 h after inoculation in the first group, samples were simultaneously collected from each group to observe the incidence of the two materials at three stages after inoculation (12 HPI, 24 HPI, and 48 HPI) compared with the control group (0 HPI). Apple leaves were collected from three HF plants and used for whole genome resequencing.

Leaf samples were quickly placed in liquid nitrogen and stored at -80°C for resequencing. Three parallel leaves on three trees were collected for each sample set, representing three independent biological replicates. The strand-specific transcriptome sequencing and whole genome resequencing was performed by BerryGenomics (Beijing, China).

2.2 Genome and strand-specific transcriptome sequencing

Total cellular DNA was extracted using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany), as per the manufacturer's instructions. The final libraries constructed above were sequenced using the Illumina Novaseq6000 platform, and 150-bp paired-end reads were generated with an insert size of approximately 350 bp. The raw reads obtained from sequencing were quality controlled, and the clean reads obtained after screening were compared with the HFTH1 reference genome (Zhang et al., 2019) using BWA (version 0.7.15; Li and Durbin, 2009) for variant calling.

As per the manufacturer's instructions, total RNA was isolated using cetyltrimethyl ammonium bromide method (Pavy et al., 2008). The purity of RNA was measured using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and analyzed for RNA degradation and contamination

using agarose gel electrophoresis. Strand-specific cDNA libraries were constructed using the dUTP method, as per a previous study (Parkhomchuk et al., 2009). Further, the 24 individual libraries were sequenced using the Illumina Novaseq6000 platform (BerryGenomics, Beijing, China).

2.3 Analysis of data and allele–specific expression

The raw reads of sequenced raw sequences were subjected to strict quality control to obtain valid and high-quality clean reads. The clean reads were compared with the HFTH1 apple reference genome using HISAT2 (v 2.20; Kim et al., 2015) to obtain information such as the position of the read on the compared genome and quality of the match; thus, the gene or transcript is annotated and quantified. Gene-level quantification was performed using featureCounts (Version 2.0.1; Liao et al., 2014), and gene expression levels were calculated as FPKM. Based on the comparison results, the differential expression of genes in each sample was analyzed using edgeR package (Robinson et al., 2010), and the pvalue and padj values of differential expression were calculated. The differential genes with padj < 0.05 and |log2FoldChange| > 1 were screened. GO enrichment analysis was performed using topGO software, and KEGG enrichment analysis was performed using KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/). The heatmap was created using Tbtools (Chen et al., 2020).

To identify alleles, genome and transcriptome sequencing data compared with the reference genome were called for variants using GATK (version 3.8.0; Brouard et al., 2019). In this study, only SNPs were used to distinguish alleles using VCFTools (version 0.1.14) to filter SNPs and retaining SNP sites with read depths above 8×, the remaining SNP sites were annotated using snpEff (Cingolani et al., 2012). The SNP sites located in the CDS region of the genome were used for the identification of ASE. Genes that simultaneously satisfy the following conditions were considered ASE genes (ASEGs; Tian et al., 2022): (1) read counts of the reference allele divided by total read counts were >0.75 or <0.25, (2) different SNPs on the same gene exhibited same direction of significant bias, and (3) genes were identified as ASEGs in at least two out of three replicates.

2.4 Vector construction and transient overexpression in apple leaves

Specific primers were designed to amplify *MdFLS2-1* and *MdFLS2-2* in HF based on the coding sequence of *MdFLS2* predicted from the HFTH1 reference genome (Supplementary Table S7). The CDS region of *MdFLS2* allele was inserted into the *Nde I* digest site of PRI101 vector using 2× MultiF Seamless Assembly Mix (ABclonal, Wuhan, China) according to the in-fusion cloning method. Specific primers were designed to amplify the 2058-bp promoter sequence of *MdFLS2* allele based on the HFTH1 reference genome (Supplementary Table S7) and ligated to the pESI-Blunt vector. The ligated vector was transformed into *E. coli* DH5α competent cells, and the monoclonal colonies were picked and sent to GENEWIZ (Tianjin, China) for sequencing. The allelic variants of the gene were analyzed using ClustalX 2.1.

A recombinant plasmid inserted with the CDS region of *MdFLS2-1* was transformed into *Agrobacterium tumefaciens* strain GV3101 (Bai et al., 2013). The vector-transformed *Agrobacterium* was cultured in 10 mL YEP medium and shaken at 28°C and 130 rpm until OD_{600} reached to 1. The bacterial broth was centrifuged at 5000 rpm, resuspended in 10 mL of buffer (10 mM $MgCl_2$,10 mM MES [pH 5.7], and 200 μM acetosyringone), and used after 3 h of activation at 28°C. To improve the transformation efficiency, we selected GL-3 apple plants that were favorable for transient transformation experiments. GL-3 apple plants cultured *in vitro* for 1 month were selected, and two spots on both sides of the leaf veins were selected to be infiltrated by injection. The infiltrated plants were placed in new MS medium and incubated for 4 days at 25°C in a light incubator under 16 h light/8 h dark conditions. Three *Agrobacterium*-infiltrated plants were randomly selected, and RNA from the injected leaves was extracted for gene expression analysis. After 4 days of infiltration, the remaining plants were inoculated with AAAP spore suspension at the location of the *Agrobacterium* injection, and the inoculated plants were placed back into MES medium and incubated in a light incubator for 48 h. In the control group, GL-3 apple leaves were treated in the same manner with sterile injection buffer and injection buffer containing empty vector, and the same batch of AAAP spore suspension was used to inoculate the infiltration site of the leaves. Plants with AAAP-infected leaves were examined at 48 HPI, and three independent biological replicates were collected and photographed. The differences in incidence levels were analyzed using student's t test (Zhang et al., 2018).

2.5 Luciferase reporter assay

The TRIM-like sequence from the promoter of *MdFLS2-1* and TDR2 sequence from the promoter of *MdFLS2-2* were respectively fused to the cauliflower mosaic virus 35S minimal promoter (mpCaMV) and ligated to the *HindIII-BamHI* site of pGrernII0800-*LUC* vector, to generate the construct structures TRIM:mpCaMV : *LUC* and TDR2:mpCaMV : *LUC*. Sequences were synthesized (Supplementary Table S8) by GENEWIZ (Tianjin, China). Two reporter constructs were used to simultaneously transform *Agrobacterium* GV3101. Leaves of 8-week-old *Nicotiana benthamiana* plants were permeated with a needle-free syringe using the above injection buffer. The infiltrated plants were first incubated in a phytotron at 23°C for 12 h, followed by 60 h of growth under 16 h light/8 h dark conditions.

The working solution (100 mM stock solution; 0.1% Triton X-100) was evenly spread onto the plant leaves using a high-pressure sprayer. The leaves were left in dark for 7 min, and the LUC images were captured using a low-light cooled CCD imaging apparatus (Tanon 5200Multi, China). The LUC/REN ratio was analyzed as described in a previous study (Liu et al., 2019). The experiment was independently repeated three times to obtain similar results.

2.6 Analysis of β-glucuronidase activity

The TRIM-like and TDR2 sequences fused with the 35S minimal promoter were respectively ligated to the *HindIII-BamHI* site of the pBI121 vector to generate reporter constructs and used to transform *Agrobacterium* GV3101. The GV3101 strain containing the recombinant plasmid was infiltrated into the abaxial leaf surface of *N. benthamiana* plants using the injection buffer described above. Three biological replicates were set for each infection. Infected plants were grown in a climatic chamber for 3 days, and β-glucuronidase (GUS) activity was detected as described in a previous study (Jefferson et al., 1987).

2.7 RNA extraction and RT-qPCR analysis

After freezing the plant leaves in liquid nitrogen, total RNA was extracted using Quick RNA Isolation Kit (Huayueyang Biotechnology, China) according to the manufacturer's instructions. RNA quality was detected using agarose gel electrophoresis, and RNA concentration was determined using NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). The same concentration of RNA was reverse transcribed to cDNA using PrimeScript RT Master Mix (TakaRa, Tokyo, Japan). RT-qPCR was conducted using a TB-Green PCR kit (TakaRa, Tokyo, Japan) as per the manufacturer's instructions. The relative expression of each gene was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) using *Tubulin* as the reference gene to normalize the gene expression level. Three experimental replicates were set up for each sample, and the mean standard deviation was calculated. RT-qPCR primers (Supplementary Table S7) were designed using primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast).

3 Results

3.1 HF and HFTH1 have different sensitivity to AAAP

Based on field observation for 3 years, the homozygous Hanfu line HFTH1 exhibited symptoms of severe spotted defoliation disease (Figure 1A), causing early defoliation in the wet rainy season. Meanwhile, the heterozygote HF exhibited no symptoms of susceptibility to the disease.

We performed simultaneous infection experiments on isolated leaves of HF and HFTH1 using the AAAP strain and recorded the symptoms on leaves (Figure 1B). The leaves of HFTH1 started exhibiting tissue necrosis by 12 HPI, brown spots by 24 HPI, and dark brown spots with conspicuous white hyphae on the necrotic tissue by 48 HPI. In contrast, HF leaves exhibited a later onset and milder symptoms. By 12HPI, no signs of disease were observed on HF leaves. By 24HPI, slight symptoms of the disease appeared. By 48HPI, lesions expanded and deepened to brown.

3.2 Sequencing and identification of ASE

The leaves of HF and HFTH1 after 12, 24, and 48 h of AAAP inoculation were set up in three replicates. We mixed the mock-inoculated leaves of three time points into one sample as the control (0 HPI). A total of 24 samples were sequenced using the Illumina Novaseq6000 platform for Strand-specific transcriptome sequencing. In total, 20,379,378–36,174,753 clean reads were generated from 24 RNA-seq libraries, with Clean GC% of 45.43–46.52, CleanQ20 of 97.33–98.10, and CleanQ30 of 92.80–94.50 (Supplementary Table S1). This indicated that the obtained RNA-Seq data were of high quality and suitable for further analysis. The transcriptome sequencing data were matched with the HFTH1 apple reference genome with 94.76%–97.19% similarity for the HFTH1 and 93.75%–95.19% similarity for the HF. The high matching rate ensured the utilization and accuracy of the transcriptome data.

Whole genome resequencing of heterozygous diploid HF was performed at a sequencing depth of 35×, and a total of 571,019 SNPs and 695,491 indels (Figure 2A) were identified using the HFTH1 genome as a reference. To improve the accuracy of ASE analysis, 52,765–46,854 SNP sites (Supplementary Table S2) located in the CDS region of genes and shared in both genome and transcriptome with transcriptome sequencing depths > 8 were retained in combination with transcriptome data of HF at each period post AAAP inoculation and control (Gao et al., 2021). These SNPs were used to analyze the expression levels between alleles. Alleles located in the HFTH1 reference genome with expression ratios > 0.75 or < 0.25 were considered to have ASE, whereas the presence of ASE in at least two of the three replicate samples with the same orientation was identified as an ASE gene (ASEG). Finally, a total of 2406 ASEGs were identified in the transcriptome, and 899–1086 ASEGs were identified in each of the four time periods (Figure 2B). Most ASEGs were biased toward reference genome expression during at each time period after AAAP inoculation and mock inoculation, and more genes exhibited ASE as the time of AAAP infection progressed (Supplementary Table S3). These results suggested that AAAP infection may have altered the allele expression pattern in HF leaves.

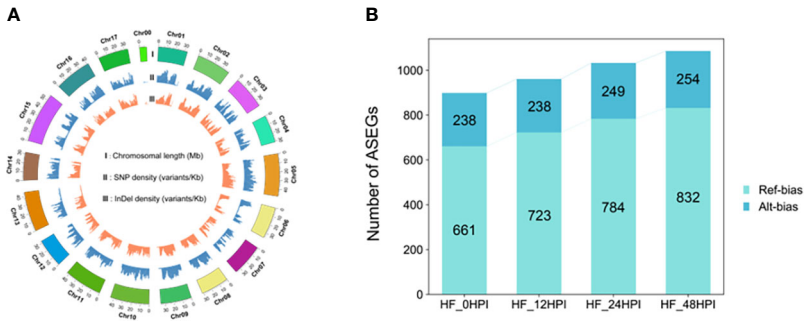


FIGURE 2
Identification of ASEGs. **(A)** Circos plot of HF resequencing data, chromosome length, SNP density, and indel density. **(B)** The number of ASEGs in HF leaves at 12, 24, and 48 h after AAAP inoculation and mock inoculation (0 HPI).

3.3 Two modes of ASEGs at different time periods before and after AAAP infection

A comparative analysis of ASEGs at different time periods before and after AAAP infection revealed that 132 genes exhibited ASE with the same direction of expression bias before and after the infection. Of them, 114 genes were biased toward the reference genome expression and 18 genes were biased toward the alternative genome expression (Figure 3A). Several studies have suggested that genes that exhibit consistent ASE across conditions may have dominant effects (Shao et al., 2019). GO enrichment analysis was used to analyze the biological functions of these ASEGs (Supplementary Table S4), and 89 genes were grouped into three main categories of GO: biological process, cellular component, and molecular function (Figure 3B). In the category of biological processes, more genes were involved in "oxidation-reduction process," "protein phosphorylation," and "proteolysis." In the category of cellular components, most genes were classified as "integral component of membrane." In the category of molecular functions, most genes were involved in "ATP binding."

The remaining 2274 genes of the 2406 ASEGs exhibited inconsistent patterns of allelic bias expression at different time periods before and after the infection (Figure 3A). The expression pattern of this allele may cause an overdominance effect in heterozygotes (Zhang et al., 2021). GO enrichment analysis of these genes (Supplementary Table S5) revealed that in the category of biological processes, most of the genes were involved in "oxidation-reduction process," "protein phosphorylation," and "regulation of transcription, DNA-templates" (Figure 3C). In the category of cellular components, the vast majority of genes were classified as "integral component of membrane." Among the category of molecular functions, most of the genes were enriched in "ATP binding," "metal ion binding," and "zinc ion binding." These results were similar with the enrichment results of biased consistent ASEGs, suggesting that ASEGs may function mainly in some specific subcategories of cellular components and biological processes. These biological processes are closely related to plants' resistance to pathogens (Turra et al., 2014; David et al., 2019), predicting that ASEGs may play an important role in the process of plant immune response.

3.4 Differential expression analysis and KEGG functional annotation

To identify defense genes that affect the difference in sensitivity of HF and HFTH1 to AAAP, we performed a comparative analysis of the transcriptomic data of HF and HFTH1 at 12, 24, and 48 h after the infection and mock inoculation (control; 0 HPI). Gene expression was compared between sample sets at different time periods after AAAP infection and mock inoculation, and differentially expressed genes (DEGs) that were up- and downregulated were counted. The results revealed that the up- and downregulated DEGs after AAAP infection in HF and HFTH1 increased with the advancement of infection time, compared with the control (Supplementary Figures S1A-D). This suggested that more genes were involved in the response to AAAP as the time of infection advanced. Meanwhile, the number of DEGs, both up- and downregulated, was higher in HF than in HFTH1 at each post-infection time compared with the control (Supplementary Figures S1A, B). This indicated that more genes were involved in disease-resistant varieties in response to AAAP stress. The transcriptome data of HF and HFTH1 in the same time after infection and mock inoculation were analyzed, and the DEGs that were up- and downregulated by HF compared with HFTH1 at each time period were separately counted. Due to the similar genetic background between the two species, relatively few DEGs were obtained. Moreover, and the number of DEGs upregulated in HF relative to HFTH1 was higher than that of DEGs

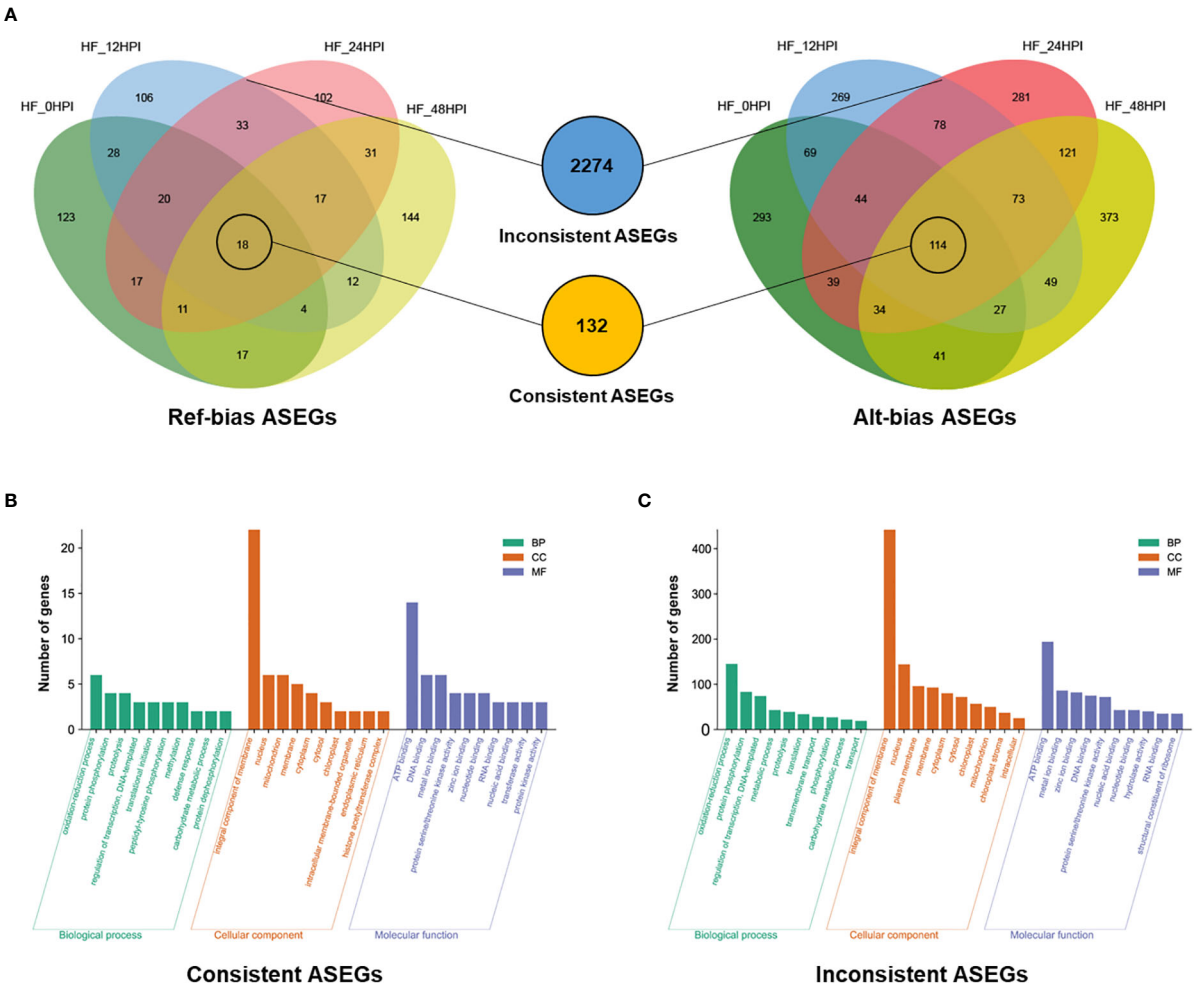


FIGURE 3
Characteristics of the two models of ASEGs. **(A)** Venn diagram analysis of reference genomic bias (Ref-bias) and alternative genomic bias (Alt-bias) at various times before and after AAAP infection. The middle blue and orange circles indicate inconsistent and consistent ASEGs, respectively. GO enrichment analysis of **(B)** consistent and **(C)** inconsistent ASEGs.

downregulated in all four time periods after infection and mock inoculation (Figure 4A).

Based on orthology (KO) terminology mapping of the upregulated DEGs of HF compared with HFTH1 at the same period after pathogenic inoculation to the KEGG pathway, the top 20 significantly enriched pathways at each time period were selected for analysis (Figures 4B-D). Multiple KEGG pathways were significantly simultaneously enriched in the comparative combinations at each time period; they were "Tyrosine metabolism," "Biosynthesis of secondary metabolites," "Isoquinoline alkaloid biosynthesis," "Metabolic pathways," "Plant-pathogen interaction," "Amino sugar and nucleotide sugar metabolism," "Alanine, aspartate, and glutamate metabolism," "MAPK signaling pathway-plant," and "Cyanoamino acid metabolism." Among them, "Plant-pathogen interaction" and "MAPK signaling pathway-plant" are directly related to the defense response of plants against pathogens (Dodds and Rathjen, 2010). MAPK signaling is a central pathway in immune response transduction and plays a crucial role in phytohormone-induced immune responses (Meng and Zhang, 2013). DEGs enriched in "Plant-pathogen interaction" and "MAPK signaling pathway-plant" at three time periods after pathogenic infection comprised a dataset containing 51 "core" immune genes (Supplementary Table S6). This dataset provides candidate genes for further analysis and functional identification.

3.5 Identification of *MdFLS2* and expression analysis of related genes

Combining the above "core" immune genes dataset with the 2406 ASEGs, the Venn diagram indicated that six genes were

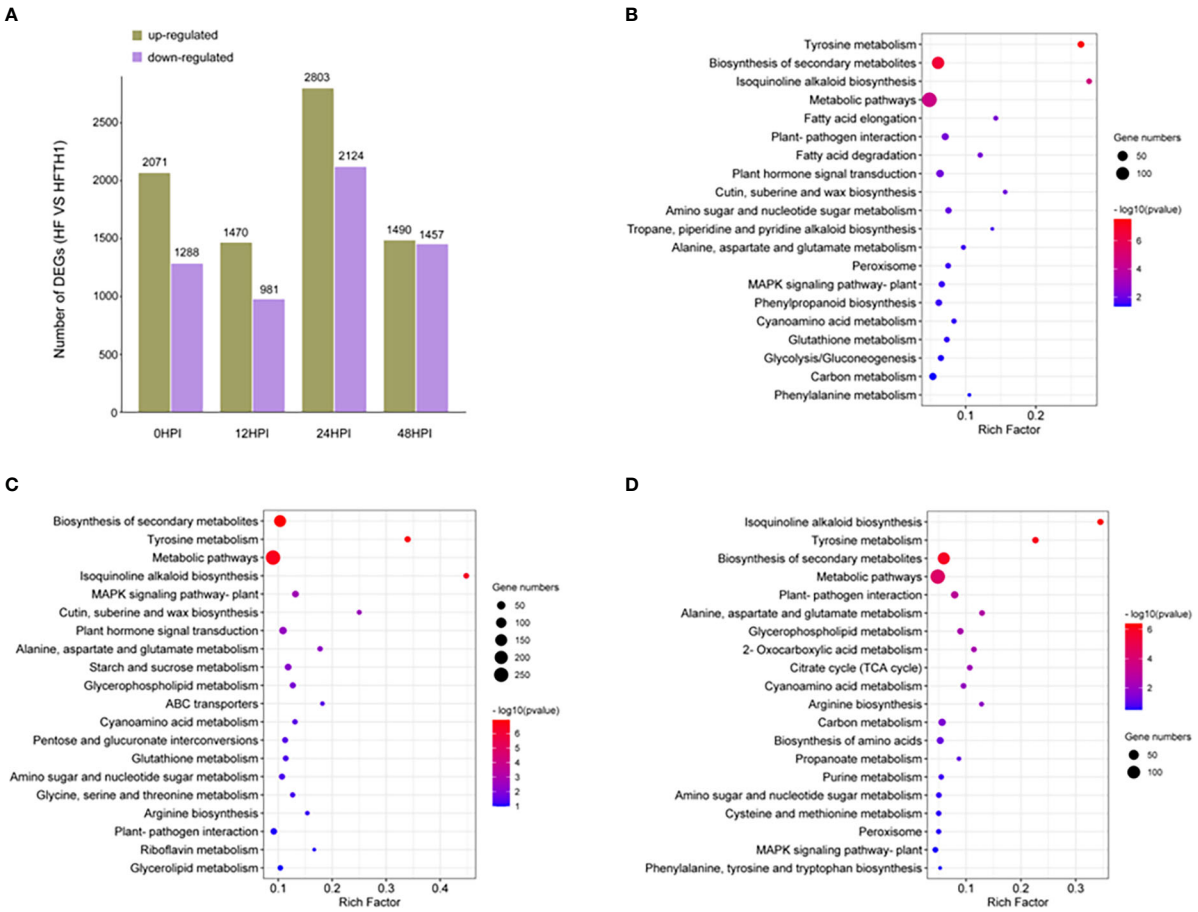


FIGURE 4
Comparative transcriptomic analysis of HFTH1 and HF before and after AAAP infection. **(A)** DEGs in HF compared with HFTH1 at the same time before and after AAAP infection. **(B–D)** Compared with HFTH1, HF up-regulated the KEGG pathway with significant enrichment of DEGs at 12 HPI **(B)**, 24 HPI, and **(C)** and 48 HPI **(D)**.

simultaneously present in both datasets (Figure 5A). Heatmaps of the expression of these genes between HF and HFTH1 leaves indicated that *MdFLS2* consistently exhibited significant differences in expression levels at three time periods after AAAP infection (Figure 5B). Meanwhile, *MdFLS2* was consistently biased toward alternative genomic ASE after AAAP infection and mock inoculation. Previous studies have reported that overexpression of *MdFLS2* in *Arabidopsis* enhances salicylic acid signaling and improves resistance to *Botryosphaeria dothidea* in *Arabidopsis* (Liu et al., 2018). There are some similarities in the disease resistance mechanisms of apples to AAAP and *Botryosphaeria dothidea*; for example, sorbitol can enhance the resistance of apples to both AAAP and *Botryosphaeria dothidea* (Meng et al., 2018; He et al., 2022).

To further clarify the differential expression of genes from *MdFLS2*-related pathway between the two plant materials, we analyzed the differential expression of related pathway genes in HFTH1 and HF, including those from "Plant-pathogen interaction," "MAPK signaling pathway-plant," and salicylic acid pathway in plant hormone signal transduction. Heatmap analysis revealed that the expression levels of *MdFLS2*-related pathway genes were basically higher in HF than in HFTH1; particularly, expression levels of genes related to salicylic acid pathway were significantly higher in disease-resistant HF than disease-susceptible HFTH1 at the early stage of pathogenic infection, i.e., at 12 HPI and 24 HPI after AAAP inoculation (Figure 5C). These genes may play an essential role in the early defense against pathogenic infection.

3.6 *MdFLS2-1* is a positive regulator of resistance to AAAP

The *MdFLS2* allele located in the HF alternative genome was named *MdFLS2-1*, and that located in HFTH1 was named

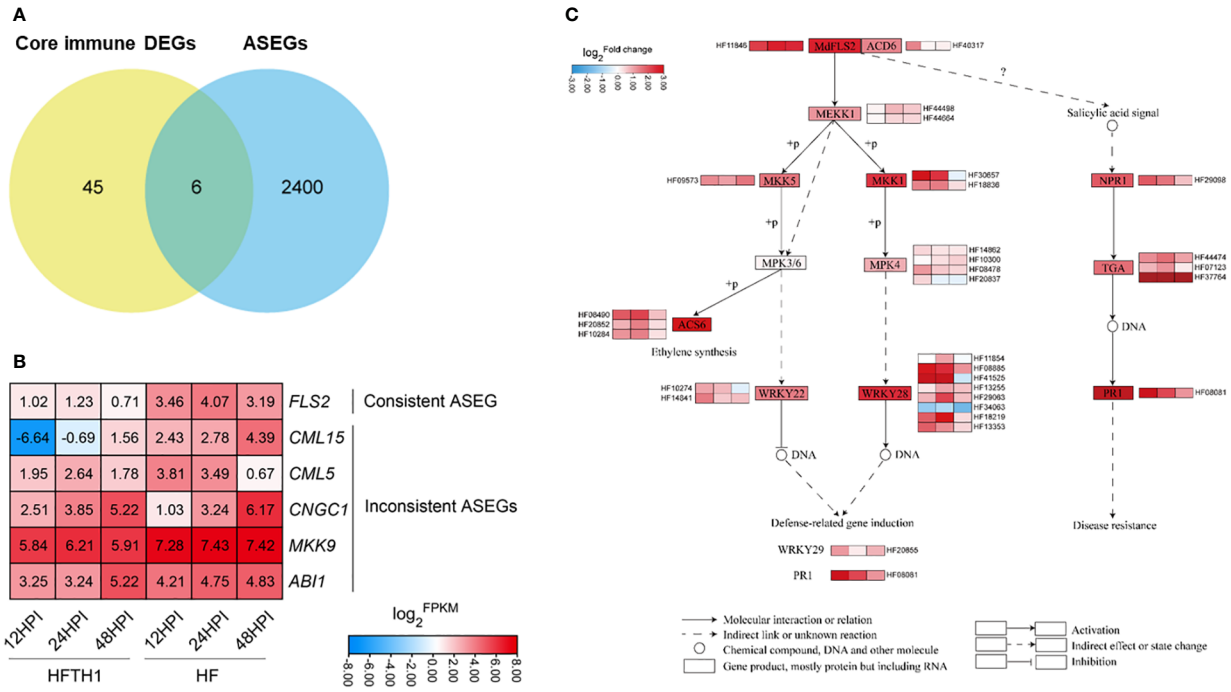


FIGURE 5
The ASE gene *MdFLS2* may affect the defense response of apple leaves to AAAP. **(A)** Venn diagram analysis of "core" immune DEGs and ASEGs. **(B)** Heatmap showing log_2^{FPKM} of six "core" immune ASEGs in HFTH1 and HF at 12, 24, and 48 h after AAAP inoculation. **(C)** Heatmap sequentially shows the $log_2^{Fold\ change}$ of genes from *MdFLS2*-related pathway in HF compared with HFTH1 at 12, 24, and 48 h after AAAP inoculation.

MdFLS2-2. ASE analysis revealed that *MdFLS2-1* exhibited a significant expression advantage compared with *MdFLS2-2* in HF after AAAP infection and mock inoculation. To identify the allelic variants located in the CDS region of the *MdFLS2*, we cloned *MdFLS2-1* and *MdFLS2-2* in HF leaves. *MdFLS2-1* and *MdFLS2-2* had 10 SNPs in the CDS region and caused difference in five amino acids. By comparing the protein sequence of FLS2 with that of *A. thaliana* (Gomez-Gomez and Boller, 2000), *Citrus sinensis*, *Prunus dulcis*, *Rosa chinensis*, *Trifolium pratense*, *Medicago truncatula*, *Glycine max*, and *P. persica*, we discovered that the amino acid sites of the variants were not conserved in several species (Supplementary Figure S2). However, these results are insufficient to indicate whether amino acid variants of MdFLS2-1 and MdFLS2-2 cause functional differences.

We observed that the previously reported *MdFLS2* is actually *MdFLS2-2* (Liu et al., 2018). To determine the function of *MdFLS2-1* and its role in AAAP infection, we constructed the PRI01-*MdFLS2-1* vector and performed transient expression experiments in GL-3 apple leaves. RT-qPCR analysis of transgenic leaves revealed that *MdFLS2* expression levels were significantly increased in OE-MdFLS2-1 transgenic leaves compared with those in untransformed control leaves (WT) with empty vector (EV) (Figure 6A). AAAP inoculation tests on transgenic leaves and controls revealed that apple leaves overexpressing *MdFLS2-1* had significantly less disease incidence than WT and EV plants at 48 HPI (Figures 6C, D). These results suggested that *MdFLS2-1* positively regulates resistance to AAAP in apple. RT-qPCR analysis of genes related to salicylic acid pathway in transgenic leaves revealed that the expression levels of marker genes of salicylic acid signaling pathway (*MdNPR1*, *MdTGA3*, and *MdPR1*) were significantly higher in OE-MdFLS2-1 plants compared with WT and EV plants (Figure 6B). These results suggested that the expression of both *MdFLS2-1* and *MdFLS2-2* is positively correlated with the salicylic acid signaling pathway, and *MdFLS2-2* positively regulates the resistance of apple leaves to AAAP. Therefore, we speculated that *MdFLS2-1* and *MdFLS2-2* play at least similar positive roles in the resistance of apples to fungi.

3.7 TRIM-like sequence as an enhancer of *MdFLS2-1* expression

Since in heterozygous diploids, alleles are placed in the same *trans* regulation background and cellular environment, the effect of *trans* regulation on allele expression can be temporarily shielded, thus prioritizing *cis* regulation of variants in the genome (Hill et al., 2021). To investigate the *cis*-regulatory mechanism causing specific

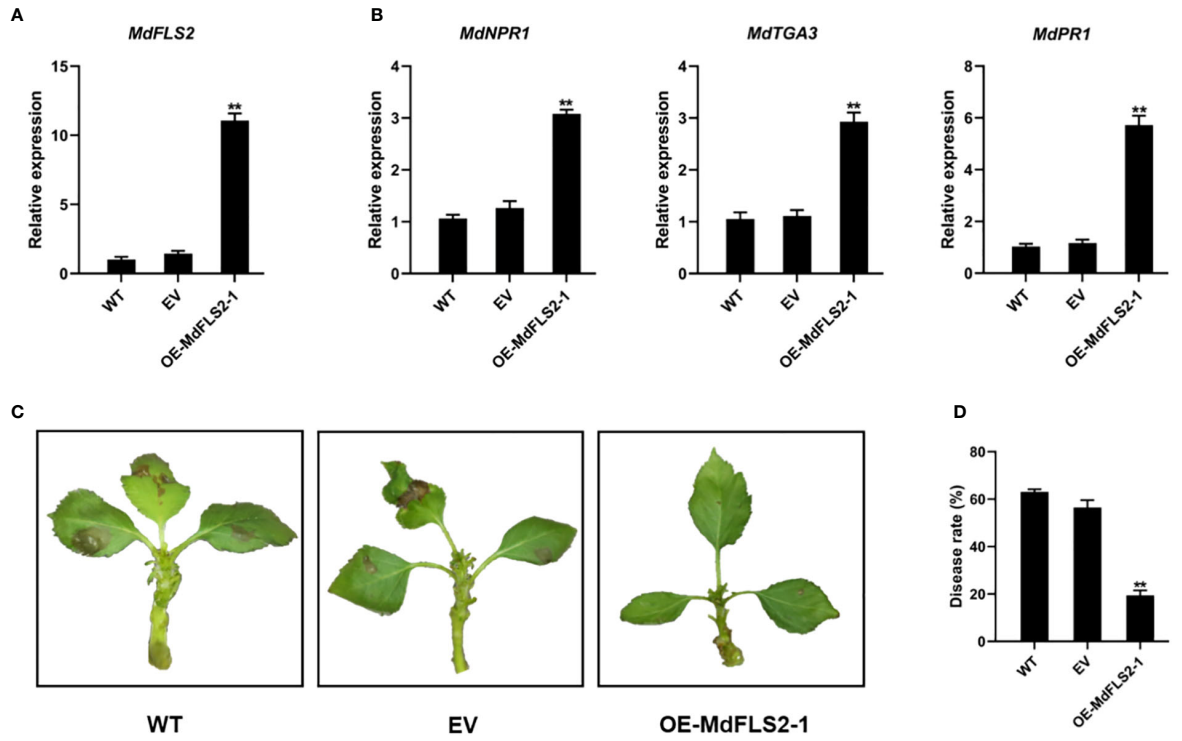


FIGURE 6
MdFLS2-1 positively regulates resistance of apple to AAAP. RT-qPCR analysis of the expression level of **(A)** MdFLS2 and **(B)** salicylate pathway marker genes MdNPR1, MdTGA3, and MdPR1 in leaves of WT, EV, and OE-MdFLS2-1 plants. Asterisks denote Student's t-tests, significance: **P < 0.01. The mean ± SD of biological triplicates were taken for every value. **(C)** Symptoms of infection in leaves of WT, EV, and OE-MdFLS2-1 plants at 48 h after AAAP inoculation. **(D)** Incidence statistics of leaves of WT, EV, and OE-MdFLS2-1 plants at 48 h after AAAP inoculation. Values are the means ± SD of three biological replicates (Student's t-tests, **P < 0.01).

expression of the *MdFLS2* allele, we searched the promoter region of the *MdFLS2* allele using whole-genome resequencing data of HF. Based on resequencing data and *MdFLS2* allele promoter cloning, we identified an 85-bp TRIM-like sequence insertion in the promoter of *MdFLS2-1*, which was located 235-bp upstream of the translation start codon (Figure 7A). The TRIM-like sequence consists of two almost identical 32-bp TDRs and a 21-bp intermediate sequence (Figure 7A). The TRIM-like sequence is flanked by 3-bp target site duplication that was generated after insertion. The two TDRs of the TRIM-like element have two base differences, named TDR1 and TDR2, respectively. In the promoter region of *MdFLS2-2*, the TRIM-like sequence undergoes truncation, retaining only a TDR2 sequence (Figure 7A).

Several studies have reported that transposon insertion in the promoter region can provide enhancer activity and thus increase the expression level of genes (Wei and Cao, 2016). *MdFLS2* alleles with intact transposon insertion in the promoter region had higher expression levels, which is consistent with the pattern of transposon insertions enhancing gene expression. To determine whether the insertion of the TRIM-like element provided enhancer activity, we ligated an 85-bp TRIM-like sequence inserted in the promoter of *MdFLS2-1* to the 35S minimal promoter and performed transient experiments in leaves of *N. benthamiana* to assess the effect of the TRIM-like element on the expression of firefly luciferase gene and β-glucuronidase gene. The results revealed that the TRIM-like sequence led to a significant increase in reporter gene expression, relative to the use of only a TDR2 sequence from the promoter of *MdFLS2-2*-containing structure (Figures 7B-D). This suggested that the insertion of the TRIM-like sequence in the promoter region of *MdFLS2-1* increased the expression level of the gene, causing differential expression of the *MdFLS2* allele in HF leaves. This natural variation in the promoter region of the *MdFLS2* allele resulted in its significantly higher expression level in HF than in the homozygote HFTH1.

4 Discussion

ASE analysis in heterozygous organisms provides an unprecedented method to compare the effects of *cis*-regulatory variants on gene expression in the same *trans*-regulatory background and cellular environment (Bao et al., 2019). Genome-wide ASE analysis was performed by RNA-seq of leaves of HF, an important disease-resistant cultivar, after

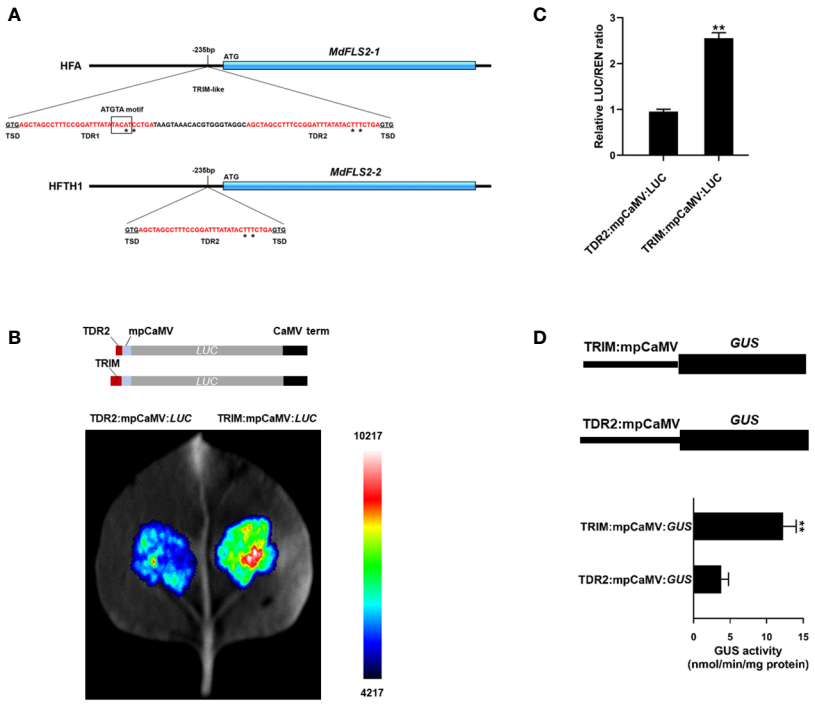


FIGURE 7
Insertion of the TRIM-like sequence affects the expression level of *MdFLS2* allele. **(A)** Molecular structure of *MdFLS2-1* from HF alternative genome (HFA) and of *MdFLS2-2* from HFTH1 reference genome with flanking sequence. The TRIM-like sequence contains the following sequence features: TSD (black underlined), TDRs (red font), and intermediate sequence (black font). Base differences between TDR1 and TDR2 are marked with asterisks. The black box indicates the ATGTA motif. **(B)** Transient expression analysis revealed that TRIM-like sequence insertion could significantly increase the expression level of luciferase. Upper panel represents the TDR2:mpCaMV:*LUC* (up) and TRIM:mpCaMV:*LUC* (down) constructs, both containing the minimal promoter from the cauliflower mosaic virus (mpCaMV), luciferase ORF, and cauliflower mosaic virus terminator (CaMV term), to which the TDR2 sequence and TRIM-like sequence are respectively added. Lower panel indicates luciferase assay of Nicotiana benthamiana leaves 72 h after infection with Agrobacterium strains containing TDR2:mpCaMV:LUC (left) and TRIM:mpCaMV:*LUC* (right). **(C)** The relative ratio of LUC/REN for both TDR2:mpCaMV:*LUC* and TRIM:mpCaMV:*LUC*. Values are the means ± SD of three biological replicates (Student's t-tests, $^{**}P < 0.01$). **(D)** GUS activity in the transiently transformed N. benthamiana leaves with constructs TRIM:mpCaMV:GUS and TDR2:mpCaMV:GUS under normal conditions. The mean ± SD of biological triplicates were taken for every value. Asterisks denote significant differences by Student's t-test ($^{**}P < 0.01$).

AAAP infection and mock infection. A large number of ASE genes were identified. These genes form a validated dataset to rapidly identify the *cis*-regulatory mechanisms affecting the expression levels of these alleles by comparing genomes.

4.1 *MdFLS2* is involved in the resistance of apple leaves against AAAP

PTI is an important mode of plant immunity (Bigeard et al., 2015). Some receptor kinases are involved in the resistance of plants to fungi by linking hormone signals (Tang et al., 2017). By comparative transcriptomic analysis, we identified an important receptor kinase, *MdFLS2*, that exhibited significant differences in expression between HF and HFTH1 after AAAP inoculation. There was a consistent bias of ASE for *MdFLS2* in HF after AAAP infection and mock inoculation, and the expression level of the HF-derived allele *MdFLS2-1* was significantly higher than that of the HFTH1-derived allele *MdFLS2-2*. Previous studies have reported that *MdFLS2-2* is involved in the defense response of apple to fungi by enhancing the expression levels of the genes related to salicylic acid pathway. In the transcriptome data, we observed an association of *MdFLS2* with the genes related to salicylic acid pathway, with their higher expression levels in the AAAP-resistant HF compared with AAAP-susceptible HFTH1 at the early stage of AAAP infection; the genes down-regulated at a later stage. Related studies have indicated that an intact salicylic acid pathway is essential in the defense of *Solanum tuberosum* against *Alternaria solani* (Brouwer et al., 2020). Studies on the defense of *Chrysanthemum morifolium* against *Alternaria* (Liu et al., 2020; Zhao et al., 2020) have reported that the salicylic acid pathway plays an important role in the early response of plants against pathogenic attacks. Transient overexpression of *MdFLS2-1* could improve resistance of apple leaves to AAAP (Figures 6C, D). At the same time, we observed that overexpression of *MdFLS2-1* was accompanied by increased

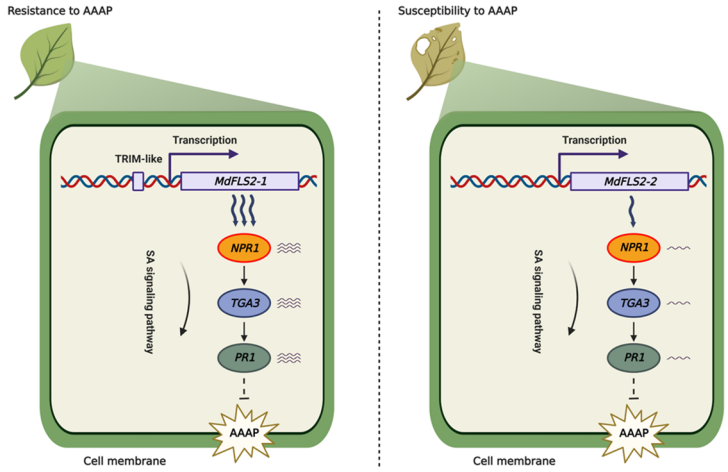


FIGURE 8
Model for the role of a TRIM-like sequence insertion in the *MdFLS2-1* promoter region in regulating resistance of apple to Alternaria blotch disease. *MdFLS2* allele is involved in the resistance of apple to Alternaria blotch disease *via* increasing the expression level of genes related to salicylic acid pathway (*NPR1, TGA3,* and *PR1*). Insertion of a TRIM-like sequence into the *MdFLS2-1* promoter region enhances *MdFLS2-1* expression to improve resistance of apple to Alternaria blotch disease. Arrow: activation; Bar: repression.

expression levels of genes related to salicylic acid pathway (Figures 6A, B). These results predicted a conserved and active role of salicylic acid pathway in the defense of various plants against various pathogenic variants of *Alternaria*. On the other hand, *MdFLS2-1* and *MdFLS2-2* could regulate the level of salicylic acid signaling in apple leaves during resistance to fungi.

Studies on *Arabidopsis* indicated that AtFLS2 receptor kinase can function in complex with various proteins, including AtACD6, which can form a complex with AtFLS2 to enhance salicylic acid signaling (Tateda et al., 2014), and AtACD6 and AtFLS2 can promote the activation of salicylic acid signaling even in the absence of PAMPs (Tateda et al., 2015). Significant differences in *AtACD6* homolog *MdACD6* were reflected at 12 HPI in susceptible varieties, which is consistent with the expression trend of genes related to salicylic acid pathway (Figure 5C). We speculated that a similar mechanism may exist in apple, where *MdFLS2* regulates salicylic acid signaling through *MdACD6*. However, further studies are needed to elucidate this mechanism.

4.2 A TRIM-like insertion in *MdFLS2* allele promoter causes its ASE

It is widely reported that insertion of transposable elements in the promoter region affects expression level of the genes (Mao et al., 2015; Xia et al., 2017; Niu et al., 2022). In this study, we identified a TRIM-like sequence inserted in the promoter region of *MdFLS2-1*, which was experimentally demonstrated to have enhancer activity. In contrast, promoter region of the other allele *MdFLS2-2* only had a TDR2 sequence of the TRIM-like element; thus, the expression level of *MdFLS2-2* was significantly lower than that of *MdFLS2-1*. Based on these results, we developed a model to explain the difference in the sensitivity of HF and HFTH1 to AAAP. After AAAP infection, HF containing *MdFLS2-1* exhibited high expression levels of genes related to salicylic acid pathway, which in turn led to resistance of HF to AAAP. In contrast, HFTH1 with lower expression levels of *MdFLS2-2* exhibited susceptibility to AAAP (Figure 8).

Some studies have reported that insertion of transposable elements can provide effective binding sites for transcription factors (Sahebi et al., 2018). An ATGTA motif is present in the TDR1 sequence of the TRIM-like element. Due to mutations during evolution, the ATGTA motif is not present in the TDR2 sequence. Previous studies have identified the ATGTA motif as a binding site for the *MdEIL1* transcription factor in apple (An et al., 2018). Interestingly, *AtEIL1*, the homolog of *MdEIL1* in *Arabidopsis*, can bind to the promoter region of *AtFLS2* to regulate its expression level (Boutrot et al., 2010). In apples, this regulatory mechanism is expected to be further revealed in future studies. Overall, our study provided new insights into the molecular mechanism of resistance to AAAP by apple and provided a reference for breeding with disease resistance 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: National Center for Biotechnology Information (NCBI) BioProject database under BioProject ID: PRJNA896728 and PRJNA897033.

Author contributions

LZ, PC and ZL initiated the project and designed the study. ZL, KL, CJ, AY, JY performed biological experiments. ZL performed data analysis and wrote the manuscript. XH, CZ optimized experimental protocol. LZ, and PC revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-RIP-02) and the National Key Research and Development Program of China (2019YFD1001403).

Acknowledgments

We thank Dr. Zhihong Zhang from Shenyang Agricultural University for providing tissue-cultured GL-3 plants, and Dr. Jia-Long Yao from The New Zealand Institute for Plant & Food Research Limited for providing the pGreenII0800-LUC vector.

Conflict of interest

The handling editor LC declared a shared affiliation with the Author/s ZL & LZ at the time of review Chinese Academy of Agricultural Sciences CAAS, Beijing, China

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1090621/full#supplementary-material

References

An, J. P., Wang, X. F., Li, Y. Y., Song, L. Q., Zhao, L. L., You, C. X., et al. (2018). EIN3-LIKE1, MYB1, and ETHYLENE RESPONSE FACTOR3 act in a regulatory loop that synergistically modulates ethylene biosynthesis and anthocyanin accumulation. *Plant Physiol.* 178, 808–823. doi: 10.1104/pp.18.00068

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in arabidopsis innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a

Bai, T. T., Xie, W. B., Zhou, P. P., Wu, Z. L., Xiao, W. C., Zhou, L., et al. (2013). Transcriptome and expression profile analysis of highly resistant and susceptible banana roots challenged with fusarium oxysporum f. sp. cubense tropical race 4. *PLoS One* 8, e73945. doi: 10.1371/journal.pone.0073945

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., and Wendel, J. F. (2019). Unraveling cis and trans regulatory evolution during cotton domestication. *Nat. Commun.* 10, 5399. doi: 10.1038/s41467-019-13386-w

Bigeard, J., Colcombet, J., and Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Mol. Plant* 8, 521–539. doi: 10.1016/j.molp.2014.12.022

Boutrot, F., Segonzac, C., Chang, K. N., Qiao, H., Ecker, J. R., Zipfel, C., et al. (2010). Direct transcriptional control of the arabidopsis immune receptor FLS2 by the ethylene-dependent transcription factors EIN3 and EIL1. *Proc. Natl. Acad. Sci. United States America* 107, 14502–14507. doi: 10.1073/pnas.1003347107

Brouard, J. S., Schenkel, F., Marete, A., and Bissonnette, N. (2019). The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J. Anim. Sci. Biotechnol.* 10. doi: 10.1186/s40104-019-0359-0

Brouwer, S. M., Odilbekov, F., Burra, D. D., Lenman, M., Hedley, P. E., Grenville-Briggs, L., et al. (2020). Intact salicylic acid signalling is required for potato defence against the necrotrophic fungus alternaria solani. *Plant Mol. Biol.* 104, 1–19. doi: 10.1007/s11103-020-01019-6

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen-Hui, Q., Xian-Yan, Z., Han, J., Hai-Tao, L., Yong-Xu, W., Da-Gang, H., et al. (2018). Molecular cloning and functional identification of an apple flagellin receptor MdFLS2 gene. *J. Integr. Agric.* 17, 2694–2703. doi: 10.1016/S2095-3119(18)62009-X

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Danna, C. H., Millet, Y. A., Koller, T., Han, S. W., Bent, A. F., Ronald, P. C., et al. (2011). The arabidopsis flagellin receptor FLS2 mediates the perception of xanthomonas Ax21 secreted peptides. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9286–9291. doi: 10.1073/pnas.1106366108

David, L., Kang, J., and Chen, S. (2019). Targeted metabolomics of plant hormones and redox metabolites in stomatal immunity. *Methods Mol. Biol.* 2085, 79–92. doi: 10.1007/978-1-0716-0142-6_6

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812

Filajdic, N., and Sutton, T. B. (1991). Identification and distribution of alternaria mali on apples in north Carolina and susceptibility of different varieties of apples to alternaria blotch. *Plant Dis.* 75, 1045–1048. doi: 10.1094/PD-75-1045

Gao, Z. Y., Li, H., Yang, X. Y., Yang, P. F., Chen, J. M., and Shi, T. (2021). Biased allelic expression in tissues of F1 hybrids between tropical and temperate lotus (Nelumbo nuicfera). *Plant Mol. Biol.* 106, 207–220. doi: 10.1007/s11103-021-01138-8

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in arabidopsis. *Mol. Cell* 5, 1003–1011. doi: 10.1016/s1097-2765(00)80265-8

He, X., Meng, H., Wang, H., He, P., Chang, Y., Wang, S., et al. (2022). Quantitative proteomic sequencing of f 1 hybrid populations reveals the function of sorbitol in apple resistance to botryosphaeria dothidea. *Hortic. Res.* 9, uhac115. doi: 10.1093/hr/uhac115

Hill, M. S., Vande Zande, P., and Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* 22, 203–215. doi: 10.1038/s41576-020-00304-w

Hou, Y., Yu, X., Chen, W., Zhuang, W., Wang, S., Sun, C., et al. (2021). MdWRKY75e enhances resistance to alternaria alternata in malus domestica. *Hortic. Res.* 8, 225. doi: 10.1038/s41438-021-00701-0

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: betaâ glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6, 3901–3907. doi: 10.1002/j.1460-2075.1987.tb02730.x

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kong, Q., Sun, T., Qu, N., Ma, J., Li, M., Cheng, Y. T., et al. (2016). Two redundant receptor-like cytoplasmic kinases function downstream of pattern recognition receptors to regulate activation of SA biosynthesis. *Plant Physiol.* 171, 1344–1354. doi: 10.1104/pp.15.01954

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Li, H., and Durbin, R.. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Liu, X., Liang, Y., Zhang, W., Hou, Y., Feng, S., Qiu, H., et al. (2018). MdFLS2 recognizes bacterial flagellin flg22 and enhances immune resistance against apple ring rot causal fungi in arabidopsis fls2 mutant. *Acta Hortic. Sin.* 45 (5), 827–844. doi: 16420/j.issn.0513-353x.2017-0687

Liu, Y., Xin, J., Liu, L., Song, A., Guan, Z., Fang, W., et al. (2020). A temporal gene expression map of chrysanthemum leaves infected with alternaria alternata reveals different stages of defense mechanisms. *Hortic. Res.* 7, 23. doi: 10.1038/s41438-020-0245-0

Liu, Y., Yang, T., Lin, Z., Gu, B., Xing, C., Zhao, L., et al. (2019). A WRKY transcription factor PbrWRKY53 from pyrus betulaefolia is involved in drought tolerance and AsA accumulation. *Plant Biotechnol. J.* 17, 1770–1787. doi: 10.1111/pbi.13099

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

Lu, D., Lin, W., Gao, X., Wu, S., Cheng, C., Avila, J., et al. (2011). Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity. *Science* 332, 1439–1442. doi: 10.1126/science.1204903

Mao, H. D., Wang, H. W., Liu, S. X., Li, Z., Yang, X. H., Yan, J. B., et al. (2015). A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* 6 8326. doi: 10.1038/ncomms9326

Meng, D., Li, C., Park, H. J., Gonzalez, J., Wang, J., Dandekar, A. M., et al. (2018). Sorbitol modulates resistance to alternaria alternata by regulating the expression of an NLR resistance gene in apple. *Plant Cell* 30, 1562–1581. doi: 10.1105/tpc.18.00231

Meng, X., and Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annu. Rev. Phytopathol.* 51, 245–266. doi: 10.1146/annurev-phyto-082712-102314

Moriya, S., Terakami, S., Okada, K., Shimizu, T., Adachi, Y., Katayose, Y., et al. (2019). Identification of candidate genes responsible for the susceptibility of apple (Malus x domestica borkh.) to alternaria blotch. *BMC Plant Biol.* 19, 132. doi: 10.1186/s12870-019-1737-7

Ngou, B. P. M., Ahn, H. K., Ding, P., and Jones, J. D. G. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* 592, 110–115. doi: 10.1038/s41586-021-03315-7

Ngou, B. P. M., Jones, J. D. G., and Ding, P. (2022). Plant immune networks. *Trends Plant Sci.* 27, 255–273. doi: 10.1016/j.tplants.2021.08.012

Niu, C., Jiang, L., Cao, F., Liu, C., Guo, J., Zhang, Z., et al. (2022). Methylation of a MITE insertion in the MdRFNR1-1 promoter is positively associated with its allelic expression in apple in response to drought stress. *Plant Cell* 34, 3983–4006. doi: 10.1093/plcell/koac220

Nomura, K., Mecey, C., Lee, Y. N., Imboden, L. A., Chang, J. H., and He, S. Y. (2011). Effector-triggered immunity blocks pathogen degradation of an immunity-associated vesicle traffic regulator in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10774–10779. doi: 10.1073/pnas.1103338108

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., et al. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37, e123. doi: 10.1093/nar/gkp596

Pavy, N., Boyle, B., Nelson, C., Paule, C., Giguere, I., Caron, S., et al. (2008). Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* 180, 766–786. doi: 10.1111/j.1469-8137.2008.02615.x

Reim, S., Rohr, A.-D., Winkelmann, T., Weiß, S., Liu, B., Beerhues, L., et al. (2019). Genes involved in stress response and especially in phytoalexin biosynthesis are upregulated in four malus genotypes in response to apple replant disease. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01724

Ren, J., Mao, J., Zuo, C., Calderón-Urrea, A., Dawuda, M. M., Zhao, X., et al. (2017). Significant and unique changes in phosphorylation levels of four phosphoproteins in two apple rootstock genotypes under drought stress. *Mol. Genet. Genomics* 292, 1307–1322. doi: 10.1007/s00438-017-1348-7

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

Sahebi, M., Hanafi, M. M., Van Wijnen, A. J., Rice, D., Raffi, M. Y., Azizi, P., et al. (2018). Contribution of transposable elements in the plant's genome. *Gene* 665, 155–166. doi: 10.1016/j.gene.2018.04.050

Sarowar, S., Alam, S. T., Makandar, R., Lee, H., Trick, H. N., Dong, Y., et al. (2019). Targeting the pattern-triggered immunity pathway to enhance resistance to fusarium graminearum. *Mol. Plant Pathol.* 20, 626–640. doi: 10.1111/mpp.12781

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5653–5658. doi: 10.1073/pnas.1820513116

Shi, Q., Febres, V. J., Jones, J. B., and Moore, G. A. (2016). A survey of FLS2 genes from multiple citrus species identifies candidates for enhancing disease resistance to xanthomonas citri ssp. citri. *Hortic. Res.* 3, 16022. doi: 10.1038/hortres.2016.22

Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., et al. (2020). Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* 52, 1423–1432. doi: 10.1038/s41588-020-00723-9

Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., et al. (2013). Structural basis for flg22-induced activation of the arabidopsis FLS2-BAK1 immune complex. *Science* 342, 624–628. doi: 10.1126/science.1243825

Tang, D., Wang, G., and Zhou, J. M. (2017). Receptor kinases in plant-pathogen interactions: More than pattern recognition. *Plant Cell* 29, 618–637. doi: 10.1105/tpc.16.00891

Tateda, C., Zhang, Z. Q., and Greenberg, J. T. (2015). Linking pattern recognition and salicylic acid responses in arabidopsis through ACCELERATED CELL DEATH6 and receptors. *Plant Signaling Behav.* 10. doi: 10.1080/15592324.2015.1010912

Tateda, C., Zhang, Z., Shrestha, J., Jelenska, J., Chinchilla, D., and Greenberg, J. T. (2014). Salicylic acid regulates arabidopsis microbial pattern receptor kinase levels and signaling. *Plant Cell* 26, 4171–4187. doi: 10.1105/tpc.114.131938

Tian, Y., Thrimawithana, A., Ding, T., Guo, J., Gleave, A., Chagne, D., et al. (2022). Transposon insertions regulate genome-wide allele-specific expression and underpin flower colour variations in apple (Malus spp.). *Plant Biotechnol. J.* 20, 1285–1297. doi: 10.1111/pbi.13806

Turra, D., Segorbe, D., and Di Pietro, A. (2014). Protein kinases in plant-pathogenic fungi: Conserved regulators of infection. *Annu. Rev. Phytopathol.* 52, 267–288. doi: 10.1146/annurev-phyto-102313-050143

Wang, H., Dong, B., Jiang, J., Fang, W., Guan, Z., Liao, Y., et al. (2014). Characterization of *in vitro* haploid and doubled haploid chrysanthemum morifolium plants *via* unfertilized ovule culture for phenotypical traits and DNA methylation pattern. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00738

Wei, L., and Cao, X. (2016). The effect of transposable elements on phenotypic variation: insights from plants to humans. *Sci. China Life Sci.* 59, 24–37. doi: 10.1007/s11427-015-4993-2

Xia, C., Zhang, L., Zou, C., Gu, Y., Duan, J., Zhao, G., et al. (2017). A TRIM insertion in the promoter of Ms2 causes male sterility in wheat. *Nat. Commun.* 8, 15407. doi: 10.1038/ncomms15407

Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., et al. (2021). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 592, 105–109. doi: 10.1038/s41586-021-03316-6

Zhang, X., Chen, S., Shi, L., Gong, D., Zhang, S., Zhao, Q., et al. (2021). Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant camellia sinensis. *Nat. Genet.* 53, 1250–1259. doi: 10.1038/s41588-021-00895-y

Zhang, M., Huang, S., Gao, Y., Fu, W., Qu, G., Zhao, Y., et al. (2020a). Fine mapping of a leaf flattening gene bralcm through BSR-seq in Chinese cabbage (Brassica rapa l. ssp. pekinensis). *Sci. Rep.* 10, 13924. doi: 10.1038/s41598-020-70975-2

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* 10, 1494. doi: 10.1038/s41467-019-09518-x

Zhang, Q., Ma, C., Zhang, Y., Gu, Z., Li, W., Duan, X., et al. (2018). A single-nucleotide polymorphism in the promoter of a hairpin RNA contributes to alternaria alternata leaf spot resistance in apple (Malus x domestica). *Plant Cell* 30, 1924–1942. doi: 10.1105/tpc.18.00042

Zhang, N., Pombo, M. A., Rosli, H. G., and Martin, G. B. (2020b). Tomato wall-associated kinase SlWak1 depends on Fls2/Fls3 to promote apoplastic immune responses to pseudomonas syringae. *Plant Physiol.* 183, 1869–1882. doi: 10.1104/pp.20.00144

Zhang, C. X., Tian, Y., and Cong, P. H. (2015). Proteome analysis of pathogen-responsive proteins from apple leaves induced by the alternaria blotch alternaria alternata. *PLoS One* 10, e0122233. doi: 10.1371/journal.pone.0122233

Zhao, X., Song, L., Jiang, L., Zhu, Y., Gao, Q., Wang, D., et al. (2020). The integration of transcriptomic and transgenic analyses reveals the involvement of the SA response pathway in the defense of chrysanthemum against the necrotrophic fungus alternaria sp. *Hortic. Res.* 7, 80. doi: 10.1038/s41438-020-0297-1

Zhu, L., Ni, W., Liu, S., Cai, B., Xing, H., and Wang, S. (2017). Transcriptomics analysis of apple leaves in response to alternaria alternata apple pathotype infection. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00022

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. *Curr. Opin. Immunol.* 20, 10–16. doi: 10.1016/j.coi.2007.11.003

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jones, J. D., Felix, G., et al. (2004). Bacterial disease resistance in arabidopsis through flagellin perception. *Nature* 428, 764–767. doi: 10.1038/nature02485

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fpls.2023.1132017

OPEN ACCESS

EDITED BY
Guo-Fei Tan,
Guizhou Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Xun Wu,
Guizou Drought Grain Sorghum Research Institute, China
Pengcheng Li,
Yangzhou University, China
Huiyong Li,
Henan Academy of Agricultural Sciences, China

*CORRESPONDENCE
Jie Guo
nxgj1115326@sxau.edu.cn
Jiafa Chen
chenjiafa@henau.edu.cn
Chunhui Li
lichunhui@caas.cn

†These authors have contributed equally to this work

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

RECEIVED 26 December 2022
ACCEPTED 13 January 2023
PUBLISHED 07 February 2023

CITATION
Wang K, Zhang Z, Sha X, Yu P, Li Y, Zhang D, Liu X, He G, Li Y, Wang T, Guo J, Chen J and Li C (2023) Identification of a new QTL underlying seminal root number in a maize-teosinte population. *Front. Plant Sci.* 14:1132017. doi: 10.3389/fpls.2023.1132017

Identification of a new QTL underlying seminal root number in a maize-teosinte population

Kailiang Wang[1†], Zhen Zhang[2†], XiaoQian Sha[3], Peng Yu[4,5], Yongxiang Li[3], Dengfeng Zhang[3], Xuyang Liu[3], Guanhua He[3], Yu Li[3], Tianyu Wang[3], Jie Guo[1*], Jiafa Chen[2*] and Chunhui Li[3*]

[1]College of Agriculture, Shanxi Agricultural University, Jinzhong, China, [2]College of Life Sciences, Henan Agricultural University, Zhengzhou, China, [3]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [4]Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany, [5]Emmy Noether Group Root Functional Biology, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany

Seminal roots play an important role in acquisition of water and nutrients by maize seedlings. Compared with its teosinte ancestor, maize underwent a change in seminal root number (SRN). Although several key genes controlling SRN have been cloned, identification and utilization of new genes from teosinte would be useful for improving maize root architecture. In this study, a maize-teosinte BC_2F_6 population containing 206 individuals genotyped by resequencing was used to conduct high-resolution quantitative trait locus (QTL) mapping of SRN. A new major QTL on chromosome 7 (*qSRN7*) was identified. Differentially expressed genes (DEGs) based on RNA-Seq were identified between two inbred lines with no SRN and multiple SRN at two periods of seminal roots primordia formation. A total of 116 DEGs detected in at least one period were identified within the *qSRN7* interval. Three DEGs (Zm00001d021572, Zm00001d021579 and Zm00001d021861) associated with SRN were identified through regional association mapping. When compared with reported domestication-related selective sweeps, Zm00001d021572 was selected during maize domestication. Our findings provide important insights into the genetic basis of SRN and identify a promising candidate gene for further studies on SRN.

KEYWORDS
domestication, RNA-Seq, root development, *Zea mays*, QTL

1 Introduction

Maize (*Zea mays* ssp. *mays* L.) is one of the most important crops worldwide. It was domesticated ~9,000 years ago from Mexican lowland annual teosinte (*Zea mays* ssp. *parviglumis*). Maize has several major morphological differences from its ancestor (Matsuoka et al., 2002). In regard to aboveground morphological traits, teosinte has multiple tillers and lateral branches, small ears and few seeds with hard cupulate fruit cases, whereas maize is usually unbranched and has few tillers, but large ears with many rows

of seeds that are not covered by hard cupulate fruit cases (Doebley, 2004). For underground root traits, Burton et al. (2013) documented that on average, the number of seminal roots in teosinte is about 0.5, whereas the number in cultivated maize varies from 0 to 8. Roots play crucial functions such as anchoring the plant to the ground, acquiring resources from the soil, and providing mechanical support to the stem. A root ideotype would be resilient against biotic and abiotic stress (Lynch, 2019). The maize root system is usually composed of one embryonic primary root, several seminal roots, and many crown, brace and lateral roots (Hochholdinger, 2009). The seminal root is essential for seedling survival and development during the first 2-3 weeks of seedling growth (Hochholdinger et al., 2004; Lynch, 2013). Seminal root primordia are initiated at the scutellum node about 25 days post pollination (Erdelska and Vidovencova, 1993). Previous studies reported that seminal roots penetrate the soil earlier than postembryonic roots (Weaver, 1926), and that increased seminal root number (SRN) not only contributed to total phosphorus and nitrogen uptake in the first 25 days of seedlings growth (Perkins and Lynch, 2021), but also enhanced drought tolerance in maize seedlings (Guo et al., 2020).

Variation in SRN is typically quantitative. Based on linkage mapping, previous studies identified more than 54 QTLs in different populations (Guo et al., 2018). For example, Zhu et al. (2006) used a recombinant inbred line population derived from a cross between B73 and Mo17 to detect six QTLs for SRN under high and low phosphorus conditions. Pestsova et al. (2016) identified a QTL for SRN on chromosome 8 under conditions of nitrogen deficiency in a doubled haploid line (DHL) population. Salvi et al. (2016) used an introgression library (IL) derived from landrace Gaspe as the donor parent and B73 as the recipient parent, to detect three QTLs. Wang et al. (2019) conducted a Genome-wide association study (GWAS) of SRN in 297 maize inbred lines, and identified QTLs related to SRN on chromosomes 2 and 8. Ma et al. (2021) identified a quantitative trait nucleotide (QTN) associated with SRN in an association panel containing 362 inbred lines. Although many QTLs/QTNs for SRN were identified using modern maize inbred lines, an understanding of the genetic basis of SRN and its evolution remains inadequate.

Few genes underlying SRN have been cloned following genetic analysis of mutants. The *RTCS* gene encodes a lateral organ boundaries domain (LBD) protein involving in seminal and crown root primordium formation and maintenance (Hetz et al., 1996; Taramino et al., 2007). The *RUM1* gene encodes a monocot-specific Aux/IAA protein that controls formation of seminal and lateral roots at the position of the primary root (Woll et al., 2005; von Behrens et al., 2011). The *BIGE1* gene cloned following isolation of a *bige1*-UMU5 mutant, encodes a multidrug and toxin extrusion (MATE) transporter, and regulates the timing and rate of initiation of seminal roots and plant lateral organs in both seed and plant development (Suzuki et al., 2015). Previous studies reported that elite alleles of several key genes controlling morphological traits derived from teosinte play important roles in increasing grain yield and enhancing biotic and abiotic stress (Tian et al., 2019; Wang et al., 2021; Barnes et al., 2022; Chen et al., 2022). Hence, identification of candidate genes and their alleles underlying SRN from teosinte might facilitate the use of novel genetic diversity in maize breeding.

In this study, we performed high-resolution QTL mapping for SRN, using a BC_2F_6 population (hereafter, TP population) derived from a cross between teosinte and inbred line PH4CV and genotyped by resequencing. Two inbred lines with large differences in SRN from an association panel were used for RNA-Seq analysis to identify DEGs related to the formation and development of SRN. Regional association mapping of SRN was carried out in an association panel containing 351 inbred lines. Combining the results from the two approaches, candidate genes underlying SRN were identified, and would provide promising genetic resources for improving maize root architecture.

2 Materials and methods

2.1 Plant materials

A TP population containing 206 individuals was derived from a single F_1 seed from a cross between inbred line PH4CV as female parent and *Zea mays ssp. mexicana* (CIMMYT accession number 249743). In addition, an association panel including 351 elite inbred lines derived from a large association mapping population (Li et al., 2022) was used to conduct regional association mapping.

2.2 Genotyping and recombination bin map construction

Genomic DNA from individuals in the TP population was extracted by the CTAB method. Genome sequencing libraries were sequenced using the Illumina Hiseq X platform, yielding a total of 1.18 Tb raw sequences (average depth 2.5×) with 150-bp paired-end reads. After checking and filtering sequence quality, the remaining sequences were mapped to the B73 reference genome (B73_V4). A total of 62,217,878 single nucleotide polymorphisms (SNPs) were identified in the population. To obtain high-quality SNPs, we removed low-quality SNPs with missing rate >60%, minor allele frequency (MAF) > 0.175 or MAF <0.075, and heterozygosity rate >5%. Finally, 138,208 high-quality SNPs were used to construct a recombination bin map.

The R package "binmapr" was used to fix genotypes using a windows size set to 15. Redundant markers were removed by the "bin" function in IciMapping software. The genetic distances between the final 1,951 bin markers were determined using a Kosambi method in IciMapping software with the "map" function. The linkage map and recombination bin map were drawn in R software with the "LinkageMapView" package and "plot" function, respectively.

2.3 Phenotypic data collection and statistical analysis

To phenotype SRN, fifteen seeds of each line from the TP population and association panel were surface sterilized with 6% hypochlorite for 10 min, followed by three washes in distilled water, transferred to wet filter paper, rolled up and placed in plastic buckets

containing distilled water. The buckets were incubated in darkness at 25°C. Three days later, the buckets were transferred to a new incubator at 23°C\8h darkness and 28°C\16h light. The distilled water in the buckets was replaced for every two days. After five days, the number of seminal roots for 10 healthy plants of each line was manually counted and average values of SRN were subsequently used in analysis.

2.4 QTL mapping for SRN

QTL analysis of SRN based on linkage map and phenotypic values from the TP population was conducted using WinQTL Cartographer v2.5 software by composite interval mapping (CIM) (Wang et al., 2012), with a 1,000 permutations test at 95% confidence level to determine the optimal logarithm of odds (LOD) threshold values. A 1.5-LOD drop method was applied for defining the QTL confidence interval. Names were assigned to QTL following the nomenclature proposed by McCouch (1997), which combines the letter "q" for QTL, an abbreviation for the name of the trait, and a number for the chromosome.

2.5 RNA sequencing and data analysis

Inbred lines IA2132 and PHW30 from the association panel were grown in a culture room. The number of seminal roots was determined from the seminal root primordia, which are initiated at the scutellar node of kernel about 25 days after pollination (Erdelska and Vidovencova, 1993). Samples from each inbred line for RNA-Seq were collected at the kernels scutellar node at the 20^{th} and 30^{th} days after pollination, respectively. Four biological replicates for each line were collected at each sampling time. All samples were immediately frozen in liquid nitrogen and stored at -80°C for RNA isolation. RNA extraction and library preparation for each sample were performed by Novogene Corporation. Sixteen libraries were sequenced using the Illumina HiSeq 2000 platform (Illumina, CA, USA), and 150 bp paired-end reads were generated. The raw reads were filtered to remove sequencing adapters as well as low quality reads (base number of Qphred ≤ 20 accounting for more than 50% of the read length). HISAT2 software was used to compare clean reads quickly and accurately with the reference genome to obtain the locations of reads on the AGv4 reference genome (Mortazavi et al., 2008). The FPKM (expected number of Fragments Per Kilobase of transcript per Million base pair sequences) was used to estimate gene expression levels (Trapnell et al., 2010). Differential expression analysis between the two inbred lines was performed *via* the pairwise comparison algorithm DESeq (Anders et al., 2013). DEGs were screened according to the following criteria: |log2 foldchange (FC)| ≥ 1, false discovery rate (FDR), and an adjusted P-value < 0.05 (Robinson et al., 2010). Gene Ontology (GO) enrichment analysis of the DEGs was conducted to identify the enriched biological functions between the two genotypes (Yu et al., 2012).

2.6 Regional association mapping underlying QTL *qSRN7*

The genotype data from the association panel of 351 inbred lines in the *qSRN7* interval were extracted from the resequencing data of 1,604 maize inbred lines reported by Li et al. (2022). After filtering using missing rates < 20%, MAF >0.05, a total of 33,164 high-quality SNPs were obtained. Principal component analysis (PCA) and kinship (*K*) matrix were calculated using 43,252 SNPs identified by the 50K SNP chip in the association panel. A mixed linear model (MLM) with PCA and *K* matrix was used to conduct regional association mapping in TASSEL 5.0 software (Bradbury et al., 2007). To determine the significant threshold for regional association results, we estimated the number of independent SNPs by pruning SNPs in the PLINK (window size 100, step size 50 and $r^2 \geq 0.2$). After pruning, the number of independent SNPs in the *qSRN7* interval was determined to be 6,070. We then selected 1.65×10^{-4} (1/6070) as the threshold of association signals.

3 Results

3.1 Construction of a recombination bin map

The TP population was genotyped using whole genome resequencing technology. A total of 138,208 high-quality SNPs were identified in the TP population and used to determine bin markers. Finally, a total of 1,951 recombination bins were obtained by using the "binmapr" package in R software (Figure 1). The average physical interval of adjacent bins was 1,104.7 kb, with a maximum of 65,441.9

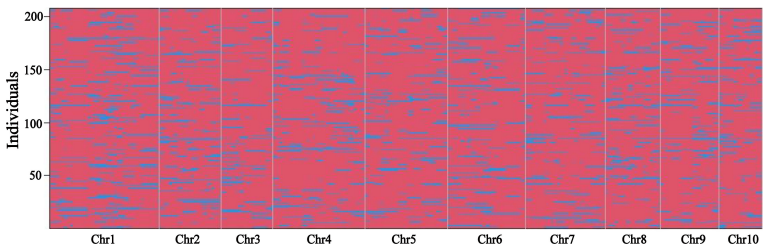


FIGURE 1
Recombination bin map of the TP population. Bin map consists of 1,951 markers. The physical position of markers is based on the B73 RefGen_v4 sequence. Red: PH4CV genotype; Blue: teosinte genotype.

kb and minimum of 0.4 kb. Bin markers were mapped to maize B73 RefGen_V4 to assess the quality and accuracy of the map. The scatter plot of physical position of bin markers on all 10 chromosomes aligned well with the B73 reference genome, which indicated good collinearity between the maize B73 reference genome and the bin markers. The genetic length of the linkage map constructed using the 1,951 bin markers (Supplementary Figure 1) was 1,142.1 cM with an average distance of 0.59 cM between adjacent markers. The number of markers per chromosome ranged from 102 (chromosome 9) to 329 (chromosome 4), with an average of about 195 bin markers per chromosome.

3.2 QTL mapping for SRN

We observed that the SRN showed a large phenotypic difference between the PH4CV and teosinte parents (Figure 2A). The average SRN of the female parent PH4CV was 3.2, whereas the teosinte parent had no seminal roots. In the TP population developed from the cross between PH4CV and teosinte, SRN ranged from 0.3 to 4.5, with a coefficient of variation (C.V.) of 26.3% (Figure 2B and Supplementary Table 1). Two QTLs were detected on chromosomes 1 and 7 (Figure 2C), explaining 6.18% and 6.25% of phenotypic variation, respectively. The known *RTCS* gene that has been cloned and regulates SRN, was located near the QTL peak on chromosome 1. The new QTL on chromosome 7 was named *qSRN7*, and a total of 345 candidate genes were located within the QTL interval.

3.3 Transcriptome sequencing analysis

We conducted a comparative transcriptome profiling of two inbred lines with extreme phenotypic differences in SRN. After filtration of low-quality sequences and adaptor sequences, a total of 670 million clean reads were obtained for the 16 RNA libraries. On average, 86.7% of reads were mapped uniquely to the B73 reference genome (AGPv4) (Supplementary Table 2). Pearson correlation coefficients among the different biological replicates of the same genotype varied from 0.92 to 0.99 (Supplementary Figure 2A), suggesting high quality of the replicates. In the 16 samples, 59.8% of expressed genes were expressed at low levels ($0 \leq FPKM1 < 1$), 37.2% were expressed at medium levels ($1 \leq FPKM < 60$), and 2.96% were expressed at high levels ($FPKM \geq 60$) (Supplementary Figure 2B). To determine which expressed genes correlated with SRN development, DEGs were analyzed between line IA2132 with no SRN and PHW30 with multiple SRN at both sampling times (Figure 3A). Totals of 6,682 (3,114 down-regulated and 3,568 up-regulated) and 8,620 DEGs (3,397 down-regulated and 5,223 up-regulated) genes were identified at the 20th and 30th days after pollination, respectively (Figure 3B), with 4,241 DEGs shared in

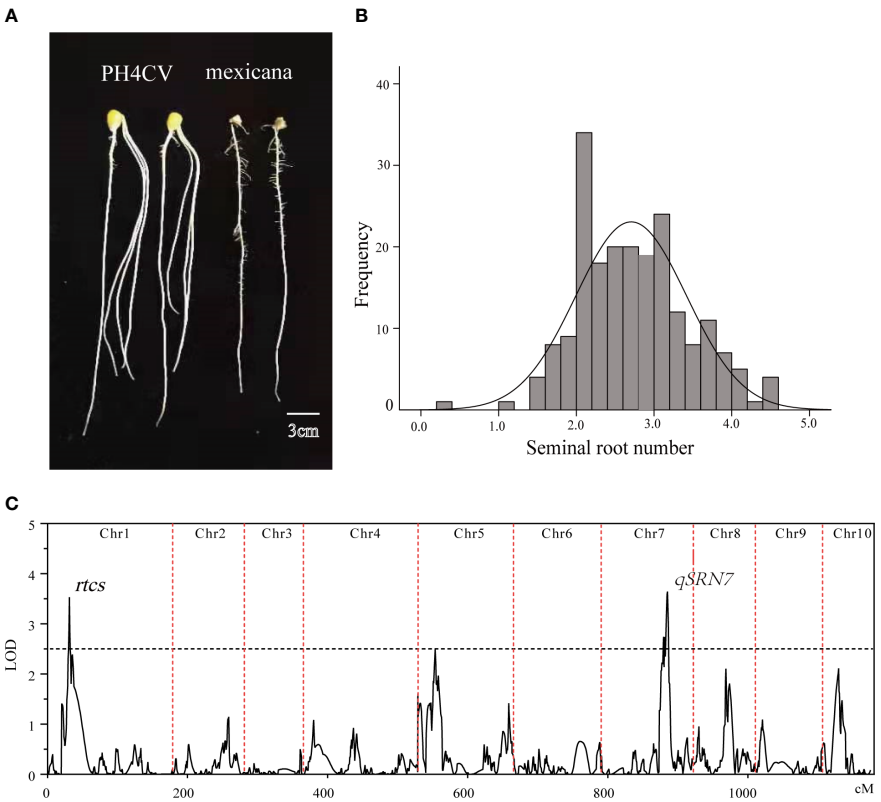


FIGURE 2
QTL mapping of SRN. **(A)** SRN in PH4CV and teosinte. **(B)** Frequency distribution of SRN in the TP population. **(C)** QTLs detected for SRN in the TP population.

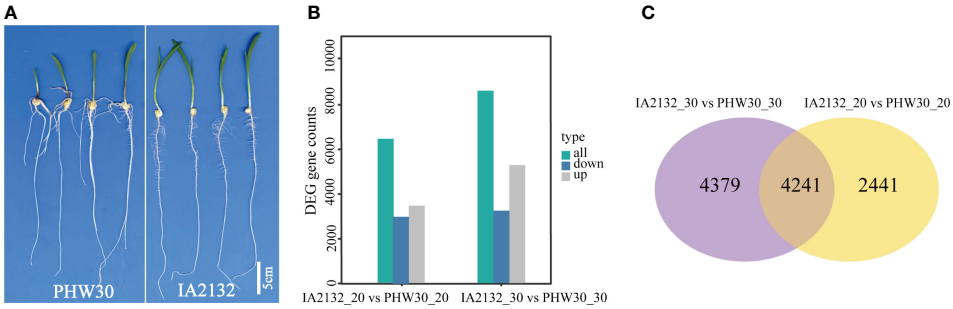


FIGURE 3
Transcriptome sequencing of two inbred lines with extreme difference in SRN. **(A)** SRN difference between PHW30 and IA2132. **(B)** DEGs between PHW30 and IA2132 at two sampling times. IA2132_20 vs PHW30_20 and IA2132_30 vs PHW30_30 represent DEGs at 20th and 30th days after pollination, respectively. **(C)** Venn diagram shows overlapping DEGs between IA2132_30 vs PHW30_30 and IA2132_20 vs PHW30_20.

both samplings (Figure 3C). When compared with the *qSRN7*, 116 DEGs detected at least once were located within the *qSRN7* interval (Supplementary Table 3).

Among the DEGs, we successfully identified *RTCS* (Zm00001d027679) and *BIGE1* (Zm00001d012883) at 30th days after pollination, both genes regulate initiation of seminal roots in maize. We also identified promising DEGs associated with root development and growth. For example, *ZmEXPB2* (Zm00001d029899) encoding an expansin-B4 protein, which increases primary root length (Zainab et al., 2021), was identified at both sampling times. Zm00001d029592 encoding a root hair defective (RHD3) protein, homologous to *Arabidopsis RHD3* involved in regulating cell expansion and normal root hair development (Zheng and Chen, 2011), was also identified at .both times. GO analysis of shared DEGs identified at both sampling times revealed enrichment of genes in protein homodimerization activity, protein heterodimerization activity, heme binding, and tetrapyrrole binding (Supplementary Figures 3A, B). Previous studies documented that the interaction between RTCS and its paralogue RTCS-LIKE (RTCL) could form heterodimerization to regulate shoot-borne root initiation in maize (Majer et al., 2012), and RUM1 and its homolog RUM1-LIKE1 (RUL1) regulating seminal and lateral root formation can homo and heterodimerize *in vivo* (Zhang et al., 2016). These results suggested that those DEGs might be considered important candidate genes regulating seminal root formation in maize.

3.4 Regional association mapping underlying QTL *qSRN7*

To identify candidate genes underlying *qSRN7*, we conducted regional association mapping in the association panel of 351 inbred

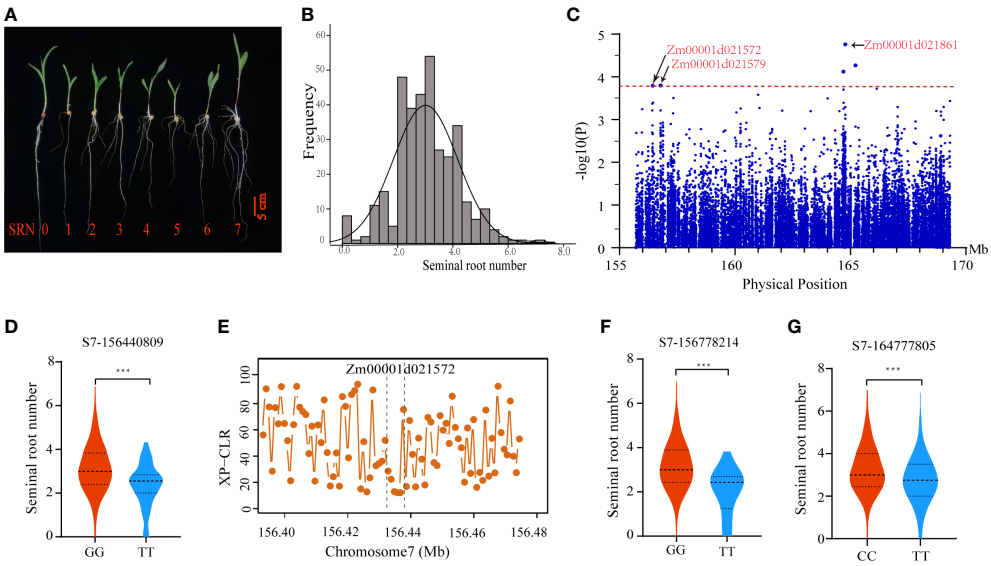


FIGURE 4
Identification of candidate genes underlying SRN. **(A)** SRN phenotypes of representative inbred lines in the association panel. **(B)** Frequency distribution of SRN in the association panel. **(C)** Manhattan plot of regional association mapping. **(D)** Phenotypic analysis of inbred lines with different haplotypes for associated SNPs (S7_156440809). **(E)** XP-CLR scores of Zm00001d021572 selected during maize domestication. The two dashed lines represent the physical position of Zm00001d021572. **(F, G)** Phenotypic analysis of inbred lines with different haplotypes for associated SNPs (S7_156778214 and S7_164777805). Center lines indicate the medians and lines above and below the median are 25th and 75th percentiles, and the whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. Statistical significance was determined using two-sided *t*-tests. ***, *P* <0.001.

lines. The panel showed large phenotypic differences in SRN ranging from 0 to 7 (Figures 4A, B and Supplementary Table 4). Using the MLM model in TASSEL 5.0 software, we performed an association analysis of SRN within the *qSRN7* interval. Five associated SNPs were detected at a threshold of $-\log10(P) > 3.78$ (Figure 4C). We detected five candidate genes based on the physical location of those associated SNPs, including Zm00001d021861, Zm00001d021857, Zm00001d021572, Zm000001d021579 and Zm00001d021874.

3.5 Identification of high-confidence candidate genes for SRN

To identify high-confidence candidate genes underlying *qSRN7*, we integrated DEGs obtained by RNA-Seq and regional association results. Three of the 5 candidate genes obtained by regional association analysis were differentially expressed between the inbred line with no SRN and the inbred line with multiple SRN; they were Zm00001d021572, Zm00001d021579 and Zm00001d021861. For Zm00001d021572 (encoding an anthocyanidin 3-O-glucosyltransferase), an associated SNP (S7_156440809) located within the UTR3 region, could form two haplotypes. Inbred lines with the GG haplotype had significantly more seminal roots than those with TT haplotype (Figure 4D). Morphologically, SRN of modern maize inbred lines differ substantially from their ancestors. To ascertain whether Zm00001d021572 underwent selection during maize domestication, we searched for this gene in selected regions by comparing maize and *parviglumis* using the XP-CLR method reported by Chen et al. (2022). We found that Zm00001d021572 was selected during maize domestication (Figure 4E). For Zm00001d021579 (encoding a DUF869 domain-containing family protein), an associated SNP (S7_156778214) located within the exon, could form two haplotypes. Inbred lines with TT haplotype had significantly more seminal roots than those with GG haplotype (Figure 4F). Zm00001d021861 encodes a vacuole membrane protein KMS1, which is involved in root development in *Arabidopsis* (Wang et al., 2011). An associated SNP (S7_164777805) located near the physical position of Zm00001d021861, could form two haplotypes. Inbred lines with CC haplotype had significantly more seminal roots than those with TT haplotype (Figure 4G). The expression data of these three candidate genes (Zm00001d021572, Zm00001d021579 and Zm00001d021861) in embryo, endosperm, seed, silks, tassel, cob, leaf, internode and root tissues were downloaded from MaizeGDB. The results showed that these genes had relatively high expression in the root tissues (Supplementary Figure 4). These results suggested that all three candidate genes might play important roles in regulating SRN in maize.

4 Discussion

Maize and teosinte show large differences in morphology and environmental adaptation (Doebley, 2004). For example, teosinte has many tillers and lateral branches, small ears and fewer seeds than modern cultivated maize, and several key genes controlling these morphologic differences between teosinte and modern maize have been cloned (Clark et al., 2006; Stitzer and Ross-Ibarra, 2018). For underground traits, teosinte has no or few seminal roots compared with modern cultivated maize (Burton et al., 2013). A similar pattern was reported for wild and cultivated barley (Grando and Ceccarelli, 1995). Thus domestication appears to have increased SRN in some crop species. This may be due to the growth of teosinte seedlings requiring the seeds to provide carbohydrates and nutrients before initiation of photosynthesis, resulting in teosinte needs having insufficient seed carbohydrates to guarantee the growth of the radicle and coleoptile. For modern maize, the larger seeds might have facilitated an increase in SRN through providing additional resources for early plant development and growth (Perkins and Lynch, 2021), and improved efficiency of water and nutrient acquisition (Hufford et al., 2012).

Climate change (e.g., heat, drought, floods and disease outbreaks) is adversely affecting crop yields worldwide. To facilitate adaptation to climate change, breeders will need to produce new varieties displaying both higher yield as well as improved adaptation to different environments. Hence, it is important to use more and new genetic resources to develop stress-resilient varieties for adaptation to extreme environments. Teosinte, the closest wild relative of maize is adapted to a diverse range of environments (Hufford et al., 2012), and has diverse alleles that are absent in modern maize (Mammadov et al., 2018). In this study, a maize-teosinte TP population was constructed and genotyped by resequencing, and used to conduct QTL mapping for SRN. Two major QTLs were detected. The QTL on chromosome 1 overlapped with physical location of the cloned *RTCS* gene controlling SRN, indicating the accuracy of QTL mapping based on this TP population. By comparing the physical positions of reported QTLs for SRN, we found that the other QTL-*qSRN7* on chromosome 7 was new. The teosinte allele reduced the number of seminal roots. The TP population could be used in the future to phenotype abiotic and biotic stress and disease traits to identify elite teosinte alleles.

RNA-Seq has become an effective technology to detect expressed genes related to development and growth traits in maize, such as kernel development (Li et al., 2022) and drought tolerance (Liu et al., 2020). To our knowledge, there have been few RNA-Seq studies on seminal root primordia, which ultimately determine the number of seminal roots. We used RNA-Seq to identify DEGs underlying SRN between two inbred lines with no and multiple SRN at two sampling dates. Totals of 6,682 and 8,620 DEGs were identified at the 20th and 30th days after pollination, respectively, with 4,241 DEGs shared between both stages. Genes *RTCS* and *RUM1* underlying SRN were identified previously. The *RTCS* gene encodes an LBD protein involved in auxin signal transduction (Taramino et al., 2007). Expression of *RTCS* is activated by binding of auxin response factor (ARF) 34 to LBD elements in the promoter (Xu et al., 2015). Among the 4,241 DEGs obtained in this study, a group of DEGs encoding LBD was detected, including Zm00001d038197, Zm00001d043036 and Zm00001d033347. Zm00001d043036 was reported to positively regulate lateral root formation in *Arabidopsis* (Cho et al., 2019). The *RUM1* gene encodes a canonical Aux/IAA protein that is a central regulator of auxin signaling. Interaction between RUM1 and

downstream ARF in the pericycle is involved in the initiation of seminal roots (von Behrens et al., 2011). In this study, many genes encoding plant root initiation and development ARF were identified among the 4,241 DEGs, such as Zm00001d031522, Zm00001d032683 and Zm00001d045026. These results provide valuable information for further studies of those DEGs and their effects on SRN.

One hundred and sixteen DEGs were detected within the *qSRN7* interval. Functional annotations indicated that many DEGs were involved in root development. For example, Zm00001d021861 encodes a vacuole membrane protein KMS1 and an *Arabi*dopsis KMS1 RNAi line exhibited shorter roots than the wild type (Wang et al., 2011). Zm00001d021745 encodes D-type cyclin, which is mainly expressed in CEI (cortex/endodermal initials) and CEID (CEI daughter) cells during early root development in *Arabidopsis* (Di Laurenzio et al., 1996; Wysocka-Diller et al., 2000; Sozzani et al., 2010; Yang et al., 2022). Zm00001d021724 encodes an F-box domain-containing protein, and overexpression of F-box protein gene *MAIF1* in rice promotes root growth and reduces abiotic stress tolerance (Yan et al., 2011). Thus, further investigation of DEGs located within the *qSRN7* interval could be helpful in identifying candidate genes underlying SRN in maize.

Previous studies suggested that regional association mapping was an effective method to identify candidate genes underlying agronomic traits. In this study, we used regional association analysis to identify SNPs significantly associated with SRN within the *qSRN7* interval (about 13.6 Mb). Five associated SNPs were identified. When combining associated SNPs and DEGs within the QTL, we identified three SNPs significantly associated with three DEGs (Zm00001d021572, Zm00001d021579 and Zm00001d021861). When genes selected during maize domestication reported by Chen et al. (2022) were compared with these three DEGs, we found that Zm00001d021572 underwent selection during domestication. The function of these three candidate genes will be validated by knockout and overexpression in future studies. Collectively, the combined use of RNA-Seq, regional association analysis and selective sweeps during domestication will be beneficial in identifying high-confidence candidate genes associated with SRN.

5 Conclusion

This study dissected the genetic basis of SRN in maize, through conventional QTL mapping in a maize-teosinte population genotyped by resequencing, RNA-Seq and regional association mapping. A new QTL underlying SRN was identified. A total of 4,241 DEGs shared in two sampling times were identified. Among those DEGs, we identified known genes controlling maize SRN and many candidate genes involved in root development and growth. Combining data on DEGs and associated SNPs obtained by regional association mapping within the *qSRN7* interval, three high-confidence candidate genes underlying SRN were identified. Of the three genes, Zm00001d021572 reported in a previous study was selected during maize domestication. The results of this study provide insight into the genetic basis of SRN and potential candidate genes for improving seminal root system in maize.

Data availability statement

The data presented in the study are deposited in the Genome Sequence Archive, accession number PRJCA013887.

Author contributions

CL, JC and JG conceived and designed the experiments. KW, ZZ and PY collected the samples to conduct resequencing and RNA-Seq. KW, ZZ and XS performed phenotypic measurement and conducted data analysis. KW, ZZ and CL wrote the manuscript. CL, JG and JC provided experimental support and participated in revising the manuscript. YXL, DZ, XL and GH participated in data analysis. TW and YL supervised the project. The authors read and approved the final manuscript.

Funding

This work was supported by grants from National Key Research and Development Program of China (2021YFD1200700), Natural Science Foundation of Henan (212300410046), Ministry of Science and Technology of China (2020YFE0202300), China Agriculture Research System (CARS-02-03) and the CAAS Innovation Program.

Acknowledgments

The authors would like to thank the team of Jiansheng Li and Xiaohong Yang, College of Agriculture, China Agricultural University, for providing XP-CLR data.

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.2023.1132017/full#supplementary-material

References

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., et al. (2013). Count-based differential expression analysis of RNA sequencing data using r and bioconductor. *Nat. Protoc.* 8, 1765–1786. doi: 10.1038/nprot.2013.099

Barnes, A. C., Rodríguez-Zapata, F., Blöcher-Juárez, K. A., Gates, D. J., Janzen, G. M., Kur, A., et al. (2022). An adaptive teosinte mexicana introgression modulates phosphatidylcholine levels and is associated with maize flowering time. *bioRxiv* 119, e2100036119. doi: 10.1073/pnas.2100036119

Bradbury, P. J., Zhang, Z., 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

Burton, A. L., Brown, K. M., and Lynch, J. P. (2013). Phenotypic diversity of root anatomical and architectural traits in *Zea* species. *Crop Sci.* 53, 1042–1055. doi: 10.2135/cropsci2012.07.0440

Chen, W. K., Chen, L., Zhang, X., Yang, N., Guo, J. H., Wang, M., et al. (2022). Convergent selection of a WD40 protein that enhances grain yield in maize and rice. *Science* 375, eabg7985. doi: 10.1126/science.abg7985

Chen, L., Luo, J. Y., Jin, M. L., Yang, N., Liu, X. G., Peng, Y., et al. (2022). Genome sequencing reveals evidence of adaptive variation in the genus *Zea*. *Nat. Genet.* 54, 1736–1174. doi: 10.1038/s41588-022-01184-y

Cho, C., Jeon, E., Pandey, S. K., Ha, S. H., and Kim, J. (2019). LBD13 positively regulates lateral root formation in *Arabidopsis*. *Planta* 249, 1251–1258. doi: 10.1007/s00425-018-03087-x

Clark, R. M., Wagler, T. N., Quijada, P., and Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38, 594–597. doi: 10.1038/ng1784

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J. E., Pysh, L., Helariutta, Y., Freshour, G., et al. (1996). The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86, 423–433. doi: 10.1016/S0092-8674(00)80115-4

Doebley, J. (2004). The genetics of maize evolution. *Annu. Rev. Genet.* 38, 37–59. doi: 10.1146/annurev.genet.38.072902.092425

Erdelska, O., and Vidovencova, Z. (1993). Development of adventitious seminal root primordia of maize during embryogenesis. *Biologia. Ser. A. (Slovak Republic)* 48, 85–88.

Grando, S., and Ceccarelli, S. (1995). Seminal root morphology and coleoptile length in wild (Hordeum vulgare ssp spontaneum) and cultivated (Hordeum vulgare ssp vulgare) barley. *Euphytica* 86, 73–80. doi: 10.1007/BF00035941

Guo, J., Chen, L., Li, Y. X., Shi, Y. S., Song, Y. C., Zhang, D. F., et al. (2018). Meta-QTL analysis and identification of candidate genes related to root traits in maize. *Euphytica* 214, 223. doi: 10.1007/s10681-018-2283-3

Guo, J., Li, C., Zhang, X., Li, Y., Zhang, D., Shi, Y., et al. (2020). Transcriptome and GWAS analyses reveal candidate gene for seminal root length of maize seedlings under drought stress. *Plant Sci.* 292, 110380. doi: 10.1016/j.plantsci.2019.110380

Hetz, W., Hochholdinger, F., Schwall, M., and Feix, G. (1996). Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. *Plant J.* 10, 845–857. doi: 10.1046/j.1365-313X.1996.10050845.x

Hochholdinger, F. (2009). "The maize root system: Morphology, anatomy, and genetics," in *Handbook of maize* (New York: Springer), 145–160. doi: 10.1007/978-0-387-79418-1_8

Hochholdinger, F., Park, W. J., Sauer, M., and Woll, K. (2004). From weeds to crops: Genetic analysis of root development in cereals. *Trends Plant Sci.* 9, 42–48. doi: 10.1016/j.tplants.2003.11.003

Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E., and Tenaillon, M. I. (2012). Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS One* 7, e47659. doi: 10.1371/journal.pone.0047659

Li, C. H., Guan, H. H., Jing, X., Li, Y. Y., Wang, B. B., Li, Y. X., et al. (2022). Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. *Nat. Plants* 8, 750–763. doi: 10.1038/s41477-022-01190-2

Li, Y. X., Lu, J. W., He, C., Wu, X., Cui, Y., Chen, L., et al. (2022). Cis-regulatory variation affecting gene expression contributes to the improvement of maize kernel size. *Plant J.* 111, 1595–1608. doi: 10.1111/tpj.15910

Liu, S. X., Li, C. P., Wang, H. W., Wang, S. H., Yang, S. P., Liu, X. H., et al. (2020). Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biol.* 21, 1–22. doi: 10.1186/s13059-020-02069-1

Lynch, J. P. (2013). Steep, cheap and deep: An ideotype to optimize water and n acquisition by maize root systems. *Ann. Bot.* 112, 347–357. doi: 10.1093/aob/mcs293

Lynch, J. P. (2019). Root phenotypes for improved nutrient capture: An underexploited opportunity for global agriculture. *New Phytol.* 223, 548–564. doi: 10.1111/nph.15738

Majer, C., Xu, C. Z., Berendzen, K. W., and Hochholdinger, F. (2012). Molecular interactions of rootless concerning crown and seminal roots, a LOB domain protein regulating shoot-borne root initiation in maize *(Zea mays* l.). *Philos. Trans. R. Soc Lond. B Biol. Sci.* 367, 1542–1551. doi: 10.1098/rstb.2011.0238

Mammadov, J., Buyyarapu, R., Guttikonda, S. K., Parliament, K., Abdurakhmonov, I. Y., and Kumpatla, S. P. (2018). Wild relatives of maize, rice, cotton, and soybean: Treasure troves for tolerance to biotic and abiotic stresses. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00886

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, G. J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084. doi: 10.1073/pnas.052125199

Ma, P., Zhang, X., Luo, B. W., Chen, Z., He, X., Zhang, H. Y., et al. (2021). Transcriptomic and genome-wide association study reveal long noncoding RNAs responding to nitrogen deficiency in maize. *BMC Plant Biol.* 21, 1–19. doi: 10.1186/s12870-021-02847-4

McCouch, S. (1997). Report on QTL nomenclature. *Rice Genet. Newsl.* 14, 11–13.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Perkins, A. C., and Lynch, J. P. (2021). Increased seminal root number associated with domestication improves nitrogen and phosphorus acquisition in maize seedlings. *Ann. Bot.* 128, 453–468. doi: 10.1093/aob/mcab074

Pestsova, E., Lichtblau, D., Wever, C., Presterl, T., Bolduan, T., Ouzunova, M., et al. (2016). QTL mapping of seedling root traits associated with nitrogen and water use efficiency in maize. *Euphytica* 209, 585–602. doi: 10.1007/s10681-015-1625-7

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

Salvi, S., Giuliani, S., Ricciolini, C., Carraro, N., Maccaferri, M., Presterl, T., et al. (2016). Two major quantitative trait loci controlling the number of seminal roots in maize co-map with the root developmental genes *rtcs* and *rum1*. *J. Exp. Bot.* 67, 1149–1159. doi: 10.1093/jxb/erw011

Sozzani, R., Cui, H., Moreno-Risueno, M. A., Busch, W., Van Norman, J. M., Vernoux, T., et al. (2010). Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. *Nature* 466, 128–132. doi: 10.1038/nature09143

Stitzer, M. C., and Ross-Ibarra, J. (2018). Maize domestication and gene interaction. *New Phytol.* 220, 395–408. doi: 10.1111/nph.15350

Suzuki, M., Sato, Y., Wu, S., Kang, B. H., and McCarty, D. R. (2015). Conserved functions of the MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation rate. *Plant Cell* 27, 2288–2300. doi: 10.1105/tpc.15.00290

Taramino, G., Sauer, M., Stauffer, J. L., Multani, D., Niu, X. M., Sakai, H., et al. (2007). The maize (*Zea mays* l.) *RTCS* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *Plant J.* 50, 649–659. doi: 10.1111/j.1365-313X.2007.03075.x

Tian, J. G., Wang, C. L., Xia, J. L., Wu, L. S., Xu, G. H., Wu, W. H., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* 365, 658–664. doi: 10.1126/science.aax5482

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

von Behrens, I., Komatsu, M., Zhang, Y. X., Berendzen, K. W., Niu, X. M., Sakai, H., et al. (2011). Rootless with undetectable meristem 1 encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. *Plant J.* 66, 341–353. doi: 10.1111/j.1365-313X.2011.04495.x

Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL cartographer 2.5 department of statistics* (Raleigh: North Carolina State University).

Wang, H. Z., Hou, J. B., Ye, P., Hu, L., Huang, J. S., Dai, Z. K., et al. (2021). A teosinte-derived allele of a MYB transcription repressor confers multiple disease resistance in maize. *Mol. Plant* 14, 1846–1863. doi: 10.1016/j.molp.2021.07.008

Wang, P. W., Hummel, E., Osterrieder, A., Meyer, A. J., Frigerio, L., Sparkes, I., et al. (2011). KMS1 and KMS2, two plant endoplasmic reticulum proteins involved in the early secretory pathway. *Plant J.* 66, 613–628. doi: 10.1111/j.1365-313X.2011.04522.x

Wang, H. M., Wei, J., Li, P. C., Wang, Y. Y., Ge, Z. Z., Qian, J. Y., et al. (2019). Integrating GWAS and gene expression analysis identifies candidate genes for root morphology traits in maize at the seedling stage. *Genes* 10, 773. doi: 10.3390/genes10100773

Weaver, J. E. (1926). *Root development of field crops* (McGraw-Hill Book Company: New York). doi: 10.2134/agronj1926.00021962001800060007x

Woll, K., Borsuk, L. A., Stransky, H., Nettleton, D., Schnable, P. S., and Hochholdinger, F. (2005). Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiol.* 139, 1255–1267. doi: 10.1104/pp.105.067330

Wysocka-Diller, J. W., Helariutta, Y., Fukaki, H., Malamy, J. E., and Benfey, P. N. (2000). Molecular analysis of *SCARECROW* function reveals a radial patterning mechanism common to root and shoot. *Development* 127, 595–603. doi: 10.1242/dev.127.3.595

Xu, C. Z., Tai, H. H., Saleem, M., Ludwig, Y., Majer, C., Berendzen, K. W., et al. (2015). Cooperative action of the paralogous maize lateral organ boundaries (LOB) domain

proteins RTCS and RTCL in shoot-borne root formation. *New Phytol.* 207, 1123–1133. doi: 10.1111/nph.13420

Yan, Y. S., Chen, X. Y., Yang, K., Sun, Z. X., Fu, Y. P., Zhang, Y. M., et al. (2011). Overexpression of an f-box protein gene reduces abiotic stress tolerance and promotes root growth in rice. *Mol. Plant* 4, 190–197. doi: 10.1093/mp/ssq066

Yang, L., Zhu, M., Yang, Y., Wang, K., Che, Y., Yang, S., et al. (2022). CDC48B facilitates the intercellular trafficking of SHORT-ROOT during radial patterning in roots. *J. Integr. Plant Biol.* 64, 843–858. doi: 10.1111/jipb.13231

Yu, G. C., Wang, L. G., Han, Y. Y., and He, Q. Y. (2012). clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zainab, R., Ayesha, B., Ghulam, S., Kashif, A., Muhammad, S., Sidra, R., et al. (2021). Drought stress induces differential DNA methylation shift at symmetric and asymmetric cytosine sites in the promoter region of *ZmEXPB2* gene in maize. *Int. J. Agric. Biol.* 25, 319–326. doi: 10.17957/IJAB/15.1671

Zhang, Y., Marcon, C., Tai, H., von Behrens, I., Ludwig, Y., Hey, S., et al. (2016). Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. *J. Exp. Bot.* 67, 1137–1147. doi: 10.1093/jxb/erv519

Zheng, H., and Chen, J. (2011). Emerging aspects of ER organization in root hair tip growth: lessons from RHD3 and atlastin. *Plant Signal* 6, 1710–1713. doi: 10.4161/psb.6.11.17477

Zhu, J. M., Mickelson, S. M., Kaeppler, S. M., and Lynch, J. P. (2006). Detection of quantitative trait loci for seminal root traits in maize (*Zea mays* l.) seedlings grown under differential phosphorus levels. *Theor. Appl. Genet.* 113, 1–10. doi: 10.1007/s00122-006-0260-z

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 March 2023
DOI 10.3389/fpls.2023.1145327

OPEN ACCESS

EDITED BY
Hifzur Rahman,
International Center for Biosaline Agriculture (ICBA), United Arab Emirates

REVIEWED BY
Deepmala Sehgal,
Syngenta, United Kingdom
Namrata Dhaka,
Central University of Haryana, India

*CORRESPONDENCE
Lin Chen
chenlin@caas.cn
Zehui Chen
chenzh907@sina.com

†These authors have contributed equally to this work

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

RECEIVED 16 January 2023
ACCEPTED 10 March 2023
PUBLISHED 22 March 2023

CITATION
Wu X, Liu Y, Lu X, Tu L, Gao Y, Wang D, Guo S, Xiao Y, Xiao P, Guo X, Wang A, Liu P, Zhu Y, Chen L and Chen Z (2023) Integration of GWAS, linkage analysis and transcriptome analysis to reveal the genetic basis of flowering time-related traits in maize.
Front. Plant Sci. 14:1145327.
doi: 10.3389/fpls.2023.1145327

Integration of GWAS, linkage analysis and transcriptome analysis to reveal the genetic basis of flowering time-related traits in maize

Xun Wu[1†], Ying Liu[1†], Xuefeng Lu[1†], Liang Tu[1], Yuan Gao[1], Dong Wang[1,2], Shuang Guo[1,2], Yifei Xiao[1,2], Pingfang Xiao[1,2], Xiangyang Guo[1], Angui Wang[1], Pengfei Liu[1], Yunfang Zhu[1], Lin Chen[3*] and Zehui Chen[1*]

[1]Institute of Upland Food Crops, Guizhou Academy of Agricultural Sciences, Guiyang, Guizhou, China, [2]College of Agriculture, Guizhou University, Guiyang, Guizhou, China, [3]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

Maize (*Zea mays*) inbred lines vary greatly in flowering time, but the genetic basis of this variation is unknown. In this study, three maize flowering-related traits (DTT, days to tasselling; DTP, days to pollen shed; DTS, days to silking) were evaluated with an association panel consisting of 226 maize inbred lines and an $F_{2:3}$ population with 120 offspring from a cross between the T32 and Qi319 lines in different environments. A total of 82 significant single nucleotide polymorphisms (SNPs) and 117 candidate genes were identified by genome-wide association analysis. Twenty-one quantitative trait loci (QTLs) and 65 candidate genes were found for maize flowering time by linkage analysis with the constructed high-density genetic map. Transcriptome analysis was performed for Qi319, which is an early-maturing inbred line, and T32, which is a late-maturing inbred line, in two different environments. Compared with T32, Qi319 showed upregulation of 3815 genes and downregulation of 3906 genes. By integrating a genome-wide association study (GWAS), linkage analysis and transcriptome analysis, 25 important candidate genes for maize flowering time were identified. Together, our results provide an important resource and a foundation for an enhanced understanding of flowering time in maize.

KEYWORDS

maize, genome-wide association study, flowering time, RNA-Seq, candidate gene

1 Introduction

Maize, one of the most important crops in the world (Yang and Yan, 2021), originated in the lowland tropics of South America. Since it was first domesticated 9000 years ago, its planting area has expanded (Matsuoka et al., 2002), and this spread to various parts of the globe has resulted in rich phenotypic and genetic variation (Li et al., 2016). Among the phenotypic variations in maize, the variation in flowering time is an important factor in determining adaptation to local environments and is also a key selection standard for maize breeding and germplasm innovation (Shi et al., 2022).

In recent decades, the main genetic factors controlling flowering time in the model plant *Arabidopsis thaliana* have begun to be determined (Song et al., 2013; Johansson and Staiger, 2015; Freytes et al., 2021). Moreover, some important flowering time genes from other plant species have been cloned, such as *FLOWERING LOCUS T* (*FT*), *CONSTANS* (*CO*), and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS* (*SOC1*), which has been helpful for understanding the plant molecular regulatory network controlling flowering time (Wickland and Hanzawa, 2015; Liu et al., 2008; Shim et al., 2017). In maize, flowering time shows rich variation, with the earliest flowering time being 35 days and the latest being 120 days (Colasanti and Muszynski, 2009). Recently, many genes that regulate maize flowering time have been cloned (Kozaki et al., 2004; Hung et al., 2012; Yang et al., 2013; Guo et al., 2018). For example, *ZmCCT* is a homologous gene of the rice photoperiod response regulator *Ghd7* (Hung et al., 2012; Yang et al., 2013), *ZCN8* is homologous to the Arabidopsis *FT* gene (Guo et al., 2018), and *ID1* has a zinc finger domain (Kozaki et al., 2004). These findings are of great significance for understanding the genetic regulatory network of maize flowering time. However, maize flowering time is a typical quantitative trait and is jointly regulated by multiple genes (Buckler et al., 2009). Therefore, it is necessary to further analyse the genetic basis of flowering time regulation in maize and to explore new genetic regulatory loci.

Linkage analysis and genome-wide association studies (GWASs) have proven to be effective methods for mining quantitative trait loci (QTLs) for important traits, including maize flowering time. Buckler et al. (2009) systematically studied the genetic basis of maize flowering time by using the nested association mapping (NAM) population, which contains 5000 recombinant inbred lines (RILs). On this basis, an additive genetic model was proposed to explain the flowering time of maize: maize flowering time is controlled by multiple micro-QTLs (Buckler et al., 2009). In subsequent studies, Li et al. (2016) identified nearly 1000 significant single nucleotide polymorphisms (SNPs), which were associated with 220 candidate genes related to maize flowering time, by using an extensive association mapping population of more than 8000 lines. Notably, due to the high diversity among maize inbred lines, it is difficult to determine all the factors regulating flowering time (Hufford et al., 2021). Moreover, the significant loci identified for different populations are not consistent (Li et al., 2016). Therefore, it is necessary to investigate additional populations to discover new genetic regulatory loci and candidate genes for flowering time to provide a basis for subsequent maize germplasm improvement.

In this study, we used two maize germplasm panels, (i) a natural association panel with 226 inbred lines and (ii) a linkage mapping population with 120 $F_{2:3}$ offspring obtained from the cross between the T32 and Qi319 maize lines, to identify the genetic loci associated with maize flowering time. Three flowering time-related traits (DTT, days to tasselling; DTP, days to pollen shed; DTS, days to silking) were analyzed in different environments. The GWAS approach, QTL analysis and transcriptomic analysis were combined in this study to identify new loci and reveal candidate genes for maize flowering time.

2 Materials and methods

2.1 Plant materials

A linkage population with 120 $F_{2:3}$ offspring developed from parents of the maize T32 (renamed Ki32) and Qi319 lines and the association mapping panel with 226 inbred lines were selected as the plant experimental materials in this study. T32 is a tropical maize line derived from the Suwan population and is widely used for breeding in southern China. Qi319 is a temperate line and has been widely used in temperate regions, especially in northern China (Wu et al., 2019).

2.2 Field experiment and flowering time-related trait evaluation

The association panel materials were planted and phenotyped in Guiyang (GY, 106.7°N, 26.5°E), Guizhou Province; in Sanya (SA, 18.36°N, 109.16°E), Hainan Province; and in Zhangye (ZY, 38.93N, 100.45°E), Gansu Province in 2020. The linkage population was planted and phenotyped in Sanya (2019) and Guiyang (2020). Each line was planted in a single row 3 m in length with 12 individual plants per row. The fertilization, irrigation, pest control and weed management for all field trials were the same as those of the local field. Three flowering time-related traits (DTT, DTS, DTP) were recorded when 50% of plants exhibited the corresponding traits.

2.3 DNA extraction and genotyping

Genomic DNA was extracted from young leaves of F_2 plants using the cetyltrimethylammonium bromide (CTAB) procedure based on our previously described methods (Wu et al., 2015). DNA quality testing and genotyping by sequencing (GBS) assessments were completed by the Beijing Compass Biotechnology Company by using previously described methods (Elshire et al., 2011). The high-quality SNPs between parents were identified by alignment with B73

RefGen_v4 using the BWA package and GATK (Lai et al., 2010; McKenna et al., 2010). The calling and annotation of SNPs were accomplished using SAMTOOLS software (Li and Durbin, 2009). In addition, genotyping, population structure detection, kinship determination, and principal component analysis (PCA) of the association panel had already been completed in our previous research (Wu et al., 2019).

2.4 QTL mapping and genome-wide association study

A total of 169,108 high-quality SNPs were selected to construct a genetic map using the ordering algorithm. QTL analyses were conducted using QTL IciMapping software Version 4.1 (Li et al., 2008). A total of 43,252 SNPs were selected to perform a phenotype–genotype GWAS by using TASSEL v5.2.80 software, with a mixed linear model (MLM) in which population structure and pairwise kinship were treated as covariates (Yu et al., 2006). The significant cut-off value was defined as a logarithm of odds (LOD) score >4.

2.5 Transcriptome data analysis

T32 and Qi319 were planted at the Sanya and Zhangye sites. Leaves were collected from three replicates of each inbred line at the V9 stage. A total of 42 samples were collected for total RNA extraction. Total RNA was collected by using TRIzol reagent, and the construction of cDNA libraries and RNA sequencing were performed by Biomarker Technologies (Beijing, China) with the Illumina HiSeq 2000 platform. The clean reads were mapped to the maize B73 reference genome assembly V4 by using TopHat2 (Trapnell et al., 2012). The gene expression level was estimated by using the fragments per kilobase per million reads (FPKM) value. The differentially expressed genes were obtained by using the R statistical software package DESeq with $P_{adj} < 0.05$ and $|\log_2(\text{fold change [FC]})| \geq 1$ (Anders and Huber, 2010).

2.6 Candidate gene detection and qRT-PCR analysis

Based on the maize B73 reference genome assembly V4, genes located within two times the linkage disequilibrium distance of one quantitative trait nucleotide were determined to be candidate genes for flowering time-related traits. Functional annotations of these candidate genes were completed using the Protein–Protein Basic Local Alignment Search Tool (BlastP) and conserved domain search tools. The qRT-PCR primers were designed by Primer5 software and are listed in Table S1. The GAPDH gene was used for data normalization, and three biological replicates were used for each sample. To analyse the data, the $2^{-(\triangle\triangle CT)}$ method was used (Livak and Schmittgen, 2001).

3 Results

3.1 Phenotypic variation in different environments

The results showed that the average DTT was 59.03 days in Sanya, 76.56 days in Guiyang, and 93.69 days in Zhangye (Figure 1A, Table S2). The average DTP was 60.11 days in Sanya, 77.57 days in Guiyang and 94.97 days in Zhangye (Figure 1B, Table S2). The average DTS was 60.91 days in Sanya, 77.98 days in Guiyang and 95.93 days in Zhangye (Figure 1C, Table S2). The population at the Zhangye site had a higher DTT, DTP and DTS than those at the Sanya and Guiyang sites (Figure 1). There were significant correlations between the three traits in different site-specific environments (Table S3). Significant effects of genotype and genotype × environment ($G \times E$) were found for flowering time-related traits in the association panel (Table S4). The H^2 for DTT, DTP and DTS was 0.75, 0.74 and 0.74, respectively, as calculated as described in a previous study. These results showed that flowering time was influenced by the environment.

Based on our previous results, the association mapping panel could be divided into seven subgroups (HCL645 subgroup, T32

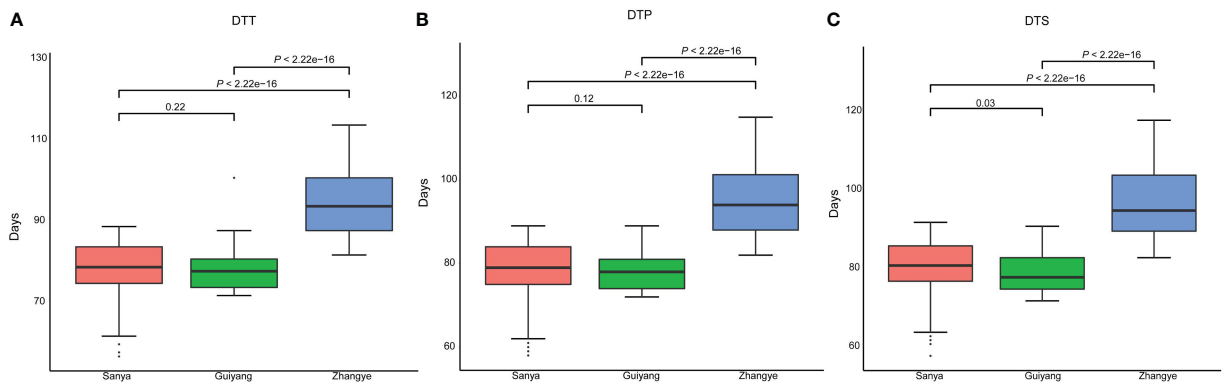


FIGURE 1
Flowering time-related traits for association mapping in three different site-specific environments (in Sanya, Guiyang and Zhangye). **(A)** days to tasselling (DTT); **(B)** days to pollen shed (DTP); **(C)** days to silking (DTS). A *t* test was used for analysis.

subgroup, QR273 subgroup, Mo17 subgroup, A801 subgroup, B73 subgroup and mixed subgroup). Based on the statistical analysis, population structure was not significantly different for the three traits (Table S5). Among the seven subgroups, the A801 subgroup had the greatest values for DTT, DTP and DTS, and the HCL645 subgroup had the smallest value for maize flowering time.

3.2 GWAS of maize flowering time

To identify the significant loci (LOD >4) associated with maize flowering time, an MLM analysis was performed on the association panel (Figure 2, Table S6). For DTT, 48, 11 and 4 significant loci were found in the populations at the Zhangye, Guiyang and Sanya sites, respectively (Figure 2A, Table S6). For DTP, a total of 41, 22, and 4 significant loci were identified at the Zhangye, Guiyang and Sanya sites, respectively (Figure 2B, Table S6). Thirty-three, 25 and 3 loci were found to be significantly associated with DTS at the Zhangye, Guiyang and Sanya sites, respectively (Figure 2C, Table S6). The amount of phenotypic variation explained by these significant loci ranged from 9.2% to 14.0% (Table S6). PZE-106004147, which was found to be associated with DTT at the Sanya site and was located on Chr6, explained the least phenotypic variation, and PZE-108068611, which was found to be associated with DTS at Guiyang and located on Chr8, explained the most.

In addition, these significant loci related to maize flowering time were detected only in certain site-specific environments. In the three different site-specific environments (in Zhangye, Guiyang and Sanya), DTT, DTP and DTS were significantly correlated with only 2 (PZE-104072142 and PZE-104096936), 3 (PZE-101256915, PZE-104072142 and PZE-104096936) and 3 (PZE-108090522, PZE-104072142 and PZE-104096936) loci, respectively (Figures 2D–F). Here, a total of 43 loci were found to be significantly associated with DTT, DTP and DTS simultaneously (Figure 2G). There were also 9 loci that regulated DTT and DTP simultaneously and 7 loci that regulated DTP and DTS simultaneously. These results indicated that there is a significant genetic correlation between flowering time-related traits and that these traits are highly susceptible to site-specific environmental impacts.

A total of 117 candidate genes were found around the 82 significant SNPs (Table S6). Among these candidate genes, some genes known to be related to flowering time in plants were detected. For example, Zm00001d050018 (*bzip68*) encodes the ABI5 protein, and its homologue in Arabidopsis delays flowering (Chang et al., 2019). Zm00001d044272 (*bhlh94*) encodes the bHLH transcription factor, and homologous genes can upregulate the expression level of *FT* to regulate flowering time in Arabidopsis (Li et al., 2017). Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted for the 117 candidate genes. The GO results showed that eight genes were involved in GTP binding and six genes were involved in the carbohydrate metabolic process (Figure 3A). The KEGG results showed that these genes were involved in four main processes (genetic information process, metabolism, organismal systems cellular process and environmental information process), six genes were involved in the amino acid biosynthesis process, and five genes were involved in the plant hormone signal transduction process (Figure 3B).

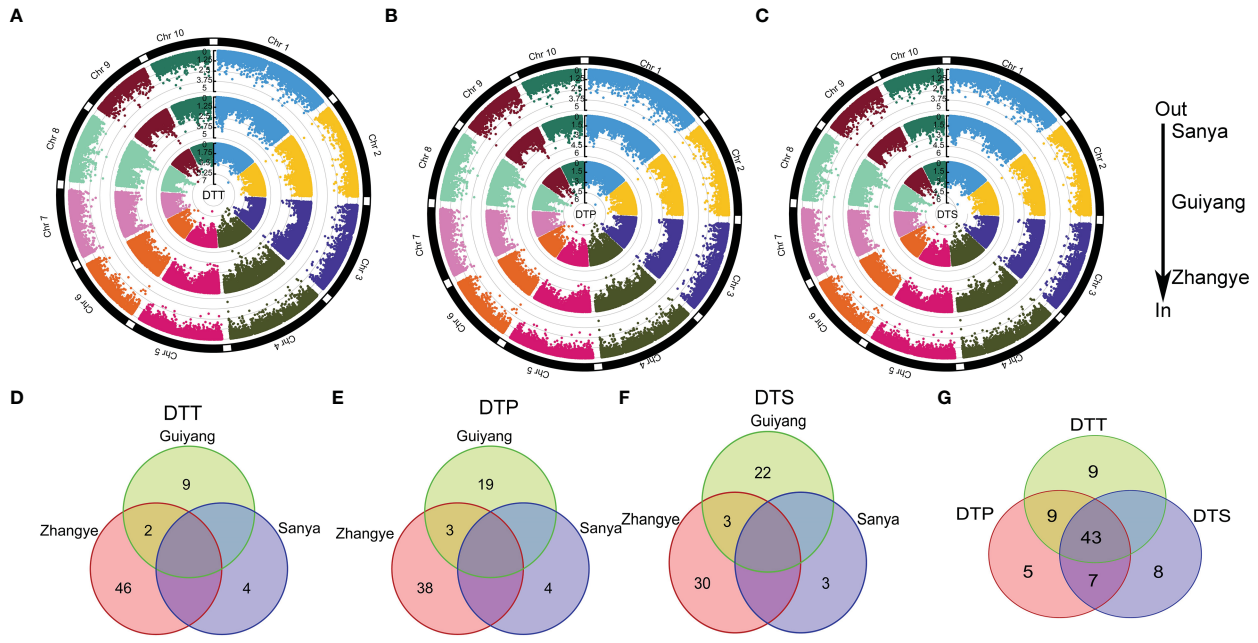


FIGURE 2
The GWAS for maize flowering time. **(A)** DTT; **(B)** DTP; **(C)** DTS. The three circles from outer to inner indicate three different local environments (in Sanya, Guiyang and Zhangye). **(D–F)** show the common significant SNPs in the three local environments for DTT, DTP and DTS, respectively. **(G)** shows that the common significant SNPs were simultaneously related to DTT, DTP and DTS.

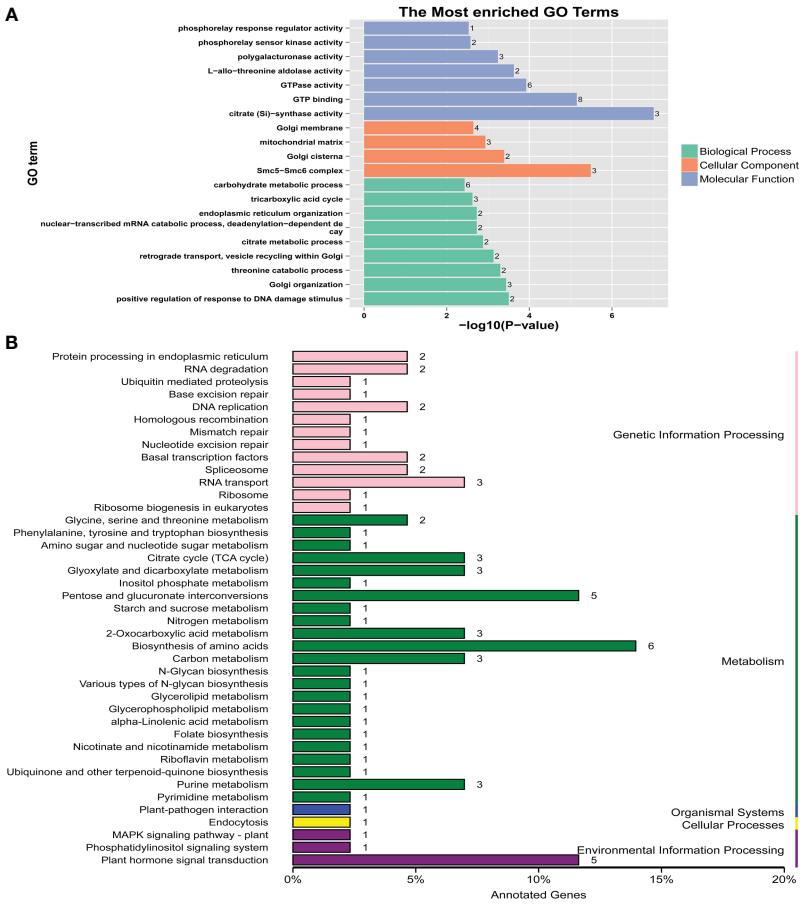


FIGURE 3
GO and KEGG analyses for the candidate genes identified by GWAS. **(A)** GO analysis. **(B)** KEGG analysis.

3.3 QTL analysis

T32 and Qi319 exhibited significant differences in flowering time in different environments (Figure 4A). T32 had a longer flowering time than Qi319 in Guiyang and Zhangye (Table S7). For DTT, DTP and DTS, this $F_{2:3}$ population showed more variation at the Zhangye site than at the Guiyang site (Figure 4B, Table S7). In this population, the H^2 values for DTT, DTP and DTS were 0.559, 0.557 and 0.558, respectively, which suggests that the environment plays an important role in maize flowering time. A total of 169,108 SNP markers were used to construct the genetic linkage map (Figure S1). The SNP number for each chromosome ranged from 12,112 (Chr9) to 23,145 (Chr1).

For DTT, a total of eleven QTLs were found in the Zhangye (eight QTLs) and Guiyang (three QTLs) environments (Figure 4C, Table S8). The phenotypic variation in DTT explained by each locus ranged from 11.4% (*qZYDTT5*) to 32.5% (*qGYDTT1*). For DTP, four QTLs (one QTL found at the Zhangye site and three QTLs found in Guiyang) were identified in this population (Figure 4C, Table S8). The amount of phenotypic variation explained ranged from 22.4% (*qZYDTP1*) to 35.4% (*qGYDTP2*). For DTS, six QTLs were found in the Zhangye (three QTLs) and Guiyang (three QTLs) environments (Figure 4C, Table S8). The phenotypic variation ranged from 7.8% (*qZYDTS2*) to 32.6% (*qGYDTS1*). Among the QTLs related to DTT, DTP and DTS, two genetic regions exhibited pleiotropic effects: *qGYDTT1* and *qGYDTP1* were both located at 304.2 Mb on Chr1, and *qGYDTT3* and *qGYDTP3* were both located at 180.1 Mb on Chr8 (Table S8). These results suggested that these two genomic regions can simultaneously regulate DTT and DTP in maize.

Sixty-five candidate genes were detected in these QTL interval regions, and some important genes related to plant flowering time were found. For example, Zm00001d011669, which is located in the *qZYDTT7* region, encodes an MYB transcription factor; Zm00001d029584, which is located in the *qZYDTS1* region, encodes a zinc finger protein; and Zm00001d003293 encodes an NAC transcription factor.

3.4 Transcriptome analysis

To identify the genes involved in maize flowering time, the differentially expressed genes (DEGs) between T32 and Qi319 in different environments (Sanya and Zhangye) were identified

10.3389/fpls.2023.1145327

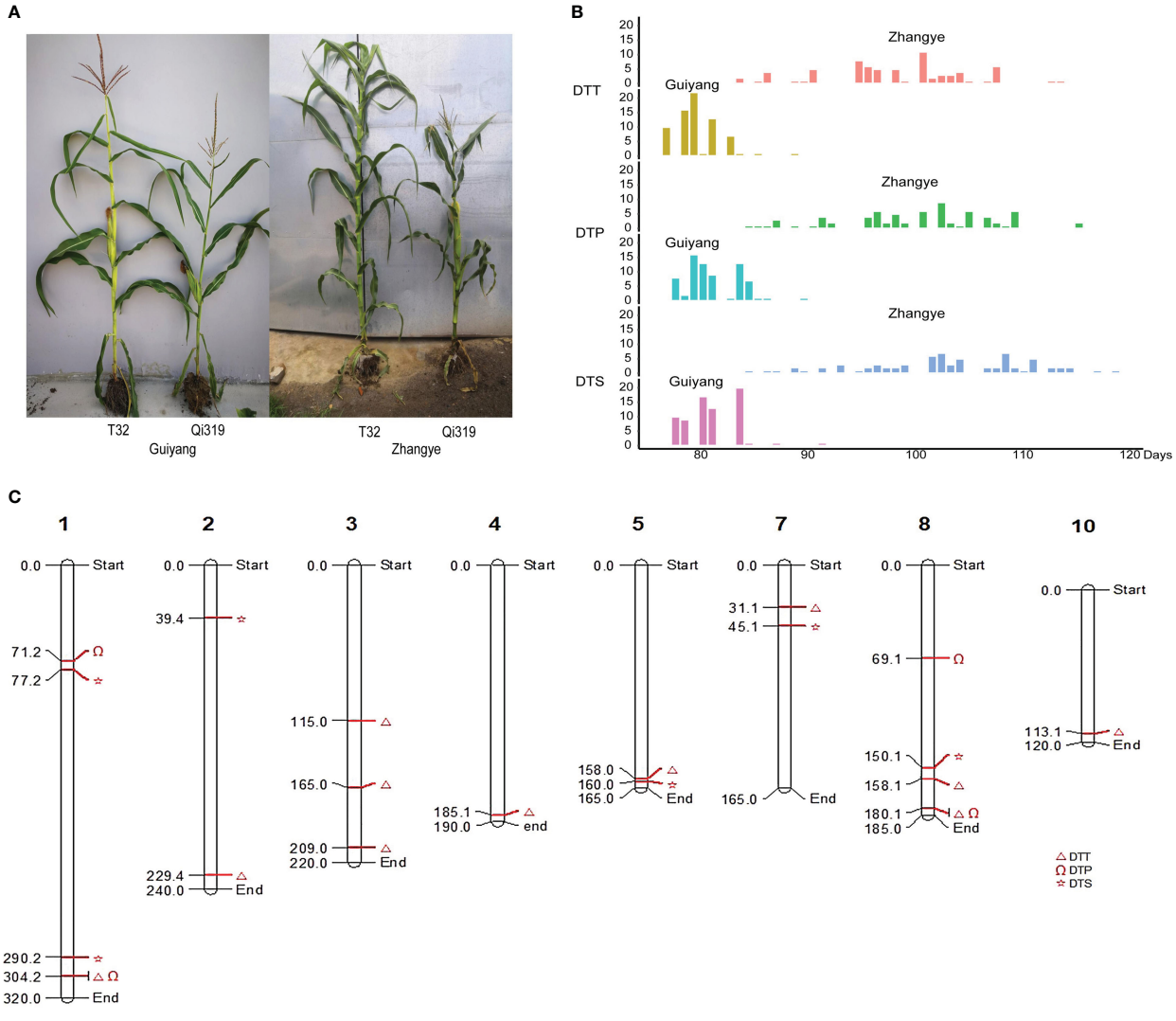


FIGURE 4
Linkage analysis for flowering time in the $F_{2:3}$ population. **(A)** Performance of the parents (T32 and Qi319) in the different site-specific environments (Guiyang and Zhangye). **(B)** The distribution of DTT, DTP and DTS in the $F_{2:3}$ population at the Zhangye site and in Guiyang. **(C)** The QTLs identified in this population.

(Figure 5A, Tables S9, S10). In Sanya, 2776 genes were upregulated and 2098 genes were downregulated in Qi319 compared with T32. In Zhangye, 2586 genes were upregulated and 3139 genes were downregulated in Qi319. Among these DEGs, 1477 common genes were upregulated and 1331 common genes were downregulated in the two different environments (Figure 5B). To identify the biological functions of these DEGs, GO enrichment analyses were conducted. There was a significant difference among the upregulated and downregulated genes. Among the upregulated genes, 58 genes were involved in oxidoreductase activity, 25 genes were related to light stimulus, 23 genes were involved in polysaccharide binding, and 15 genes were involved in photosynthesis (Figure 5C). Among the downregulated genes, 113 genes had protein serine/threonine kinase activity, 45 genes were involved in the protein folding process, and 28 genes were involved in the ABA process (Figure 5D).

3.5 Combining linkage analysis, the GWAS approach and transcriptome analysis to identify candidate genes for maize flowering time

To identify the candidate genes for maize flowering time, the results of GWASs and linkage and transcriptome analyses were integrated. Among the 117 candidate genes identified by using GWAS, 16 genes were differentially expressed between T32 and Qi319 (Figure 6A). Among these 16 genes, six genes were differentially expressed in both environments. For example, the expression levels of Zm00001d003058, which encodes a threonine aldolase protein (Figure 6B), and Zm00001d053684, which encodes a citrate synthase protein, were higher in T32 than in Qi319 in Zhangye and Sanya (Figure 6C). The expression levels of Zm00001d040569 and Zm00001d049023 were higher in Qi319

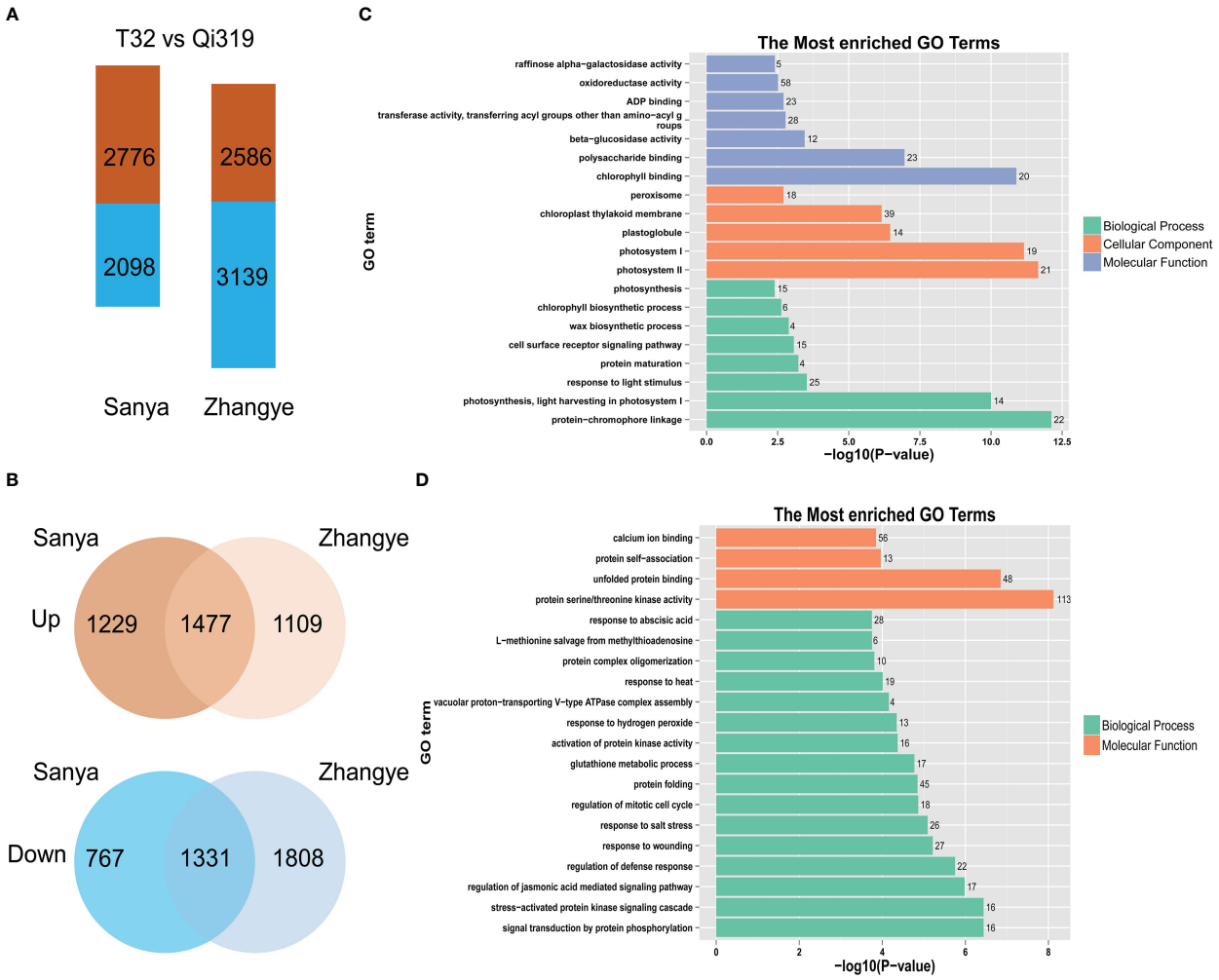


FIGURE 5
The DEGs identified between T32 and Qi319 in different site-specific environments (in Zhangye and Sanya). **(A)** The number of DEGs in different environments. **(B)** The common DEGs in different environments. **(C)** The most enriched GO terms for the upregulated genes. **(D)** The most enriched GO terms for the downregulated genes.

than in T32 in the two environments (Figures 6D, E). In addition, we also found that four genes had different expression levels only in Zhangye. For example, Zm00001d010635, which encodes a zinc finger protein, differed in expression between T32 and Qi319 only in Zhangye (Figure 6F). Five genes were found to have different expression levels in Sanya, such as Zm00001d044272, which encodes a bHLH transcription factor (Figure 6G).

Among the 65 candidate genes identified by QTL analysis, 9 genes were differentially expressed between T32 and Qi319 (Figure 7A). Three genes were differentially expressed in the two environments: Zm00001d003294 was downregulated in Qi319 at the Zhangye and Sanya sites (Figure 7B), and Zm00001d007345 was upregulated in Qi319 at the Zhangye and Sanya sites (Figure 7C). Four genes were downregulated in Qi319 only at the Zhangye site, such as Zm00001d029448, which encodes a TIFY 10B protein (Figure 7D). Two genes were differentially expressed only in Sanya. The results of qRT-PCR were consistent with the results of transcriptome analysis (Figures 7B–E).

To further narrow the genomic regions related to maize flowering time, the results of QTL analysis and the GWAS approach were integrated in this study. Interestingly, one major QTL related to DTT, which was found in plants at the Zhangye site (*qZYDTT7*, LOD = 4.91, R^2 =26.85%), was identified by GWAS (Figures 8A, B). PZE-108104613, which is located in the interval of *qZYDTT7*, was significantly (LOD = 5.16, R^2 =12.77%) related to DTT in plants at the Zhangye site. This SNP has two alleles (A and G), and the average DTT associated with the A allele (97.8 days) was significantly different from that associated with the G allele (92.1 days) (Figure 8C). Based on the B73 reference genome, a total of 17 genes were found in this interval (Chr8, 158-159 Mb). Among these genes, Zm00001d011666, which encodes a calcium-dependent protein kinase family protein, and Zm00001d011668, which encodes the DNAJ family protein, had different expression levels between T32 and Qi319 in Zhangye (Figure 8D, E), and Zm00001d011673, which is also named *fps2* and is related to the development of maize leaves, had different expression level between T32 and Qi319 at the Sanya site (Figure 8F).

A

No.	Gene ID	Zhangye	Sanya	Annotation
1	Zm00001d003058	down	down	threonine aldolase family protein
2	Zm00001d020152	down	down	uncharacterized protein
3	Zm00001d051069	down	down	calcium binding atopy-related autoantigen 1
4	Zm00001d053684	down	down	citrate synthase family protein
5	Zm00001d027788	up	up	uncharacterized protein
6	Zm00001d040569	up	up	zinc finger family protein
7	Zm00001d049023	up	up	uncharacterized protein
8	Zm00001d010635	down	NA	RING zinc finger domain superfamily protein
9	Zm00001d032557	down	NA	uncharacterized protein
10	Zm00001d011700	up	NA	receptor-like serine/threonine-protein kinase ALE2
11	Zm00001d028537	up	NA	nucleic acid binding protein precursor
12	Zm00001d044272	NA	down	transcription factor bHLH041
13	Zm00001d015591	NA	up	uncharacterized protein
14	Zm00001d038819	NA	up	Non-structural maintenance of chromosomes element 4 homolog A
15	Zm00001d050018	NA	up	ABSCISIC ACID-INSENSITIVE 5-like protein 5
16	Zm00001d051316	NA	up	trihelix transcription factor GTL1

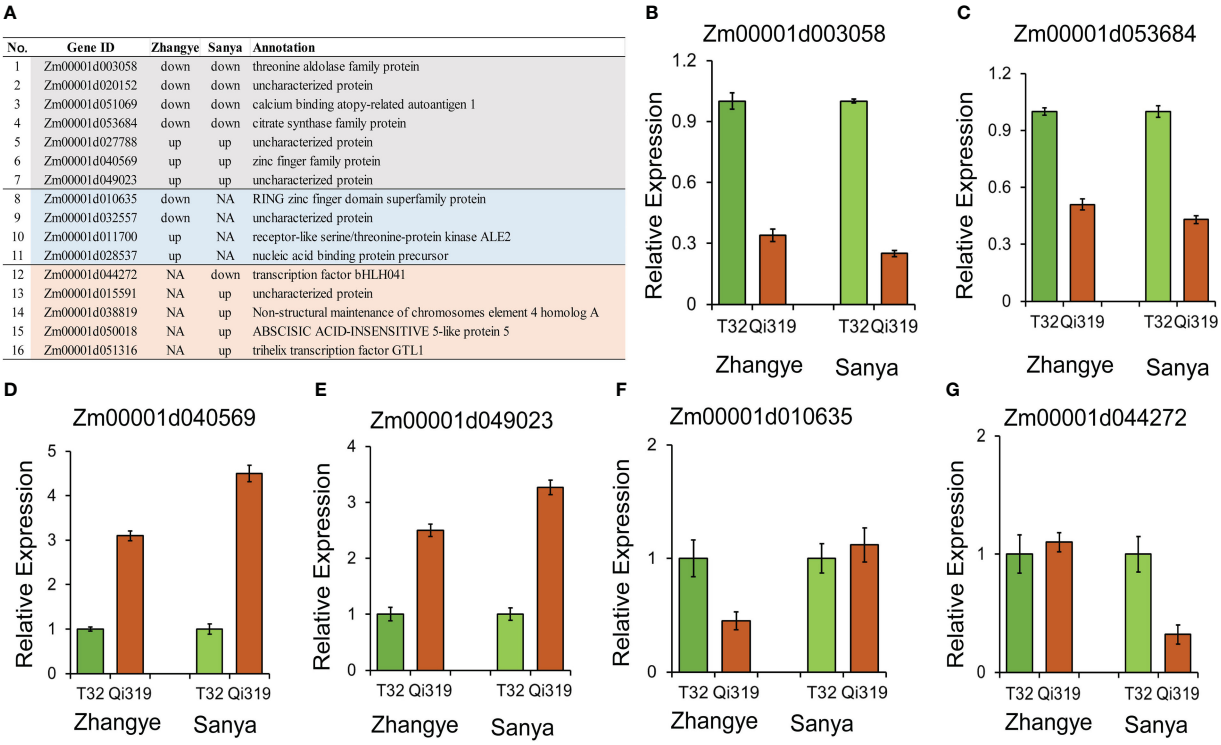


FIGURE 6
Candidate genes identified by GWAS and transcriptome analysis. **(A)** Information on sixteen DEGs. B-G, qRT-PCR results for selected genes. **(B)** Zm00001d003058; **(C)** Zm00001d053684; **(D)** Zm00001d040569; **(E)** Zm00001d049023; **(F)** Zm00001d010635; **(G)** Zm00001d044272.

4 Discussion

Compared with teosinte, which can only grow in tropical environments, maize has become one of the most widely planted crops in the world (Yang and Yan, 2021). The main reason is that maize adapts to different geographical environments *via* flowering time regulation (Li et al., 2016). During Chinese maize breeding practices, T32 was a foundation parental line derived from the Suwan germplasm, which showed a high combining ability but a longer reproductive period. It matured later than another foundation parental line, Qi319, derived from temperate maize germplasm in tropical regions. T32 cannot flower normally in temperate environments based on breeding experience, but Q319 can flower normally in different environments. Therefore, maize flowering time is an important characteristic that determines local environmental adaptation and is easily affected by the local environment. Similar to the findings of a previous study (Shi et al., 2022), the interaction effect of genotype and the environment on maize flowering time-related traits was significant in this study. Revealing the potential genetic basis of flowering time will aid in the selection of stable varieties in different local environments and will improve maize yields.

The GWAS approach has been shown to be an effective strategy for mining genetic loci for flowering time in maize (Buckler et al., 2009; Li et al., 2016; Shi et al., 2022). For example, a total of 18 SNPs and 19 candidate genes involved in maize flowering time were found in an association panel that consisted of 252 inbred lines (Vanous et al., 2018), and Li et al. (2016) identified nearly 1000 SNPs and 220 candidate genes using an extremely large association panel. In this study, a total of 82 SNPs and 117 candidate genes for maize flowering time-related traits were found in plants growing in different site-specific environments. Compared with the results of Li et al. (2016), 14 common SNPs were found in our study (Table S6). Among these candidate genes, some important candidate genes for flowering time were found. For example, Zm00001d007191 encodes an MYB transcription factor, and many previous studies have shown that MYB transcription factors, such as *MYB30* (Liu et al., 2014), *MYB106* (Hong et al., 2021), and *CmMYB2* (Zhu et al., 2020), play an important role in the development of flowers. Zm00001d044272 encodes a *bHLH* transcription factor. In Arabidopsis, the *bHLH* transcription factors *MYC2*, *MYC3*, and *MYC4* delay flowering time *via* the jasmonate pathway (Wang et al., 2017). In rice, two bHLH transcription factors (HBP1 and POH1) control flowering time by regulating the expression level of *Hd1* (Yin et al., 2023). In addition, Zm00001d050018, which encodes the ABI5 protein, was also found to be an important candidate gene for maize flowering time. In Arabidopsis, AtU2AF65b functions in abscisic acid (ABA)-mediated flowering by regulating the precursor messenger RNA splicing of ABI5 (Xiong et al., 2019).

QTL analysis is another effective method for mining the genetic loci for quantitative traits (Maschietto et al., 2017; Su et al., 2017; Wang et al., 2018). The resolution of QTL analysis can be enhanced by using a high-density genetic map (Maschietto et al., 2017; Su et al., 2017; Wang et al., 2018). In this study, a high-density genetic

A

No.	Gene ID	Zhangye	Sanya	Annotation
1	Zm00001d003294	down	down	Pentatricopeptide repeat-containing protein mitochondrial
2	Zm00001d029573	down	down	Protein argonaute 7
3	Zm00001d007345	up	up	11-beta-hydroxysteroid dehydrogenase-like 5
4	Zm00001d007341	down	NA	HVA22-like protein e
5	Zm00001d011666	down	NA	Putative calcium-dependent protein kinase family protein
6	Zm00001d011668	down	NA	DNAJ heat shock N-terminal domain-containing protein
7	Zm00001d029448	down	NA	Protein TIFY 10B
8	Zm00001d029580	NA	down	3-oxo-Delta(45)-steroid 5-beta-reductase
9	Zm00001d029583	NA	up	Polygalacturonase inhibitor 1

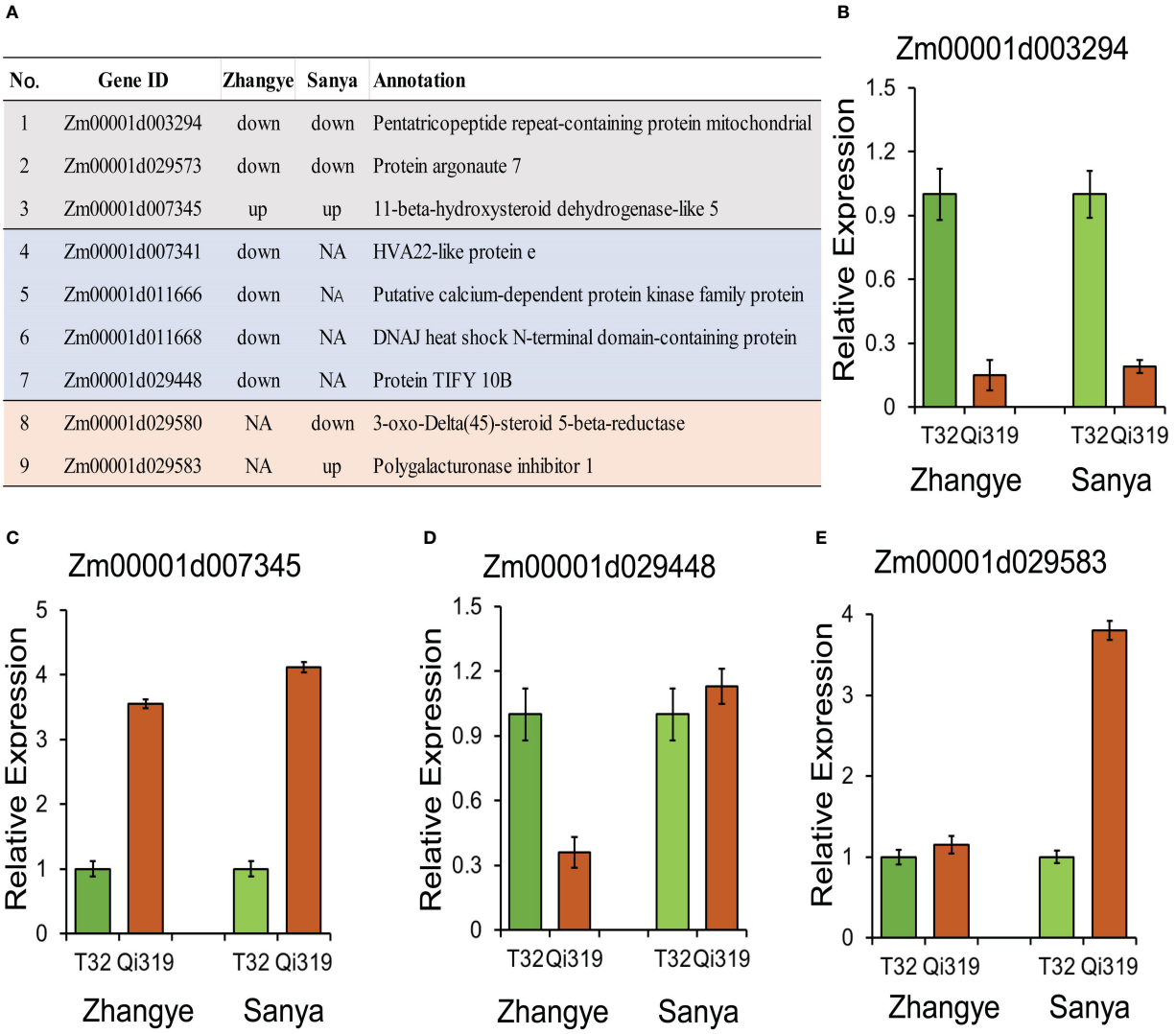


FIGURE 7
Candidate genes identified by QTL mapping and transcriptome analysis. **(A)** Information on nine DEGs. **(B–E)** qRT-PCR results for selected genes. **(B)** Zm00001d003294; **(C)** Zm00001d007345; **(D)** Zm00001d029448; **(E)** Zm00001d029583.

map was constructed with 169,108 markers by using the GBS method. In this study, we found nine QTL regions that were also found in a previous study (Table S8). For example, *qZYDTT1* and *qZYDTT3* were also found in Li et al. (2016) as *CN_QTL_17* and *CN_QTL_21*, respectively. These results showed that the nine QTL regions may be hotspot regions related to maize flowering time. Among the 65 candidate genes, some genes have been identified as important for plant flowering time. For example, Zm00001d023424 encodes the bZIP transcription factor, and Zm00001d047250 encodes the PLATZ transcription factor. In rice, *OsbZIP62*, which is a functional orthologue of FLOWERING LOCUS D, regulates the floral transition and panicle development (Kaur et al., 2021). In grapevine, *VviPLATZ1* is a major factor that controls female flower morphology determination (Iocco-Corena et al., 2021).

The combination of the GWAS approach, linkage analysis and transcriptome analysis can help us quickly identify candidate genes. For example, *ZmWRKY14*, which is a regulator of maize leaf number, flowering time and biomass yield, has been identified based on GWAS and linkage analysis (Li et al., 2021). By using the same strategy, seventeen candidate genes significantly associated with maize flowering time and leaf number have been found (Shi et al., 2022). In this study, 25 important candidate genes were found by integrating the results of GWASs, QTL mapping, and transcriptome analysis. Among the 25 DEGs, some genes had an important role in plant flowering time, such as Zm00001d044272, which encodes the bHLH transcription factor, and Zm00001d05008, which is the ABI5 gene. In addition, we found that one genome region located on Chr 8 (158 Mb) was significantly associated with DTT, which we were able to identify simultaneously by linkage analysis and GWAS. After combining these results with the results of transcriptome analysis, three important candidate genes were found. Zm00001d011666 encodes a calcium-dependent protein kinase (CPK) family protein. A previous study showed that the CPK32 gene can control pollen tube growth in tobacco and maize (Zhou et al., 2014; Li et al., 2018). Zm00001d011668 encodes the DNAJ family protein, and previous studies have shown that the

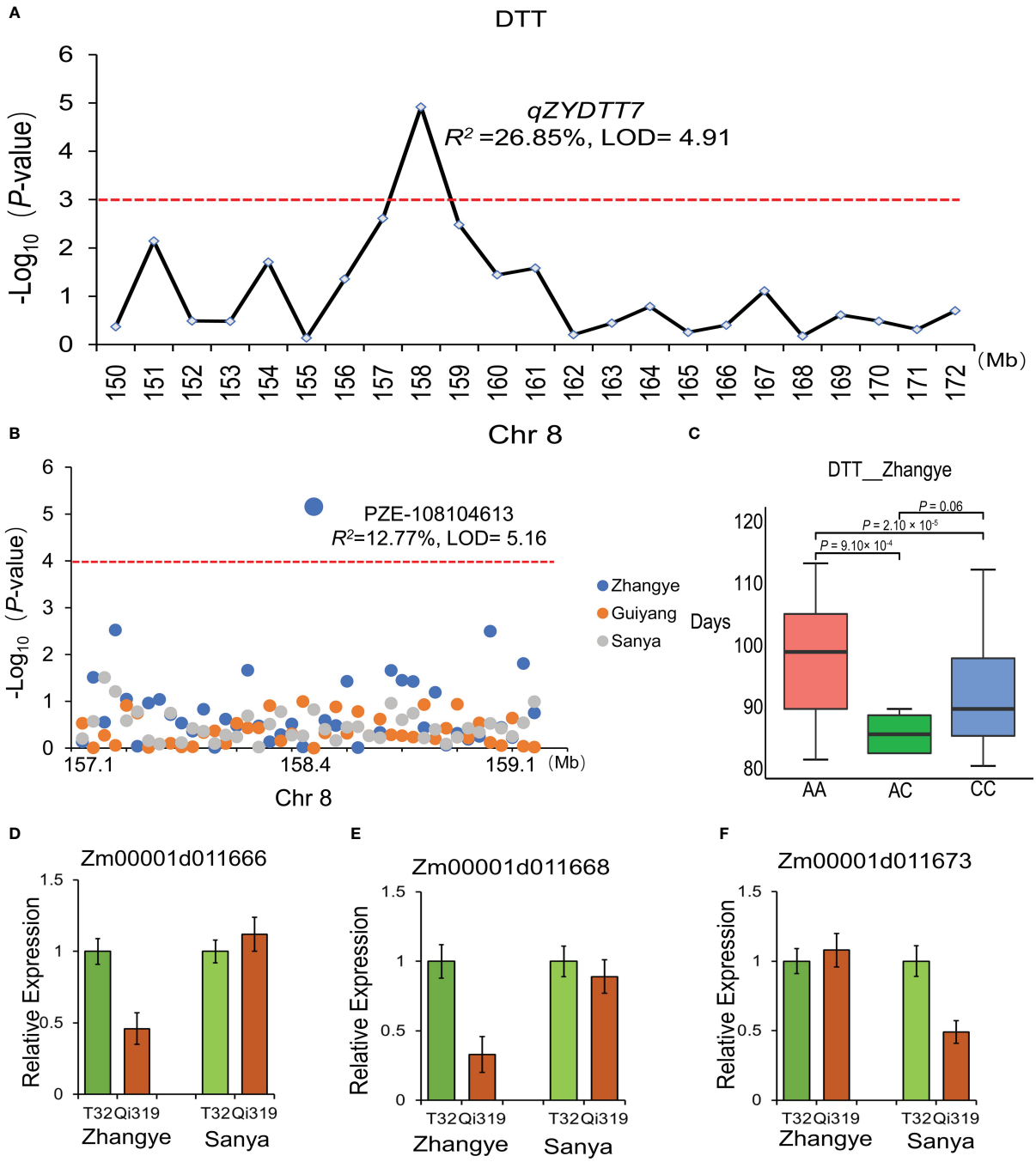


FIGURE 8
The candidate genes identified by the combination of QTL analysis, the GWAS approach and transcriptome analysis. **(A)** *qZYDTT7*, which was found to be significantly related to DTT at the Zhangye site, is located on Chr8 (158 Mb). **(B)** PZE-105104613, which was found to be related to DTT at the Zhangye site and is located in the same genomic region. **(C)** Haplotype analysis of PZE-105104613 with DTT. T tests were used for analysis. **(D–F)** The expression of the three candidate genes in the T32 and Qi319 lines in different site-specific environments. **(D)** Zm00001d011666; **(E)** Zm00001d011668; **(F)** Zm00001d011673.

DNAJ family protein plays an important role in seed filling and abiotic stress response (Hajduch et al., 2010). Zm00001d011673, which is also named *fps2*, encodes the farnesyl diphosphate synthase 2 protein. A previous study showed that Zm00001d011673 can interact with ZmIPT2, which can regulate leaf senescence and grain yield in maize (Song et al., 2022).

In conclusion, in this study, a total of 82 significant SNPs with 117 candidate genes and 21 QTLs with 65 candidate genes associated with maize flowering time were found by using the GWAS approach and QTL analysis, which can be evaluated and used in molecular-assisted breeding practices in the future. By combining the GWAS, QTL and transcriptome analysis results,

25 DEGs were found, and among these genes, three important candidate genes (Zm00001d011666, Zm00001d011668 and Zm00001d011673) were inferred as regulators of flowering time in maize. Our results provide an important gene resource for maize breeding to improve flowering time.

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

XW: Conception and design, Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - original draft. YL: Contributed to some of the experiments, Formal analysis, Data curation. XL: Conceptualization, Formal analysis, Validation, Data curation. LT: Contributed to some of the experiments. YG: Contributed to some of the experiments. DW: Contributed to some of the experiments. SG: Contributed to some of the experiments. YX: Contributed to some of the experiments. PX: Contributed to some of the experiments. XG: Contributed to some of the experiments. AW: Contributed to some of the experiments. PL: Contributed to some of the experiments. YZ: Contributed to some of the experiments. LC: Conception and design, Conceptualization, Methodology, Validation, Data curation, Writing - review & editing. ZC: Funding acquisition, Supervision, Methodology, Project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation (32171981, 32160451 and 32060460), the Guizhou Natural Science Foundation (ZK[2022]yiban236), the Guizhou Provincial Science and Technology Plan Project (Qian Kehe Support [2022] key 025,026, and 029), the Innovation Capacity Construction of Breeding Scientific Research Platform in Guizhou Province (QianKeHeFuQi [2022] 014), and the Construction of Genetic Transformation Platform for Dryland Grain Crops in Guizhou Province (QKZYD[2022]4011).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1145327/full#supplementary-material

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science* 325, 714–718. doi: 10.1126/science.1174276

Chang, G., Yang, W., Zhang, Q., Huang, J., Yang, Y., and Hu, X. (2019). ABI5-BINDING PROTEIN2 coordinates CONSTANS to delay flowering by recruiting the transcriptional corepressor TPR2. *Plant Physiol.* 179, 477–490. doi: 10.1104/pp.18.00865

Colasanti, J., and Muszynski, M. (2009). "The maize floral transition," in *Handbook of maize: Its biology*. Eds. J. L. Bennetzen and S. C. Hake (New York, NY: Springer), 41–55. doi: 10.1007/978-0-387-79418-1_3

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6, e19379. doi: 10.1371/journal.pone.0019379

Freytes, S. N., Canelo, M., and Cerdán, P. D. (2021). Regulation of flowering time: when and where? *Curr. Opin. Plant Biol.* 63, 102049. doi: 10.1016/j.pbi.2021.102049

Guo, L., Wang, X., Zhao, M., Huang, C., Li, C., Li, D., et al. (2018). Stepwise cis-regulatory changes in ZCN8 contribute to maize flowering-time adaptation. *Curr. Biol.* 28, 3005–3015.e4. doi: 10.1016/j.cub.2018.07.029

Hajduch, M., Hearne, L. B., Miernyk, J. A., Casteel, J. E., Joshi, T., Agrawal, G. K., et al. (2010). Systems analysis of seed filling in arabidopsis: Using general linear modeling to assess concordance of transcript and protein expression. *Plant Physiol.* 152, 2078–2087. doi: 10.1104/pp.109.152413

Hong, L., Niu, F., Lin, Y., Wang, S., Chen, L., and Jiang, L. (2021). MYB106 is a negative regulator and a substrate for CRL3BPM E3 ligase in regulating flowering time in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 63, 1104–1119. doi: 10.1111/jipb.13071

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373, 655–662. doi: 10.1126/science.abg5289

Hung, H. Y., Shannon, L. M., Tian, F., Bradbury, P. J., Chen, C., Flint-Garcia, S. A., et al. (2012). ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1913–E1921. doi: 10.1073/pnas.1203189109

Iocco-Corena, P., Chaïb, J., Torregrosa, L., Mackenzie, D., Thomas, M. R., and Smith, H. M. (2021). VviPLATZ1 is a major factor that controls female flower morphology determination in grapevine. *Nat. Commun.* 12, 6995. doi: 10.1038/s41467-021-27259-8

Johansson, M., and Staiger, D. (2015). Time to flower: interplay between photoperiod and the circadian clock. *J. Exp. Bot.* 66, 719–730. doi: 10.1093/jxb/eru441

Kaur, A., Nijhawan, A., Yadav, M., and Khurana, J. P. (2021). OsbZIP62/OsFD7, a functional ortholog of FLOWERING LOCUS d, regulates floral transition and panicle development in rice. *J. Exp. Bot.* 72, 7826–7845. doi: 10.1093/jxb/erab396

Kozaki, A., Hake, S., and Colasanti, J. (2004). The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic Acids Res.* 32, 1710–1720. doi: 10.1093/nar/gkh337

Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., et al. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* 42, 1027–1030. doi: 10.1038/ng.684

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, Y. X., Li, C., Bradbury, P. J., Liu, X., Lu, F., Romay, C. M., et al. (2016). Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J.* 86, 391–402. doi: 10.1111/tpj.13174

Li, J., Li, Y., Deng, Y., Chen, P., Feng, F., Chen, W., et al. (2018). Acalcium-dependent protein kinase, ZmCPK32, specifically expressed in maize pollen to regulate pollen tube growth. *PloS One* 13, e0195787. doi: 10.1371/journal.pone.0195787

Li, Z., Li, K., Yang, X., Hao, H., and Jing, H. C. (2021). Combined QTL mapping and association study reveals candidate genes for leaf number and flowering time in maize. *Theor. Appl. Genet.* 134, 3459–3472. doi: 10.1007/s00122-021-03907-x

Li, H., Ribaut, J. M., Li, Z., and Wang, J. (2008). Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor. Appl. Genet.* 116, 243–260. doi: 10.1007/s00122-007-0663-5

Li, Y., Wang, H., Li, X., Liang, G., and Yu, D. (2017). Two DELLA-interacting proteins bHLH48 and bHLH60 regulate flowering under long-day conditions in arabidopsis thaliana. *J. Exp. Bot.* 68, 2757–2767. doi: 10.1093/jxb/erx143

Liu, C., Chen, H., Er, H. L., Soo, H. M., Kumar, P. P., Han, J. H., et al. (2008). Direct interaction of AGL24 and SOC1 integrates flowering signals in *Arabidopsis*. *Development* 135, 1481–1491. doi: 10.1242/dev.020255

Liu, L., Zhang, J., Adrian, J., Gissot, L., Coupland, G., Yu, D., et al. (2014). Elevated levels of MYB30 in the phloem accelerate flowering in arabidopsis through the regulation of FLOWERING LOCUS T. *PloS One* 9 (2), e89799. doi: 10.1371/journal.pone.0089799

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle CT}$ method, methods: A companion to methods in enzymology. *Methods: A Companion to Methods in Enzymology* 25, 402–408. doi: 10.1006/meth.2001.1262

Maschietto, V., Colombi, C., Pirona, R., Pea, G., Strozzi, F., Marocco, A., et al. (2017). QTL mapping and candidate genes for resistance to *Fusarium* ear rot and fumonisin contamination in maize. *BMC Plant Biol.* 17, 20. doi: 10.1186/s12870-017-0970-1

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, G. J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6080–6084. doi: 10.1073/pnas.052125199

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Shi, J., Wang, Y., Wang, C., Wang, L., Zeng, W., Han, G., et al. (2022). Linkage mapping combined with GWAS revealed the genetic structural relationship and candidate genes of maize flowering time-related traits. *BMC Plant Biol.* 22, 328. doi: 10.1186/s12870-022-03711-9

Shim, J. S., Kubota, A., and Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in arabidopsis: CONSTANS is a hub for signal integration. *Plant Physiol.* 173, 5–15. doi: 10.1104/pp.16.01327

Song, Y. H., Ito, S., and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends Plant Sci.* 18, 575–583. doi: 10.1016/j.tplants.2013.05.003

Song, Y. F., Li, C. X., Zhu, Y., Guo, P., Wang, Q., Zhang, L., et al. (2022). Overexpression of ZmIPT2 gene delays leaf senescence and improves grain yield in maize. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.963873

Su, C., Wang, W., Gong, S., Zuo, J., Li, S., and Xu, S. (2017). High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00706

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Vanous, A., Gardner, C., Blanco, M., Martin-Schwarze, A., Lipka, A. E., Flint-Garcia, S., et al. (2018). Association mapping of flowering and height traits in germplasm enhancement of maize doubled haploid (GEM-DH) lines. *Plant Genome* 11, e170083. doi: 10.3835/plantgenome2017.09.0083

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

Wang, B., Zhu, Y., Zhu, J., Liu, Z., Liu, H., Dong, X., et al. (2018). Identification and fine-mapping of a major maize leaf width QTL in a re-sequenced large recombinant inbred lines population. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00101

Wickland, D. P., and Hanzawa, Y. (2015). The flowering locus T/Terminal flower1 gene family: functional evolution and molecular mechanisms. *Mol. Plant* 8, 983–997. doi: 10.1016/j.molp.2015.01.007

Wu, X., Li, Y., Li, X., Li, C., Shi, Y., Song, Y., et al. (2015). Analysis of genetic differentiation and genomic variation to reveal potential regions of importance during maize improvement. *BMC Plant Biol.* 15, 256. doi: 10.1186/s12870-015-0646-7

Wu, X., Wang, A., Guo, X., Liu, P., Zhu, Y., Li, X., et al. (2019). Genetic characterization of maize germplasm derived from suwan population and temperate resources. *Hereditas* 156, 2. doi: 10.1186/s41065-018-0077-1

Xiong, F., Ren, J. J., Yu, Q., Wang, Y. Y., Lu, C. C., Kong, L. J., et al. (2019). AtU2AF65b functions in abscisic acid mediated flowering *via* regulating the precursor messenger RNA splicing of ABI5 and FLC in *Arabidopsis*. *New Phytol.* 223, 277–292. doi: 10.1111/nph.15756

Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., et al. (2013). CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16969–16974. doi: 10.1073/pnas.1310949110

Yang, N., and Yan, J. (2021). New genomic approaches for enhancing maize genetic improvement, curr. opin. *Plant Biol.* 60, 101977. doi: 10.1016/j.pbi.2020.11.002

Yin, Y. B., Yan, Z. Q., Guan, J. N., Huo, Y. Q., Wang, T. Q., Li, T., et al. (2023). Two interacting basic helix-Loop-Helix transcription factors control flowering time in rice. *Plant Physiol.*, kiad077. doi: 10.1093/plphys/kiad077

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

Zhou, L., Lan, W., Jiang, Y., Fang, W., and Luan, S. (2014). A calcium-dependent protein kinase interacts with and activates a calcium channel to regulate pollen tube growth. *Mol. Plant* 7, 369–376. doi: 10.1093/mp/sst125

Zhu, L., Guan, Y. X., Liu, Y. N., Zhang, Z. H., Jaffar, M. A., Song, A., et al. (2020). Regulation of flowering time in chrysanthemum by the R2R3 MYB transcription factor CmMYB2 is associated with changes in gibberellin metabolism. *Hortic. Res.* 7, 96. doi: 10.1038/s41438-020-0317-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 May 2023
DOI 10.3389/fpls.2023.1132307

OPEN ACCESS

EDITED BY
Lin Chen,
Institute of Animal Sciences (CAAS), China

REVIEWED BY
Peng Wu,
Yangzhou University, China
Nanshan Du,
Henan Agricultural University, China
Lijun Ou,
Hunan Agricultural University, China

*CORRESPONDENCE
Haoru Tang
htang@sicau.edu.cn

†These authors have contributed equally to this work

RECEIVED 27 December 2022
ACCEPTED 10 April 2023
PUBLISHED 08 May 2023

CITATION
Li M, Zhang R, Zhou J, Du J, Li X, Zhang Y, Chen Q, Wang Y, Lin Y, Zhang Y, He W, Wang X, Xiong A, Luo Y and Tang H (2023) Comprehensive analysis of HSF genes from celery (*Apium graveolens* L.) and functional characterization of *AgHSFa6-1* in response to heat stress.
Front. Plant Sci. 14:1132307.
doi: 10.3389/fpls.2023.1132307

Comprehensive analysis of HSF genes from celery (*Apium graveolens* L.) and functional characterization of *AgHSFa6-1* in response to heat stress

Mengyao Li[1†], Ran Zhang[1†], Jin Zhou[1†], Jiageng Du[1], Xiaoyan Li[1], Yong Zhang[1], Qing Chen[1], Yan Wang[1,2], Yuanxiu Lin[1,2], Yunting Zhang[1,2], Wen He[1,2], Xiaorong Wang[1,2], Aisheng Xiong[3], Ya Luo[1] and Haoru Tang[1,2*]

[1]College of Horticulture, Sichuan Agricultural University, Chengdu, China, [2]Institute of Pomology and Olericulture, Sichuan Agricultural University, Chengdu, China, [3]College of Horticulture, Nanjing Agricultural University, Nanjing, China

High temperature stress is regarded as one of the significant abiotic stresses affecting the composition and distribution of natural habitats and the productivity of agriculturally significant plants worldwide. The HSF family is one of the most important transcription factors (TFs) families in plants and capable of responding rapidly to heat and other abiotic stresses. In this study, 29 *AgHSFs* were identified in celery and classified into three classes (A, B, and C) and 14 subgroups. The gene structures of *AgHSFs* in same subgroups were conserved, whereas in different classes were varied. AgHSF proteins were predicted to be involved in multiple biological processes by interacting with other proteins. Expression analysis revealed that *AgHSF* genes play a significant role in response to heat stress. Subsequently, *AgHSFa6-1*, which was significantly induced by high temperature, was selected for functional validation. *AgHSFa6-1* was identified as a nuclear protein, and can upregulate the expression of certain downstream genes (*HSP98.7*, *HSP70-1*, *BOB1*, *CPN60B*, *ADH2*, *APX1*, *GOLS1*) in response to high-temperature treatment. Overexpression of *AgHSFa6-1* in yeast and *Arabidopsis* displayed higher thermotolerance, both morphologically and physiologically. In response to heat stress, the transgenic plants produced considerably more proline, solute protein, antioxidant enzymes, and less MDA than wild-type (WT) plants. Overall, this study revealed that *AgHSF* family members perform a key role in response to high temperature, and *AgHSFa6-1* acts as a positive regulator by augmenting the ROS-scavenging system to maintain membrane integrity, reducing stomatal apertures to control water loss, and upregulating the expression level of heat-stress sensitive genes to improve celery thermotolerance.

KEYWORDS

celery, heat shock transcription factor, heat stress, function annotation, transgenic, transcriptional regulation

1 Introduction

Celery (*Apium graveolens* L.) is a valuable edible and medical vegetable crop belonging to the Apiaceae family that originated in the Mediterranean and the Middle East and is now extensively cultivated and consumed worldwide (Li et al., 2014; Li et al., 2018). However, celery cannot endure high temperatures; its optimal germination and growth temperatures are 15-20 °C and 15-25 °C, respectively (Richard, 2004). Because of this, heat stress negatively affects celery growth and yield. Increasing the heat tolerance of celery can materialize its annual supply and significantly improve its quality and production. Plants have developed a variety of defense strategies to relieve abiotic stresses, including morphological mechanisms (Zhao et al., 2022), plant hormone regulation (Waadt et al., 2022), and modulating the expression of stress-responsive genes (Zhou et al., 2019). Previous studies have shown that some transcription factors (TFs), such as bZIP, DREB, MYB and HSF, are the key regulators of several horticultural plants response to various abiotic and biotic stress (Kumar et al., 2021; Li et al., 2022b).

Heat shock transcription factors (HSFs) are one of the plant's most important transcription factors (Fan et al., 2021). The HSF family members can respond to heat stress and other abiotic stresses, including cold, salt, and drought, by binding to the reverse repeat region of heat shock elements (HSEs) and promoting transcription of heat shock proteins (HSPs) (Guo et al., 2016; Jacob et al., 2017). Although the members of the HSF family have different functions, their structures are highly similar (Chen et al., 2018). Typically, plant HSF protein has a conserved patterned structure, including a DNA-binding domain (DBD) at the N-terminus, two hydrophobic 7-peptide repeats (HR-A/B region) oligomerization domain (OD), nuclear localization signal (NLS), nuclear export signal (NES), activation structure (AHA motif), and repressor domain (RD) (Guo et al., 2016; Jiang et al., 2021). The HSF family in plants has been further classified into three classes (HSFA, HSFB, and HSFC) based on the length of their basic amino acid (aa) sequences between the DBD and HR-A/B regions and the number of amino acid residues inserted into the HR-A/B regions (Nover et al., 2001; Liu et al., 2018).

The study demonstrated that HSFA members played a primary role in responding to heat stress, while HSFB assists HSFA in its response to heat (Chan-Schaminet et al., 2009). Moreover, study has found that the HSFA genes tends to regulate the expression of downstream related HSP genes to improve plant thermotolerance, while the HSFB genes negatively regulates the expression of HSP genes (Wang et al., 2021). In recent years, several studies have supported that HSFA genes exhibit similar role in heat stress. For example, overexpression of *AtHsfA1a* in *Arabidopsis* can enhance heat stress tolerance in transgenic plants (Qian et al., 2014). The overexpression of maize *ZmHsfA2* in the *Arabidopsis hsfa2* mutant restored heat tolerance and raised the survival rate after heat stress (Li et al., 2019). However, much less is known about the function of HSFB and HSFC genes. It is demonstrated that HSFB inhabited the growth of roots and aerial organs in transgenic *Arabidopsis*, and could help plants acquire thermotolerance (Tan et al., 2021). The studies presented above suggest that some HSFs played critical roles in enhancing heat tolerance in plant.

Although HSF transcription factors have been previously found in celery, this study only identified 17 HSF genes and also did not determine their functions (Li et al., 2018). The response network studies on HSF in celery are still limited. In the present study, 29 *AgHSF* genes were identified in celery, and *AgHSFa6-1* was isolated for further investigation of its activities under heat stress by transferring into *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. Transgenic lines displayed higher thermotolerance than wild-type (WT)in morphologically and physiologically, as shown by increased antioxidant enzyme activity, proline content, solute protein, etc. These results revealed that overexpression of *AgHSFa6-1* enhanced the tolerance of transgenic plants mainly through boosting the ROS-scavenging system, osmoregulation, and a specific molecular mechanism. In conclusion, this study established a response network of *AgHSFa6-1* in maintaining heat tolerance, which provide a theoretical basis for celery breeding and cultivating.

2 Materials and methods

2.1 Identification of HSF family in celery and phylogenetic relationship analysis

Genome sequences were downloaded from published celery genome database (http://apiaceae.njau.edu.cn/celerydb and http://celerydb.bio2db.com/) for constructing the local BLAST (Li et al., 2022b). The hidden Markov model profile of the HSF domain (PF00447) from the Pfam database (http://pfam.xfam.org/) was downloaded and utilized to identify the HSF members in celery using the HMMER 3.0 program with default parameters. Candidate proteins were further submitted to NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd) to confirm the conserved domains. The physicochemical properties of the HSF genes identified from celery were analyzed using online ExPASy software (http://web.expasy.org/protparam). The subcellular localization was predicted with WOLF PSORT (https://www.genscript.com/tools/wolf-psort) and SoftberryProtComp 9.0 (http://www.softberry.com). The HSF sequences of *Arabidopsis* were downloaded from TAIR (https://www.arabidopsis.org/). ClustalX (v2.1) was used to align multiple sequences of all HSFs with default parameters, and phylogenetic tree was constructed by MEGA7.0 using neighbor-joining (NJ) method with 1000 bootstrap replicates.

2.2 Chromosomal distribution and gene structure, and conserved motif analysis of AgHSFs

The chromosomal location of *AgHSFs* was visualized through TBtools software (Chen et al., 2020). The exon/intron structure of *AgHSFs* was analyzed by online tool Gene Structure Displayer

Server v2.0 (http://gsds.gao-lab.org/). Conserved motifs were analyzed using MEME Suite v5.4.1 with the parameters set as follows: the maximum number of motifs, 20; and the optimum width of each motif, between 10 and 50 residues.

2.3 Promoter region and interaction network analysis

The *cis*-regulatory elements in the 1.5 kb promoter regions of *AgHSFs* were predicted using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) (Lescot et al., 2002). The STRING software was used to output the protein interaction value data between *AgHSFs* and other proteins in celery genome, and established protein interaction network. Results were visualized with Cytoscape software v3.7 (Shannon et al., 2003).

2.4 Gene cloning and subcellular localization

The heat-tolerant variety 'Jinnan Shiqin' and the non-heat-tolerant variety 'Liuhe Huangxinqin' were used as plant materials. Celery seeds were placed in an incubator at 20 °C in dark for germination, and then transferred them into growth chambers at 25 °C/20 °C (14 h light/10 h dark cycles). When the plants reached the stage of about 10 leaves, they were transferred into 38 °C for 24 h heat stress. The celery leave samples were collected after at 0 h (QC1), 4 h (QC2), 12 h (QC3), and 24 h (QC4) of treatment. Plant RNA samples were extracted using a total RNA kit (Tiangen, Beijing, China), and then converted into cDNA using PrimeScript RT Kit (TaKaRa, Dalian, China). The sequences of *AgHSF* genes were amplified for RT-PCR using *EasyTaq* DNA Polymerase Kit (TransGen Biotech, Beijing, China). The PCR product was ligated with *pEASY*-T1 vector and sequenced in Tsingke Biotechnology Co., Ltd. (Beijing, China).

The full-length of *AgHSFa6-1* was amplified using specific primers (Table S1), and inserted into green fluorescent protein (GFP)-fusion expression vector pA7. The expression vector *35S*:*AgHSFa6-1*:*GFP* was constructed with two *Bam*H I restriction enzyme sites. The empty vector containing 35S::GFP protein served as a control. The fusion expression and empty vector were transformed in tobacco leaves by DNA particle bombardment system (PDS-1000, Bio-Rad, USA). The tobacco was firstly disposed in dark for 24 h, fluorescence was then observed using a laser confocal microscope (LSM800, Zeiss, Germany).

2.5 Transformation in yeast and Arabidopsis and stress tolerance analysis

The full-length open reading fragments of *AgHSFa6-1* were amplified by PCR using specific primers, and inserted into the pYES2 vector (Table S1). Then the recombinant vector was transformed into *Saccharomyces cerevisiae* strain *INVSC1*, and positive transformants were selected in the SD-Ura medium at 30 °C. For heat stress treatment, the positive yeasts were cultured on the SG-Ura medium for 3 d at 30 °C, 39 °C, 41 °C and 45 °C, respectively.

For the overexpression of *AgHSFa6-1*, the double-cauliflower mosaic virus 35S (CaMV 35S) promoter and NOS terminator from pSAT1-cEYFP-N1 were amplified and subcloned into the vector pCAMBIA1301. To obtain transgenic *Arabidopsis*, the amplification product was inserted downstream of the double CaMV 35S promoter into the modified pCAMBIA1301 vector between *BamH* I and *Sac* I sites. The construct contains the CaMV 35S promoter driving *AgHSFa6-1* and the CaMV 35S promoter driving the *Hyg* gene for hygromycin resistance as a selectable marker. The pCAMBIA1301 vector also contained a GUS reporter gene following CaMV 35S promoter. The recombinant vector was introduced into the *Agrobacterium tumefaciens* strain GV3101 through electroporation. Then, it was transformed into *Arabidopsis* Columbia (Col-0) wild-type (WT) plants by floral-dip method (Zhang et al., 2006). The transgenic T_0 lines were selected using hygromycin on MS medium, and the hygromycin-resistant plants were selected out to obtain T_1 seeds for PCR assays. Then, consistently selected twice to harvest T_2 seeds of transgenic *Arabidopsis*, and the T_3 homozygous lines were used for further verification.

Seeds of *Arabidopsis* were placed on 1/2 MS medium to germination. The seeds were vernalized in a dark incubator at 4 °C for three days and then changed the growth condition at 25 °C/15 °C (16 h light/8 h dark cycles) and 70% relative humidity. Ten-day-old seedlings were subjected to heat stress for three days at 38 °C/25 °C (16 h light/8 h dark cycles), and then measured the primary root length. Two-month-old seedlings were subjected to 38 °C/25 °C heat stress for 24 h. They were sampled at 0 h, 4 h, 12 h, and 24 h for RNA extraction utilized for subsequent RT-qPCR analysis. Moreover, leaf-samples of 4 h were collected for stomata analysis and samples of 24 h were collected for physiological measurements. Then plants were domesticated at 42 °C for two days to compare their phenotypic difference.

2.6 Measurements of physiological parameters

The free proline content was measured according to the acidic-ninhydrin-based colorimetric method (Bates et al., 1973). The activities of superoxide dismutase (SOD) and peroxidase (POD) were measured using the nitrogen blue tetrazolium (NBT) and the guaiacol methods, respectively (Macadam et al., 1992; Cao et al., 2018). The concentration of malondialdehyde (MDA) in leaves was measured by the TBA colorimetric method (Li et al., 2020). For stomata analysis, the plant samples were immersed into FAA fix solution, dehydrated, embedded in paraffin, sectioned, and stained with toluidine blue. The width and length of the stomatal aperture were analyzed using ImageJ software v1.8.0.112 (Graphics software; NationalInstitutes of Health: Wayne Rasband, USA, 1997).

During above experiment, three independent replicates of each transgenic lines were performed. All of the data were analyzed using SPSS 26.0 software (SPSS Inc., Chicago, IL, USA), and the significance levels were based on $P < 0.05$ (*) and $P < 0.01$ (**).

2.7 Gene expression and reverse transcription-quantitative PCR (RT-qPCR) analysis

The transcriptome data of celery under variable high temperature was obtained by this research group (Li et al., 2022a), and the transcript abundance of *AgHSF* was calculated in log2(FPKM) to generated cluster heatmaps.

RT-qPCR was performed using SYBR Premix Ex Taq (TaKaRa, Dalian, China). All the steps were followed the manufacturer's instruction (CFx384TM Real-Time System, Bio-Rad, USA), and the expression level was calculated by the $2^{-\triangle\triangle Ct}$ method (Pfaffla, 2001). *AgTUB* and *AtACT2* were served as reference genes for celery and *Arabidopsis* genes, respectively. All gene primer sequences used in this study were designed using Primer Premier software (version 6.0; Premier Biosoft International: Palo Alto, CA), and the sequences are listed in Table S1.

3 Results

3.1 Identification of HSF family genes in celery and phylogenetic analysis

A total of 29 *AgHSF* genes were identified and re-named from *AgHSFa6-1* to *AgHSFc1-1* (Table S2). The results demonstrated that almost all the genes were localized in the nucleus. The polypeptides of AgHSF proteins ranged from 232 (AgHSFa9-2)-503 aa (AgHSFa1-3), with the molecular weight ranging from 26450.87 (AgHSFa9-2)-55586.82 Da (AgHSFa1-3). The aliphatic amino acids accounted for 18%, whereas the aromatic amino acids accounted for only 8%. The theoretical isoelectric point (*pI*) of proteins belongs to class A ranged between 4.83-6.25, indicating that they were acidic. However, except for B2, the *pI* values of class B proteins were more than 7, suggesting they were alkaline. All AgHSF proteins had negative GRAVY values, which means that all proteins were hydrophilic.

A phylogenetic tree was created between 29 celery and 20 *Arabidopsis* HSF proteins to determine their categorization and explore their evolutionary relationship (Figure 1). The result revealed that the relationships of HSF genes between the two

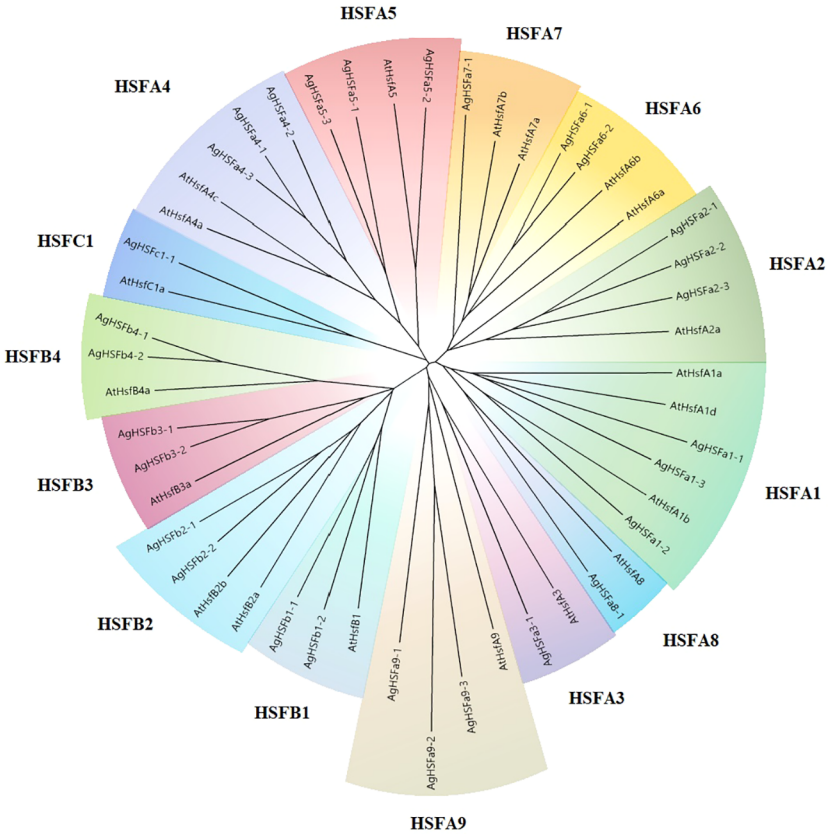


FIGURE 1
The phylogenetic tree of the HSF proteins in celery and *Arabidopsis*. The phylogenetic tree was generated using the NJ method with 1000 bootstrap replicates. Different colors and letters represent various subgroups.

species were highly homologous. The AgHSFs in celery were classified into three classes and 14 subgroups. The class A had the most abundant proteins, which were clustered into nine subgroups, corresponding to HSFA1-HSFA9, and class B were clustered into four subgroups, corresponding to HSFB1-HSFB4. In contrast, the class C only had one subgroup for one gene, corresponding to HSFC1.

3.2 Gene structure analysis

The evolution of gene family members was mainly determined by the variation of their conserved motifs and the diversity of gene structures (Hu et al., 2010). Results revealed that the longest intron was nearly 4 Kb (*AgHSFa5-2*), and all *AgHSF* genes had between two and four exons and one to three introns (Figure 2A). However, the types and numbers of motifs varied considerably (Figure 2B). A total of 20 motifs were detected in AgHSF proteins; the sequence of each motif is listed in the supplementary table (Table S3). All 29 AgHSFs shared only one motif type (motif 4), and the number of motifs varied from two to 11. Similarities were more prevalent among the same groups or classes. AgHSFs belongs to Class A usually had more motif types, and except for AgHSFa9-2, they had motifs 1 and 3 in common, and members belongs to class B had motifs 1, 4, 5, and 8 in common. Moreover, motifs were always arranged in a particular order, such as motifs 2- 1- 4- 3 and 9. The highly conserved exon-intron distribution and motif divergence indicated that AgHSFs evolved according to a group, establishing the basis for the functional diversity of different classes.

3.3 Chromosomal distribution and cis-regulatory elements of AgHSFs promoter regions

To identify the localization of *AgHSF* genes, the TBtools software was used to visualize the location of these genes on different chromosomes in the celery genome (Figure 3). Result demonstrated that except for Chr01, 28 *AgHSF* genes were unevenly distributed between Chr02 and Chr11, whereas *AgHSFa8-1* was located at an unanchored contig. The Chr04 and Chr10 had the largest number of genes (four genes), followed by Chr02, 06, 09, and 11 with three genes each. Genes belongs to class A were located on each chromosome from Chr02 to 11, while the class B genes were only present on Chr02, 03, 04, 05, and 10, and the C class genes were only located on Chr10. Some *AgHSF* genes demonstrated high cross collinearity, clustering at similar sites on the same chromosome. For instance, *AgHSFa3-1* and *AgHSFa4-3* were localized at a similar site on the same chromosome, revealing that they might have sequence similarity.

Transcription factors can regulate gene expression by binding to cis-regulatory elements in response to different biotic or abiotic stress signals (Li et al., 2021). Various cis-regulatory elements were predicted in promoter sequences of all the *AgHSF* genes, and they were divided into three groups based on their functions: plant growth and development, phytohormone responsive, and abiotic and biotic stress (Figure 4B). Result revealed that some elements were widely distributed in promoter regions of *AgHSFs*, with light response elements (Box 4, G-box, and GT1-motif) being the most frequently distributed, followed by antioxidant response elements (ARE) and stress-response element (STRE). Furthermore, abscisic

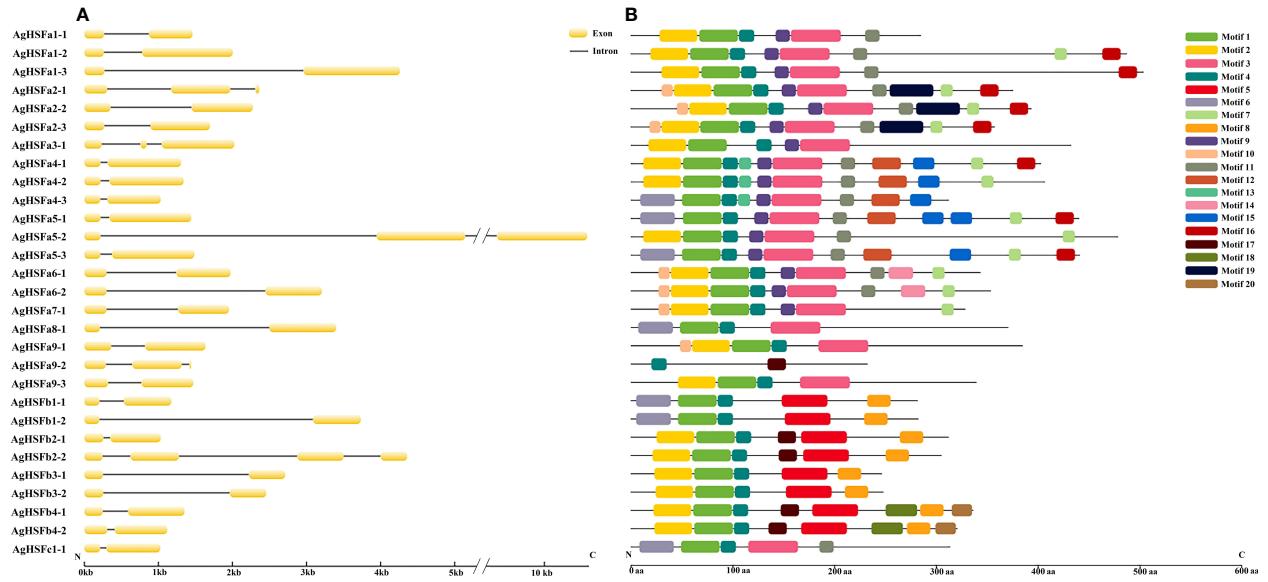


FIGURE 2
Analysis of gene structure and conserved motifs of *AgHSF* genes. **(A)** The exon and intron structure of *AgHSF* genes. The nucleotide sequence of these genes was used for analysis, and box represented exons and line represented introns. **(B)** Display of conserved motifs in AgHSF proteins. The protein sequences were subjected to Motif Search based on homology, and boxes represented motifs and different colors represent different types of motifs.

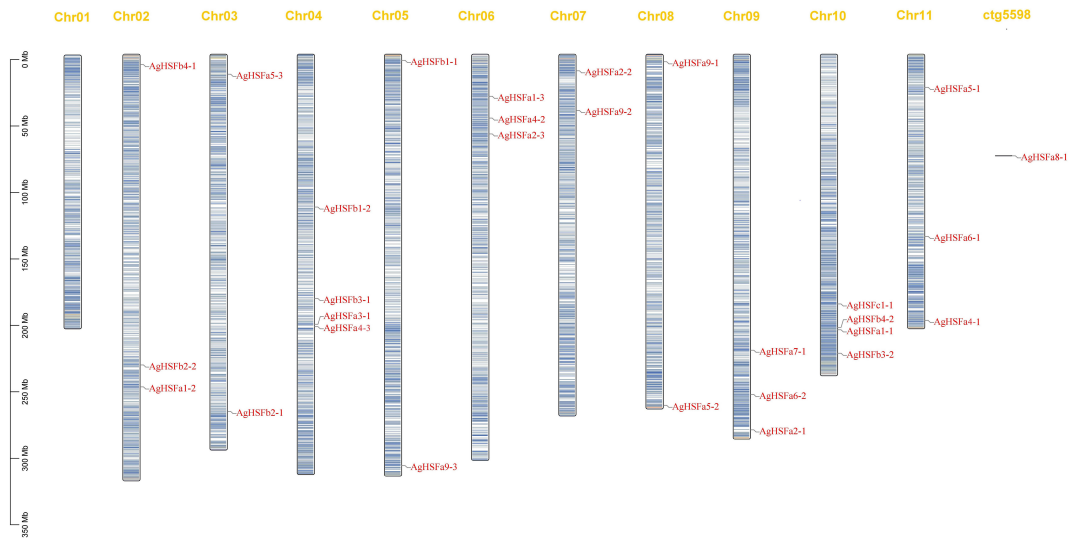


FIGURE 3
Chromosome location of *AgHSF* genes. The bars represented chromosomes, and chromosome numbers was indicated at the top of each bar with length marked at the left of whole figure.

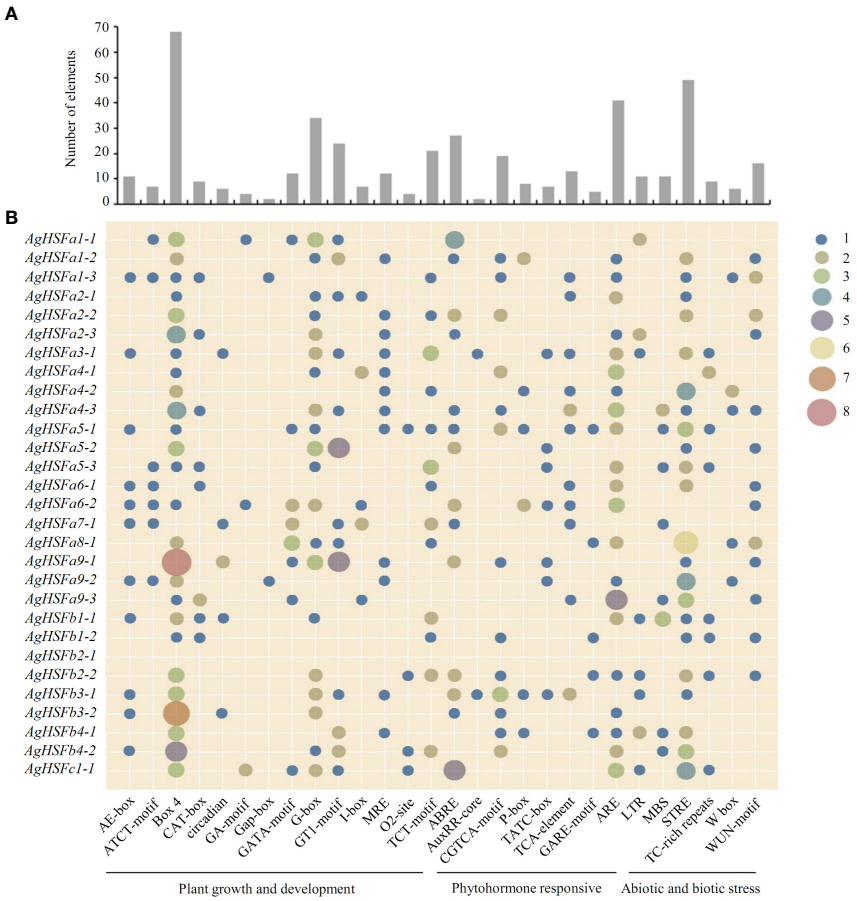


FIGURE 4
Number and distribution of cis-regulatory elements in the promoter region of *AgHSF* genes. **(A)** The total number of each element. **(B)** The distribution of elements in 29 *AgHSFs* promoter regions. The numbers (from 1 to 8) of each element were shown as different types of dots, and the bigger dot meant bigger number.

acid response elements (ABRE) were also detected in promoter regions of several *AgHSFs* (Figure 4A), indicating these genes can regulate the abscisic acid (ABA) secretion or synthesis process in response to heat shock.

3.4 Protein interaction network analysis of AgHSFs proteins

To further investigate the regulatory mechanism of HSF genes in celery, a protein interaction network comprising the homologous proteins in *Arabidopsis* was constructed (Figure 5). Except for AgHSFa9-2, 28 AgHSF proteins and 15 other proteins in the celery genome were present in this interaction network. Out of these, three HSP proteins constituted the central node and were connected with multiple AgHSF proteins. AgHSF proteins also exhibited significant relationships with DREB, PBS, MPK, and SQN proteins. HSF genes were proved to respond to heat shock by regulating the expression of HSP (Jacob et al., 2017). In our study, members of HSF family and HSP formed an expansive regulation network, indicating that these genes may play important roles in sensing and responding to heat stress.

3.5 Transcript abundance and expression pattern analysis of AgHSFs under heat stress treatments

According to the transcriptome data, all the 29 *AgHSF* genes were expressed after 0 h, 4h, 12 h, and 24 h of heat stress, referring to QC1, QC2, QC3, and QC4, respectively. Previous studies demonstrated that *HSFA2* genes in *Arabidopsis* and tomato could be induced by heat stress, resulting in enhanced permeability and antioxidant capacity of plants (Meiri et al., 2010; Fragkostefanakis et al., 2016). Several *AgHSFs*, including *AgHSFa2-2*, *AgHSFa2-3*, and *AgHSFa6-1*, displayed a typical trend of initially increasing and declining. The expression levels of *AgHSFa6-1* increased significantly at QC2 and reached their peak at this stage but decreased during QC3 and QC4 (Figure 6A). Despite this, their expression levels were significantly higher than any other *AgHSF* genes.

Nine highly expressed *AgHSFs* (*AgHSFa2-1*, *AgHSFa2-2*, *AgHSFa2-3*, *AgHSFa6-1*, *AgHSFa6-2*, *AgHSFb1-1*, *AgHSFb1-2*, *AgHSFb2-1*, *AgHSFb2-2*) were further detected by RT-qPCR to validate their expression levels in two celery cultivars (heat-tolerant celery variety 'Jinnan Shiqin' and the non-heat-tolerant celery variety 'Liuhe Huangxinqin') (Figure 6B). The expression levels of

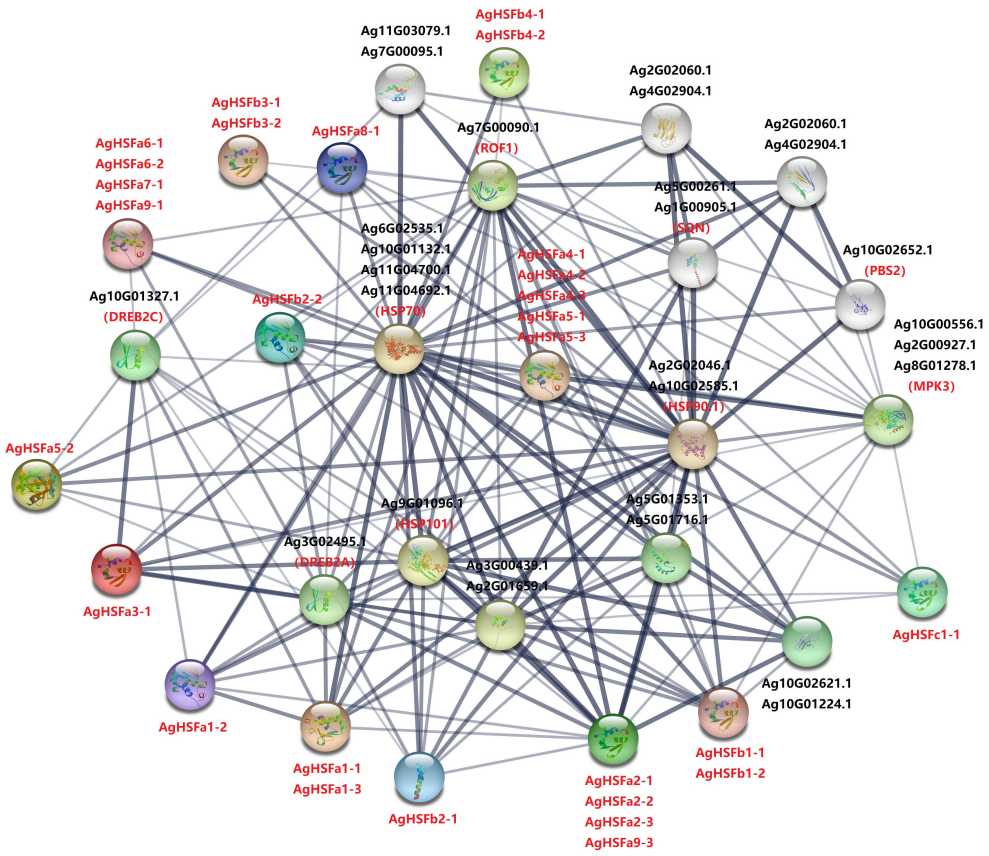


FIGURE 5
Interaction network of AgHSF in celery according to the orthologs in *Arabidopsis*. Every circle represents different protein, and all the letters are the name of proteins. The width and shade of lines represent the correlation strength of each proteins.

10.3389/fpls.2023.1132307

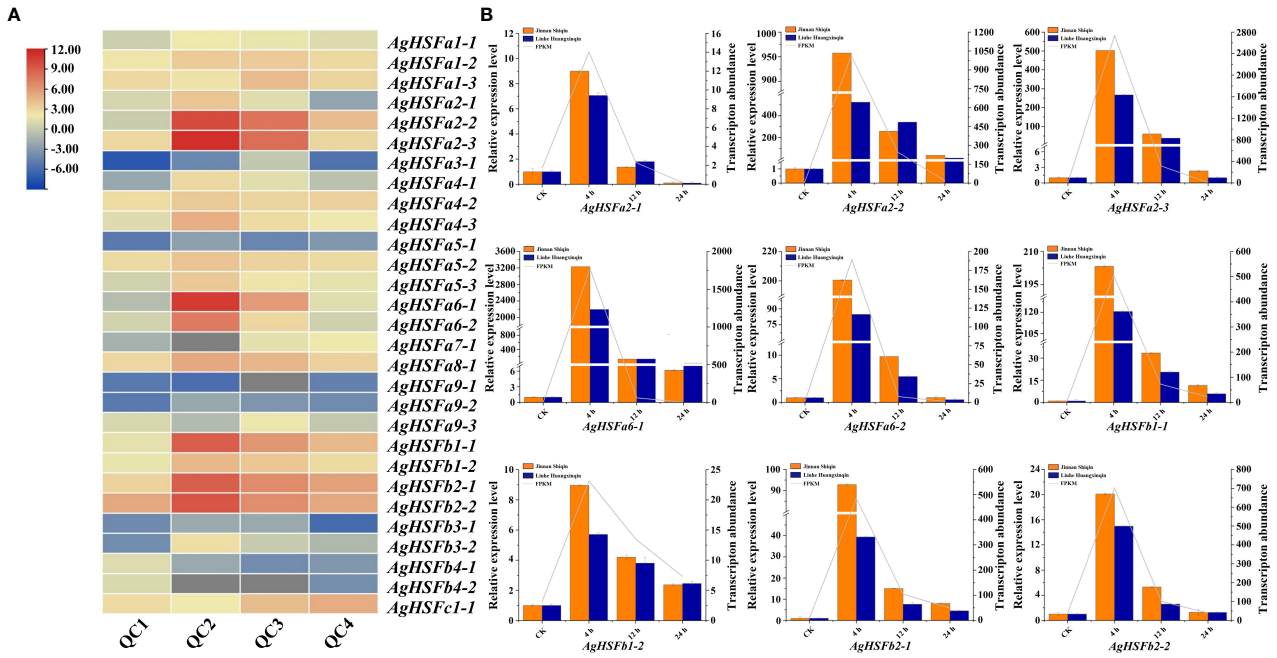


FIGURE 6
Expression pattern and expression level analysis of *AgHSF* genes. **(A)** Expression pattern of 29 *AgHSFs* under heat stress treatment based on RNA-seq data. QC1, QC2, QC3, and QC4 represented the samples treated at high temperature for 0 h, 4 h, 12 h, and 24 h, respectively. The expression values were normalized to reads per kilobase per million mapped reads (FPKM), and the data was calculated using log2 (FPKM). **(B)** Expression level analysis of nine highly expressed *AgHSFs*. Different colors represented different celery varieties, and line represented transcription abundance.

nine *AgHSFs* were all exhibited higher expression level in 'Jinnan Shiqin' after 4, 12, 24 h heat stress, especially *AgHSFa6-1* exhibited a thousand times higher expression level than CK after 4 h heat stress. This result revealed that *AgHSFa6-1* were involved in the response to heat stress, and the heat-tolerant celery variety could response to heat stress more rapidly to adapt to high temperature. Overall, similar expression trends in gene expression were detected from transcriptome data and RT-qPCR analyses.

3.6 Gene clone and subcellular localization of AgHSFa6-1

Result demonstrated that the control was strongly expressed in the cytomembrane and nucleus, whereas 35S:AgHSFa6-1:GFP was detected only in the nucleus, indicating that the AgHSFa6-1 protein was targeted in the nucleus (Figure 7). The result was consistent with previous subcellular localization predictions based on protein

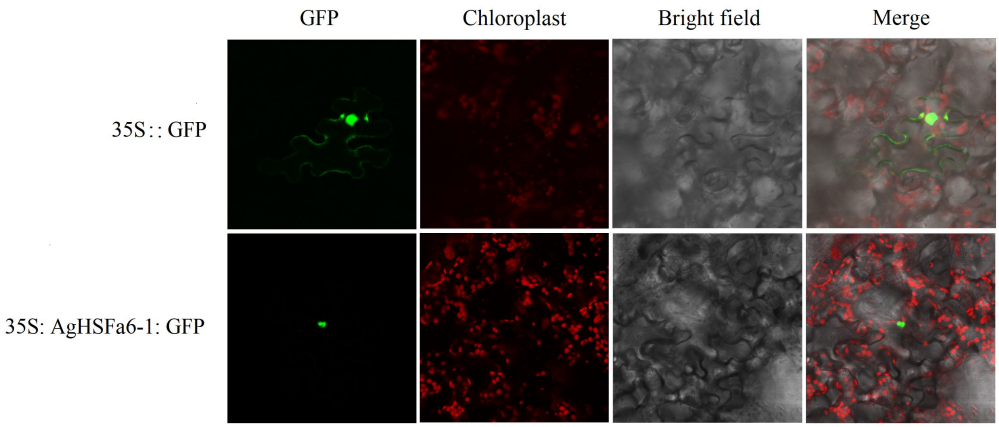


FIGURE 7
Subcellular localization of AgHSFa6-1 proteins in tobacco cells. Green signal represented the localization of AgHSFa6-1proteins.

structures and demonstrated that AgHSFa6-1 could execute the function of a transcription factor.

3.7 Analysis of AgHSFa6-1 overexpression in yeast and A. thaliana

The yeasts were cultured on SG-Ura medium at 30, 39, 41, and 45 °C for three days (Figure 8A). Result demonstrated that the control strain only grew at 30 °C, while the transgenic strains grew significantly at 39 and 41 °C. However, at the temperature of 45 °C, neither the control nor transgenic strains grew, suggesting that the *AgHSFa6-1* gene responds to temperature stresses in yeast.

The transgenic *Arabidopsis* plants were obtained using hygromycin selection, and their validity was verified by PCR and GUS staining (Figure S1). A total of 25 transgenic lines were obtained and designated as from OE-1 to OE-25. Two lines of transgenic lines (OE-17 and OE-19) and WT plants were selected to verify the *AgHSFa6-1* function in response to heat stress. The morphological and root-length differences were measured using ten-days-old and two-month-old seedlings. As illustrated in Figures 8B, C, under normal growth conditions, OE-17, OE-19, and WT grew consistently, whereas after 24 h of continuous heat stress at 38/25 °C, the WT plants visibly withered to death, but the transgenic seedlings were still alive. Additionally, the growth of WT seedlings was stalled by heat stress, whereas OE-17 and OE-19 seedlings survived, and the average root length of transgenic plants was significantly longer than that of WT (Figures 8C, G).

Stomata are the main channels for gas exchange, and plants normally adjust the stomatal conductance to vary their temperature (Wang et al., 2022). When exposed to 4 h of heat stress, the conductance of WT plants was more open than that of transgenic plants, and both the aperture size and length of WT stomata were significantly larger than those of transgenic stomata (Figures 8D–F, H, I). These findings demonstrated that overexpression *AgHSFa6-1* enhanced the heat-endurance ability of transgenic plants by regulating stomatal conductance and opening to control water loss.

3.8 Physiological and gene expression analysis of transgenic and wild-type A. thaliana plants under heat stress

Leaf samples from two-month-old *Arabidopsis* plants were used to measure all physiological parameters. The physiological indices of WT and transgenic plants under heat stress were determined,

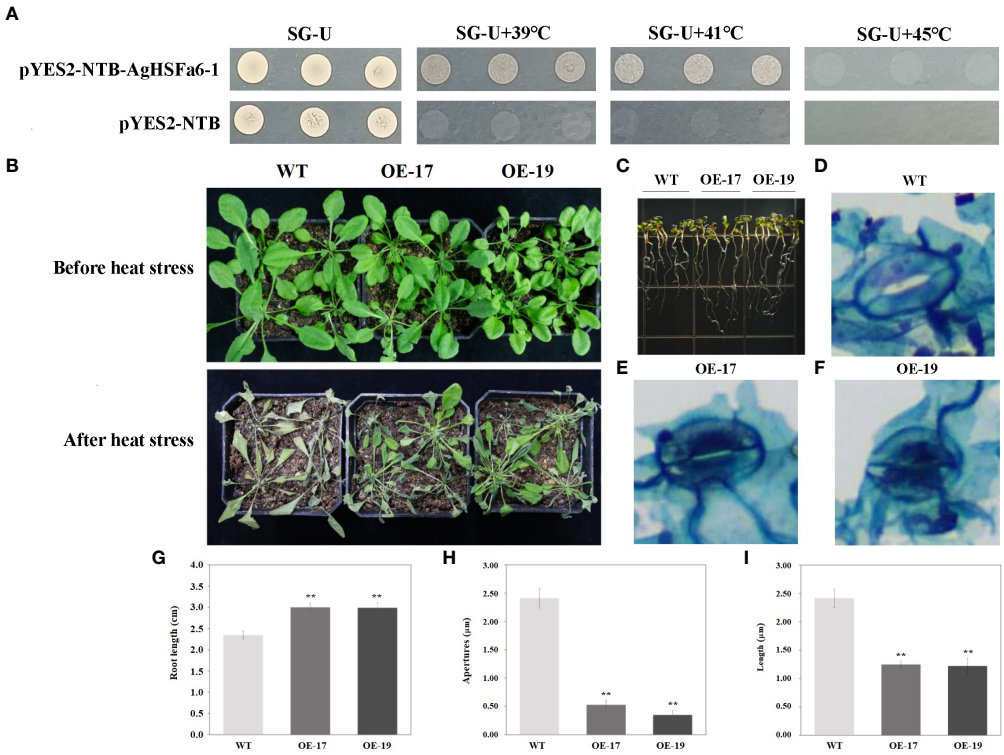


FIGURE 8
The transgenic results of overexpression *AgHSFa6-1* in yeast and *Arabidopsis*. **(A)** The growth pattern of control strain (pYES2-NBT) and recombinant strain (pYES2-NBT-AgHSFa6-1) on SG-Ura medium. The yeasts grew at different temperatures for 3 **(D) (B)** Morphological differences between transgenic and WT plants under heat stress. **(C, G)** Root-length differences between transgenic and WT plants. **(D–F, H, I)** Electron microscope photos of stomata after heat stress, and statistics of stomata apertures and length. Average root-length, stomata apertures and length of transgenic and WT plants analyzed by SPSS 26.0. ** represented the differences were significant.

including proline, MDA, soluble protein, and antioxidant enzymes (SOD and POD). Results demonstrated that the content of proline in transgenic plants were significantly higher than in WT plants, while the MDA content displayed the opposite trend (Figures 9A, B). Proline is a signal molecule that regulates osmotic pressure, and MDA reflects the degree of membrane damage (Liu et al., 2016; Fu et al., 2018). Examining soluble protein content revealed that the average values of OE-17 and OE-19 were higher than those of WT plants, consistent with the proline level (Figure 9C). Studies indicated that heat stress could promote the secretion of ROS, and antioxidant enzymes like SOD and POD have ROS-scavenging mechanisms (Wang et al., 2003; Gechev and Petrov, 2020). SOD and POD in transgenic plants were significantly higher than in WT plants following a heat treatment (Figures 9D, E). These results suggest that transgenic plants are more tolerant to high-temperature stress than WT plants due to overexpression of *AgHSFa6-1*.

To understand the regulatory function of *AgHSFa6-1* in wild-type and transgenic *Arabidopsis* plants under heat stress, the expressions of *AgHSFa6-1* and seven stress-responsive genes in two lines of transgenic *Arabidopsis* were measured using RT-qPCR (Figures 9F–N). Both *AgHSFa6-1* and stress-responsive genes showed similar expression trend in OE-17 and OE-19. The expression levels of seven stress-responsive genes showed that *AtHSP98.7* and *AtAPX1* at all stages of heat stress were higher than those of CK. However, expression levels of *BOB1*, *AtCPN60B*, *AtHSP70-1*, *AtADH2*, and *AtGOLS1* were insignificantly higher than WT during 0-4 h, but expression levels of *AtCPN60B*, *AtHSP70-1*, *AtADH2* were even lower than WT after 12 h. The expression levels of *AtHSP98.7* initially increased between 0 and 4 h, then decreased between 4 and 8 h, and peaked at 24 h. On the contrary, the expression levels of *AtCPN60B* and *AtADH2* were initially decreased and then increased. These genes were significantly upregulated by high-temperature stress and reached their peak expression at 24 h. Moreover, their expression levels were much higher in the transgenic plants than in WT plants. Expression trend of *AgHSFa6-1* in OE-17 and OE-19 was consistent with both transcriptome data and expression analysis in celery, and transcription and translation were relatively faster compared to the other seven genes. This revealed that *AgHSFa6-1* improved the resistance of transgenic *Arabidopsis* to heat stress by positively regulating the expression of abiotic stress-sensitive genes.

4 Discussion

Plant HSF TFs are key regulatory network elements and influence the responses to abiotic and biotic stresses. In this study, 29 *AgHSF* genes were identified in celery. HSFs are divided into three classes and could be divided into subgroups based on different protein structures (Kotak et al., 2010).

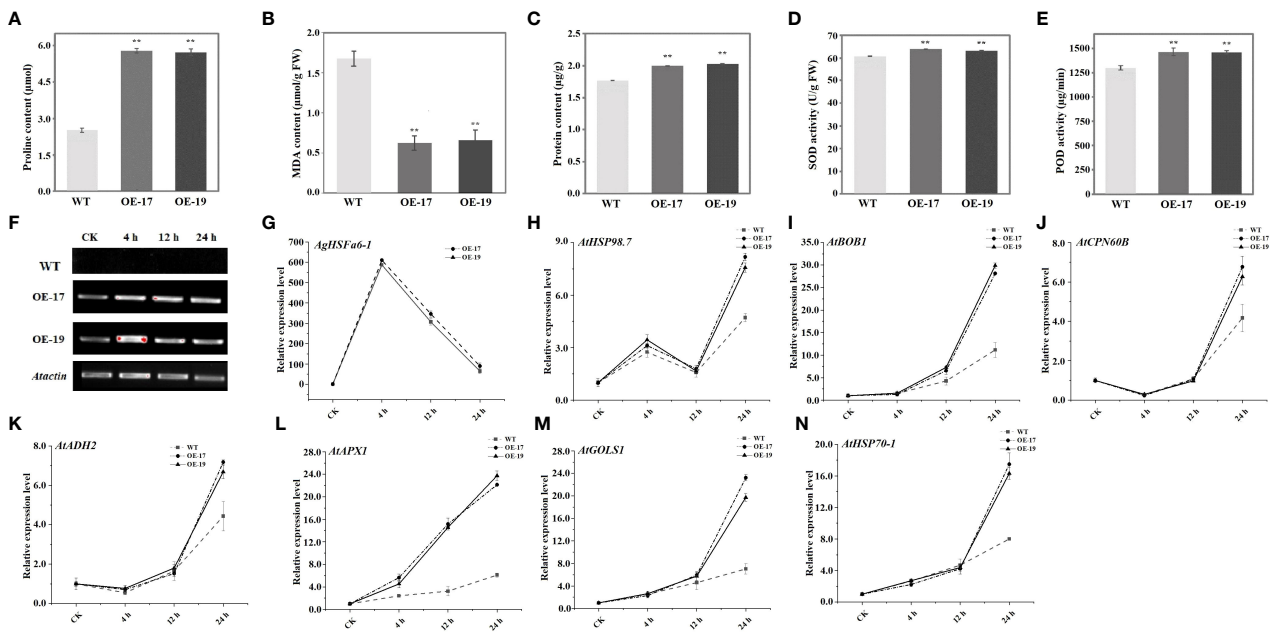


FIGURE 9
Analysis of physiological traits in transgenic and WT plants and expression profiles of heat-related genes under heat stress treatment. ** represented the differences were significant. **(A)** Proline contents in two transgenic and WT plants measured after 24 h heat stress. **(B)** MDA contents in in two transgenic and WT plants measured after 24 h heat stress. **(C)** Soluble protein contents in two transgenic and WT plants measured after 24 h heat stress. **(D, E)** Activity of SOD and POD enzyme in in two transgenic and WT plants measured after 24 h heat stress. **(F, G)** The expression levels of *AgHSFa6-1* in WT, OE-17 and OE-19. **(H–N)** The expression levels of stress-responsive genes in WT, OE-17 and OE-19. Vertical axis represented expression levels of each genes, and horizontal axis represented stages (CK was control).

According to the HSF sequences of *Arabidopsis*, the phylogenetic tree formed three distinct classes: A, B, and C. Previous studies revealed that the HSFs belongs to class A were the main regulatory factors during heat stress, whereas the classes B and C members had no transcriptional activity (Takumi et al., 2011; Hu et al., 2015). Numerous studies proved that HSFA2 and HSFA7 are the most heat-sensitive TFs in the *HSF* gene family (Zha et al., 2020); they respond to heat stress by regulating plant hormone induction (Garima et al., 2021), protein synthesis (Wang et al., 2010), and ROS signaling pathway (Chiara et al., 2012). According to the phylogenetic relationships, related genes are always distributed on the same branch of the phylogenetic tree. In this study, phylogenetic analysis revealed that the HSF genes formed 14 subgroups, and the A2, A6, and A7 HSF genes formed an individual branch. The expression level of *AgHSFa6-1* was induced by heat stress, and the transcript abundance trend was similar with *AgHSFa2-2* and *AgHSFa2-3*. However, wether genes of A2 and A6 have similar functions needs further verification. Promoters are one of the most important regulatory elements, determining the temporal and spatial expression of developmentally and physiologically significant genes (Mohsen and Huang, 2022). Several cis-regulatory elements were identified in the *AgHSFa6-1* promoter sequence, especially ARE and STRE elements, that are related to abiotic stress.

High temperature mainly causes the breakdown of the plant's membrane system, causing abnormal metabolism of water, ions, and organic solutes (Franklin, 2009). Considering that stomata are one of the most critical channels in plants for exchanging carbon dioxide (CO_2) and water vapors (H_2O), it is vital to regulate stomatal response to limit plant water loss (Lin et al., 2022; Djemal and Khoudi, 2021). It was reported that stomatal closure is generally triggered *via* the accumulation of ABA and proline (Lim et al., 2015; Maria et al., 2021). Overexpression of *VaHsfC1* in grapes enhances the heat tolerance of transgenic plants by increasing ABA content and decreasing stomatal aperture (Jiao et al., 2022). The present study also revealed that the stomatal apertures of transgenic *Arabidopsis* overexpressing *AgHSFa6-1* under heat stress was significantly lower than that of WT plants, and the proline content was significantly higher than that of WT plants. Previous research demonstrated that proline participated in the ABA metabolism pathway during abiotic stress (Sripinyowanich et al., 2013), indicating that overexpression of the *AgHSFa6-1* gene could enhance thermo-tolerance by promoting proline synthesis and ABA to regulate stomata, although the underlying mechanism remains unknown.

High temperature could increase cellular antioxidant enzyme activity, proline, etc. Plants adapted to a high-temperature environment by improving the process of water metabolism and membrane stability (Liu et al., 2020; Wu et al., 2022). Excessive ROS, particularly H_2O_2, is the primary cause of membrane damage, and MDA content is a widely used marker to assess membrane damage under abiotic stress (Zhang et al., 2022). In our research, all the transgenic strains and plants exhibited better thermo-tolerance in physiologically. The transgenic yeast remained alive at 41 °C, but the control yeast stopped growing under 39 °C. Additionally, after 24 h of heat stress, transgenic plants suffered less membrane injury due to the lower level of MDA than the WT plants. High levels of solute protein content and activity of SOD and POD enzymes were detected in transgenic plants, indicating that overexpression of *AgHSFa6-1* could increase the ability of the plant's ROS-scavenging system to enhance heat tolerance.

Previous studies have revealed that in response to heat stress, the synthesis of certain amino acids (such as serine and proline) and certain protein like HSPs were upregulated, indicating that these synthesis-upregulated amino acids and proteins may be a response to high temperature in plants (Wang et al., 2018; Ren et al., 2019). Numerous studies clarified these functional proteins. Overexpression of the *APX1* gene in *Arabidopsis* can induce the expression of ROS-scavenging related gene for response against abiotic stress (Jiang et al., 2017). *CPN60B* encodes RuBisCO and is involved in photosynthesis (Cheng et al., 2020). BOB1 is a noncanonical heat shock protein that regulates development and heat response, whereas the ABA signal can induce ADH2 (Jurkuta et al., 2009; Nie et al., 2022). Plant HSF transcription factors are downstream components of the signal transduction pathway and maintain regulatory roles for stress-related gene expression (Fragkostfanakis et al., 2015). It was hypothesized that nearly all *AgHSFs* localize to the nucleus so that they may perform transcriptional functions. Expression analysis of seven genes (*AtHSP98.7*, *AtHSP70-1*, *AtBOB1*, *AtCPN60B*, *AtADH2*, *AtAPX1*, and *AtGOLS1*) was carried out to define their interaction with *AgHSFa6-1*. Genes like *AtAPX1* exhibited a sensitive expression profile in response to heat stress compared to WT plants. The expression profiles demonstrated that as the duration of heat stress increased, the expression level of *AgHSFa6-1* increased rapidly in initial heat stress stages, and with heat stress time processing, the expression levels of stress-responsive genes maintained at a high expression level. At the same time, other proteins represented significantly higher expression levels after 12 h of heat stress, suggesting that *AgHSFa6-1* could induce the expression of downstream genes to combat heat stress.

5 Conclusion

In summary, we characterized the HSF gene from celery and functional analysis the *AgHSFa6-1* transcription factor. The results revealed that *AgHSFa6-1* acted as a positive regulator of heat tolerance. Cognizant of this, a mode of *AgHSFa6-1* regulation network was established in response to heat stress (Figure 10). As plants receive heat shock signals, *AgHSFa6-1* regulates the plant's resistance to high temperature through three mechanisms. First, overexpression *AgHSFa6-1* induced material like antioxidant enzymes (SOD, POD) to clear excessive ROS, and reduced MDA content to protected membrane from ROS damage, enhancing membrane stability. Second, through osmoregulation, the synthesis of some typical osmotic regulatory substances was upregulated and thus reduced the plant's stomatal apertures to control water loss. Third, *AgHSFa6-1* promoted the expression of downstream heat-resistance-related genes in plants. Recent

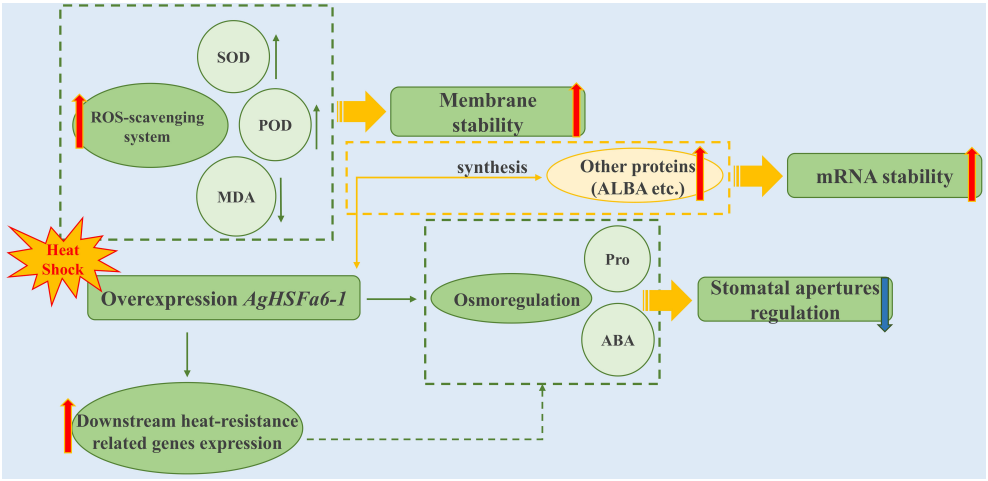


FIGURE 10
Module-centric mode of the transcriptional regulatory network of *AgHSFa6-1* in response to heat stress.

research on the HSFA2 and HSFA7 TFs has revealed a novel pathway for plants to enhance thermotolerance, as ALBA protein could bind with HSF mRNAs to maintain stability (Tong et al., 2022). As mentioned earlier, HSFA6 expressed a reasonably close relationship with A2 and A7 HSF TFs. In light of this, *AgHSFa6-1*, as a member of HSFA6, may have the same function. This work classified the AgHSF family and defined the function of *AgHSFa6-1* in response to heat shock. The results above help us understand the molecular mechanisms of celery adaptation to high temperature, and also provide reliable gene resources for molecular breeding of new heat-tolerant cultivars in celery.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, ML, YL, and HT. Methodology, RZ, JZ, JD, XL, and YXL. Resources, YL and XW. Software, YTZ and YZ. Data curation and visualization, YW, QC, and WH. Writing-original draft preparation, ML, RZ, and JZ. Project administration, AX and HT. Funding acquisition, ML and HT. All authors contributed to the article and approved the submitted version.

Funding

The research was supported by the National Natural Science Foundation of China (Grant Number 32002027) and the National Natural Science Foundation of Sichuan Province (Grant Number 2022NSFSC1647).

Acknowledgments

We would like to thank Home for Researchers (www.home-for-researchers.com) for English language editing.

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.2023.1132307/full#supplementary-material

SUPPLEMENTARY TABLE 1
Primer sequences using in this study.

SUPPLEMENTARY TABLE 2
Basic information of HSF gene family members in *A. graveolens*.

SUPPLEMENTARY TABLE 3
Protein sequences of 20 motifs in AgHSFs.

References

Bates, L. S., Waldren, R. P., and Teare, I. D. (1973). Rapid determination of free proline for water-stress studies. *Plant Soil.* 39, 205–207. doi: 10.1007/BF00018060

Cao, K., Yu, J., Xu, D. W., Ai, K. Q., Bao, E. C., and Zou, Z. R. (2018). Exposure to lower red to far-red light ratios improve tomato tolerance to salt stress. *BMC Plant Biol.* 18 (1), 92. doi: 10.1186/s12870-018-1310-9

Chan-Schaminet, K. Y., Baniwal, S. K., Bublak, D., Nover, L., and Scharf, K. D. (2009). Specific interaction between tomato HsfA1 and HsfA2 creates hetero-oligomeric superactivator complexes for synergistic activation of heat stress gene expression. *J. Biol. Chem.* 284 (31), 20848–20857. doi: 10.1074/jbc.M109.007336

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. G., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, S. S., Jiang, J., Han, X. J., Zhang, Y. X., and Zhuo, R. Y. (2018). Identification, expression analysis of the hsf family, and characterization of class A4 in *Sedum Alfredii* hance under cadmium stress. *Int. J. Mol. Sci.* 19 (4), 1216. doi: 10.3390/ijms19041216

Cheng, X. Q., Fang, T. Y., Zhao, E. H., Zheng, B. G., Huang, B. R., An, Y., et al. (2020). Protective roles of salicylic acid in maintaining integrity and functions of photosynthetic photosystems for alfalfa *(Medicago sativa* l.) tolerance to aluminum toxicity. *Plant Physiol. Biochem.* 155, 570–578. doi: 10.1016/j.plaphy.2020.08.028

Chiara, P., Valeria, B., and Pierdomenico, P. (2012). ROS signaling as common element in low oxygen and heat stress. *Plant Physiol. Biochem.* 59, 3–10. doi: 10.1016/j.plaphy.2012.02.016

Djemal, R., and Khoudi, H. (2021). The barley SHN1-type transcription factor HvSHN1 imparts heat, drought and salt tolerances in transgenic tobacco. *Plant Physiol. Biochem.* 164, 44–53. doi: 10.1016/j.plaphy.2021.04.018

Fan, K., Mao, Z. J., Ye, F. T., Pan, X. F., Li, Z. W., Lin, W. W., et al. (2021). Genome-wide identification and molecular evolution analysis of the heat shock transcription factor (HSF) gene family in four diploid and two allopolyploid *Gossypium* species. *Genomics.* 113 (5), 3112–3127. doi: 10.1016/j.ygeno.2021.07.008

Fragkostefanakis, S., Mesihovic, A., Simm, S., Paupière, M. J., Hu, Y., Paul, P., et al. (2016). HsfA2 controls the activity of developmentally and stress-regulated heat stress protection mechanisms in tomato male reproductive tissues. *Plant Physiol.* 170 (4), 2461–2477. doi: 10.1104/pp.15.01913

Fragkostefanakis, S., Roth, S., Schleiff, E., and Scharf, K. D. (2015). Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks. *Plant Cell Environ.* 38 (9), 1881–1895. doi: 10.1111/pce.12396

Franklin, K. A. (2009). Light and temperature signal crosstalk in plant development. *Curr. Opin. Plant Biol.* 12 (1), 63–68. doi: 10.1016/j.pbi.2008.09.007

Fu, Y. L., Ma, H. L., Chen, S. Y., Gu, T. Y., and Gong, J. M. (2018). Control of proline accumulation under drought *via* a novel pathway comprising the histone methylase CAU1 and the transcription factor ANAC055. *J. Exp. Bot.* 69 (3), 579–588. doi: 10.1093/jxb/erx419

Garima, S., Neelam, K. S., and Anil, G. (2021). Tango between ethylene and HSFA2 settles heat tolerance. *Trends Plant Sci.* 26 (5), 429–432. doi: 10.1016/j.tplants.2021.03.003

Gechev, T., and Petrov, V. (2020). Reactive oxygen species and abiotic stress in plants. *Int. J. Mol. Sci.* 21 (20), 7433. doi: 10.3390/ijms21207433

Guo, M., Liu, J. H., Ma, X., Luo, D. X., Gong, Z. H., and Lu, M. H. (2016). The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00114

Hu, Y., Han, Y. T., Wei, W., Li, Y. J., Zhang, K., Gao, Y. R., et al. (2015). Identification, isolation, and expression analysis of heat shock transcription factors in the diploid woodland strawberry fragaria vesca. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00736

Hu, R. B., Qi, G., Kong, Y. Z., Kong, D. J., Gao, Q., and Zhou, G. K. (2010). Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. *BMC Plant Biol.* 10, 145. doi: 10.1186/1471-2229-10-145

Jacob, P., Hirt, H., and Bendahmane, A. (2017). The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnol. J.* 15, 405–414. doi: 10.1111/pbi.12659

Jiang, L. Y., Hu, W. J., Qian, Y. X., Ren, Q. Y., and Zhang, J. (2021). Genome-wide identification, classification and expression analysis of the hsf and Hsp70 gene families in maize. *Gene.* 770, 145348. doi: 10.1016/j.gene.2020.145348

Jiang, L., Wang, W. Y., Chen, Z. P., Gao, Q. C., Xu, Q. X., and Cao, H. M. (2017). A role for APX1 gene in lead tolerance in *Arabidopsis thaliana*. *Plant Sci.* 256, 94–102. doi: 10.1016/j.plantsci.2016.11.015

Jiao, S. Z., Guo, C., Yao, W. K., Zhang, N. B., Zhang, J. Y., and Xu, W. R. (2022). An amur grape *VaHsfC1* is involved in multiple abiotic stresses. *Sci. Hortic.* 295, 110785. doi: 10.1016/j.scienta.2021.110785

Jurkuta, R. J., Kaplinsky, N. J., Spindel, J. E., and Barton, M. K. (2009). Partitioning the apical domain of the *Arabidopsis* embryo requires the BOBBER1 NudC domain protein. *Plant Cell.* 21 (7), 1957–1971. doi: 10.1105/tpc.108.065284

Kotak, S., Port, M. A., Bicker, F., and Pascal, K. D. (2010). Characterization of c-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class a hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant J.* 39 (1), 98–112. doi: 10.1111/j.1365-313X.2004.02111.x

Kumar, A., Changwal, C., Thapa, B., Tanpure, R. S., Hada, A., Singh, P. K., et al. (2021). Transcription factors: a tool box for countering the effect of abiotic stresses. *Stress Tolerance Hort Crops.*, 169–192. doi: 10.1016/B978-0-12-822849-4.00019-X

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

Li, D. L., He, Y. J., Li, S. H., Shi, S. L., Li, L. Z., Liu, Y., et al. (2021). Geneme-wide characterization and expression analysis of *AP2/ERF* genes in eggplant *(solanum melongena* l.). *Plant Physiol. Biochem.* 167, 492–503. doi: 10.1016/j.plaphy.2021.08.006

Li, M. Y., Hou, X. L., Wang, F., Tan, G. F., Xu, Z. S., and Xiong, A. S. (2018). Advances in the research of celery, an important apiaceae vegetable crop. *Crit. Rev. Biotechnol.* 38 (2), 172–183. doi: 10.1080/07388551.2017.1312275

Li, T., Huang, Y., Khadr, A., Wang, Y. H., Xu, Z. S., and Xiong, A. S. (2020). *DcDREB1A*, a DREB-binding transcription factor from *Daucus carota*, enhances drought tolerance in transgenic *Arabidopsis thaliana* and modulates lignin levels by regulating lignin-biosynthesis-related genes. *Environ. Exp. Bot.* 169, 103896. doi: 10.1016/j.envexpbot.2019.103896

Li, M. Y., Li, J., Zhang, R., Lin, Y. X., Xiong, A. S., Tan, G. F., et al. (2022a). Combined analysis of the metabolome and transcriptome to explore heat stress responses and adaptation mechanisms in celery *(Apium graveolens* L.). *Int. J. Mol. Sci.* 23 (6), 3367. doi: 10.3390/ijms23063367

Li, M. Y., Li, X. Y., Zhou, J., Sun, Y., Du, J. G., Wang, Z., et al. (2022b). Genome-wide identification and analysis of terpene synthase *(TPS)* genes in celery reveals their regulatory roles in terpenoid biosynthesis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1010780

Li, M. Y., Wang, F., Jiang, Q., Ma, J., and Xiong, A. S. (2014). Identification of SSRs and differentially expressed genes in two cultivars of celery (*Apium graveolens* L.) by deep transcriptome sequencing. *Hortic. Res.* 1, 10. doi: 10.1038/hortres.2014.10

Li, G. L., Zhang, H. N., Shao, H., Wang, G. Y., Zhang, Y. Y., Zhang, Y. J., et al. (2019). *ZmHsf05*, A new heat shock transcription factor from *Zea mays* L. improves thermotolerance in *Arabidopsis thaliana* and rescues thermotolerance defectes of the *athsfa2* mutant. *Plant Sci.* 284, 375–384. doi: 10.1016/j.plantsci.2019.03.002

Lim, C. W., Baek, W., Jung, J., Kim, J. H., and Lee, S. C. (2015). Function of ABA in stomatal defense against biotic and drought stresses. *Int. J. Mol. Sci.* 16 (7), 15251–15270. doi: 10.3390/ijms160715251

Lin, A. P., Chen, Y. T., Ponce, G., Acevedo, F. E., Anderson, C. T., Ali, J. G., et al. (2022). Stomata-mediated interactions between plants, herbivores, and the environment. *Trends Plant Sci.* 27 (3), 287–300. doi: 10.1016/j.tplants.2021.08.017

Liu, G. T., Chai, F. M., Wang, Y., Jiang, J. Z., Duan, W., Wang, Y. T., et al. (2018). Genome-wide identification and classification of hsf family in grape, and their transcriptional analysis under heat acclimation and heat stress. *Hortic. Plant J.* 4 (4), 133–143. doi: 10.1016/j.hpj.2018.06.001

Liu, Z. L., Li, L., Luo, Z. S., Zeng, F. F., Jiang, L., and Tang, K. C. (2016). Effect of brassinolide on energy status and proline metabolism in postharvest bamboo shoot during chilling stress. *Postharvest Biol. Technol.* 111, 240–246. doi: 10.1016/j.postharvbio.2015.09.016

Liu, X. S., Liang, C. C., Hou, S. G., Wang, X., Chen, D. H., Shen, J. L., et al. (2020). The LRR-RLK protein HSL3 regulates stomatal closure and the drought stress response by modulating hydrogen peroxide homeostasis. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.548034

Macadam, J. W., Nelson, C. J., and Sharp, R. E. (1992). Peroxidase activity in the leaf elongation zone of tall fescue: i. spatial distribution of ionically bound peroxidase activity in genotypes differing in length of the elongation zone. *Plant Physiol.* 99 (3), 879–885. doi: 10.1104/pp.99.3.879

Maria, L. D., Christian, J. S., Teresa, C. M., Vincente, M., Barbara, B. U., and Rosa, M. R. (2021). Synchronization of proline, ascorbate and oxidative stress pathways under the combination of salinity and heat in tomato plants. *Environ. Exp. Bot.* 183, 104351. doi: 10.1016/j.envexpbot.2020.104351

Meiri, D., Tazat, K., Cohen-Peer, R., Farchi-Pisanty, O., Aviezer-Hagai, K., Avni, A., et al. (2010). Involvement of *Arabidopsis* ROF2 (FKBP65) in thermotolerance. *Plant Mol. Biol.* 72 (1-2), 191–203. doi: 10.1007/s11103-009-9561-3

Mohsen, H., and Huang, S. S. (2022). Elucidating the biology of transcription factor-DNA interaction for accurate identification of cis-regulatory elements. *Curr. Opin. Plant Biol.* 68, 102232. doi: 10.1016/j.pbi.2022.102232

Nie, X. Z., Mohammed, M., Abir, U. I., Robert, D. H., and Claudio, S. (2022). Anaerobiosis modulation of two phytoglobins in barley *(Hordeum vulgare* L.), and their regulation by gibberellin and abscisic acid in aleurone cells. *Plant Physiol. Biochem.* 182, 174–181. doi: 10.1016/j.plaphy.2022.04.014

Nover, L., Bharti, K., Doring, P., Mishra, S., Ganguli, A., and Scharf, K. (2001). *Arabidopsis* And the heat stress transcription factor world: how many heat stress transcription factors do we need. *Cell Stress Chaperones.* 6 (3), 177–189. doi: 10.1379/1466-1268(2001)006<0177:aathst>2.0.co;2

Pfaffla, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29 (9), e45. doi: 10.1093/nar/29.9.e45

Qian, J., Chen, J., Liu, Y. F., Yang, L. L., Li, W. P., and Zhang, L. M. (2014). Overexpression of *Arabidopsis HsfA1a* enhances diverse stress tolerance by promoting stress-induced hsp expression. *Genet. Mol. Res.* 13 (1), 1233–1243. doi: 10.4238/2014.february.27.8

Ren, S. X., Ma, K. B., Lu, X. G., Chen, G., Cui, J. W., Tong, P. X., et al. (2019). Transcriptomic and metabolomic analysis of the heat-stress response of *Populus tomentosa* Carr. *Forests.* 10 (5), 383. doi: 10.3390/f10050383

Richard, N. R. (2004). Celery diseases and their management. *Dis. fruits vegetables.* 1, 441–453. doi: 10.1007/1-4020-2606-4_11

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Sripinyowanich, S., Klomsakul, P., Boonburapong, B., Bangyeekhun, T., Asami, T., Gu, H., et al. (2013). Exogenous ABA induces salt tolerance in indica rice (*Oryza sativa* L.): the role of *OsP5CS1* and *OsP5CR* gene expression during salt stress. *Environ. Exp. Bot.* 86, 94–105. doi: 10.1016/j.envexpbot.2010.01.009

Takumi, Y., Naohiko, O., Jun, N., Satoshi, K., Junya, M., Kazuo, N., et al. (2011). *Arabidopsis HsfA1* Transcription factors function as the main positive regulators in heat shock-responsive gene expression. *Mol. Genet. Genom.* 286 (5-6), 321–332. doi: 10.1007/s00438-011-0647-7

Tan, B., Yan, L., Li, H. N., Lian, X. D., Cheng, J., Wang, W., et al. (2021). Genome-wide identification of HSF family in peach and functional analysis of *PpHSF5* involvement in root and aerial organ development. *PeerJ.* 9, e10961. doi: 10.7717/peerj.10961

Tong, J. J., Ren, Z. T., Sun, L. H., Zhou, S. X., Yuan, W., Hui, Y. F., et al. (2022). ALBA proteins confer thermotolerance through stabilizing HSF messenger RNAs in cytoplasmic granules. *Nat. Plants.* 8 (7), 778–791. doi: 10.1038/s41477-022-01175-1

Waadt, R., Seller, C. A., Hsu, P. K., Takahashi, Y., Munemasa, S., and Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* 23 (10), 680–694. doi: 10.1038/s41580-022-00479-6

Wang, L. J., Fan, L., Loescher, W., Duan, W., Liu, G. L., Cheng, J. S., et al. (2010). Salicylic acid alleviates decreases in photosynthesis under heat stress and accelerates recovery in grapevine leaves. *BMC Plant Biol.* 10, 34. doi: 10.1186/1471-2229-10-34

Wang, J. Q., Hasegawa, T., Li, L. Q., Lam, S. K., Zhang, X. H., Liu, X. Y., et al. (2018). Changes in grain protein and amino acids composition of wheat and rice under short-term increased [CO_2] and temperature of canopy air in a paddy from East China. *New Phytol.* 222 (2), 726–734. doi: 10.1111/nph.15661

Wang, R., Mao, C. J., Jiang, C. H., Zhang, L., Peng, S. Y., Zhang, Y., et al. (2021). One heat shock transcription factor confers high thermal tolerance in clematis plants. *Int. J. Mol. Sci.* 22 (6), 2900. doi: 10.3390/ijms22062900

Wang, W. S., Rong, Y., Wang, X. W., Wang, C. Z., Zhang, C. L., Huo, Z. L., et al. (2022). Estimating sunflower canopy conductance under the influence of soil salinity. *Agric. For Meteorol.* 314 (4), 108778. doi: 10.1016/j.agrformet.2021.108778

Wang, W. X., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta.* 218 (1), 1–14. doi: 10.1007/s00425-003-1105-5

Wu, J. T., Gao, T., Hu, J. N., Zhao, L., Yu, C., and Ma, F. (2022). Research advances in function and regulation mechanisms of plant small heat shock proteins (sHSPs) under environmental stresses. *Sci. Total Environ.* 825, 154054. doi: 10.1016/j.scitotenv.2022.154054

Zha, Q., Xi, X. J., He, Y. N., and Jiang, A. L. (2020). Transcriptomic analysis of the leaves of two grapevine cultivars under high-temperature stress. *Sci. Hortic.* 265 (4), 109265. doi: 10.1016/j.scienta.2020.109265

Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., and Chua, N. H. (2006). Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* 1 (2), 641–646. doi: 10.1038/nprot.2006.97

Zhang, X. Y., Zhou, Y., Li, J., Gu, X. Y., Zhao, L. N., Li, B., et al. (2022). Pichia caribbica improves disease resistance of cherry tomatoes by regulating ROS metabolism. *Biol. Control.* 169 (5), 104870. doi: 10.1016/j.biocontrol.2022.104870

Zhao, X. G., Sui, X. Y., Zhao, L. X., Gao, X. X., Wang, J. X., Wen, X. Z., et al. (2022). Morphological and physiological response mechanism of lettuce *(Lactuca Sativa* L.) to consecutive heat stress. *Sci. Hortic.* 301, 111112. doi: 10.1016/j.scienta.2022.111112

Zhou, M., Zheng, S. G., Liu, R., Lu, J., Lu, L., Zhang, C. H., et al. (2019). Genome-wide identification, phylogenetic and expression analysis of the heat shock transcription factor family in bread wheat *(Triticum aestivum* L.). *BMC Genom.* 20 (1), 505. doi: 10.1186/s12864-019-5876-x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 July 2023
DOI 10.3389/fpls.2023.1215044

OPEN ACCESS

EDITED BY
Guo-Fei Tan,
Guizhou Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Guo Jian,
Yangzhou University, China
Zhenhui Wang,
Jilin Agriculture University, China

*CORRESPONDENCE
Lin Ruan
alinche@126.com
Seping Dai
daiseping@126.com
Yixun Yu
yuyixun@scau.edu.cn

RECEIVED 01 May 2023
ACCEPTED 06 July 2023
PUBLISHED 27 July 2023

CITATION
Li H, Wang W, Liu R, Tong B, Dai X, Lu Y, Yu Y, Dai S and Ruan L (2023) Long non-coding RNA-mediated competing endogenous RNA regulatory network during flower development and color formation in *Melastoma candidum*. *Front. Plant Sci.* 14:1215044. doi: 10.3389/fpls.2023.1215044

Long non-coding RNA-mediated competing endogenous RNA regulatory network during flower development and color formation in *Melastoma candidum*

Hui Li[1,2], Wei Wang[1], Rui Liu[3], Botong Tong[3,4], Xinren Dai[3], Yan Lu[5], Yixun Yu[2]*, Seping Dai[1]* and Lin Ruan[1]*

[1]Department of Botany, Guangzhou Institute of Forestry and Landscape Architecture, Guangzhou, China, [2]College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China, [3]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China, [4]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China, [5]Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Institute of Botany, Chinese Academy of Sciences, Nanjing, Jiangsu, China

M. candidum, an evergreen shrubby flower known for its superior adaptation ability in South China, has gained increased attention in garden applications. However, scant attention has been paid to its flower development and color formation process at the non-coding RNA level. To fill this gap, we conducted a comprehensive analysis based on long non-coding RNA sequencing (lncRNA-seq), RNA-seq, small RNA sequencing (sRNA-seq), and widely targeted metabolome detection of three different flower developmental stages of *M. candidum*. After differentially expressed lncRNAs (DElncRNAs), differentially expressed mRNAs (DEmRNAs), differentially expressed microRNAs (DEmiRNAs), and differentially synthesized metabolites (DSmets) analyses between the different flower developmental stages, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were conducted to identify some key genes and metabolites in flavonoid, flavone, anthocyanin, carotenoid, and alkaloid-related GO terms and biosynthetic pathways. Three direct-acting models, including antisense-acting, cis-acting, and trans-acting between lncRNAs and mRNAs, were detected to illustrate the direct function of lncRNAs on target genes during flower development and color formation. Based on the competitive endogenous RNA (ceRNA) regulatory theory, we constructed a lncRNA-mediated regulatory network composed of DElncRNAs, DEmiRNAs, DEmRNAs, and DSmets to elucidate the indirect role of lncRNAs in the flower development and color formation of *M. candidum*. By utilizing correlation analyses between DERNAs and DSmets within the ceRNA regulatory network,

alongside verification trials of the ceRNA regulatory mechanism, the study successfully illustrated the significance of lncRNAs in flower development and color formation process. This research provides a foundation for improving and regulating flower color at the lncRNA level in *M. candidum*, and sheds light on the potential applications of non-coding RNA in studies of flower development.

KEYWORDS

lncRNA, ceRNA regulatory mechanism, flower development, flower color formation, metabolites, *Melastoma candidum*

1 Introduction

LncRNAs are an influential class of molecules with a length of more than 200 nt (Wang and Chang, 2011). They originate from exonic, intronic, intragenic, and intergenic promoter regions, as well as 3' and 5' UTR enhancer sequences, and are transcripted in either a sense or antisense direction (Zhang et al., 2013). Traditionally, lncRNAs were thought to be meaningless molecules. In recent years, research has proven that lncRNAs regulate many biological processes in organisms such as gene expression adjustment (Guttman et al., 2009), post-transcription, post-translation (Crick et al., 1961), and chromosome modification (Bertone et al., 2004; Yanofsky, 2007) by combining with corresponding proteins. Unlike protein-coding genes, most lncRNAs are usually expressed at low levels and lack strong sequence conservation between species (Cabili et al., 2011; Necsulea et al., 2014). A number of studies have provided evidence that lncRNAs play a significant role in response to stress (Wunderlich et al., 2014), male sterility (Ding et al., 2012), phosphate homeostasis (Franco-Zorrilla et al., 2007), flowering time regulation (Heo and Sung, 2011), and flower and pollen development (Liu et al., 2010; Kang and Liu, 2015) in plants.

An increasing body of evidence indicating that lncRNAs may affect gene expression by either *cis*-acting on their chromosomes or *trans*-acting protein-encoding genes to carry out their functions (Wu et al., 2019). It is possible that long non-coding RNAs located upstream of a gene may associate with the promoter or other *cis*-acting elements of co-expressed genes in order to regulate gene expression at the transcriptional or post-transcriptional level. A large number of long non-coding RNAs could overlap with transcription factor binding sites, potentially preventing transcription factors from binding to the corresponding sites (de la Fuente, 2010; Chen, 2016). In some cases, lncRNAs are found to enhance the binding rate of transcription factors to nearby binding sites (de la Fuente, 2010; Yang et al., 2019). A study by Li et al. (2022) found that *cis*-acting LNC_002115 regulates hickory female floral development by influencing both *PHO2* and *SVP*. Kang and Liu (2015) identified a significant number of lncRNAs from 35 different flower and fruit tissues of diploid strawberries. They inferred a *cis* or *trans*-acting relationship between lncRNAs and their targets based on correlation analysis of lncRNAs and their target gene expression trend. In addition, they discovered that lncRNAs are not well conserved between species of plants.

As endogenous target mimics, lncRNAs can bind with miRNAs on mRNA response elements (MREs) and mitigate miRNAs' cleavage effect on target genes by sponge-like actions (Wu et al., 2019). Due to their similar structural characteristics, lncRNAs may also be negatively regulated by miRNAs through a similar mechanism to mRNAs (Chaofeng et al., 2013). In *Arabidopsis*, a ceRNA regulatory relationship showed that lncRNA IPS1 could influence the expression level of *PHO2* by binding to miR399 (Franco-Zorrilla et al., 2007). LncRNAs, *COOLAIR* and *COLDAIR*, could repress the *FLC* gene expression through an epigenetic silencing mechanism to regulate flowering time (Heo and Sung, 2011; Kim and Sung, 2017). Compared with the wild type, overexpression of lncRNA npc48 increased the rosette diameter and leaf serration and delayed flowering time (Ben Amor et al. 2009). Besides, ceRNA regulatory relations has also been demonstrated in other plants, such as cucumber (He et al., 2020), pepper (Zuo et al., 2019), rice (Xu et al., 2016), maize (Zhu et al., 2017), and tomato (Yang et al., 2019). In rice, lncRNA osa-eTM160 attenuated the repression of osa-miR160 on *ARF18* throughout the embryonic anther by target mimicry ways (Wang et al., 2017). Fang et al. (2019) found that a rice lncRNA Ef-dc transcribed from the antisense strand of the flowering activator *SOC1* locus can positively regulate the expression of *SOC1* and balance yield with maturity duration. Wang et al. (2017) found that lncRNA osa-eTM160 could attenuate the repression of osa-miR160 on osa-*ARF18* during the early developmental stage of the anther. A genome-wide association study (GWAS) analysis and ceRNA network by Xu et al. (2021) identified the key regulatory mechanism of the LTCONS_00034157- miRNA167h- PsTPS1 in the later flowering process of *Prunus sibirica*.

Flower color is an important trait that determines the ornamental quality and landscaping application value (Zhu et al., 2019b). It is the result of pigment metabolite accumulation in the vacuoles of flower epidermal cells (Mol et al., 1998; Deng et al., 2013). Flower color is predominantly due to the production of flavonoids carotenoids or betalains. (Holton, 1995). Betalains, one type of alkaloid, are water-soluble nitrogenous pigments derived from the amino acid L-tyrosine, mainly classified as red-violet betacyanins and yellow betaxanthins. In Caryophyllales, betalains

exclusively replace anthocyanin (Tanaka et al., 1998; Deng et al., 2013) to attract pollinators and seed dispersers (Sunnadeniya et al., 2016). Some researchers have demonstrated that the relevant enzymes for the production of anthocyanins are not expressed in betalains-producing plants at the biochemical level (Shimada et al., 2007; Brockington et al., 2011). However, enzymes, genes and biosynthetic pathways involved in betalain production are much less well-studied than those of flavonoids and carotenoids.

The change in color both within flowers and in isolated pigments involves a range of biochemical mechanisms. Some of the factors influencing color are temperature, co-pigments, pH, metals, sugars anthocyanin stacking, and cell shape (Bowles et al., 2006; Shoji et al., 2007). The earliest research suggested that pH, metal-complex theory, and Metalloanthocyanins could be the key factors to determine flower color. In 1913, Willstätter and Everest proposed the pH theory based on the observation of a pigment from blue cornflowers and rose, cyanin, which could display red color under acidic media and blue color under alkaline solutions (Willstätter and Everest, 1913). The metal complex theory states that anthocyanin can form complexation with metal ions such as Mg^{2+}(Mitsui et al., 1959), Fe^{3+}, Al^{3+}(Bayer et al., 1966) Ga^{3+}, In^{3+}, Co^{3+}, Mn^{2+}, Zn^{2+} and Cd^{2+} to show different colors in plants flower (Kondo et al., 1998). Metalloanthocyanins theory holds that anthocyanins, flavones, and metal ions fix at 6: 6: 2 in blue flowers (Takeda, 2006). With biological development, more and more flower color can be explained by genes or metabolite levels. Nowadays, scientists believe that flower color is predominantly controlled by the production of flavonoids, carotenoids, and alkaloid-related compounds, such as betalains (Holton, 1995). For example, researchers have found that yellow flower petals often contain yellowish xanthophylls, β-carotenoids, and chrysanthemum. All of them belong to the class of carotenoids (Nielsen et al., 2003; Kishimoto et al., 2004). Some species such as roses, and carnations are lacking blue because of the absence of Flavonoid3',5'-hydroxylase, a key enzyme catalyzing the hydroxylation reaction between dihydrokaempferol and dihydromyricetin (Holton and Cornish, 1995; Kondo et al., 1998).

M. candidum belongs to the Melastomataceae family which is mainly centered in Southeast Asia and extends to India, South China, and Northern Australia (Liu et al., 2014). Although many species of Melastomataceae have a relatively high degree of overlap in geographic distributions and flower periods, some members still face reproductive isolation problems. Additionally, many members of the Melastomataceae are dull colors, which limits their use in garden industries. Here, we took *M. candidum* as a research object and conducted lncRNA-seq, sRNA-seq, mRNA-seq, and a widely targeted metabolome for three development stages of the flowers. Based on differentially expressed RNA analyses, we developed three functional acting models, including antisense, *cis*, and trans - acting models. Two networks of antisense and *cis* models were constructed in flavonoid, anthocyanin, carotenoid, and alkaloid-related pathways. According to the ceRNA theory, a lncRNA-mediated regulatory network was also built in the aforementioned pathways. A correlation among all kinds of RNAs and metabolites was conducted to illustrate the relationship among them. This research aims to unveil the role of lncRNA in flower development in *M. candidum*.

2 Materials and methods

2.1 RNA extraction, cDNA library construction, and sequencing

Three developmental stages of the flower of *M. candidum* including closed buds with white petals (McI), closed buds with pink petals (McII), and opened buds with pink petals (McIII) were collected and immediately frozen in liquid nitrogen. Three replicates of each stage sample were taken from three seedlings. Total RNA was extracted by using an OminiPlant RNA Kit (DNase I) (CW2598, CWBIO, Taizhou, China) according to the manufacturer's protocol. An Agilent 2100 Bioanalyzer was utilized to assess the quality of the RNA (Agilent Technologies, Palo Alto, CA, USA). To generate a cDNA library, we fragmented the mRNA using a fragmentation buffer and reverse-transcribed the resulting small fragments into cDNA using random primers. The second strand cDNA was synthesized by employing DNA polymerase I, RNaseH, dNTP, and buffer. After purification with the poly(A) and PCR extraction kit (Qiagen), the synthesized products were further purified using the QiaQuick PCR extraction kit (Qiagen, Venlo, The Netherlands) and then ligated to Illumina sequencing adapters. Afterward, the second-strand cDNA was digestedusing the enzyme UNG (Uracil-N-Glycosylase), and the resulting products were separated by size on an agarose gel before being amplified by PCR. The PCR products were then sequenced on Illumina HiSeq TM 4000 platforms.

2.2 Filter of the raw data and alignment against the genome

After obtaining sequence data, raw reads consisting of adapters or low-quality bases were filtered by the fastp software (version 0.18.0) (Chen et al., 2018b) with steps: 1) removing reads containing adapters; 2) removing reads containing more than 10% of unknown nucleotides (N); 3) removing low quality reads containing more than 50% of low quality (Q-value ≤ 20) bases. After filtering, the clean short reads were aligned to the ribosome RNA (rRNA) database using Bowtie2 (version 2.2.8) (Langmead and Salzberg, 2012) to eliminate the rRNA-mapped reads. After building an index of the *M. candidum* genome, the paired-end clean reads were mapped to the reference genome by using HISAT2 (version 2.1.0) (Kim et al., 2015) with a default parameter "-rna-strandedness RF".

2.3 Transcripts reconstruction and annotation

The reconstruction of transcripts was carried out with the software Stringtie (version 1.3.4) (Pertea et al., 2015). To identify the novel transcripts among the reconstructed transcripts, all the reconstructed transcripts were aligned to the reference genome and were then categorized into 12 categories using the Cuffcompare program (Trapnell et al., 2010). A novel transcript was defined if it had one of the following class codes - u, i, j, x, c, e, or o. To further

identify new genes, we used the following parameters: length of transcript > 200 bp, number of exons > 1 (Lu et al., 2019). All the novel transcripts were then aligned to the Nr, KEGG, and GO databases to obtain protein functional annotations.

2.4 lncRNA prediction and classification

Software including Coding-Non-Coding Index (CNCI, version2) (Sun et al., 2013) and Coding Potential Calculator 2 (CPC2, version 0.9-r2) (Kong et al., 2007) (http://cpc.cbi.pku.edu.cn/) were used to assess the protein-coding potential of novel transcripts. Only transcripts which meet the protein-coding-score criteria (CNCI sequence-score < 0 and CPC2 coding probability < 0.5) were considered as lncRNAs. Based on the location relative to protein-encoding genes (PCG), lncRNAs are categorized into five classes: intergenic lncRNAs (located between two PCGs), bidirectional lncRNAs (situated on the opposite strand but within 1 kb of the promoter on the sense strand), intronic lncRNAs (located within an intron of a PCG on the sense strand), antisense lncRNAs (transcribed from the opposite strand of a PCG), and sense lncRNA (spanning multiple introns or exons within a PCG) (Ahmad et al., 2021).

2.5 Quantification of transcripts and differentially expressed analysis

The abundance of the transcript was quantified using StringTie (Pertea et al., 2015). An FPKM (fragment per kilobase of transcript per million mapped reads) value was calculated for each transcription region to mitigate the effect of varying transcript lengths and sequencing data amounts on expression. DEGs of coding RNAs and lncRNAs were analyzed separately by using DESeq2 software (Li and Dewey, 2011) between two versus groups by the following steps: 1) normalization of read counts; 2) calculation of p-value; 3) correction of value to get a false discovery rate (FDR) value. Differentially expressed lncRNAs were screened with the threshold: Fold change ≥ 2, FDR ≤ 0.05.

2.6 Differential analyses of miRNAs, mRNAs, and metabolites

For miRNAs, expression level was calculated and normalized by transcripts per million (TPM) methods. EdgeR (Robinson et al., 2010) were used to conducted differential analysis. DEmiRNAs were identified with the absolute threshold $\log_2$ (fold change) ≥ 0.585, p-value ≤ 0.05. For mRNAs, HISAT2 (Kim et al., 2015) were used to align to the genome of *M. candidum* (http://evolution.sysu.edu.cn/Sequences.html). The FPKM value was used to quantify expression abundance and variations. DEmRNAs were identified by DESeq2 (Love et al., 2014) package with the absolute threshold $\log_2$ (fold change) ≥ 1 and p-value ≤ 0.05. Metabolites were analyzed by using an LC-ESI-MS/MS system (UPLC, Shim-pack UFLCSHIMADZU CBM30A, http://www.shimadzu.com.cn/; and MS/MS (Applied Biosystems 6500 QTRAP, http://www.appliedbiosystems.com.cn/) (Chen et al., 2013) and variable importance in projection (VIP) score of (O)PLS model was applied to rank the metabolites and distinguish two versus groups. DSmets were identified with the absolute threshold $\log_2$ (fold change) ≥ 0.585, vip ≤ 1.

2.7 LncRNA-mRNA association analysis

In order to identify three direct-acting models of lncRNAs and mRNAs including antisense-regulation, *cis*-regulation, and trans-regulation, the software RNAplex (version 0.2) (Shen et al., 2014) (http://www.tbi.univie.ac.at/RNA/RNAplex.1.html) was used to predict the complementary correlation and detect the interaction between antisense lncRNAs and mRNAs. Antisense lncRNAs can form complementary base pairing with mRNAs. The ViennaRNA package (Lorenz et al., 2011) in R software was used to predict the best base pairing based on the calculation of minimum free energy through thermodynamics structure. *Cis* lncRNAs could regulate neighboring genes on the same allele. LncRNAs with the unknown region were annotated again to identify *cis*-regulators. Trans lncRNAs could regulate co-expressed genes far from them. The correlation of expression between lncRNAs and protein-coding genes was used to identify a trans-regulation relationship with a Pearson correlation of more than 0.999.

2.8 Validation of lncRNA-seq result by qRT-PCR

In order to verify the lncRNA-seq result, we chose 12 lncRNAs to perform qRT-PCR. The primers of lncRNAs used in qRT-PCR were listed in Table S1. 0.5 μg total RNA of three stages flowers was reverse-transcribed into first-strand cDNA using the PrimeScript RT reagent Kit gDNA Eraser (Takara, Dalian, China). Then, the SYBR @Premix Ex Taq TMII (Takara, Dalian, China) was used according to the manufacturer's instruction for qRT-PCR of lncRNAs on the Illumina Eco real-time PCR system (Illumina, USA). The α -tubulin gene of *M. candidum* was used as the internal reference gene. Ct values were then calculated by the $2^{-\Delta\Delta Ct}$ algorithm. Primers sequences, length of PCR products and PCR amplification efficiency for each pair of primers are listed in Table S1. Relative expression levels of selected lncRNAs at different stages of flowers were analyzed using One-way ANOVA with multiple comparison by GraphPad Prism 9. The graphs were visualized by GraphPad Prism 9 and Adobe Illustrator 2020.

2.9 Construction and validation of lncRNA-mediated ceRNA network

The ceRNA network was constructed based on the following rules: 1) negative correlation between miRNA and lncRNAs, as well as miRNA and mRNAs with calculation of Spearman correlation coefficient (SCC); 2) positive correlation between lncRNAs and mRNAs with calculation of Spearman correlation coefficient (SCC);

3) enrich degree of lncRNA binding to miRNAs and miRNAs binding to mRNAs. Relationships among ceRNA members and metabolites were visualized by the ggalluvial package in R3.6.1 software.

To verify the lncRNA-miRNA-mRNA regulatory chains, the transient coexpression experiments were performed in leaves of *Nicotiana benthamiana* according to Lu et al. (Lu et al., 2019). Precursors of miRNA (pre-miRNA), lncRNAs, and mRNAs were cloned from mixture of cDNA of three-stages flowers. For miRNAs that could not obtain pre-miRNAs in the cDNA, we performed overlap PCR to ligate these mature miRNAs into the vector. Pre-miRNAs and lncRNAs were ligated into pCAMBIA2300 at *Sma*I and *Xba*I sites. mRNAs were ligated into pCAMBIA1300 at *Xba*I and *Kpn*I sites. The vector harboring pre-miRNAs, lncRNAs, and mRNAs were then transformed into *Agrobacterium tumefaciens* GV3101 strain. Equal amounts of agrobacterial cell cultures containing miRNAs, lncRNAs, and mRNAs were mixed respectively. Then, the mixture was infiltrated into leaves of *N. benthamiana*. After being incubated in dark for two days, the infiltrated tobacco leaves were observed under laser scanning confocal microscopy (LSCM) with EGFP = 780, and collected for total RNA extractions (CW2598, CWBIO, Taizhou, China) and qRT-PCR (Takara, Dalian, China). Tobacco L23 was used as an internal reference (Liu et al., 2012). The primers were listed in Table S1.

3 Results

3.1 Statistics information of sequenced lncRNAs

Prior to conducting differentially expressed analyses, we performed basic statistical analyses on the sequencing data, such as examining correlations between samples, evaluating expression distributions, and identifying types of lncRNAs. Sample correlation was determined based on the expression level of lncRNAs (Figure 1A). Correlations among samples of each stage had a relatively high coefficient, indicating these samples meet our analysis requirements. A relatively high average expression level was observed at the late stage (Figure 1B). In order to increase the reliability of the results, we used CNCI and CPC2 to identify lncRNAs in sequenced samples. A total of 4,508 lncRNAs were identified by CNCI and a total of 6,567 lncRNAs were identified by CPC2. To ensure greater accuracy, we took the intersection of the results from both software and obtained 3,955 lncRNAs

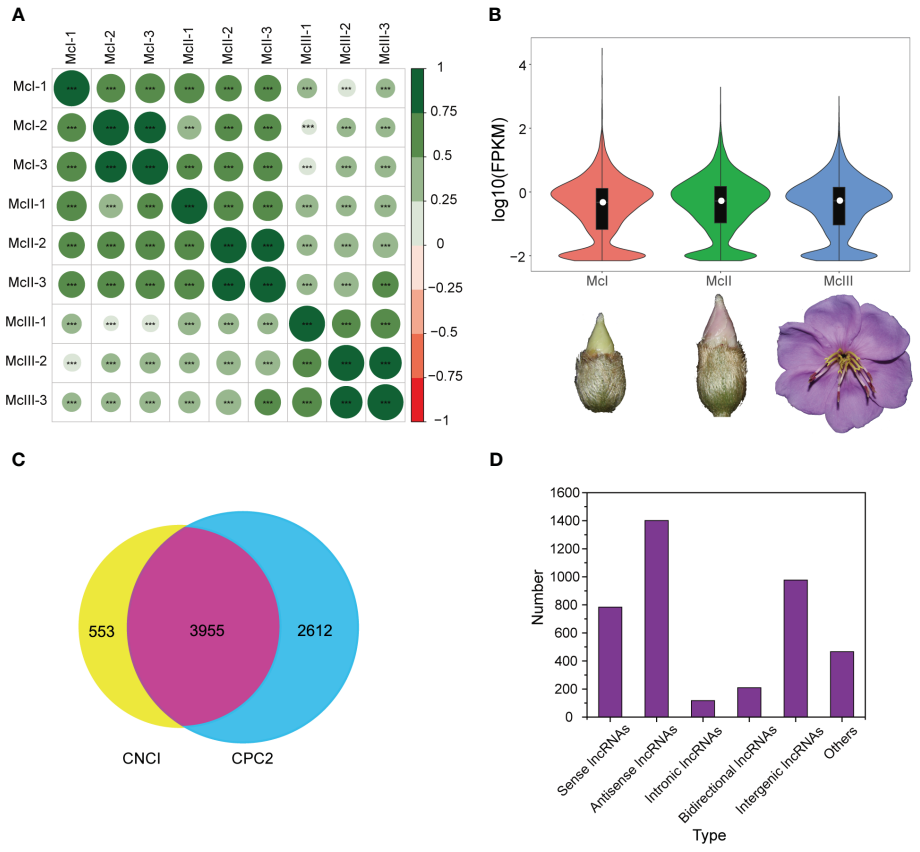


FIGURE 1
Basic statistics information of lncRNAs sequencing. **(A)** Pearson correlation analysis among different samples. Green and big size mean high correlation coefficient. "***" means a P value less than 0.001. **(B)** Violin graph reflecting expression distribution of samples in the different developmental stages. White dots in the middle of the violins represent the median value of the expression level. **(C)** Venn graph of two lncRNA prediction software. **(D)** Detected numbers of different lncRNAs categories.

(Figure 1C). Five types of lncRNA were identified, including sense lncRNAs, antisense lncRNAs, intronic lncRNAs, bidirectional lncRNAs, and intergenic lncRNAs. It should be noted that some lncRNAs in the graph cannot be assigned a precise category. Among the identified lncRNA, the proportion of antisense lncRNAs was the highest, while that of intronic lncRNAs was the lowest (Figure 1D). Overall, the lncRNAs were suitable for further analysis.

3.2 Differentially expressed analyses of lncRNAs

For all lncRNAs identified, we performed differential expression analyses among different stages (Figure 2). There were 190 up-regulated and 228 down-regulated lncRNAs in the McII vs McI group (Figure 2A; Table S2). In the comparison between the McIII and McI group, we identified 552 up-regulated and 602 down-regulated lncRNAs (Figure 2B; Table S3). 426 lncRNAs were up-regulated and 454 lncRNAs were down-regulated in the McIII vs McII group (Figure 2C; Table S4). Differential expression analyses results indicated huge differences of lncRNA expression levels between McIII and the other two stages. Then, we gave a statistics among the three comparison groups using Venn method which yieled 67 common lncRNAs (Figure 2D). As lncRNAs play a variety of roles during the different stages of development, we used the union of three comparison groups to conduct further analysis and identify the roles of lncRNAs during the different stages of flower development.

3.3 Antisense, *cis*, and trans-acting model of lncRNAs and functional annotation of their targeted mRNAs

It has been reported that lncRNAs interact with mRNAs in three distinct ways including *cis*, trans, and antisense interactions (Nie et al., 2012). *Cis* model means that lncRNAs could induce reconstruction and histone modification to influence the combination of transcript factors with promoter and enhancer or combine with transcription elements directly to affect the expression of the protein-encoding genes. In the antisense model, it is assumed that lncRNAs could directly bind to mRNAs to

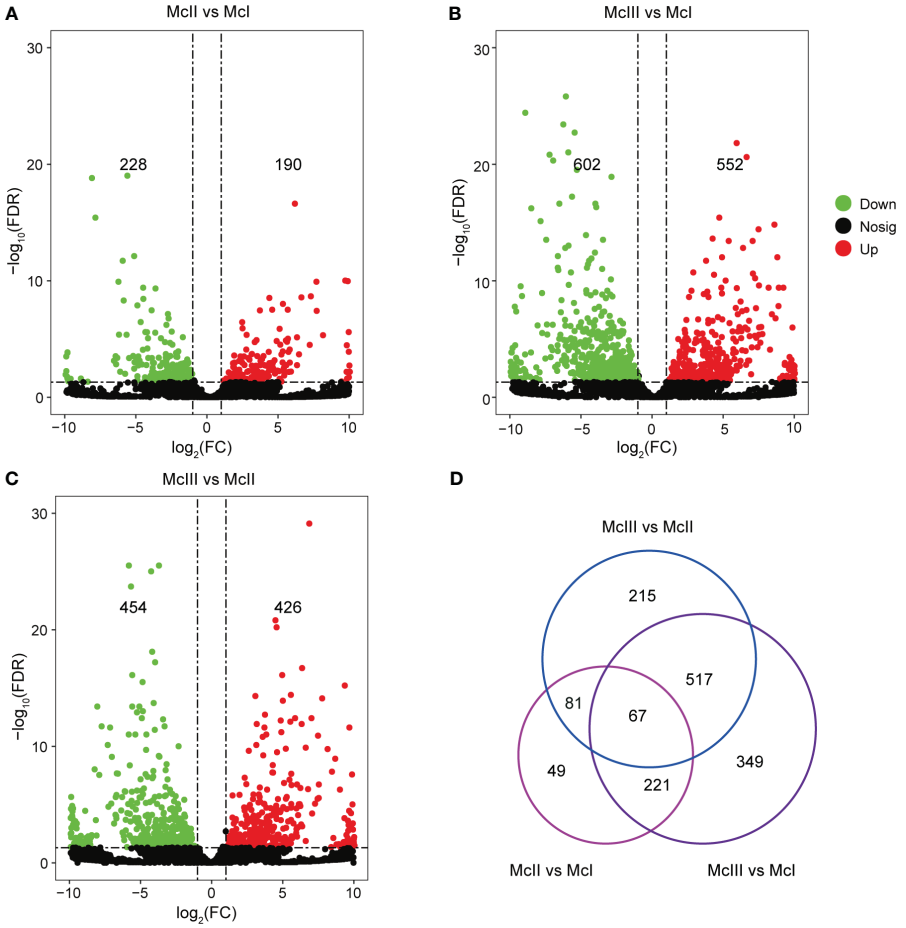


FIGURE 2
Differentially expressed lncRNAs (DElncRNAs) in different comparing groups. **(A)** DElncRNAs in stageII (McII) versus stageI (McI) group. **(B)** DElncRNAs in stageIII (McIII) versus stageI (McI) group. **(C)** DElncRNAs in stageIII (McIII) versus stageII (McII) group. Red dots mean up-regulated lncRNAs, and green dots mean down-regulated lncRNAs. **(D)** Venn graphs of different versus groups.

10.3389/fpls.2023.1215044

influence alternative splicing of mRNAs or to decay mRNAs into siRNAs, which is closely related to the stability of mRNAs. The trans model is often able to be judged by the opposite expression trends between lncRNAs and their mRNA counterparts. LncRNAs could sometimes bind to proteins translated by mRNAs, altering their activity, structure, or position to influence their corresponding mRNAs. Here we also confirmed three functional models for differentially expressed lncRNAs and then GO and KEGG analyses were performed for their corresponding mRNAs (Figures 3A, B; Tables S5-S10). In Figures 3A, B, some GO terms associated with flower development and color formation were chosen for visualization, including alkaloids, pigments, flavonoids, hormones, and flower development-related terms or path. As a result of the three acting models, the trans model enriched more genes than the other two functional models. However, we could not find the antisense models of these DElncRNAs in the flavone biosynthetic process of the result of GO analysis as well as the isoflavonoid biosynthetic process, the folate biosynthesis, and the brassinosteroid biosynthetic process of the results of KEGG analysis (Figures 3A, B). Similarly, there was also lack of *cis* models of these DElncRNAs in the Tropane Alkaloid biosynthetic process and the Alkaloid metabolic process of the results of GO analysis, and the isoflavonoid biosynthesis and Anthocyanin biosynthesis of the results of KEGG analysis (Figures 3A, B).

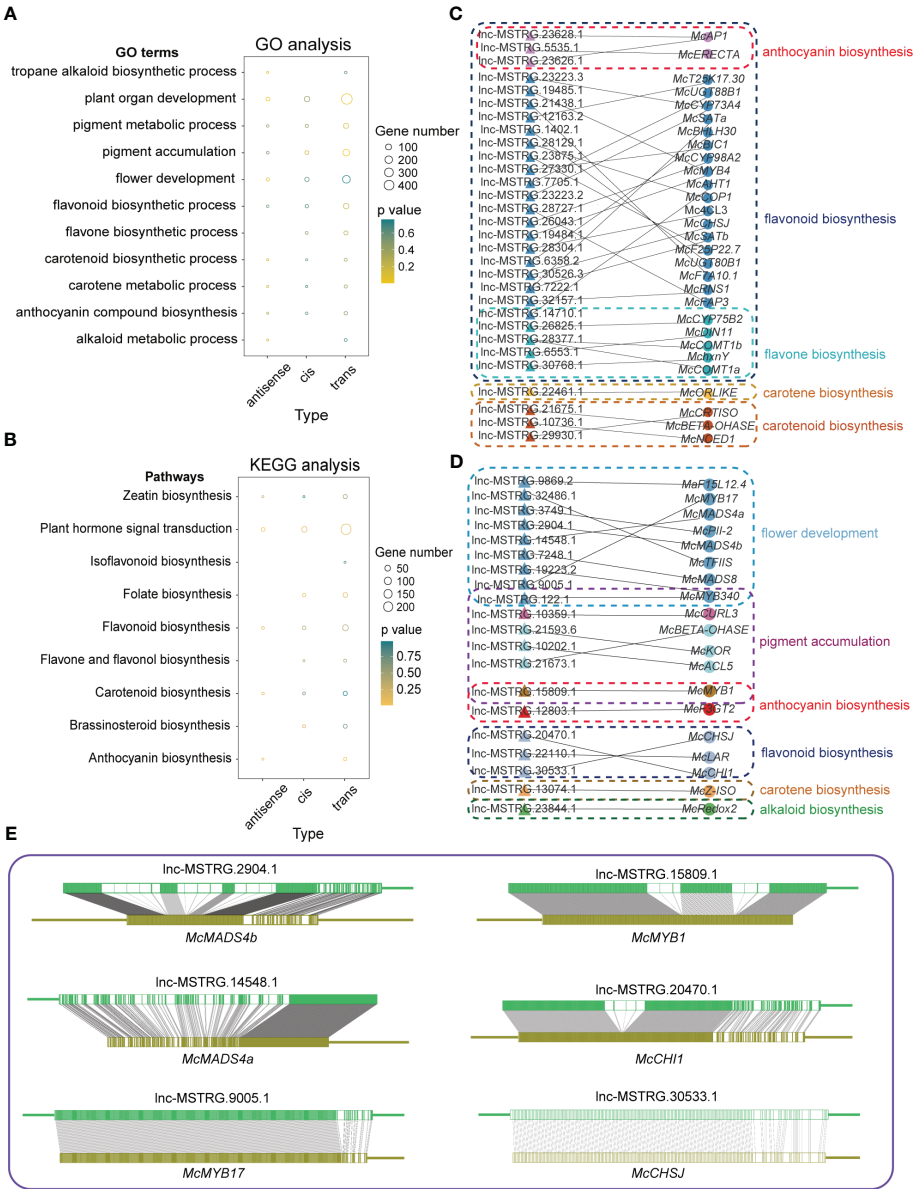


FIGURE 3
Different interacting models of lncRNAs with mRNAs. **(A)** GO analysis of three interacting models. Only biological process terms are shown in this graph. **(B)** KEGG analysis of three interacting models. In both graphs, A,B the size of the circles means enriched gene numbers in different terms or pathways, and orange means significantly enriched. **(C)** A network of *cis* model between lncRNAs and mRNAs in flower development and flower color formation related pathways. **(D)** A network of antisense model between lncRNAs and mRNAs in flower development and flower color formation related pathways. **(E)** The complementary relationship between lncRNAs and mRNAs in the antisense model.

To further investigate the potential roles of lncRNAs in regulating gene expression by antisense and *cis* models, we built lncRNAs-target genes networks (Figures 3C, D). In the *cis* model, we identified 26 lncRNAs that function on genes involved in flavonoid biosynthesis. Of these, three lncRNAs, lnc-MSTRG.23628.1, lnc-MSTRG.5535.1, and lnc-MSTRG.23626.1, were related to anthocyanin biosynthesis, while five lncRNAs, lnc-MSTRG.14710.1, lnc-MSTRG.26825.1, lnc-MSTRG.28377.1, lnc-MSTRG.6553.1, and lnc-MSTRG.30768.1 were related to flavone biosynthesis. In addition to flavonoid biosynthesis, we found one lncRNA in carotene biosynthesis and three lncRNAs in carotenoid biosynthesis (Figure 3C). In contrast to the *cis* model, we observed fewer lncRNAs involved in flavonoid biosynthesis in the antisense model, with only three lncRNAs, lnc-MSTRG.15809.1, lnc-MSTRG.12803.1, and lnc-MSTRG.30533.1, were detected. No lncRNAs involved in flavone biosynthesis process were identified (Figure 3D). Overall, our findings suggest that there are more *cis* models than antisense models involved in regulating flower development and color formation in *M. candidum*. The antisense model involves lncRNAs forming complementary relationships with genes. To investigate this further, we examined several genes, *McMADS4a*, 4b, *McMYB1,17*, *McCHSJ*, and *McCHI1*, along with their corresponding lncRNAs to assess their complementary relationship. Our results showed a high degree of complementarity between the lncRNAs and their target genes (Figure 3E).

3.4 Validation and expression model of lncRNAs in three development stages of *M. candidum*

In order to confirm the accuracy of our lncRNA-seq data, we performed qRT-PCR assays on 12 selected lncRNAs. The expression patterns of most of these lncRNAs were found to be consistent with the results obtained from the sequencing analysis. Specifically, lnc-MSTRG.17619.1, lnc-MSTRG.29477.1, and lnc-MSTRG.34245.1 were highly expressed during the McI stage, while lnc-MSTRG.10215.1, lnc-MSTRG.11402.2, lnc-MSTRG.223785.1, lnc-MSTRG.24719.1, lnc-MSTRG.29558.1, and lnc-MSTRG.29930.1 were highly expressed during the McII stage Figure 4A. Additionally, lnc-MSTRG.28008.2 and lnc-MSTRG.11415.1 were highly expressed during the McIII stage, while lnc-MSTRG.24133.1 was highly expressed in both the McII and McIII stages Figure 4A. We found that the results obtained from the lncRNA-seq analysis and the qRT-PCR experiments were positively correlated, with a slope of 1.147 and an R value of 0.727 for the McII vs McI group, and a slope of 0.968 and an R value of 0.645 for the McIII vs McII group (as shown in Figure 4B). Based on these findings, we concluded that the lncRNA-seq analysis was valid and accurate.

We investigated the expression pattern of these lncRNAs during different stages of the flower development process. In the McII vs McI group, a total of six DElncRNAs were confirmed, of which four were down-regulated and two were up-regulated (Figures 5A, D). Ten DElncRNAs were differentially expressed between the McIII and McI groups, of which three were downregulated and seven were upregulated (Figures 5B, E). Five DElncRNAs related to flower color formation were found between McIII and McII, including one that was down-regulated and four that were up-regulated (Figures 5C, F). During the transition from the second to the third stage of flower development, lnc-MSTRG.10215.1 and lnc-MSTRG.24133.1 were up-regulated and exhibited an upward trend in expression, indicating their corresponding competing endonous mRNAs were key regulator in regulating this period of development. Conversely, lnc-MSTRG.34245.1, lnc-MSTRG.29477.1, and lnc-MSTRG.17619.1 showed a decline in expression from the second stage onwards (Figures 5A, B), indicating their corresponding competing endonous mRNAs were less involved in this process. It was observed that lnc-MSTRG.11415.1 and lnc-MSTRG.29558.1 began to be highly expressed in the third period (McIII), suggesting that these two lncRNAs will play a critical role in flower development after color formation (Figures 5B, C, E, F). There is only a low expression of lnc-MSTRG.11402.2 in the second stage, showing that its target gene plays a negative role in flower coloration process.

3.5 LncRNAs-mediated ceRNA regulatory network build and verification

To further clarify the regulatory mechanism of the aforementioned lncRNAs, we first examined the supplementary relation between lncRNAs and miRNAs, miRNAs and target genes (Figure S1). Subsequently, we constructed a ceRNA regulatory network, which included lncRNAs, miRNAs, mRNAs, and metabolites for flower formation-related pathways (Figure 6A). As shown in the graph, lnc-MSTRG.10215.1 and lnc-MSTRG.11402.2 were found to regulate *McF3H* through interacting with miR5207, ultimately influencing dihydromyricetin (DHM), dihydroquercetin (DHQ), and pinobanksin (PBA). Dihydroquercetin (DHQ), also known as 3, 5, 7, 3, 4 - pentahydroxy flavanone or taxifolin, is a bioactive flavonoid that is considered one of the rarest and most effective natural antioxidants (Yu et al., 2021). Both dihydroquercetin and dihydromyricetin are colorless dihydroflavones that can be reduced to leucoanthocyanidins under the action of the dihydroflavonol-4-reductase (DFR) enzyme. Subsequently, under the catalysis of downstream enzymes, the leucoanthocyanidins undergo a process of the transformation into orange pelargonidin, reddish-purple cyanidin, and violet-blue delphinidin (Tanaka and Brugliera, 2013; Liu et al., 2019). Pinobanksin, one of the most common phenolic constituents of pine heartwood, was first isolated from *Pinus Banksiana* (Erdtman, 1944). When combined with zinc ion, this compound is specific to 3-hydroxyflavanones to produce much deeper colors during the reducing process. Pinobanksin usually produces orange-red when reduced with magnesium or zinc and hydrochloric acid (Lindstedt, 1950). *P*-coumaroyl Shikimic acid is a by-product of the anthocyanin pathway, derived from 4-coumaroyl-CoA by the formation of an ester bond with shikimic acid catalyzed by Shikimate O-hydroxycinnamoyl transferase (HCT) (Paliyath et al., 2009; Li et al., 2021). In the presence of the enzymes C4H1 and C3H3, this compound can be catalyzed again to transform into caffeoyl shikimic acid (Chen et al., 2011). Both *p*-coumaroyl Shikimic acid and caffeoyl shikimic acid are key compounds in Phenylpropanoid biosynthesis during lignin formation. The Sankey graph showed that miR3704, miR480, and miR1220 had cleavage effects on *McHST*, *McAHT1*, and *McPHT3*, respectively. Additionally, we discovered that lnc-MSTRG.24133.1

10.3389/fpls.2023.1215044

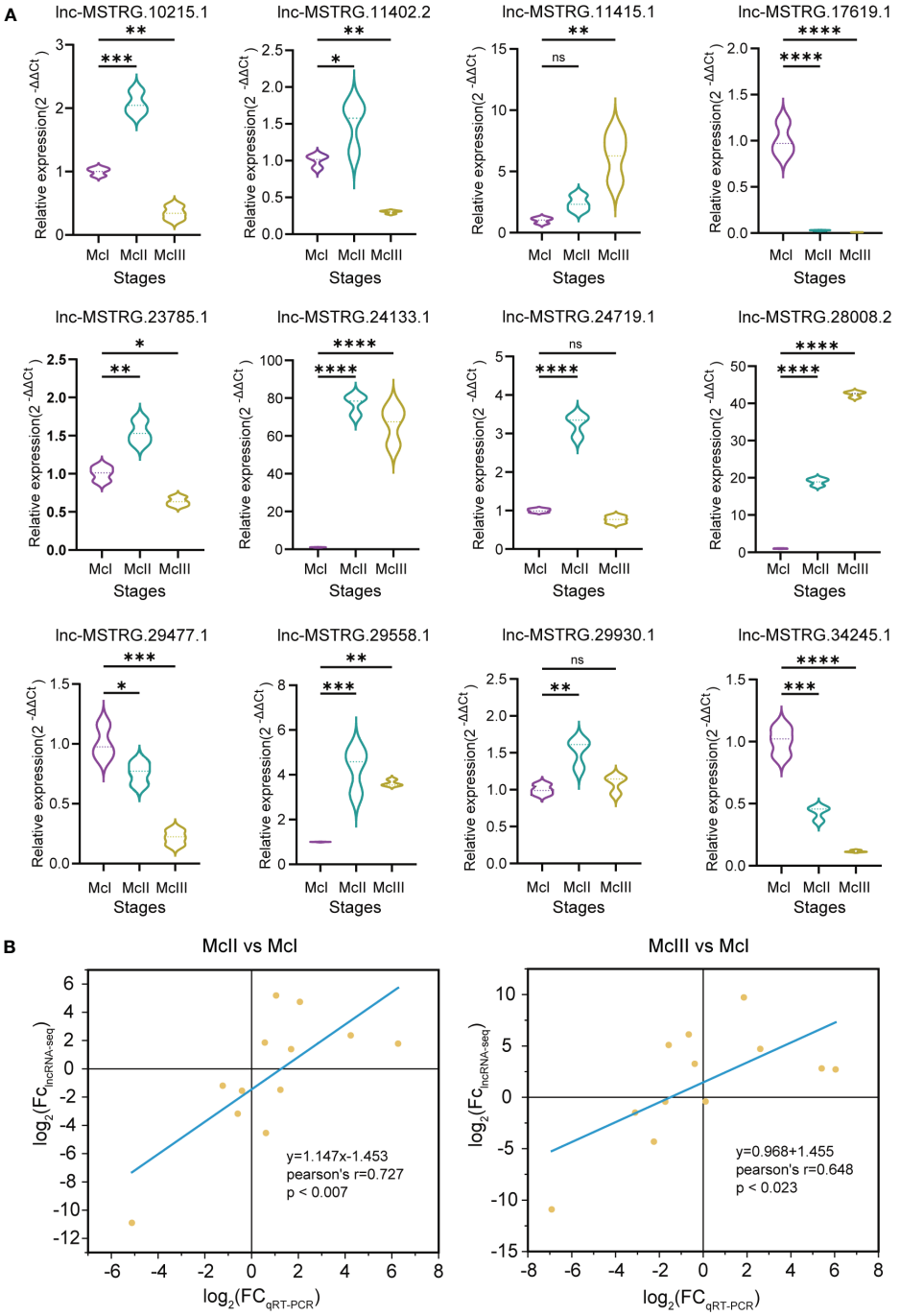


FIGURE 4
Validation of lncRAN-seq results. **(A)** qRT-PCR results of 12 selected lncRNAs. Three colors represent three stages. Dot lines within the violin graphs mean median values. "ns" means p >0.05; "*" means p ≤ 0.05; "**" means p ≤ 0.01; "***" means p ≤ 0.001; "****" means p ≤ 0.0001. **(B)** Fitting result bewtween qRT-PCR and lncRNA-seq. In both qRT-PCR and miRNA-seq results, fold change (FC) was nomorlized by log_2 algorithm.

could adsorb miR3704 to maintain a balance between miRNA and target gene. Three lncRNAs, lnc-MSTRG.17619.1, lnc-MSTRG.29477.5, and lnc-MSTRG.34245.1, were found to bind to miR480 while another set of three lncRNAs, lnc-MSTRG.23785.1, lnc-MSTRG.29558.1, and lnc-MSTRG.29930.1 were found to bind to miR1220. These results indicate that the three lncRNAs that ceRNA regulatory mechanism play an important role in the formation of *p*-coumaroyl Shikimic acid, which ultimately determine flower color. Three specific genes, namely *McUGT88B1*, *McUGT88F3*, and *McCYP98A2*, were found to have no associated with metabolites.

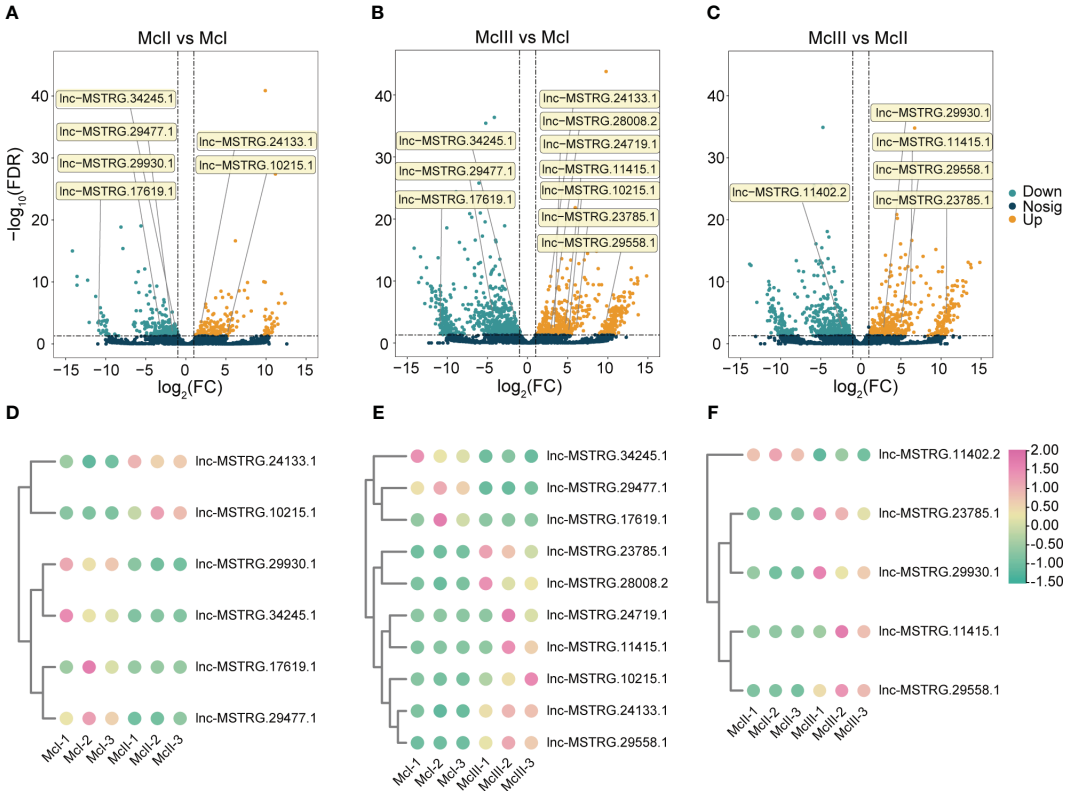


FIGURE 5
Distribution and expression heatmaps of selected lncRNAs. **(A)** Selected lncRNAs in McII vs McI group. **(B)** Selected lncRNAs in McIII vs McI group. **(C)** Selected lncRNAs in McIII vs McII group. In **(A–C)**, orange means up-regulated lncRNAs, and green means down-regulated lncRNAs. **(D)** Heatmap of selected lncRNAs in McII vs McI group. **(E)** Heatmap of selected lncRNAs in McIII vs McI group. **(F)** Heatmap of selected lncRNAs in McIII vs McII group. In graphs **(D–F)**, pink dots mean high expression level, and the light green dot means low expression level.

This may indicate that the three genes related to ceRNAs do not significant affect metabolism. It is also possible that annotation of the metabolite database is incomplete.

We conducted a correlation analysis to determine the relationship between ceRNA members and metabolites in specific pathways (Figure 6B). Each ceRNA regulatory routine showed a negative correlation between miRNAs and both lncRNAs and mRNAs, whereas a positive correlation was observed between lncRNAs and mRNAs, indicating the presence of a true competitive mechanism between lncRNAs and mRNAs. Lnc-MSTRG.28008.2 and *McASP2* were negatively correlated with L-phenylalanine, however, miR8131-3p was positively correlated with L-phenylalanine, indicating that this ceRNA regulatory chain is responsible for L-phenylalanine synthesis (Figure 6B). It is known that both the flavonoid and anthocyanin pathways originate from the amino acid L-phenylalanine, which is deaminated by the action of phenylalanine ammonia-lyase (PAL) to produce trans-cinnamic acid and ammonia (Fasoula et al., 1995). Lnc-MSTRG.29477.1, lnc-MSTRG.29558.1, *McAHT1*, and *McPHT3* were positively correlated with *p*-coumaroyl Shikimic acid, whereas, miR480 and miR1220 were negatively correlated with *p*-coumaroyl Shikimic acid, indicating that these two ceRNAs regulatory chains were key regulator in the synthesis of *p*-coumaroyl Shikimic acid. However, the rest of ceRNA regulatory networks did not display obvious putative correlations, which could be attibuted to more complex regulatory mechanisms like multiple ceRNA chains, histone modification, RNA methylation modification process, *etc*. Further studies are needed to gain a better understanding of these complex regulatory mechanisms.

To further validate the identified lncRNA-mediated ceRNA regulatory network, we employed pCAMBIA1300 harboring target genes, and pCAMBIA2300 harboring lncRNAs and miRNAs to performed tobacco injection and qRT-PCR experiment. Ordinarily, miRNAs have cleavage effects on the target genes, when lncRNAs bind with miRNAs, the cleavage effects will be weakened. We selected 7 lncRNA-mediated ceRNA regulatory routines which included 6 genes, 5 miRNAs, and 6 lncRNAs to validate the ceRNA regulatory mechanism in tobacco (Figure 7). In accordance with our anticipation, when only injecting the miRNAs and target genes into leaves of tobacco, the fluorescence signal became weak (Figure 7), and the expression level of target genes begun to decrease. When we added the lncRNAs in corresponding ceRNA chains, the fluorescence signal was regained, and the expression level of the target gene also began to increase (Figure 7). These results provided evidence for the involvement of lncRNA-mediated ceRNA regulation in *M. candidum*.

10.3389/fpls.2023.1215044

3.6 Functional model of lncRNA-mediated ceRNA during flower development of *M. candidum*

Based on the aforementioned results, we summarized some key lncRNA-mediated ceRNA regulatory relationship during flower development (Figure 8; Figure S2). LncRNAs could bind with a post-transcriptional product of genes directly to degrade them into siRNAs (small interference RNAs). Another common path of lncRNAs is to compete with mRNAs to bind to miRNAs, thereby easing the cleavage effect of miRNA and facilitating more efficient translation of mRNAs (Figure 8A). We observed some ceRNA regulatory mechanisms operating in the flavonoid pathway. Specifically, lnc-MSTRG.11402.2-miR5207-*McF3H* formed a regulatory chain that acts as a catalyst for the transformation from pinocembrin to pinobanksin (Figure 8B). Lnc-MSTRG.28008.2-miR8131-3p-*McASP2* catalyzes the conversion of phenylpyruvate into L-phenylalanine. Lnc-MSTRG.29930.1, lnc-MSTRG.23785.1, and lnc-MSTRG.29558.1 could bind with miR1220 to regulate *McPHT3* genes; lnc-MSTRG.24133.1 could affect *McHST* by interacting with miR3704; Lnc-MSTRG.34245.1, lnc-MSTRG.29477.1, and, lnc-MSTRG.17619.1 completed with miR480 to bind to *McAHT1*. All these three combinations together influenced the metabolic process from *p*-Coumaroyl-CoA to *p*-Coumaroyl shikimic acid (Figure 8B). Out of all the observed lncRNA-miRNA-mRNA-metabolites regulatory chains, lnc-MSTRG.29930.1,

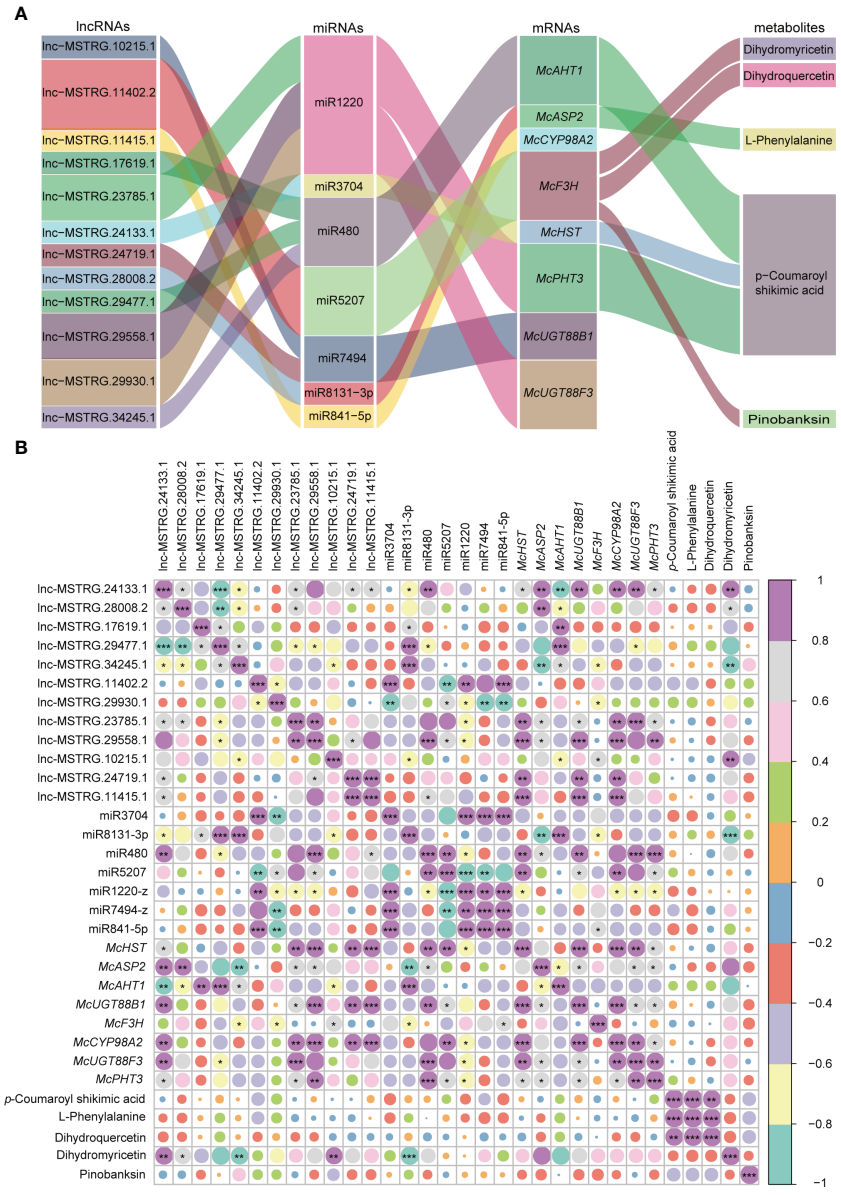


FIGURE 6
Sankey and correlation relationship map of selected lncRNAs, miRNAs, mRNAs, and metabolites. **(A)** Corresponding relationship among lncRNAs, miRNAs, mRNAs, and metabolites. **(B)** Correlation relationship among lncRNAs, miRNAs, mRNAs, and metabolites. Purple means a high correlation coefficient, and green means a low correlation coefficient. "***" means *P*-value is significant at 0.001 level, "**" means the *P*-value is significant at 0.01 level, and "*" means *P*-value is significant at 0.05 level.

10.3389/fpls.2023.1215044

McPHT3, and *p*-Coumaroyl shikimic acid were lowly expressed in the second stage, whereas miR1220 was highly expressed in the second stage, indicating the presence of a ceRNA regulatory mechanism. It was found that lnc-MSTRG.24133.1, *McHST*, and *p*-Coumaroyl shikimic acid were highly expressed in the third stage, but miR3704 was lowly expressed, suggesting that this ceRNA regulatory chain also plays an important role in the biosynthesis of *p*-Coumaroyl shikimic acid (Figure S4). We also found some irregular ceRNA regulatory chains, for example, lnc-MSTRG.11402.2-miR5207-*McF3H*, which was thought to be involved in the synthesis of dihydroquercetin, dihydromyricetin, and pinobanksin. However, we did not observe any corresponding expression trends for these metabolites; lnc-MSTRG.17619.1, lnc-MSTRG.29477.1, lnc-MSTRG.34245.1-miR480-*McAHT1*-*p*-Coumaroyl shikimic acid displayed putative trends in the first stage, whereas, in the third stage, they were inconsistent with the putative results (Figures S3, S4). These irregular phenomena may be caused by other post-transcription regulating or modifications mechanisms.

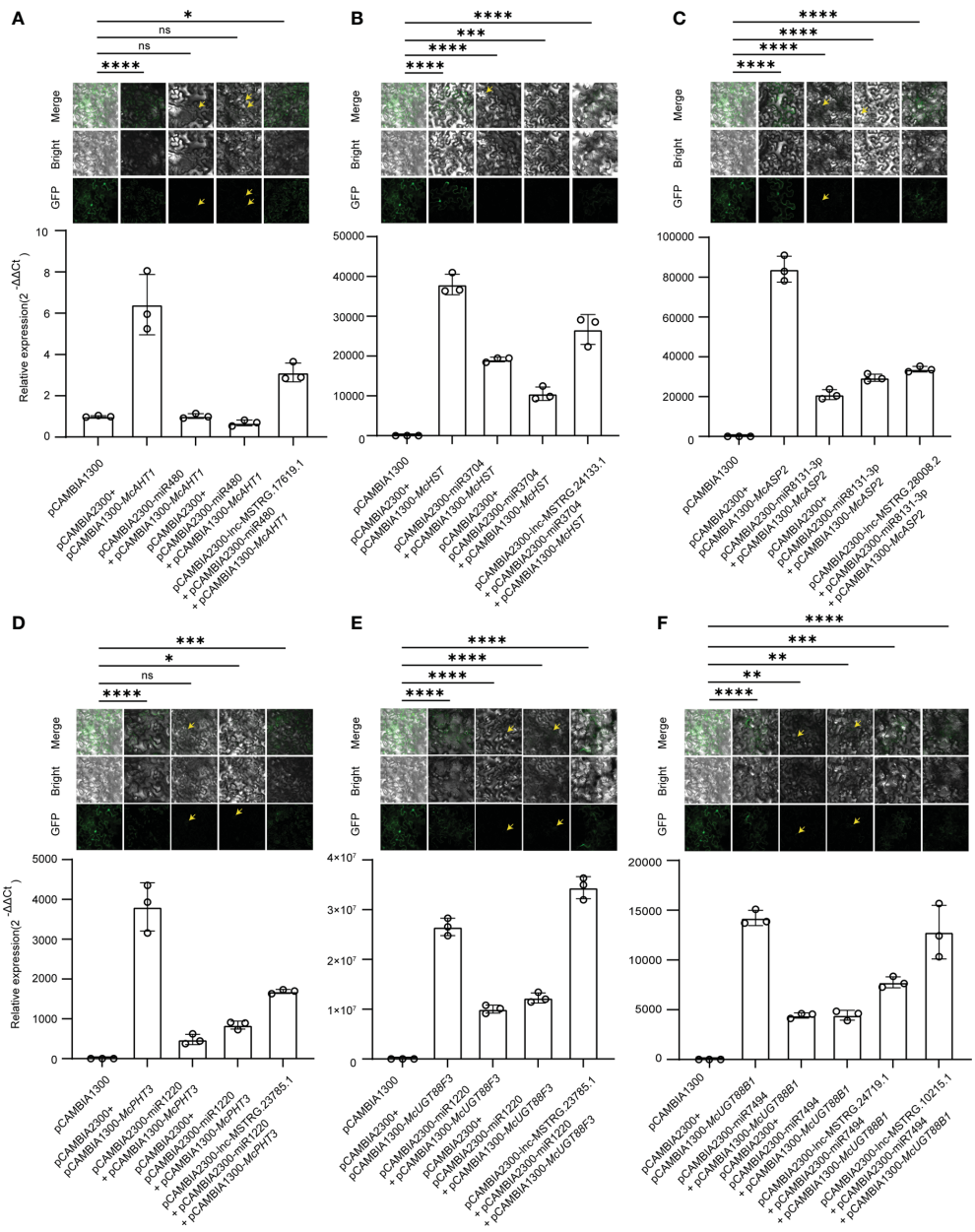


FIGURE 7
Validation of lncRNA-Mediated ceRNA regulatory network. **(A)** Validation of lnc-MSTRG.17619.1-miR480-McAHT1 regulatory chain. **(B)** Validation of lnc-MSTRG.24133.1-miR3704-McHST regulatory chain. **(C)** Validation of lnc-MSTRG.28008.2-miR8131-3p-McASP2 regulatory chain. **(D)** Validation of lnc-MSTRG.23785.1-miR1220-McPHT3 regulatory chain. **(E)** Validation of lnc-MSTRG.23785.1-miR1220-McUGT88F3 regulatory chain. **(F)** Validation of lnc-MSTRG.24719.1-miR7494-McUGT88B1 and lnc-MSTRG.10215.1-miR7494-McUGT88B1 regulatory chains. The pictures above column are GFP fluorescence graphs obtained using a laser scanning confocal microscope (LSCM). "*" means $p \leq 0.05$; "**" means $p \leq 0.01$; "***" means $p \leq 0.001$; "****" means $p \leq 0.0001$; "ns" means not significant. Yellow arrows point to weak fluorescence in the images.

4 Discussion

Early studies classified lncRNAs as transcriptional noise because of their low transcription levels and conservation. To date, lncRNAs are increasingly recognized as an important regulatory molecule in animals and plants. In plants, Numerous studies have proven lncRNAs' key roles in many biological processes, including flower time (Heo and Sung, 2011), nutrition metabolism (Borah et al., 2018), root development (Chen et al., 2018a), abiotic stress (Gai et al., 2018), etc. In this study, we conducted lncRNAs sequencing for the flower of *M. candidum*. In total, 1,499 DElncRNAs were identified. Several lncRNAs with *cis*, trans, and antisense functions on mRNA were identified, and the lncRNAs-mediated ceRNA regulatory mechanism was verified. Our research proved that lncRNAs play a pivotal role in flower development and color control in *M. candidum* via direct or indirect ways.

The statistics databases of lncRNAs include CPC, CNCI, PFAM protein structure domain (PFAM), and coding potential assessment tool (CPAT), which will lead to significant discrepancies in the number of lncRNAs detected. For example, by using CPC and CNCI databases, Zhu et al. (2019a) identified 32,036 lncRNAs in the leaf and shoot of *Camellia sinensis*. Ye et al. (2019) identified 1,860, 3,342, 6,102, and 5,543 lncRNAs by CNCI, CPC, PFAM, and CPAT databases, respectively, in eight tissues of *Ginkgo biloba*, finally, they obtained 1,270 common lncRNAs. Our sequencing results showed that CPC2 and CNC, respectively, produced 4,508 and 6,567 lncRNAs. Finally, 3,955 common lncRNAs based on Venn analysis were identified. Our lncRNA number is more than Ye et al.'s research but far less than Zhu et al.'s research. Since lncRNAs lack codon regions, they are less conserved than protein-coding genes. They may possess a conserved motif but are not easily detected by the BLAST method (Kang and Liu, 2015). Numerous studies had reported that lncRNAs also possessed tissue-specific expression profiles (Ding et al., 2019; Zong et al., 2021). Pairwise comparisons between tissues and stages revealed different proportions of lncRNA isoforms and loci among different tissues or stages in *Fragaria vesca* (Kang and Liu, 2015). So, we concluded that these differences in detected lncRNA numbers in our research are normal and they may be caused by the different detected tissues and species. Further analysis showed that our lncRNAs data was moderate and sufficient to investigate their regulatory relationships involved.

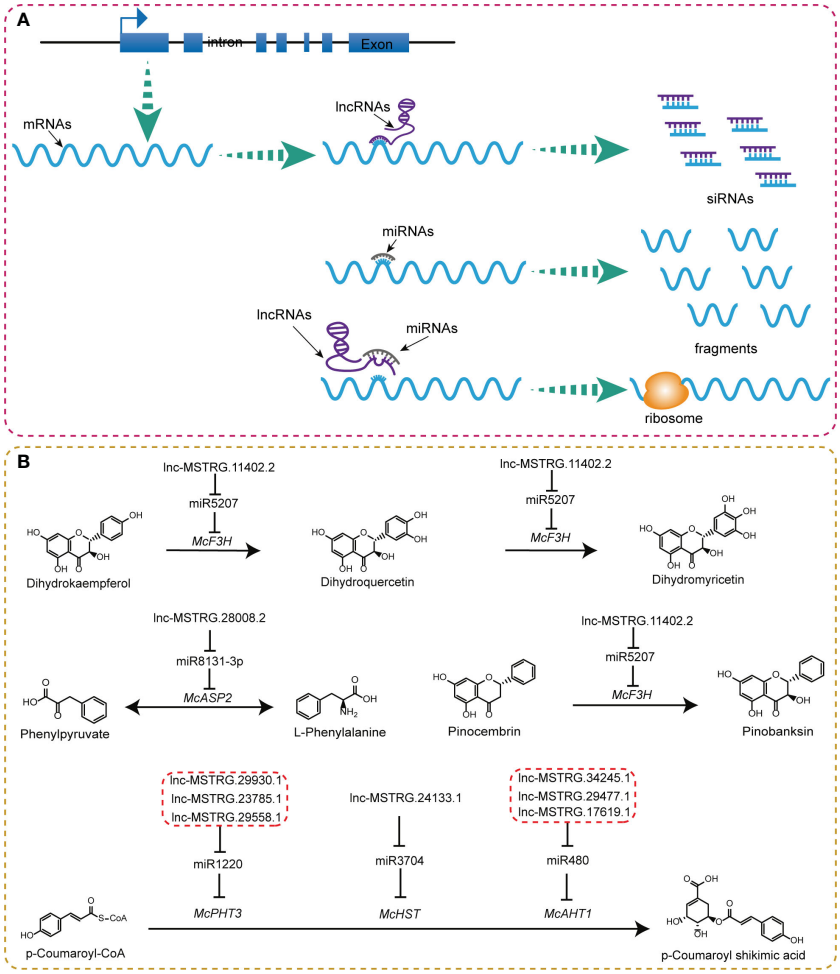


FIGURE 8
LncRNAs-mediated ceRNAs regulatory mechanism in flavonoids pathway. **(A)** Functional model of lncRNA . **(B)** Location of lncRNA-mediated ceRNA regulatory chains in corresponding compound biosynthesis process of flavonoids pathway.

The importance of lncRNAs in coloring and pigment formation has been demonstrated by several studies. For example, TCONS_01039552 and PONTK.3920.2 could regulate the expression of *F3H* in sea buckthorn fruit and *Solanum tuberosum*, respectively (Zhang et al., 2018; Bao et al., 2022). Our research identified two lncRNAs, lnc-MSTRG.10215.1 and lnc-MSTRG.11402.2, that can regulate *F3H* via ceRNA regulatory mechanism. (Figure 6A). In strawberry fruit, TRINITY_DN48515_c0_g3_i1 and TRINITY_DN1328_c0_g1_i1 could positively and negatively correlate with *CHI* and *CHS*, respectively (Lin et al., 2018); in *Solanum tuberosum* L, PONTK.2668.1 and PONTK.2668.15 could regulate *CHS* (Bao et al., 2022). In our research, we observed that Lnc-MSTRG.30526.3 and lnc-MSTRG.30533.1 could regulate *CHSJ* by both *cis* and antisense models (Figures 3C, D). Lnc-MSTRG.20470.1 could regulate *CHI1* through the antisense model (Figure 3D). Zhu et al. (2019a) found that LTCONS00054003 targets *4CL* by *cis* model in fresh leaf and shoot of *Camellia sinensis*. Similarly, our research found that lnc-MSTRG.28304.1 targeted *4CL3*, one member of *4CL*. In addition, lnc-MSTRG.28377.1 targeted on two *COMT1* members, *McCOMT1a* and *McCOMT1b*, and Lnc-MSTRG.19485.1 targeted *McMYB4* (Figure 3C). Previous studies demonstrated that *4CL3* displays a strong preference for 4-coumaric acid as substrate and is expressed at high levels in flowers but not in lignified organs suggesting that the primary function of *4CL3* is to provide activated 4-coumaric acid for the chalcone synthase (CHS) reaction that feeds the flavonoid-specific branch pathways in *Arabidopsis* (Ehlting et al., 1999; Kumar and Ellis, 2003). *AtMYB4* may affect pollen development by altering the flux of the phenylpropanoid pathway and pollen wall composition (Preston et al., 2004). COMT1 could catalyze the conversion of caffeic acid to ferulic acid and of 5-hydroxyferulic acid to sinapic acid (Gowri et al., 1991). COMT1 also methylates 5-hydroxyferuloyl CoA derivatives and flavonols with vicinal aromatic dihydroxy groups, such as quercetin (Fellenberg et al., 2012). In the antisense model, lncRNAs can pair with target genes, such as *McMYB1*, *McMYB17*, *McMADS4*, and *McMADS8* to regulate gene expression and control anthocyanin biosynthesis and pigment accumulation in floral organs. It was reported that a decrease of *MYB1* expression inhibits anthocyanin biosynthesis in bagged Chinese bayberry fruit, suggesting that *MYB1* may be involved in anthocyanin biosynthesis (Niu et al., 2010). In apples, the transcript level of *MYB17* was highly correlated with anthocyanin level, suggesting its role in pigment accumulation. The MADS-box motif has been identified in three different classes of genes in floral organs: A, B, and C. The genes of class A and C are responsible for the development of sepals and carpels, respectively. And the genes of classes B and C together control the formation of stamens (Kang et al., 1998). Additionally, we found some genes exist in both the ceRNA regulatory mechanism and the *cis* model, including *AHT1*, *UGT88B1*, and *CYP98A1*. These results indicated that lncRNAs are key factors to controll flower color formation and development through diverse mechanisms in *M. candidum*.

In the trans model of lncRNA regulation, lncRNAs are able to control gene expression from a distance by interacting with target genes at a different chromosomal locus. (Fukuda et al., 2019; Fu et al., 2020). Previous study demonstrated that DEGs regulated by lncRNAs in the trans model were significantly higher than DEGs regulated by *cis* and miRNA-mediated models, in which 413 lncRNAs were found to be capable of regulating 6060 genes in the trans model (Zong et al., 2021). Our GO and KEGG analysis results also showed a similar phenomenon. According to our results, a lot of lncRNAs have been shown to regulate protein-coding genes. In GO analysis, we found that the antisense model of lncRNAs could not be detected in the flavone biosynthesis process, or the alkaloid metabolic process; *cis* model of lncRNAs could not be detected in the tropane alkaloid biosynthetic process, and the alkaloid metabolic process. KEGG analysis results revealed that the antisense model of lncRNAs was missing in isoflavonoid biosynthesis pathways, folate biosynthesis pathways, flavone and flavonol biosynthesis pathways, as well as brassinosteroid biosynthesis pathways. The *cis* model of lncRNAs was not observed in the pathway for the synthesis of isoflavonoids and anthocyanins. In addition, the enriched gene numbers in the trans model were also significantly higher than in other acting models (Figures 3A, B). While numerous studies have demonstrated that these three acting models exist, researchers are unable to confirm their role directly due to the absence of related techniques. It is urgent to find a method to affirm the exact function of lncRNAs.

Tobacco injection assay is usually influenced by many factors such as ratio of GV3101 harboring different vectors, activity of the bacteria, injection volume of bacteria, and state of the tobacco leaves. These factors could influence experiment result directly. In our qRT-PCR assay of tobacco, we observed that the expression levels of *McUGT88F3* were higher in pCAMBIA2300-lnc-MSTRG.23785.1 + pCAMBIA2300-miR1220 + pCAMBIA1300-*McUGT88F3* combination than in pCAMBIA2300 + pCAMBIA1300-*McUGT88F3* combination. This phenomenon was not consistent with our expectation. In one hand, although we have controlled OD of bacteria to 0.7 - 1.0 and standardized the final OD to ~ 0.8 before injection, the exact states of the bacteria, such as viability of the bacteria and DNA quality in the bacteria, are still hard to keep consistent, which could affect the experiment results. Furthermore, the quality of tobacco leaves, such as the thickness, age, and texture, could affect the efficiency and volume of the injection, eventually causing different expression levels of target genes. Therefore, strict control of the experimental conditions is crucial for the accurate interpretation of the results. Another possible reason is that there may be highly conservated miRNAs in tobacco and *M. candium*. When injecting the exogenous target genes, the endogenous miRNAs of tobacco could also degrade these exogenous target genes, leading to the decrease of the expression level of them. The presence of lncRNA may contribute to the enhancement of *McUGT88F3* expression by protecting it from degradation by both endogenous and exogenous miRNAs, eventually leading to higher expression level of target genes in experiment group than in control group. Further studies are needed to confirm these hypotheses.

5 Conclusion

Collectively, the sole flower color of *M. candidum* greatly limits its application in the gardening industry, which is also a common problem for many garden plants. In our research, we identified specific lncRNAs

that regulate flower development or color formation by comparing different stages of flower development of *M. candidum*. Plant development involves a complex set of physiological and biochemical reactions, such as splicing, methylation modification, histone modification at post- transcription or translation level, which could indirectly influence the expression patterns of functional genes. Since most current research focuses mainly on the direct regulation of functional genes or enzymes, some other key factors controlling flower color may be neglected. Despite lncRNAs being a relatively new classification of non-coding RNAs, they play various roles in plants and should be more widely studied in the future. Our research has proven the pivotal role of lncRNAs during flower development and color formation, offering a new avenue for regulating flower color at non-coding RNA levels in *M. candidum*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI BioProject accession number: PRJNA884434.

Author contributions

LR, SD and YY designed this research; HL and WW collected the samples for this research; HL analysed, visualized the data, and wrote the draft; HL, BT, RL, XD and YL conducted the verification experiments; LR, SD and YY revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Funding of scientific research projects for postdoc (Grant No. 2022BSHKYZZ); project of Guangzhou Ecological Garden Science & Technology Collaborative Innovation Center (202206010058).

Acknowledgments

We greatly appreciate prof Renchao Zhou of Sun Yat-sen University for providing the genome of *M. candidum*; Guangzhou Genedenovo Biotechnology for sequencing and bioinformatics analysis; Prof Quanzi Li of Chinese Academy of Forestry for providing experimental equipment and filed for this 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.2023.1215044/full#supplementary-material

References

Ahmad, P., Bensaoud, C., Mekki, I., Rehman, M. U., and Kotsyfakis, M. (2021). Long Non-Coding RNAs and their potential roles in the vector–host–pathogen triad. *Life* 11, 56. doi: 10.3390/life11010056

Bao, Y., Nie, T., Wang, D., and Chen, Q. (2022). Anthocyanin regulatory networks in *Solanum tuberosum* L. leaves elucidated via integrated metabolomics, transcriptomics, and StAN1 overexpression. *BMC Plant Biol.* 22, 1–17. doi: 10.1186/s12870-022-03557-1

Bayer, E., Egeter, H., Fink, A., Nether, K., and Wegmann, K. (1966). Complex formation and flower colors. *Angew. Chem. Int. Ed. Engl.* 5, 791–798. doi: 10.1002/anie.196607911

Ben Amor, B., Wirth, S., Merchan, F., Laporte, P., d'Aubenton-Carafa, Y., Hirsch, J., et al. (2009). Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* 19, 57–69. doi: 10.1101/gr.080275.108

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306, 2242–2246. doi: 10.1126/science.1103388

Borah, P., Das, A., Milner, M. J., Ali, A., Bentley, A. R., and Pandey, R. (2018). Long non-coding RNAs as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. *Genes* 9, 459. doi: 10.3390/genes9090459

Bowles, D., Lim, E.-K., Poppenberger, B., and Vaistij, F. E. (2006). Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.* 57, 567–597. doi: 10.1146/annurev.arplant.57.032905.105429

Brockington, S. F., Walker, R. H., Glover, B. J., Soltis, P. S., and Soltis, D. E. (2011). Complex pigment evolution in the *Caryophyllales*. *New Phytol.* 190, 854–864. doi: 10.2307/20869113

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Chaofeng, T., Minghua, W., and Guiyuan, L. (2013). The interaction between lncRNA and microRNA contributes to tumor. *Chin. J. Biochem. Mol. Biol.* 29, 1029–1034.

Chen, L.-L. (2016). Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* 41, 761–772. doi: 10.1016/j.tibs.2016.07.003

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

Chen, H.-C., Li, Q., Shuford, C. M., Liu, J., Muddiman, D. C., Sederoff, R. R., et al. (2011). Membrane protein complexes catalyze both 4-and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21253–21258. doi: 10.1073/pnas.1116416109

Chen, L., Shi, S., Jiang, N., Khanzada, H., Wassan, G. M., Zhu, C., et al. (2018a). Genome-wide analysis of long non-coding RNAs affecting roots development at an early stage in the rice response to cadmium stress. *BMC Genomics* 19, 1–10. doi: 10.1186/s12864-018-4807-6

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018b). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1101/274100

Crick, F., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227–1232. doi: 10.1038/1921227a0

de la Fuente, A. (2010). From 'differential expression' to 'differential networking'–identification of dysfunctional regulatory networks in diseases. *Trends Genet.* 26, 326–333. doi: 10.1016/j.tig.2010.05.001

Deng, J., Chen, S., Yin, X., Wang, K., Liu, Y., Li, S., et al. (2013). Systematic qualitative and quantitative assessment of anthocyanins, flavones and flavonols in the petals of 108 lotus (*Nelumbo nucifera*) cultivars. *Food Chem.* 139, 307–312. doi: 10.1016/j.foodchem.2013.02.010

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2654–2659. doi: 10.1073/pnas.1121374109

Ding, Z., Tie, W., Fu, L., Yan, Y., Liu, G., Yan, W., et al. (2019). Strand-specific RNA-seq based identification and functional prediction of drought-responsive lncRNAs in cassava. *BMC Genomics* 20, 1–13. doi: 10.1186/s12864-019-5585-5

Ehlting, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., and Kombrink, E. (1999). Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* 19, 9–20. doi: 10.1046/j.1365-313X.1999.00491.x

Erdtman, H. (1944). The phenolic constituent of pineheartwood. *Sven. Kem. Tidskr.* 56, 26.

Fang, J., Zhang, F., Wang, H., Wang, W., Zhao, F., Li, Z., et al. (2019). Ef-cd locus shortens rice maturity duration without yield penalty. *Proc. Natl. Acad. Sci. U.S.A.* 116, 18717–18722. doi: 10.1073/pnas.1815030116

Fasoula, D. A., Stephens, P. A., Nickell, C. D., and Vodkin, L. O. (1995). Cosegregation of purple-throat flower color with dihydroflavonol reductase polymorphism in soybean. *Crop Sci.* 35, 1028–1031. doi: 10.2135/cropsci1995.0011183X003500040017x

Fellenberg, C., van Ohlen, M., Handrick, V., and Vogt, T. (2012). The role of CCoAOMT1 and COMT1 in *Arabidopsis* anthers. *Planta* 236, 51–61. doi: 10.1007/s00425-011-1586-6

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079

Fu, L., Ding, Z., Tan, D., Han, B., Sun, X., and Zhang, J. (2020). Genome-wide discovery and functional prediction of salt-responsive lncRNAs in duckweed. *BMC Genomics* 21, 1–14. doi: 10.1186/s12864-020-6633-x

Fukuda, M., Nishida, S., Kakei, Y., Shimada, Y., and Fujiwara, T. (2019). Genome-wide analysis of long intergenic noncoding RNAs responding to low-nutrient conditions in *Arabidopsis thaliana*: possible involvement of trans-acting siRNA3 in response to low nitrogen. *Plant Cell Physiol.* 60, 1961–1973. doi: 10.1093/pcp/pcz048

Gai, Y.-P., Yuan, S.-S., Zhao, Y.-N., Zhao, H.-N., Zhang, H.-L., and Ji, X.-L. (2018). A novel LncRNA, MuLnc1, associated with environmental stress in Mulberry (*Morus multicaulis*). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00669

Gowri, G., Bugos, R. C., Campbell, W. H., and Dixon, M. R. A. (1991). Stress Responses in Alfalfa (*Medicago sativa* L.) X. Molecular cloning and expression of S-Adenosyl-L-Methionine: caffeic acid3-0-methyltransferase, a key enzyme of lignin biosynthesis. *Plant Physiol.* 97, 7–14. doi: 10.1104/pp.97.1.7

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223. doi: 10.1038/nature07672

He, X., Guo, S., Wang, Y., Wang, L., and Sun, J. (2020). Systematic identification and analysis of heat: tress-responsive lncRNAs, circRNAs and miRNAs with associated coexpression and ceRNA networks in cucumber (*Cucumis sativus* L.). *Physiol. Plant* 168, 736–754. doi: 10.1111/ppl.12997

Heo, J. B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76–79. doi: 10.1126/science.1197349

Holton, T. A. (1995). Modification of flower colour via manipulation of P450 gene expression in transgenic plants. *Drug Metab. Drug Interact.* 12, 359–368. doi: 10.1515/DMDI.1995.12.3-4.359

Holton, T., and Cornish, E. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7, 1071–1083. doi: 10.26877/asset.v4i1.11659

Kang, C., and Liu, Z. (2015). Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics* 16, 1–15. doi: 10.1186/s12864-015-2014-2

Kang, H. G., Jeon, J. S., and Lee, S. (1998). "Identification of class B and class C floral organ identity genes from rice plants. *Plant Mol. Biol.* 38, 1021–1029. doi: 10.1023/A:1006051911291

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kim, D.-H. , and Sung, S. (2017). Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev. Cell* 40, 302–312. e304. doi: 10.1016/j.devcel.2016.12.021

Kishimoto, S., Maoka, T., Nakayama, M., and Ohmiya, A. (2004). Carotenoid composition in petals of chrysanthemum (*Dendranthema grandiflorum* (Ramat.) Kitamura). *Phytochemistry* 65, 2781–2787. doi: 10.1016/j.phytochem.2004.08.038

Kondo, T., Ueda, M., Isobe, M., and Goto, T. (1998). A new molecular mechanism of blue color development with protocyanin, a supramolecular pigment from cornflower, *Centaurea cyanus*. *Tetrahedron Lett.* 39, 8307–8310. doi: 10.1016/S0040-4039(98)01858-9

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, 345–349. doi: 10.1093/nar/gkm391

Kumar, A., and Ellis, B. E. (2003). 4-Coumarate: CoA ligase gene family in Rubus idaeus: cDNA structures, evolution, and expression. *Plant Mol. Biol.* 51, 327–340. doi: 10.1023/A:1022004923982

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 1–16. doi: 10.1186/1471-2105-12-323

Li, X., Fan, J., Luo, S., Yin, L., Liao, H., Cui, X., et al. (2021). Comparative transcriptome analysis identified important genes and regulatory pathways for flower color variation in *Paphiopedilum hirsutissimum*. *BMC Plant Biol.* 21, 1–17. doi: 10.1186/s12870-021-03256-3

Li, C., Jin, H., Zhang, W., Qin, T., Zhang, X., Pu, Z., et al. (2022). Whole-transcriptome analysis reveals long noncoding RNAs involved in female floral development of hickory (*Carya cathayensis Sarg.*). *Front. Genet.* 13. doi: 10.3389/fgene.2022.910488

Lin, Y., Jiang, L., Chen, Q., Li, Y., Zhang, Y., Luo, Y., et al. (2018). Comparative transcriptome profiling analysis of red-and white-fleshed strawberry (*Fragaria* × *ananassa*) provides new insight into the regulation of the anthocyanin pathway. *Plant Cell Physiol.* 59, 1844–1859. doi: 10.1093/pcp/pcy098

Lindstedt, G. (1950). Constituents of Pine heartwood. XXI. The structure of pinobanksin. *Acta Chem. Scand.* 4, 772–781. doi: 10.3891/acta.chem.scand.04-0772

Liu, T., Chen, Y., Chao, L., Wang, S., Wu, W., Dai, S., et al. (2014). Extensive hybridization and introgression between *Melastoma candidum* and *M. sanguineum*. *PloS One* 9, e96680. doi: 10.1371/journal.pone.0096680

Liu, H., Lou, Q., Ma, J., Su, B., Gao, Z., and Liu, Y. (2019). Cloning and functional characterization of dihydroflavonol 4-reductase gene involved in anthocyanidin biosynthesis of grape hyacinth. *Int. J. Mol. Sci.* 20, 4743. doi: 10.3390/ijms20194743

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3′ processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. *Science* 327, 94–97. doi: 10.1126/science.1180278

Liu, D., Shi, L., Han, C., Yu, J., Li, D., and Zhang, Y. (2012). Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. *PloS One* 7, e46451. doi: 10.1371/journal.pone.0046451

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA package 2.0. *Algorithms Mol. Biol.* 6, 1–14. doi: 10.1186/1748-7188-6-26

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/PREACCEPT-8897612761307401

Lu, Y., Deng, S., Li, Z., Wu, J., Liu, Q., Liu, W., et al. (2019). Competing endogenous RNA networks underlying anatomical and physiological characteristics of poplar wood in acclimation to low nitrogen availability. *Plant Cell Physiol.* 60, 2478–2495. doi: 10.1093/pcp/pcz146

Mitsui, S., Hayashi, K., and Hattori, S. (1959). Further studies on commelinin, a crystalline blue metallo-anthocyanin from the flowers of commelina studies on anthocyanins. *Proc. Jpn. Acad.* 35, 169–174. doi: 10.2183/pjab1945.35.169

Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends Plant Sci.* 3, 212–217. doi: 10.1016/S1360-1385(98)01242-4

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943

Nie, L., Wu, H. J., Hsu, J. M., Chang, S. S., and Hung, M. C. (2012). Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am. J. Transl. Res.* 4, 127–150. doi: 10.1155/2012/952452

Nielsen, K. M., Lewis, D. H., and Morgan, E. R. (2003). Characterization of carotenoid pigments and their biosynthesis in two yellow flowered lines of *Sandersonia aurantiaca* (Hook). *Euphytica* 130, 25–34. doi: 10.1023/A:1022328828688

Niu, S.-S., Xu, C.-J., Zhang, W.-S., Zhang, B., Li, X., Lin-Wang, K., et al. (2010). Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta* 231, 887–899. doi: 10.1007/s00425-009-1095-z

Paliyath, G., Murr, D. P., Handa, A. K., and Lurie, S. (2009). *Postharvest biology and technology of fruits, vegetables, and flowers* John Wiley & Sons, (2th Ed), USA.

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, 290–295. doi: 10.1038/nbt.3122

Preston, J., Wheeler, J., Heazlewood, J., Li, S. F., and Parish, R. W. (2004). *AtMYB32* is required for normal pollen development in *Arabidopsis thaliana*. *Plant J.* 40, 979–995. doi: 10.1111/j.1365-313X.2004.02280.x

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

Shen, S., Park, J. W., Lu, Z.-x., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5593–E5601. doi: 10.1073/pnas.1419161111

Shimada, S., Otsuki, H., and Sakuta, M. (2007). Transcriptional control of anthocyanin biosynthetic genes in the *Caryophyllales*. *J. Exp. Bot.* 58, 957–967. doi: 10.1093/jxb/erl256

Shoji, K., Miki, N., Nakajima, N., Momonoi, K., Kato, C., and Yoshida, K. (2007). Perianth bottom-specific blue color development in tulip cv. Murasakizuisho requires ferric ions. *Plant Cell Physiol.* 48, 243–251. doi: 10.1093/pcp/pcl060

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41, e166. doi: 10.1093/nar/gkt646

Sunnadeniya, R., Bean, A., Brown, M., Akhavan, N., Hatlestad, G., Gonzalez, A., et al. (2016). Tyrosine hydroxylation in betalain pigment biosynthesis is performed by cytochrome P450 enzymes in beets (*Beta vulgaris*). *PloS One* 11, e0149417. doi: 10.1371/journal.pone.0149417

Takeda, K. (2006). Blue metal complex pigments involved in blue flower color. *Proc. Jpn. Acad.* 82, 142–154. doi: 10.2183/pjab.82.142

Tanaka, Y., and Brugliera, F. (2013). Flower colour and cytochromes P450. *Phytochem. Rev.* 368, 20120432. doi: 10.1007/s11101-006-9003-7

Tanaka, Y., Tsuda, S., and Kusumi, T. (1998). Metabolic engineering to modify flower color. *Plant Cell Physiol.* 39, 1119–1126. doi: 10.1093/oxfordjournals.pcp.a029312

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, M., Wu, H.-J., Fang, J., Chu, C., and Wang, X.-J. (2017). A long noncoding RNA involved in rice reproductive development by negatively regulating osa-miR160. *Sci. Bull.* 62, 470–475. doi: 10.1016/j.scib.2017.03.013

Willstätter, R., and Everest, A. E. (1913). Untersuchungen über die Anthocyane. I. Über den Farbstoff der Kornblume. *Justus Liebigs Ann. Chem.* 401, 189–232. doi: 10.1002/jlac.19134010205

Wu, X., Shi, T., Iqbal, S., Zhang, Y., Liu, L., and Gao, Z. (2019). Genome-wide discovery and characterization of flower development related long non-coding RNAs in *Prunus mume*. *BMC Plant Biol.* 19, 1–17. doi: 10.1186/s12870-019-1672-7

Wunderlich, M., Groß-Hardt, R., and Schöffl, F. (2014). Heat shock factor HSFB2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. *Plant Mol. Biol.* 85, 541–550. doi: 10.1007/s11103-014-0202-0

Xu, W., Bao, W., Liu, H., Chen, C., Bai, H., Huang, M., et al. (2021). Insights into the molecular mechanisms of late flowering in *Prunus sibirica* using whole-genome and transcriptome analyses. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.802827

Xu, X.-W., Zhou, X.-H., Wang, R.-R., Peng, W.-L., An, Y., and Chen, L.-L. (2016). Functional analysis of long intergenic non-coding RNAs in phosphate-starved rice using competing endogenous RNA network. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep20715

Yang, Z., Yang, C., Wang, Z., Yang, Z., Chen, D., Wu, Y., et al. (2019). LncRNA expression profile and ceRNA analysis in tomato during flowering. *PloS One* 14, e0210650. doi: 10.1371/journal.pone.0210650

Yanofsky, C. (2007). Establishing the triplet nature of the genetic code. *Cell* 128, 815–818. doi: 10.1016/j.cell.2007.02.029

Ye, J., Cheng, S., Zhou, X., Chen, Z., Kim, S. U., Tan, J., et al. (2019). A global survey of full-length transcriptome of *Ginkgo biloba* reveals transcript variants involved in flavonoid biosynthesis. *Ind. Crops Prod.* 139, 111547. doi: 10.1016/j.indcrop.2019.111547

Yu, Z.-W., Zhang, N., Jiang, C.-Y., Wu, S.-X., Feng, X.-Y., and Feng, X.-Y. (2021). Exploring the genes involved in biosynthesis of dihydroquercetin and dihydromyricetin in *Ampelopsis grossedentata*. *Sci. Rep.* 11, 1–14. doi: 10.1038/s41598-021-95071-x

Zhang, G., Chen, D., Zhang, T., Duan, A., Zhang, J., and He, C. (2018). Transcriptomic and functional analyses unveil the role of long non-coding RNAs in anthocyanin biosynthesis during *sea buckthorn* fruit ripening. *DNA Res.* 25, 465–476. doi: 10.1093/dnares/dsy017

Zhang, J., Mujahid, H., Hou, Y., Nallamilli, B. R., and Peng, Z. (2013). Plant long ncRNAs: a new frontier for gene regulatory control. *Am. J. Plant Sci.* 04, 1038–1045. doi: 10.4236/ajps.2013.45128

Zhu, H.-h., Yang, J.-x., Xiao, C.-h., Mao, T.-y., Zhang, J., and Zhang, H.-y. (2019b). Differences in flavonoid pathway metabolites and transcripts affect yellow petal colouration in the aquatic plant *Nelumbo nucifera*. *BMC Plant Biol.* 19, 1–18. doi: 10.1186/s12870-019-1886-8

Zhu, C., Zhang, S., Fu, H., Zhou, C., Chen, L., Chang, X., et al. (2019a). Transcriptome and phytochemical analyses provide new insights into long non-coding RNAs and characteristic secondary metabolites of oolong tea (*Camellia sinensis*) in solar-withering. *Front. Plant Sci.* 10, 1683. doi: 10.21203/rs.2.11183/v1

Zhu, M., Zhang, M., Xing, L., Li, W., Jiang, H., Wang, L., et al. (2017). Transcriptomic analysis of long non-coding RNAs and coding genes uncovers a complex regulatory network that is involved in maize seed development. *Genes* 8, 274. doi: 10.3390/genes8100274

Zong, X., Wang, S., Han, Y., Zhao, Q., Xu, P., Yan, Q., et al. (2021). Genome-wide profiling of the potential regulatory network of lncRNA and mRNA in *Melilotus albus* under salt stress. *Environ. Exp. Bot.* 189, 104548. doi: 10.1016/j.envexpbot.2021.104548

Zuo, J., Wang, Y., Zhu, B., Luo, Y., Wang, Q., and Gao, L. (2019). Network analysis of noncoding RNAs in pepper provides insights into fruit ripening control. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-45427-1

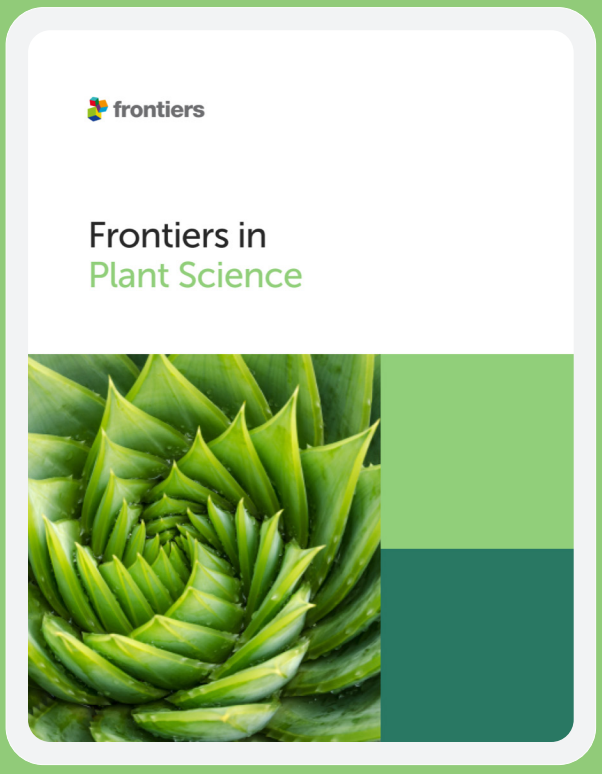
Frontiers in
Plant Science

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