GENOMICS IN FLOWER DEVELOPMENT: FROM 'OMICS' TO FUNCTIONAL CHARACTERIZATION

EDITED BY: Lin Zhang, Heping Cao, Yunpeng Cao and Liangsheng Zhang

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-996-6
DOI 10.3389/978-2-88974-996-6

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

GENOMICS IN FLOWER DEVELOPMENT: FROM 'OMICS' TO FUNCTIONAL CHARACTERIZATION

Topic Editors:
Lin Zhang, Central South University Forestry and Technology, China
Heping Cao, United States Department of Agriculture (USDA), United States
Yunpeng Cao, Chinese Academy of Sciences (CAS), China
Liangsheng Zhang, Zhejiang University, China

Citation: Zhang, L., Cao, H., Cao, Y., Zhang, L., eds. (2022).
Genomics in Flower Development: From 'Omics' to Functional Characterization.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-996-6

Table of Contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 March 2021
doi: 10.3389/fgene.2021.643552

Comparative Analysis of Complete Chloroplast Genome Sequences in *Edgeworthia* (Thymelaeaceae) and New Insights Into Phylogenetic Relationships

Shaojuan Qian[1], Yonghong Zhang[1] and Shiou Yih Lee[2]**

[1] School of Life Sciences, Yunnan Normal University, Kunming, China, [2] State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Science, Sun Yat-sen University, Guangzhou, China

The complete chloroplast genomes of three species of *Edgeworthia* namely, *Edgeworthia albiflora*, *Edgeworthia chrysantha*, and *Edgeworthia gardneri* (Thymelaeaceae), are reported and characterized. The chloroplast genomes displayed a typical quadripartite structure with conserved genome arrangement and specific divergence. The genomes ranged in length from 172,708 to 173,621 bp and displayed similar GC content of 36.5–36.7%. A total of 138–139 genes were predicted, including 92–93 protein-coding, 38 tRNAs and eight rRNAs genes. Variation in the number of short simple repeats and inverted region boundaries of the three cp genomes were observed. A mutational hotspot was detected along the nucleotide sequence from the *ndh*F to the *trn*L-UAG genes. The chloroplast genome-based and internal transcribed spacer (ITS)-based phylogenetic analyses using maximum-likelihood (ML) and Bayesian inference (BI) revealed that *E. albiflora* diverged before *E. chrysantha* and *E. gardneri* and placed the *Edgeworthia* clade at the base of the Eurasian Daphne group with strong bootstrap support. With an effective taxonomic treatment of the species of *Edgeworthia*, further molecular analyses of their intra- and interspecific genetic variation are inclined to support the treatment of *E. albiflora* and *E. gardneri* as two natural groups. The genetic information obtained from this study will provide valuable genomic resources for the identification of additional species and for deducing the phylogenetic evolution of *Edgeworthia*.

Keywords: chloroplast genome, comparative analysis, *Edgeworthia*, internal transcribed spacers, phylogenetic relationship, Thymelaeaceae

OPEN ACCESS

Edited by:
Heping Cao,
United States Department of Agriculture (USDA), United States

Reviewed by:
Ueric José Borges de Souza,
Federal University of Tocantins, Brazil
Ze Peng,
University of Florida, United States

****Correspondence:***
Yonghong Zhang
daphnecn@aliyun.com
Shiou Yih Lee
leesy3@mail.sysu.edu.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *18 December 2020*
Accepted: *18 February 2021*
Published: *10 March 2021*

Citation:
Qian SJ, Zhang YH and Lee SY (2021) Comparative Analysis of Complete Chloroplast Genome Sequences in Edgeworthia (Thymelaeaceae) and New Insights Into Phylogenetic Relationships. Front. Genet. 12:643552. doi: 10.3389/fgene.2021.643552

INTRODUCTION

The family Thymelaeaceae is composed of about 900 species in 45 genera. The most recently proposed taxonomic classification based on palynological findings divided the family into two major subfamilies, Octolepidoideae and Thymelaeoideae. The latter consists of three tribes: Aquilarieae, Daphneae, and Synandrodaphneae (Herber, 2003). The Daphneae accounts for the most genera, which were further clustered into four different groups, Daphne, Gnidia, Linostoma, and Phaleria, with the Daphne group containing the most genera.

As a member of the Daphne group, *Edgeworthia* is reported to contain five species; *E. albiflora* Nakai, *E. chrysantha* Lindl., *Edgeworthia eriosolenoides* K. M. Feng & S. C. Huang, *E. gardneri* (Wall.) Meisn., and *E. longipes* Lace. They occur naturally in China, India, and nearby regions (The Plant List, 2013; Wang and Gilbert, 2017). Studies of the phytochemical and pharmacological properties of *Edgeworthia* have received much attention among researchers (Nan et al., 2018), as the inflorescences and stems of *E. chrysantha* and *E. gardneri* are regarded as effective folk medicines for muscle relaxation and to treat rheumatism (Xiao, 2002; Che et al., 2010). Rich in low-lignin fibers and their ease of propagation, plants of *Edgeworthia* are not only cultivated as ornamentals in urban areas (Clennett et al., 2002; Wang et al., 2017), but are also the preferred raw material for high quality paper products, such as banknotes, and artificial cotton production (Lan et al., 2013). For molecular information, genetic studies on *Edgeworthia* were confined to the molecular placement of *E. chrysantha* within the Thymelaeaceae using short gene sequences (Van der Bank et al., 2002; Beaumont et al., 2009; Motsi et al., 2010; Foster et al., 2016), while genetic information for other species of *Edgeworthia* is limited.

The chloroplast (cp) genome is responsible for photosynthesis (Leister, 2003; Wicke et al., 2011). In general, the cp is maternally inherited and consists of a quadripartite circular double-stranded DNA molecule that comprises two copies of inverted repeat (IRs) regions, a large single-copy (LSC) region, and a small single copy (SSC) region (Palmer, 1985). The length of a typical cp genome ranges between 120,000 and 160,000 bp but variations can occur. Due to its relatively small size, simple structure, conserved gene content, and order, cp genome sequences have been widely used in phylogenetic studies and provide valuable data for resolving complex evolutionary relationships (Jansen et al., 2007; Moore et al., 2010).

At present, there are only 25 complete cp genomes for taxa in the family Thymelaeaceae available publicly in the GenBank database (as of 1st December 2020), with *Aquilaria sinensis* as the first taxon reported (Wang et al., 2016). Due to the lack of molecular information on the genus *Edgeworthia*, we used next-generation sequencing technology to obtain the complete cp genomes of three species of *Edgeworthia*, including *E. albiflora*, *E. chrysantha*, and *E. gardneri*. We constructed and characterized the cp genome structure of these species and performed phylogenetic analyses at the genome-scale level. In addition, to expand the genomic resources from these valuable species, we also sequenced the nuclear ribosomal DNA internal transcribed spacer (ITS) region to reveal the phylogenetic relationships of these species of *Edgeworthia* to other closely related taxa in the Daphne group.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

Fresh leaf samples from three species of *Edgeworthia* namely, *E. albiflora*, *E. chrysantha*, and *E. gardneri*, were collected from plants in their natural habitat and were stored in Ziplock bags filled with silica gel beads prior to transportation to the laboratory. Voucher specimens of the three species were deposited in the Herbarium of Yunnan Normal University (YNUB) (**Table 1**). Total genomic DNA was extracted from the silica-dried leaves using the modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987) and was further purified using Wizard® DNA Clean-Up System (Promega, United States).

Genome Sequencing, Assembly and Annotation

Next-generation sequencing was conducted on the Illumina HiSeq 2500 platform and a 350-bp paired-end library was prepared. The raw reads were filtered to obtain high-quality clean reads using NGS QC Toolkit v2.3.3 with default parameters (Patel and Jain, 2012). The cp genome was assembled using NOVOPlasty (Dierckxsens et al., 2017) with the *rbc*L gene of *Daphne kiusiana* Miq. extracted from the complete cp genome sequence (GenBank accession KY991380) as the seed sequence. The complete cp genome was annotated using Geneious v10.1.3 (Kearse et al., 2012) by referring to the cp genome sequence of *D. kiusiana*. Annotations on the protein-coding (CDS) sequences present in the genome were manually checked using the open reading frame (ORF) and the tRNA genes were verified using the online tRNAscan-SE web server with default settings (Lowe and Chan, 2016). The complete cp genome was visualized using OGDRAW v1.3.1 (Greiner et al., 2019); all cp genome sequences were deposited in the NCBI GenBank database under the accession numbers MW246180 (*E. albiflora*), MN511715 (*E. chrysantha*), MW246181 (*E. gardneri*) for future reference.

Repeats Analyses

Simple sequence repeats (SSRs) were identified using the MISA-web (Beier et al., 2017). The minimum number of repeats was set at 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively. SSRs were manually checked for redundancy. Identification of the four different types (forward, palindromic, reverse, and complement) of large repeats were conducted with REPuter (Kurtz et al., 2001). The size and identity of the large repeats were limited to not less than 30 bp and 90%, respectively; while the Hamming distance was set at 3.0.

Genome Comparison

The junctions and borders of the inverted repeat (IR) regions were visualized using IRscope (Amiryousefi et al., 2018) and further edited using Adobe Photoshop CS6 (Adobe, United States). For comparative analysis, the three sequences of the *Edgeworthia* cp genomes were compared using mVISTA (Mayor et al., 2000) using Shuffle-LAGAN mode. The cp genome sequence of *E. albiflora* was selected as the reference genome. The output image was manually edited using Adobe Illustrator 2020 (Adobe, United States). All three genome sequences of the *Edgeworthia* cp were aligned using MAFFT v7.409 (Katoh and Standley, 2013). Highly divergent regions between the species were identified using DnaSP v5.10 (Librado and Rozas, 2009). The nucleotide divergence values of the cp genome sequence alignment were analyzed using the sliding

TABLE 1 | Basic characteristics of chloroplast genomes in three species of *Edgeworthia*.

	E. albiflora	*E. chrysantha*	*E. gardneri*
Sample location	Miyi, Sichuan, China	Kunming Botanical Garden, Yunnan, China	Motuo, Tibet, China
Collector and collection number	Zhang Y.; ZYH11	Zhang Y.; RXK26	Zhang Y.; RXK53
Chloroplast genome GenBank accession number	MW246180	MN511715	MW246181
ITS GenBank accession number	MW255615	MW255616	MW255617
LSC (bp)	86,862	85,824	86,388
SSC (bp)	2,681	2,816	3,017
IR (bp)	42,039	42,034	41,952
Total (bp)	173,621	172,708	173,309
Protein-coding genes	93	93	92
tRNA genes	38	38	38
rRNA genes	8	8	8
Total	139	139	138
GC content (%)	36.5	36.7	36.5

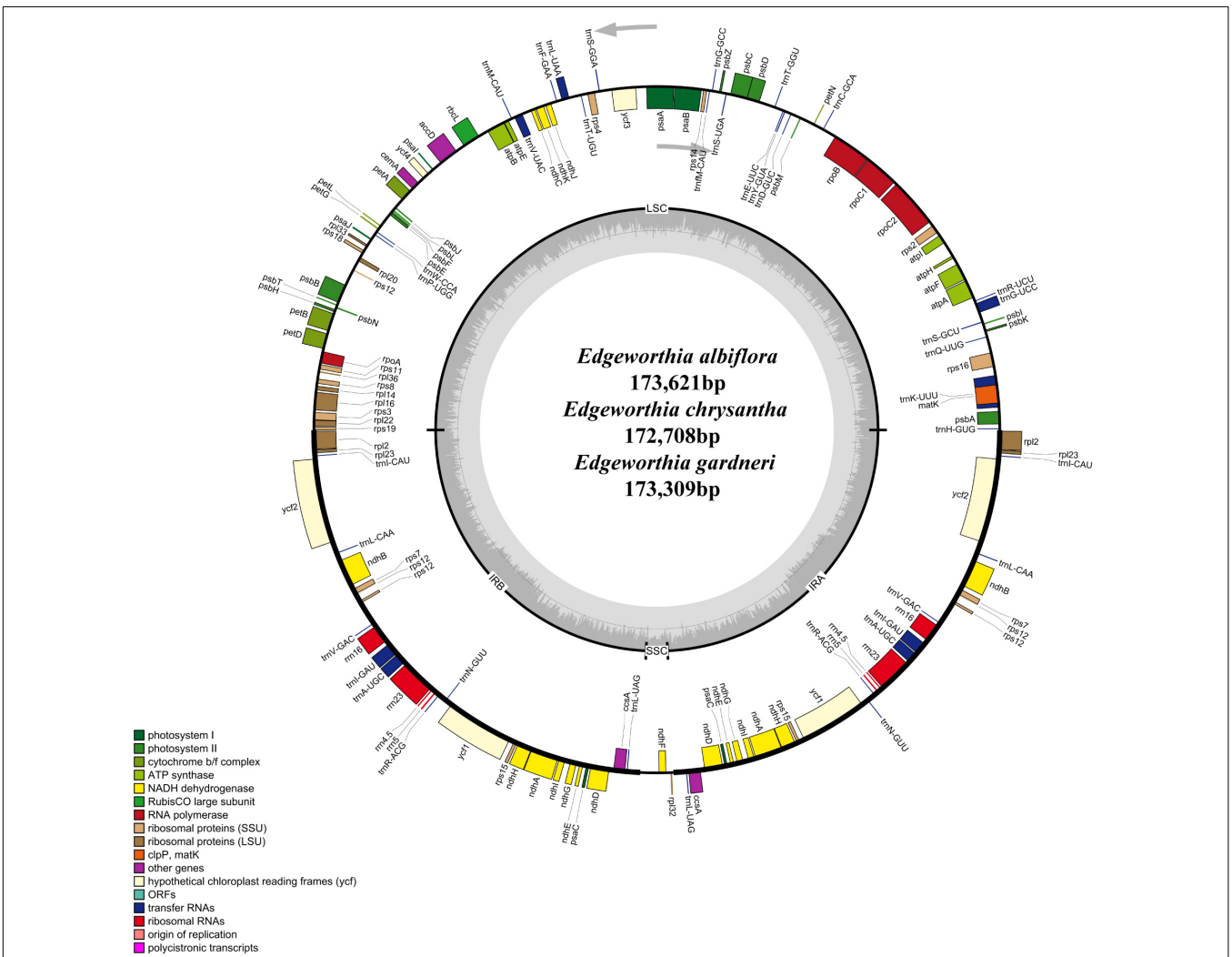


FIGURE 1 | Gene map for chloroplast genomes of three species of *Edgeworthia*. Annotated genes are colored according to functional categories. Genes placed outside the outer circle were transcribed clockwise; genes placed inside the circle were transcribed counterclockwise. Dark gray in inner circle represents GC content whereas light gray corresponds to AT content.

window method, with a window length of 1000 bp and a 500-bp step size.

Codon Use Preference Analysis

All the CDS gene sequences were manually extracted from the chloroplast genome. The codon usage frequency in each of the three species of *Edgeworthia* was analyzed for all the PCGs using MEGA5 (Kumar et al., 2008). The relative synonymous codon usage (RSCU) was conducted to determine if the plastid genes were under selection.

Polymerase Chain Reaction and Sanger Sequencing

Polymerase chain reaction (PCR) amplification was conducted on a final reaction volume, with a 20 μL volume reaction consisting of 10 μL of 2× Taq PCR StarMix with loading dye (Genstar Biosolutions, China), 1 μL of each primer, 6 μL of distilled water, and 2 μL 5 ng genomic DNA as a template. The ITS universal primer set: ITS1, 5′-TCC GTA GGT GAA CCT GCG G-3′ (forward) and ITS4, 5′-TCC TCC GCT TAT TGA TAT GC-3′ (reverse), was used to obtain the ITS region (White et al., 1990). PCR amplification was programmed with thermal settings of an initial denaturation at 94°C for 5 min; denaturation at 94°C for 60 s, annealing at 55°C for 60 s, extension at 72°C for 60 s; and a final extension at 72°C for 7 min. Upon verification via electrophoresis on a 1.0% agarose gel and documented under the UV machine, the PCR products were sent for direct Sanger sequencing at both ends using an ABI 3730 DNA Analyzer (Applied Biosystems, United States). Results acquired from the Sanger sequencing were aligned and manually edited to obtain the clean sequences of the three species of *Edgeworthia*. The ITS sequences for *E. albiflora* (MW255615), *E. chrysantha* (MW255616), and *E. gardneri* (MW255617) were deposited in the NCBI GenBank database for future reference.

Phylogenetic Analyses

Complete cp genome sequences of 15 taxa from the family Thymelaeaceae were included for phylogenetic analyses using maximum likelihood (ML) methods and Bayesian inference (BI). Multiple sequence alignment was carried out using MAFFT v7.409 (Katoh and Standley, 2013). Based on the Akaike information criterion calculated from the Modeltest 3.7 (Posada and Crandall, 1998), the generalized-time-reversible (GTR) model with gamma (+G) and invariant sites included (+I) (=GTR + G + I) was the best-fitting substitution model for both the ML and BI analyses. The ML tree was constructed using RAxML 8.2.11, under 1,000 bootstrap replicates (Stamatakis, 2014). BI analysis was conducted using MrBayes 3.2.5 (Drummond and Rambaut, 2007), in which the Markov Chain Monte Carlo analysis was performed under 1,000,000 generations and four Markov chains. Samplings were conducted at every 1,000 generations. The first 25% of the trees was discarded as burn-in; the remaining trees were estimated using the 50% majority-rule consensus tree and Bayesian posterior probabilities. Two closely related species, *Hibiscus hamabo* (Malvaceae; KR259988) and *Eugenia uniflora* (Myrtaceae; KR867678) were included as outgroups.

A total of 23 ITS sequences from the members of Thymelaeaceae, representing 21 taxa from eight genera in the Daphne group of tribe Daphneae, two taxa from tribe Aquilarieae and one taxon from the subfamily Octolepidoideae, were retrieved from the NCBI GenBank database. The latter three taxa were then selected as outgroups. Along with the ITS sequences of the three *Edgeworthia* species, the sequences were MUSCLE-aligned using MEGA 5 (Kumar et al., 2008) and trimmed using trimAL v1.2 (Capella-Gutiérrez et al., 2009) with the gappyout method in order to reduce the systematic errors produced by poor alignment. Phylogenetic analyses were carried out using both the ML and BI method. For ML analysis, the optimal DNA substitution model for

TABLE 2 | Gene contents in three *Edgeworthia* species chloroplast genome.

Classification	Genes
Self-replication	
Large ribosomal subunits	*rpl2*(×2), rpl14, rpl16*, rpl20, rpl22, rpl23(×2), rpl32, rpl33, rpl36*
Small ribosomal subunits	*rps2, rps3, rps4, rps7, rps8, rps11, rps12**[a](×2), rps14, rps15(×2), rps16*, rps18, rps19*
DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
Ribosomal RNAs	*rrn4.5(×2), rrn5(×2), rrn16(×2), rrn23(×2)*
Transfer RNAs	*trnA-UGC*(×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU(×2), trnI-GAU*(×2), trnK-UUU*, trnL-CAA(×2), trnL-UAA*, trnL-UAG(×2), trnM-CAU, trnN-GUU(×2), trnP-UGG, trnQ-UUG, trnR-ACG(×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(×2), trnV-UAC*, trnW-CCA, trnY-GUA*
Photosynthesis-related genes	
Photosystem I	*psaA, psaB, psaC(×2), psaI, psaJ*
Photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z
NAD(P)H dehydrogenase complex	*ndh*A*(×2), *ndh*B*(×2), *ndh*C, *ndh*D(×2), *ndh*E(×2), *ndh*F, *ndh*G(×2), *ndh*H(×2), *ndh*I(×2), *ndh*J, *ndhK*
F-type ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
Cytochrome b6/f complex	*pet*A, *pet*B*, *pet*D*, *pet*G, *pet*L, *pet*N
Rubisco	*rbc*L
Other genes	
Envelope membrane protein	*cem*A
Maturase	*mat*K
Cytochrome c biogenesis protein	*ccs*A(×2)
Subunit of acetyl-CoA-carboxylase	*acc*D
Unknown	
Conserved open reading frames	*ycf*1(×2), *ycf*2(×2), *ycf*3**, *ycf*4

**Gene containing one intron, **gene containing two introns,*
[a]trans-splinting gene, (×2) shows gene shave two copies.

the ML analysis calculated using the "Find Best DNA/Protein Model (ML)" function embedded in MEGA 5 (Kumar et al., 2008) was the Kimura two-parameter (K2P) with discrete Gamma model (+G) and invariant included (+I) (=K2P + G + I). Calculation was conducted with 1,000 bootstrap replicates on each branch node and all gaps and missing data were included in the analysis. For BI analysis, calculation was performed using MrBayes v3.2.5 (Drummond and Rambaut, 2007) following the same parameters and settings as mentioned above.

TABLE 3 | Locations and sizes of the intron-containing genes recorded in the chloroplast genome of the three *Edgeworthia* species.

Species	Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
Edgeworthia albiflora	*atp*F	LSC	144	864	411	na	na
	ndhA	IR	558	1165	540	na	na
	*ndh*B	IR	777	686	756	na	na
	*pet*B	LSC	6	770	698	na	na
	*pet*D	LSC	8	786	475	na	na
	*rpl*2	IR	393	724	391	na	na
	*rpl*16	LSC	9	1080	399	na	na
	*trn*A-UGC	IR	38	812	35	na	na
	*trn*G-UCC	LSC	23	720	49	na	na
	*trn*I-GAU	IR	37	955	35	na	na
	*trn*K-UUU	LSC	37	2559	35	na	na
	*trn*L-UAA	LSC	34	546	50	na	na
	*trn*V-UAC	LSC	38	615	35	na	na
	*ycf*3	LSC	126	732	228	761	153
Edgeworthia chrysantha	*atp*F	LSC	144	867	411	na	na
	*ndh*A	IR	558	1163	540	na	na
	*ndh*B	IR	777	690	756	na	na
	*pet*B	LSC	6	759	698	na	na
	*pet*D	LSC	8	790	475	na	na
	*rpl*2	IR	391	683	434	na	na
	*rpl*16	LSC	9	1066	399	na	na
	*rpo*C1	LSC	432	767	1617	na	na
	*rps*16	LSC	40	934	206	na	na
	*trn*A-UGC	IR	38	811	35	na	na
	*trn*G-UCC	LSC	23	718	49	na	na
	*trn*I-GAU	IR	37	955	35	na	na
	*trn*K-UUU	LSC	37	2555	35	na	na
	*trn*L-UAA	LSC	34	551	50	na	na
	*trn*V-UAC	LSC	38	621	35	na	na
	*ycf*3	LSC	126	732	228	761	153
Edgeworthia gardneri	*atp*F	LSC	144	856	411	na	na
	*ndh*A	IR	558	1162	540	na	na
	*ndh*B	IR	777	690	756	na	na
	*pet*B	LSC	6	766	698	na	na
	*pet*D	LSC	8	781	475	na	na
	*rpl*2	IR	393	681	434	na	na
	*rpl*16	LSC	9	1076	399	na	na
	*rpo*C1	LSC	432	757	1617	na	na
	*rps*16	LSC	40	946	200	na	na
	*trn*A-UGC	IR	38	812	35	na	na
	*trn*G-UCC	LSC	23	725	49	na	na
	*trn*I-GAU	IR	37	955	35	na	na
	*trn*K-UUU	LSC	37	2531	35	na	na
	*trn*L-UAA	LSC	34	547	50	na	na
	*trn*V-UAC	LSC	38	620	35	na	na
	*ycf*3	LSC	126	727	228	762	153

na, not available.

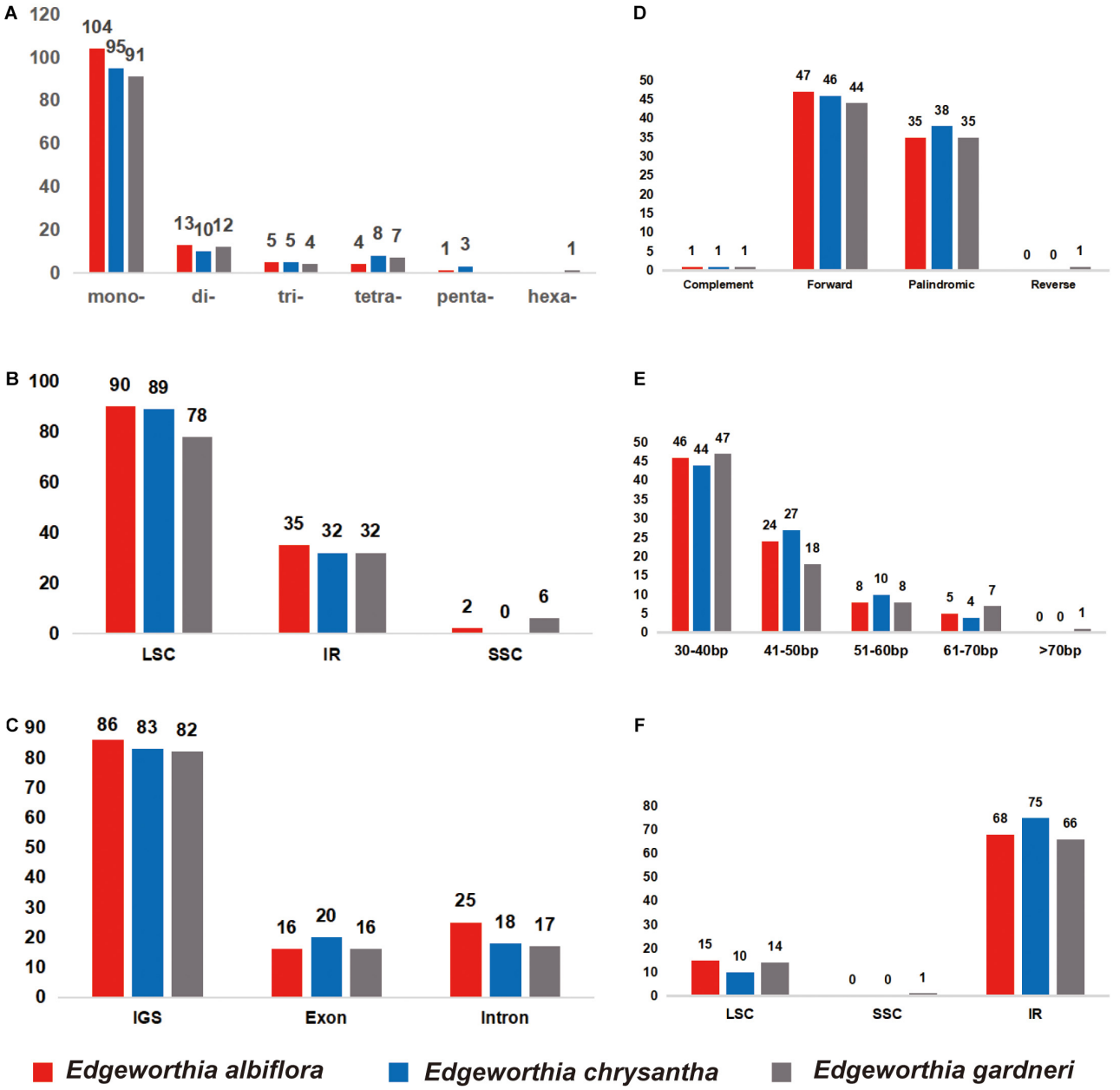


FIGURE 2 | Distribution of small sequence repeats (SSRs) in the chloroplast genomes of three species of *Edgeworthia*. **(A)** Number of different SSR types detected in chloroplast genomes of three species of *Edgeworthia*; **(B)** Frequencies of identified SSRs in large single-copy (LSC), small single-copy (SSC), and inverted repeats (IRs) regions; **(C)** Frequencies of identified SSRs in the gene coding and intergenic region. **(D)** Frequencies of four different large repeat types (complement, forward, palindromic, and reverse) in the chloroplast genomes of three species of *Edgeworthia*; **(E)** Frequencies of large repeats based on different sequence length (bp) groups; **(F)** Frequencies of large repeats in the LSC, SSC, and IRs regions.

RESULTS

Chloroplast Genome Features

A total of 23,660,708 raw reads were obtained and the raw reads were directly fed into the assembly pipeline to obtained the maximum amount of useful data. Prior to genome assembly, a total of 374,342 aligned reads were acquired, and 222,318 assembled reads of an average coverage depth of 217 times per site were incorporated in the genome assembly. Three contigs representing three species of *Edgeworthia* were obtained at the end of the assembly process.

The cp genomes of the species of *Edgeworthia* were typical quadripartite structures that ranged in size from 172,708 bp (*E. chrysantha*) to 173,621 bp (*E. albiflora*) (**Figure 1**). All genomes contained a pair of IRs (41,952–42,039 bp), separated by a large single-copy (LSC) region (85,824–86,862 bp) and a small

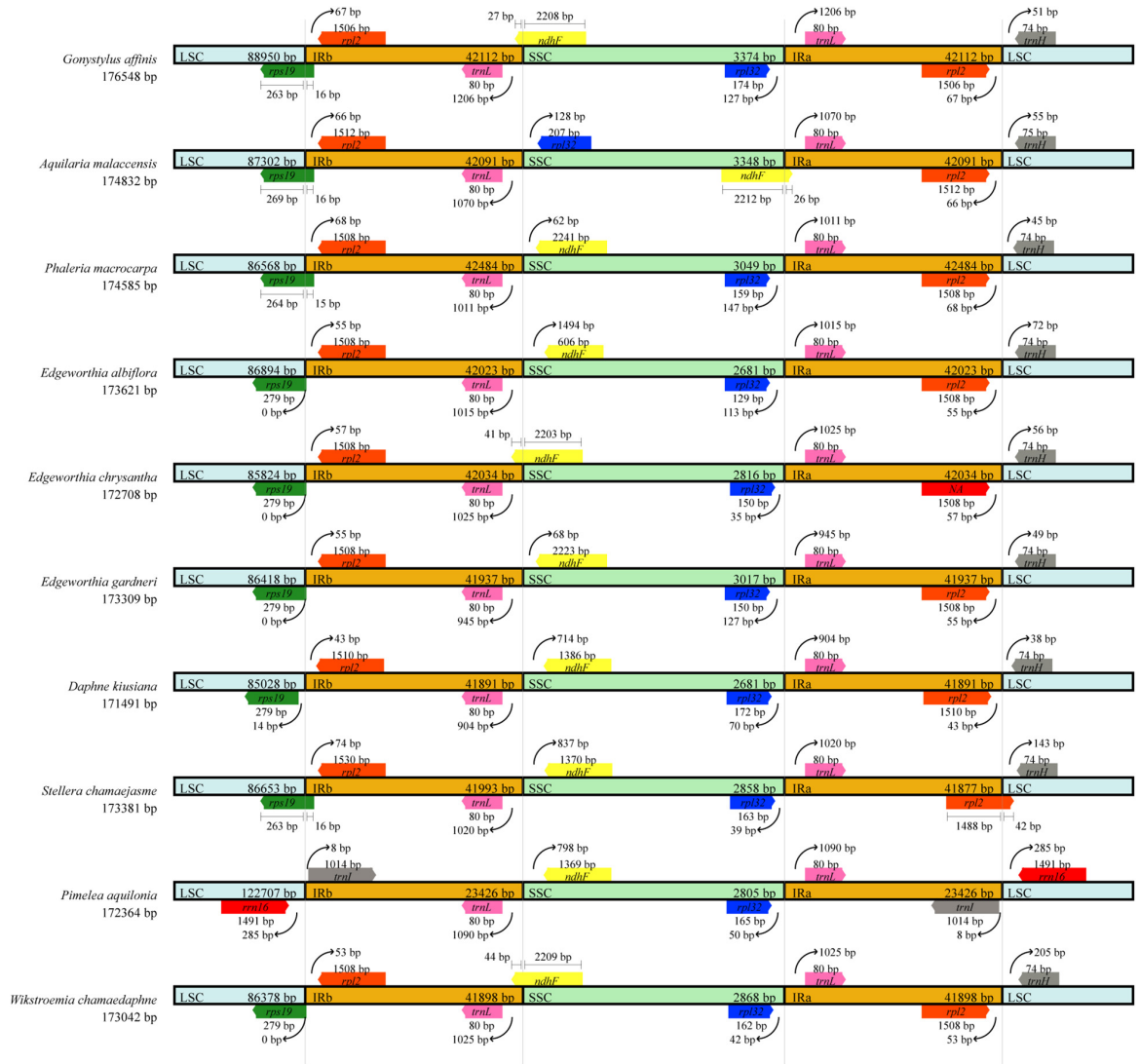


FIGURE 3 | Comparison of borders between large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions across ten chloroplast genomes of Thymelaeaceae.

single-copy (SSC) region (2,681–3,017 bp) (**Table 1**). A total of 139 genes were predicted for the cp genomes of *E. albiflora* and *E. chrysantha*, including 93 CDS genes, 38 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. However, only 138 genes were recorded in the cp genome of *E. gardneri*, in which only 92 CDS genes, 38 tRNA genes, and eight rRNA genes were predicted. In the LSC region, *E. albiflora* and *E. chrysantha* each had 61 CDS genes, but only 60 CDS genes were reported in *E. gardneri*. The CDS gene, *cem*A, was presumably a pseudogene in the cp genome of *E. gardneri*; while only two CDS genes were recorded in all three species of *Edgeworthia*. All three cp genomes of *Edgeworthia* had 27 genes replicated in both IR regions, including 15 CDS genes (*ccs*A, *ndh*A, *ndh*B, *ndh*D, *ndh*E, *ndh*G, *ndh*H, *ndh*I, *rpl*2, *rpl*23, *rps*12, *rps*15, *psa*C, *ycf*1, and *ycf*2), eight tRNA genes (*trn*A-UGC, *trn*I-CAU, *trn*I-GAU, *trn*L-CAA, *trn*L-UAG, *trn*N-GUU, *trn*R-ACG, and *trn*V-GAC), and four rRNA genes (*rrn*4.5, *rrn*5, *rrn*16, and *rrn*23) (**Table 2**). Of the 19 intron-containing genes, 17 of them (*atp*F, *ndh*A, *ndh*B, *pet*B, *pet*D, *rpl*2, *rpl*16, *rpo*C1, *rps*16, *trn*A-UGC, *trn*G-UCC, *trn*I-GAU, *trn*K-UUU, *trn*L-UAA, and *trn*V-UAC) contained one intron, whereas two of them (*rps*12 and *ycf*3) contained two introns (**Table 3**). The GC content of *E. chrysantha* and *E. albiflora* was 36.5%, while the GC content of *E. chrysantha* was 36.7%.

Sequence Repeats

Simple sequence repeat analysis detected 127, 121, and 115 SSRs in *E. albiflora*, *E. chrysantha*, and *E. gardneri*, respectively (**Figure 2A**). Most of the SSRs were in the LSC regions when compared to the SSC and IR regions (**Figure 2B**). The SSRs were more abundant in the intergenic spacer region when compared to both the intronic and exon regions; more than 80 SSRs were detected in the intergenic spacer regions of the three species of

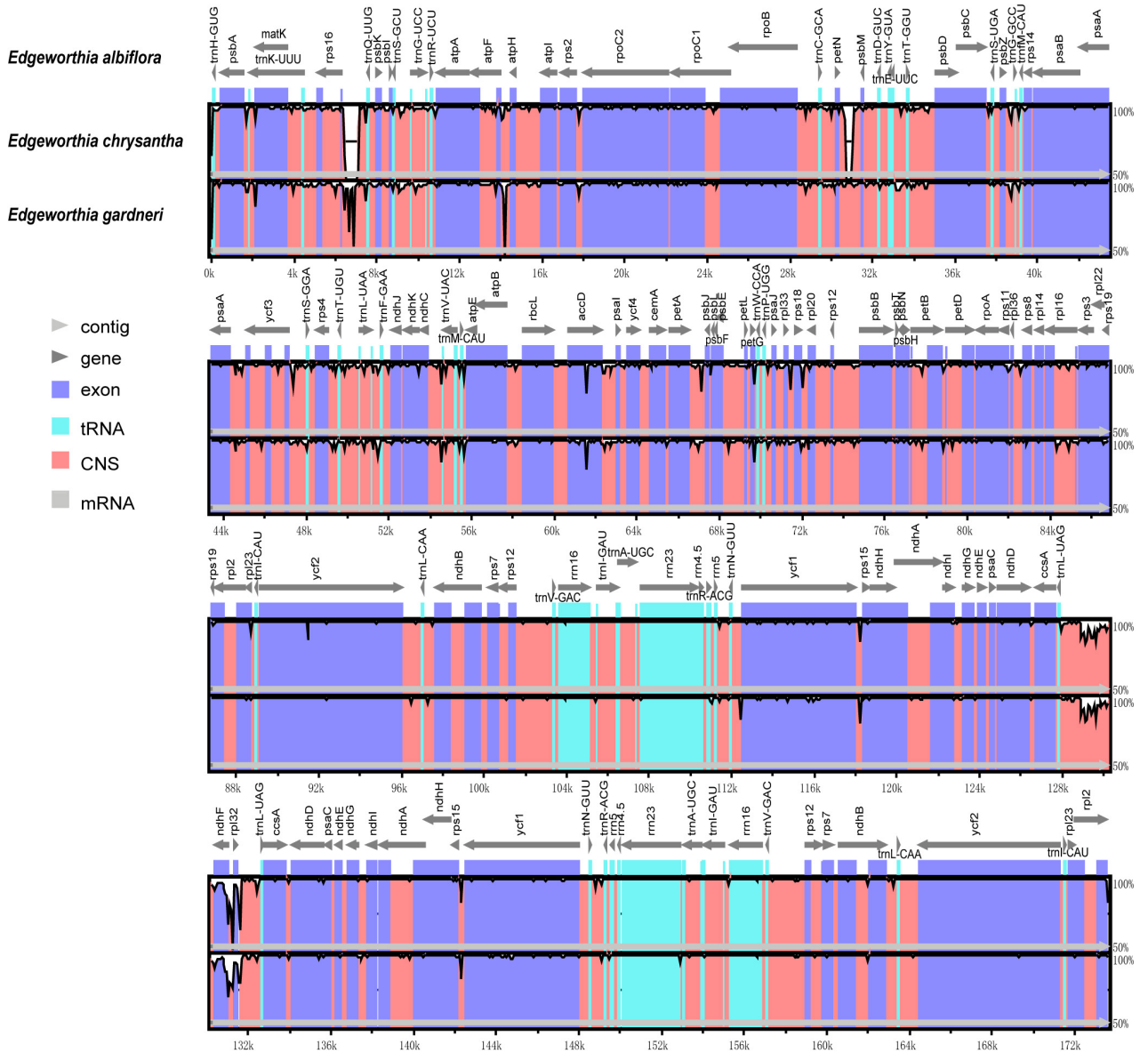


FIGURE 4 | Complete chloroplast genome comparison of three species of *Edgeworthia* using mVISTA alignment program. Genome regions are color-coded and described in the figure legend.

Edgeworthia, and the number of SSRs in the intronic and exon regions were only recorded between 16 and 25 (**Figure 2C**). Cp genomes of all three *Edgeworthia* species contained mono-, di-, tri-, and tetranucleotide SSRs; while *E. albiflora* and *E. chrysantha* had one and three pentanucleotide SSRs, respectively, none were recorded in *E. gardneri*. Yet, *E. gardneri* was recorded with one hexanucleotide SSR, which was not present in the other two species. Considering sequence complementary, eight classified repeat types were found present in all three species of *Edgeworthia* (data not shown). The repeat type C/G were only detected in *E. albiflora* and *E. chrysantha*; while the repeat type AATT/AATT were only detected in *E. chrysantha* and *E. gardneri*. The repeat types AAG/GTT and ACAT/ATGT were exclusive to *E. chrysantha*, and *E. gardneri* was the only species recorded with the repeat types AATTC/AATTG and ACCCC/GGGGT.

For large repeats, forward repeats were recorded most abundant in the cp genome of the three *Edgeworthia* species, ranging from 44 to 47, followed by the palindromic repeats that ranged from 35 to 38 (**Figure 2D**). However, there was no records for reverse repeats in *E. albiflora* and *E. chrysantha*, but one in *E. gardneri*. The large repeats were recorded mostly in sequence length of 30–40 bp (**Figure 2E**). *Edgeworthia gardneri* was recorded with one large repeat with sequence length over 70 bp, but only large repeats with length from 60 to 70 bp were recorded longest for *E. albiflora* and *E. chrysantha*. The large repeats were recorded mostly distributed in the IR region,

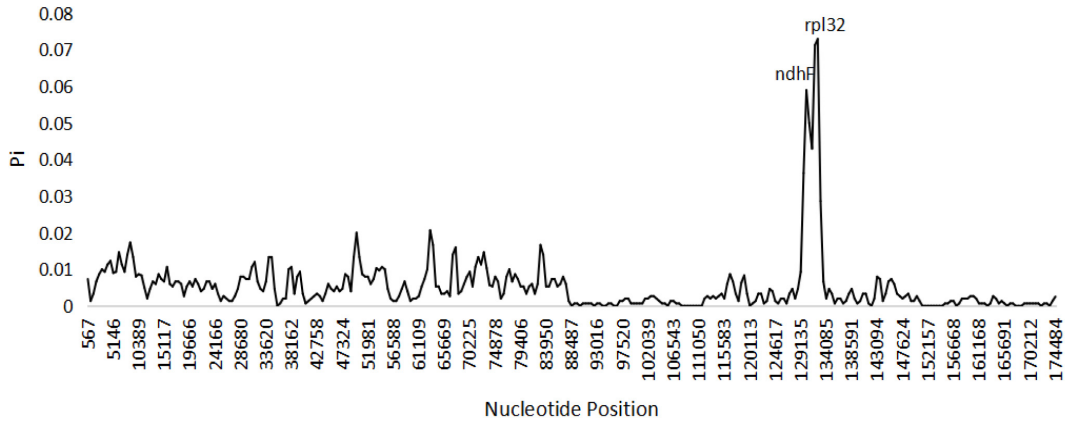


FIGURE 5 | Sliding window analysis of the complete chloroplast genome sequences among three species of *Edgeworthia* (Window length: 1000 bp; step size: 500 bp).

followed by the LSC region; at least one large repeat was recorded in the SSC region in *E. gardneri* (**Figure 2F**).

Contraction and Expansion of the IR Regions

Chloroplast genome structure and the junction positions between IR regions were well-conserved among the three species of *Edgeworthia*, but structural variation was present in the IRs/SC borders (**Figure 3**). The *ndh*F gene extended to the IRB region in the cp genome of *E. chrysantha*, but not for *E. albiflora* and *E. gardneri*. The *ndh*F gene in the latter two species was located in the SSC region. The *rps*19 gene that was located in the LSC region in *E. chrysantha* extended into the IRB region in *E. albiflora* and *E. gardneri*. When compared to other seven closely related genera in the family Thymelaeaceae, the placement of genes adjacent to the IR junctions were identical to those in the cp genome of *Gonystylus affinis*, *D. kiusiana*, *Phaleria macrocarpa*, *Stellera chamaejasme*, and *Wikstroemia chamaedaphne*.

Comparative Genomic Analysis

Based on the genome sequence alignment of the three species of *Edgeworthia*, distinct sequence variation was detected in three gene regions; *pet*N-*psb*M, *trn*L-UAG-*rpl*32, and *rps*16-*trn*Q-UUG (**Figure 4**). With a nucleotide diversity (Pi) cut-off point set at Pi $\geq$ 0.04, the sliding window analysis detected three highly variable regions in the genome sequence alignment of the three species of *Edgeworthia* (**Figure 5**). The highly variable regions were all located within the protein-coding genes, *ndh*F and mainly manifested in the SSC region, between *ndh*F and *rpl*32 of the cp genome.

Codon Use Preference Analysis

A total of 29,529–30,093 codons of the CDS genes were recorded in the three cp genomes of *Edgeworthia*. The RSCU value for each species exhibited similar codon preference in the 64 codons in the CDS genes (**Supplementary Table 1**). As a result, 30 of them exhibited greater preference (RSCU $>$ 1); 32 of them were least preferred (RSCU $<$ 1); two of them displayed no preferences (RSCU = 1). The isoleucine (Ile)-encoded codon AUU exhibited the greatest occurrence (n = 1,269); while the Ile-encoded codon UGA exhibited the least occurrence (n = 20). Among the preferred codons, 27 of them were A/U-ended. Among the three stop codons, UAA was recorded to be more abundant than UAA and UGA, thus displaying higher preferences. There were no rare codons (RSCU $<$ 0.1) found in the CDS genes of the three cp genomes of *Edgeworthia*.

Phylogenetic Analysis

Phylogenetic analyses using the complete cp genome sequence for both ML and BI methods revealed similar topological structure between the two phylogenetic trees (**Figure 6A**). Strong bootstrap support and high posterior probabilities were recorded at all branch nodes. All taxa included in this study displayed monophyletic relationships. In the Daphneae, *Edgeworthia* diverged before *Daphne*, *Stellera*, and *Wikstroemia*. The three species of *Edgeworthia* formed a monophyletic clade, with *E. albiflora* diverging before *E. chrysantha* and *E. gardneri*.

For the ITS-based ML and BI analyses, both phylogenetic trees exhibited identical tree structure and species placement (**Figure 6B**). The Daphne group displayed a paraphyletic relationship, in which the genera *Daphnopsis*, *Dirca*, and *Ovidia* formed a cluster; while *Edgeworthia* was placed at the base of the latter cluster, with strong bootstrap support and Bayesian posterior probability (ML $\geq$ 75, BI $\geq$ 0.95) that also consisted of *Daphne*, *Diarthron*, *Stellera*, *Thymelaea*, and *Wikstroemia*. In the *Edgeworthia* clade, the three *Edgeworthia* species formed a monophyletic clade, with *E. albiflora* diverging before *E. chrysantha* and *E. gardneri* under strong bootstrap support and Bayesian posterior probability.

DISCUSSION

Similar to other members of the Thymelaeaceae, the cp genome of *Edgeworthia* was rather well-conserved across the three species analyzed in this study. By comparing published cp genome reports on members of the Thymelaeaceae, it was determined that the complete cp genome sequence of *Edgeworthia* is

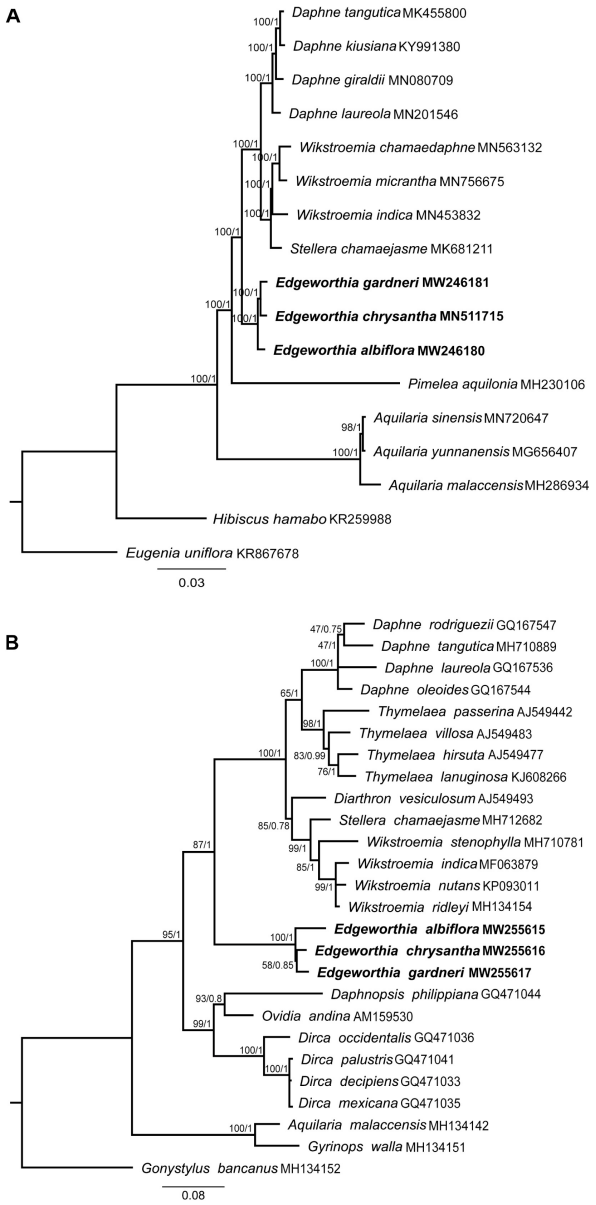


FIGURE 6 | Maximum-likelihood (ML) and MrBayes (BI) tree analyses were based on 1,000 bootstrap replicates. Bootstrap support values and Bayesian posterior probabilities are indicated at each branch nodes. Three species of Edgeworthia used in this study are highlighted in bold. **(A)** Combined phylogenetic tree of *Edgeworthia* and allied genera based on the chloroplast genome sequences of 15 taxa from Thymelaeaceae. Two related species, *Hibiscus hamabo* (Malvaceae; KR259988) and *Eugenia uniflora* (Myrtaceae; KR867678) were included as outgroups. **(B)** Phylogenetic analyses of Thymelaeaceae based on nuclear ribosomal DNA internal transcribed spacer (ITS) gene sequences. A total of 23 ITS sequences from the members of Thymelaeaceae, representing 21 taxa from eight genera in the Daphne group of tribe Daphneae, two taxa from tribe *Aquilarieae* and one taxon from subfamily Octolepidoideae.

shorter than those of *G. affinis* of subfamily Octolepidoideae (176,548 bp) and nine species of *Aquilaria* (*Aquilaria beccariana*, *Aquilaria crassna*, *Aquilaria hirta*, *Aquilaria malaccensis*, *Aquilaria microcarpa*, *Aquilaria rostrata*, *A. sinensis*, *Aquilaria subintegra*, and *Aquilaria yunnanensis*; 174,693–174,907 bp) from Aquilarieae of subfamily Thymelaeoideae (Hishamuddin et al., 2020), but was longer than the four species of *Daphne* (*Daphne giraldii*, *D. kiusiana*, *Daphne laureola*, and *Daphne tangutica*; 169,944–171,643 bp) (Cho et al., 2018; Könyves et al., 2019; Yan et al., 2019a,b), two species of *Wikstroemia* (*W. chamaedaphne* and *Wikstroemia indica*; 151,731–173,042 bp) from the Daphne group (Qian and Zhang, 2020; Qian et al., 2020), and *Pimelea aquilonia* (172,364 bp) from the Gnidia group (Foster et al., 2018), all belonging to the Daphneae of subfamily Thymelaeoideae. *Stellera chamaejasme*, from the monospecific *Stellera*, was the only species from the Daphne group with a total cp genome size within the cp genome size range of *Edgeworthia*, which was 173,381 bp (Yun et al., 2019). Eventually, rather small size of the SSC region was observed in the cp genome of most species from Thymelaeaceae. In general, the expansion, shrinkage and loss of the IR regions are some of the known reasons to variations in cp genome sizes of seed plants (Jansen and Ruhlman, 2012). Since information on complete cp genomes in members of the Thymelaeaceae is limited, we could not infer any significance between the cp genome size and the systematic position in the Thymelaeaceae.

In this study, the three species of *Edgeworthia* shared almost the same number of genes in their complete cp genomes; *E. gardneri* was recorded as having one gene fewer when compared to the other two species – the protein-coding gene, *cem*A, was not annotated. The *cem*A gene of *E. gardneri* was thought to be a pseudogene as the loci of the premature stop codons differ from the others and could be causing the gene to be non-functional; while the location of the stop codons in *cem*A could be useful for distinguishing *E. gardneri* from *E. albiflora* and *E. chrysantha*. The *cem*A gene encodes the chloroplast membrane protein, which may play an important role in plastid maintenance and intracellular communication (Sasaki et al., 1993; Sonoda et al., 1997). The *cem*A homolog is also recorded to be essential for carbon dioxide (CO_2) transfer in cyanobacteria (Katoh et al., 1996). The stomatal density of *E. gardneri* is lower than that in *E. chrysantha* (Zhang et al., 2015) and the deletion of *cem*A gene may be related to its adaptation in its habitat, which has an average altitude of above 1,000 m a.s.l.

Based on literature reports, members of the Daphne group can be divided into two natural groups due to the effects on geographical separation and by minor morphological differences (Herber, 2003). The seven American genera (*Daphnopsis*, *Dirca*, *Funifera*, *Goodallia*, *Lagetta*, *Ovidia*, and *Schoenobiblus*) have petals and/or long filaments, whereas the seven Eurasian genera (*Daphne*, *Diarthron*, *Edgeworthia*, *Rhamnoneuron*, *Stellera*, *Thymelaea*, and *Wikstroemia*) lack petals, but are reported to have sessile or subsessile anthers (Herber, 2003). Eventually, the most comprehensive phylogenetic study conducted on Thymelaeaceae using the combined *rbc*L, *trn*L-*trn*F datasets and ITS sequences revealed that members of the Daphne group were actually divisible into two different clades in which *Daphne*, *Diarthron*, *Edgeworthia*, *Stellera*, and *Wikstroemia* formed one clade; while *Ovidia* and *Dirca* were in another clade with *Peddiea* (Phaleria

group) and *Stephanodaphne* (Linostoma group) (Beaumont et al., 2009). Similarly, a non-monophyletic relationship was observed in the Daphne group in this study based on the ITS sequences. In our study, the three American genera, *Dirca*, *Ovidia*, and *Daphnopsis* formed a clade independent from the Eurasian genera (**Figure 6B**). Both the phylogenetic analyses based on cp genome sequences and the nuclear ribosomal DNA ITS sequences, or a combination of both the cp and nuclear gene sequences, placed *Edgeworthia* at the base of the Daphne group among the Eurasian genera.

Taxonomic controversy is also present in *Edgeworthia*. Based on the Flora of China, there are at least five species, whereas four species (*E. albiflora*, *E. chrysantha*, *E. eriosolenoides*, and *E. gardneri*) are in the Asia region (Wang and Gilbert, 2017). The recent revision by The Plant List committee recognizes only four species; *E. eriosolenoides*, *E. gardneri*, *E. longipes*, and *Edgeworthia tomentosa* (The Plant List, 2013). The former synonymized *E. chrysantha* with *E. tomentosa*, as the name *E. tomentosa* was considered invalid in a revision conducted on the collections of Thunberg on *Magnolia tomentosa* (Hamaya, 1955), and also *E. albiflora* with *E. gardneri* (Wang and Gilbert, 2017). *Edgeworthia albiflora*, a species treated as a distinct since it was first discovered in 1924 (Nakai, 1924; Duncan and Mellinger, 1972), was later regarded as a synonym of *E. gardneri* (Clennett et al., 2002). The synonymy was not accepted in the Flora of China, where *E. albiflora* was treated as distinct (Wang and Gilbert, 2017). It is noteworthy that we also failed to recover information to synonymize *E. albiflora* under *E. gardneri*. Based on our field observations, *E. albiflora* and *E. gardneri* can be differentiated through their morphological features and do not pose a challenge in species recognition. The interior of the calyx of *E. gardneri* is yellow, the ovary uniformly hairy, the stigma rounded; the interior of the calyx of *E. albiflora* is white, the base of the ovary is glabrous and the apex hairy and the stigma is clavate (Wang and Gilbert, 2017). Meanwhile, the leaf epidermis is entirely different in *E. albiflora* and *E. gardneri*, with paracytic stomatal types and cyclocytic stomatal types, respectively (Zhang et al., 2015). From a molecular perspective, it is generally accepted that species are delimited when the interspecific variation is greater than intraspecific variation (Lim et al., 2012). Thus, to further strengthen the case for recognizing *E. albiflora*, we compared the genetic information and found that the alignment between *E. albiflora*, *E. chrysantha*, and *E. gardneri* is consisted of greater interspecific variation (cp genome: pairwise distance = 0.0045–0.0061, containing 1038 singletons; ITS: pairwise distance = 0.0226–0.0365, containing 29 singletons) than its intraspecific variation that was based on the alignment between our collection of *E. chrysantha* and another published genome of *E. chrysantha* (cp genome: MT135125; ITS: AJ744932) (cp genome: pairwise distance = 0.0004, containing 70 singletons; ITS: pairwise distance = 0.0030, containing two singletons) (data not shown). Furthermore, the molecular placement with strong bootstrap support based on the ML and BI trees using both the complete cp genome sequences and ITS sequence, analyzed separately, revealed that *E. albiflora* and *E. gardneri* should be treated separately (**Figure 6B**). Unless there is a stronger case to synonymize the two species via morphological features, judging from the molecular evidence and personal field observation, they should be regarded as two natural groups. On the other hand, two species, *E. eriosolenoides* and *E. longipes* were not included in this study; no specimens had been collected since they were first described.

CONCLUSION

The entire cp genomes of *E. albiflora*, *E. chrysantha*, and *E. gardneri* were sequenced and analyzed in this study. We obtained such comprehensive molecular information as SSRs, IR contraction and expansion, codon usage and phylogenomic placement through explicit bioinformatic analyses of the cp genome. Furthermore, the addition of the ITS sequences for the understudied species of *Edgeworthia* provided insight for the first time on the phylogenetic relationships of the three species of *Edgeworthia* at the nuclear gene level. The data obtained from this study will likely provide a powerful genetic resource for future studies on population genetics, biological functions, molecular phylogeny, as well as evolution of *Edgeworthia*.

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/genbank/, MW246180; https://www.ncbi.nlm.nih.gov/genbank/, MN511715; https://www.ncbi.nlm.nih.gov/genbank/, MW246181; https://www.ncbi.nlm.nih.gov/genbank/, MW255615; https://www.ncbi.nlm.nih.gov/genbank/, MW255616; and https://www.ncbi.nlm.nih.gov/genbank/, MW255617.

AUTHOR CONTRIBUTIONS

SQ performed the experiments, analyzed the data, and wrote the manuscript. YZ and SL conceived the research and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31760048) and the Fundamental Research Funds for the Central Universities (33000-31611215).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.643552/full#supplementary-material

REFERENCES

Amiryousefi, A., Hyvönen, J., and Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220

Beaumont, A. J., Edwards, T. J., Manning, J., Maurin, O., Rautenbach, M., Motsi, M. C., et al. (2009). Gnidia (Thymelaeaceae) is not monophyletic: taxonomic implications for Thymelaeoideae and a partial new generic taxonomy for *Gnidia*. *Bot. J. Linn. Soc.* 160, 402–417. doi: 10.1111/j.1095-8339.2009.00988.x

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Che, C. T., Cheung, H. Y., Law, R., Luo, G. A., Man, R. Y., Tsim, K. W., et al. (2010). *Buddlejae Flos.," in Hong Kong Chinese materia medica standards*. Hong Kong, China: The Department of Health, The Government of the Hong Kong Special Administrative Region, 179–187.

Cho, W. B., Han, E. K., Choi, G., and Lee, J. H. (2018). The complete chloroplast genome of *Daphne kiusiana*, an evergreen broad-leaved shrub on Jeju Island. *Conserv. Genet. Resour.* 10, 103–106. doi: 10.1007/s12686-017-0774-5

Clennett, C., Harvey, Y., and Pearman, G. (2002). Plate 433. *Edgeworthia chrysantha*. *Curtis'S Bot. Magazine* 19, 19–27. doi: 10.1111/1467-8748.00325

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

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15. doi: 10.1016/0031-9422(80)85004-7

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Duncan, W. H., and Mellinger, M. (1972). *Edgeworthia* (Thymelaeaceae) new to the western hemisphere. *Rhodora.* 74, 436–479. doi: 10.2307/23310953

Foster, C. S. P., Cantrill, D. J., James, E. A., Syme, A. E., Jordan, R., Douglas, R., et al. (2016). Molecular phylogenetics provides new insights into the systematics of *Pimelea* and *Thecanthes* (Thymelaeaceae). *Aust. Syst. Bot.* 29, 185–196. doi: 10.1071/SB16013

Foster, C. S. P., Henwood, M. J., and Ho, S. Y. W. (2018). Plastome sequences and exploration of tree-space help to resolve the phylogeny of riceflowers (Thymelaeaceae: *Pimelea*). *Mol. Phylogenet. Evol.* 127, 156–167. doi: 10.1016/j.ympev.2018.05.018

Greiner, S., Lehwark, P., and Bock, R. (2019). Organellar genomedraw (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Hamaya, T. (1955). A dendrological monograph on the thymelaeaceae plants of japan. *Univ. Tokyo Faculty Agric. Exp. Forest Rep.* 50, 45–96.

Herber, B. E. (2003). "Thymelaeaceae," in *The Families and Genera of Vascular Plants*, eds K. Kubitzki and C. Bayer (Berlin, Heidelberg: Springer), 373–396.

Hishamuddin, M. S., Lee, S. Y., Ng, W. L., Ramlee, S. I., Lamasudin, D. U., and Mohamed, R. (2020). Comparison of eight complete chloroplast genomes of the endangered *Aquilaria* tree species (Thymelaeaceae) and their phylogenetic relationships. *Sci. Rep.* 10:13034. doi: 10.1038/s41598-020-70030-0

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., James, L. M., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104

Jansen, R. K., and Ruhlman, T. A. (2012). "Plastid genomes of seed plants," in *Genomics of Chloroplasts and Mitochondria*, eds R. Bock and V. Knoop (New York: Springer), 103–126.

Katoh, A., Lee, K. S., Fukuzawa, H., Ohyama, K., and Ogawa, T. (1996). CemA homologue essential to CO2 transport in the cyanobacterium synechocystis PCC6803. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4006–4010. doi: 10.1073/pnas.93.9.4006

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Steven, S. H., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Könyves, K., Yooprasert, S., Culham, A., and David, J. (2019). The complete plastome of *Daphne laureola* L. (Thymelaeaceae). *Mitochondrial DNA B Resour.* 4, 3364–3365. doi: 10.1080/23802359.2019.1674201

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings Bioinf.* 9, 299–306. doi: 10.1093/bib/bbn017

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

Lan, C. N., Fan, W. G., and Bao, X. M. (2013). Cultivation techniques for *Edgeworthia chrysantha* Lindl.——a wood plant with high economic value. *J. Forestry Engin.* 4, 119–121. doi: 10.3969/j.issn.1000-8101.2013.04.033

Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet.* 19, 47–56. doi: 10.1016/s0168-9525(02)00003-3

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Lim, G. S., Balke, M., and Meier, R. (2012). Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Syst. Biol.* 61, 165–169. doi: 10.1093/sysbio/syr030

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., et al. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* 16, 1046–1047. doi: 10.1093/bioinformatics/16.11.1046

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

Motsi, M. C., Moteetee, A., Beaumont, A. J., Rye, B. L., and Van der Bank, M. (2010). A phylogenetic study of *Pimlea* and *Thecanthes* (Thymelaeaceae): Evidence from plastid and nuclear ribosomal DNA sequence data. *Aust. Syst. Bot.* 23, 270–284. doi: 10.1071/SB09002

Nakai, T. (1924). Some new and noteworthy ligneous plants of eastern Asia. *J. Arnold Arbor., Harv. Univ.* 5, 72–83.

Nan, C. Y., Lu, Y. X., Zhu, J. X., Jiang, W., Zhong, G. Y., and Li, M. (2018). Advances in research on chemical constituents and pharmacological effects of Edgeworthia plants. *Chinese Trad. Patent Med.* 1, 166–171. doi: 10.3969/j.issn.1001-1528.2018.01.034

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19, 325–354. doi: 10.1146/annurev.ge.19.120185.001545

Patel, R. K., and Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PloS One.* 7:e30619. doi: 10.1371/journal.pone.0030619

Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Qian, S. J., and Zhang, Y. H. (2020). Characterization of the complete chloroplast genome of a medicinal plant, *Wikstroemia indica* (Thymelaeaceae). *Mitochondrial DNA B Resour.* 5, 83–84. doi: 10.1080/23802359.2019.1696249

Qian, S. J., Zhang, Y. H., and Li, G. D. (2020). The complete chloroplast genome of a medicinal plant, *Wikstroemia chamaedaphne* (Thymelaeaceae). *Mitochon. DNA B Resour.* 5, 648–649. doi: 10.1080/23802359.2019.1711228

Sasaki, Y., Sekiguchi, K., Nagano, Y., and Matsuno, R. (1993). Chloroplast envelope protein encoded by chloroplast genome. *Febs. Lett.* 316, 93–98. doi: 10.1016/0014-5793(93)81743-J

Sonoda, M., Katoh, H., Ohkawa, H., and Ogawa, T. (1997). Cloning of the *cot*A gene of Synechococcus PCC7942 and complementation of a *cot*A-less mutant of synechocystis PCC6803 with chimeric genes of the two strains. *Photosynth. Res.* 54, 99–105. doi: 10.1023/A:1005830303453

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

The Plant List. (2013). *The plant list*. Available Online at: http://www.theplantlist.org/ (accessed December 1st, 2020).

Van der Bank, M., Fay, M. F., and Chase, M. W. (2002). Molecular phylogenetics of thymelaeaceae with particular reference to african and australian genera. *Taxon* 51, 329–339. doi: 10.2307/1554930

Wang, X. L., Luo, W. C., Dong, Y. C., and Yan, G. F. (2017). Preliminary study on the ornamental characteristics and landscape application of *Edgeworthia chrysantha*. *Anhui Agric. Sci. Bull.* 9, 124–126. doi: 10.16377/j.cnki.issn1007-7731.2017.09.052

Wang, Y., and Gilbert, M. (2017). "Edgeworthia," in *Flora of China*, eds P. H. Raven, Z. Y. Wu, and D. Y. Hong (Beijing: Science Press), 247–248.

Wang, Y., Zhan, D. F., Jia, X., Mei, W. L., Dai, H. F., Chen, X. T., et al. (2016). Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) gilg and evolution analysis within the malvales order. *Front. Plant Sci.* 7:280. doi: 10.3389/fpls.2016.00280

White, T. J., Bruns, T. D., Lee, S. B., Taylor, J. W., Innis, M. A., Gelfand, D. H., et al. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. Guide Methods Appl.* 31, 315–322. doi: 10.1016/B978-0-12-372180-8.50042-1

Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Kai, F. M., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Xiao, P. (2002). *Modern Chinese Materia Medica*. Beijing: Chemical Industry Press.

Yan, F., Tao, X., Wang, Q. L., Zhang, Y. J., Zhang, C. M., and Yu, H. L. (2019a). The complete chloroplast genome sequence of the medicinal shrub *Daphne giraldii* Nitsche. (Thymelaeaceae). *Mitochondrial DNA B Res.* 4, 2685–2686. doi: 10.1080/23802359.2019.1644233

Yan, F., Wang, Q. L., Zhang, Y. J., Zhang, C. M., and Chen, Y. (2019b). The complete chloroplast genome sequence of medicinal plant, *Daphne tangutica* Maxim. (Thymelaeaceae). *Mitochon. DNA B Res.* 4, 1776–1777. doi: 10.1080/23802359.2019.1611397

Yun, N., Park, J., and Oh, S. H. (2019). The complete chloroplast genome of the traditional medicinal plant *Stellera chamaejasme* L. (Thymelaeaceae). *Mitochondrial DNA B Resour.* 4, 1796–1797. doi: 10.1080/23802359.2019.1612296

Zhang, Y. Z., Sun, W. G., Gao, C. Z., Cao, W. J., Zhang, Y. H., and Li, Z. M. (2015). Character of leaf epidermis and their systematic significance of *Daphne* and *Wikstroemia*. *Plant Diversity Res.* 37, 493–512. doi: 10.7677/ynzwyj201515019

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 March 2021
doi: 10.3389/fgene.2021.655843

SWEET Transporters and the Potential Functions of These Sequences in Tea (*Camellia sinensis*)

Lan Jiang[1,2], Cheng Song[3], Xi Zhu[4] and Jianke Yang[5]**

[1] Central Laboratory, Yijishan Hospital of Wannan Medical College, Wuhu, China, [2] Key Laboratory of Non-coding RNA Transformation Research of Anhui Higher Education Institution, Yijishan Hospital of Wannan Medical College, Wuhu, China, [3] College of Biological and Pharmaceutical Engineering, West Anhui University, Luan, China, [4] Department of Medicine III, University Hospital, LMU Munich, Munich, Germany, [5] School of Preclinical Medicine, Wannan Medical College, Wuhu, China

OPEN ACCESS

Edited by:
Yunpeng Cao, Central South University Forestry and Technology, China

Reviewed by:
Tianzhe Chen, Anhui Agricultural University, China
Yuanyuan Jiang, South China Agricultural University, China
Chaoju Qian, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Xi Zhu
Xi.Zhu@med.uni-muenchen.de
Jianke Yang
ajiankebc@wnmc.edu.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *19 January 2021*
Accepted: *15 February 2021*
Published: *31 March 2021*

Citation:
Jiang L, Song C, Zhu X and Yang J (2021) SWEET Transporters and the Potential Functions of These Sequences in Tea (Camellia sinensis). Front. Genet. 12:655843. doi: 10.3389/fgene.2021.655843

Tea (*Camellia sinensis*) is an important economic beverage crop. Its flowers and leaves could be used as healthcare tea for its medicinal value. SWEET proteins were recently identified in plants as sugar transporters, which participate in diverse physiological processes, including pathogen nutrition, seed filling, nectar secretion, and phloem loading. Although *SWEET* genes have been characterized and identified in model plants, such as *Arabidopsis thaliana* and *Oryza sativa*, there is very little knowledge of these genes in *C. sinensis*. In this study, 28 *CsSWEET*s were identified in *C. sinensis* and further phylogenetically divided into four subfamilies with *A. thaliana*. These identified *CsSWEET*s contained seven transmembrane helixes (TMHs) which were generated by an ancestral three-TMH unit with an internal duplication experience. Microsynteny analysis revealed that the large-scale duplication events were the main driving forces for members from *CsSWEET* family expansion in *C. sinensis*. The expression profiles of the 28 *CsSWEET*s revealed that some genes were highly expressed in reproductive tissues. Among them, *CsSWEET1a* might play crucial roles in the efflux of sucrose, and *CsSWEET17b* could control fructose content as a hexose transporter in *C. sinensis*. Remarkably, *CsSWEET12* and *CsSWEET17c* were specifically expressed in flowers, indicating that these two genes might be involved in sugar transport during flower development. The expression patterns of all *CsSWEET*s were differentially regulated under cold and drought treatments. This work provided a systematic understanding of the members from the *CsSWEET* gene family, which would be helpful for further functional studies of *CsSWEET*s in *C. sinensis*.

Keywords: *SWEET*, duplication, *Camellia sinensis*, expression, microsynteny

INTRODUCTION

SWEET is a novel kind of low-affinity sugar transporter, which does not depend on the pH value of the environment and transports in two directions along the concentration gradient (Yuan and Wang, 2013). The members of the *SWEET* family are widely distributed, which are ubiquitous in higher eukaryotes and also distributed in

protozoa, metazoa, fungi, bacteria, and archaea (Hamada et al., 2005; Saier et al., 2006; Xuan et al., 2013). The membrane proteins encoded by the *SWEET* have a certain number of conserved transmembrane domains, which are named MtN3/saliva (Hamada et al., 2005). This domain was first found in nodulin from the roots of *Medicago sativa* (Gamas et al., 1996). Recently, phylogenetic analysis has been performed for the main protein members of the SWEET family, which are divided into three branches: (i) the plant-like SWEET, most of which contain two MtN3/saliva transmembrane domains, (ii) the animal-like SWEET, which has two specific MtN3/saliva domains, and (iii) some SWEET proteins from bacteria to Archaea (cocci) and nematodes, which consist of an MtN3/saliva domain containing three transmembrane alpha helices (Yuan and Wang, 2013). A large number of *SWEET* family members have been found in *Arabidopsis thaliana* (Wipf et al., 2020), *Vitis vinifera* (Chong et al., 2014), *Solanum lycopersicum* (Shammai et al., 2018), *Manihot esculenta* (Cao et al., 2019), *Eucalyptus grandis* (Yin et al., 2020), etc., and are involved in many physiological processes. For instance, *RPG1* (*ATSWEET8*) plays a vital role in microspore outer wall formation in *Arabidopsis* (Guan et al., 2008); in *Oryza sativa*, the *SWEET* family members are also involved in reproductive development (Wang et al., 2010); *SAG29* (*AtSWEETl5*), located on the plasma membrane in *Arabidopsis*, can regulate cell activity in a hypersaline environment (Seo et al., 2011). The *SWEET*s also participate in the regulation of the aging process. The overexpression of *OsSWEET5* can cause growth delay and premature aging in rice seedlings, while no phenotypic changes were observed in the knockout ones (Zhou et al., 2014). Chu et al. (2006) found that the rice harboring mutant *OsSWEET11* (*Xa13*) can resist powdery mildew, suggesting that *SWEET*s not only served as a glucose transporter but also participated in the interaction between host and pathogen. However, the members of the *CsSWEET* gene family have not yet been identified, and the roles of *CsSWEET*s remain unclear in tea plant (*Camellia sinensis*).

Camellia sinensis, which originated from the tropical area of Southwest China, is an important economic beverage crop in China (Wei et al., 2018; Zhang et al., 2020b). Sugar transport and homeostasis contribute to plant growth and development (May et al., 1998; Lastdrager et al., 2014; Rahimi et al., 2019; Pan et al., 2020; Saddhe et al., 2020). Some published papers focused on the genes that participated in sugar metabolism, such as hexose kinase, invertase, and galactinol synthase, in *C. sinensis* (Yue et al., 2015; Zhou et al., 2017; Samarina et al., 2020). Due to the economic interest in *C. sinensis* as a beverage crop, its genome, proteome, and transcriptome were recently sequenced and released (Wei et al., 2018; Wu et al., 2018; Liu et al., 2020; Xia et al., 2020; Zhang et al., 2020b), which help us to analyze the *SWEET* genes in *C. sinensis* systematically. In the present study, 26 *CsSWEET*s in *C. sinensis* were identified, and their gene structures, phylogenetic, microsynteny, and expression were analyzed. Our study revealed the functions of these *CsSWEET*s and provided candidate genes for further research.

MATERIALS AND METHODS

Database Search for *C. sinensis* Genome

The *C. sinensis* genome with GFF annotation, CDS, and protein files were downloaded from the Tea Plant Information Archive[1] (TPIA) database (Xia et al., 2019). The *AtSWEET*s of *A. thaliana* and *HsSWEET1* of *Homo sapiens* were obtained from TAIR[2] and Genbank[3], respectively. The HMM profile of MtN3/saliva domain (PF03083) was obtained from the Pfam database[4] (Mistry et al., 2020), and the HMMER (version 3.1) software (Mistry et al., 2013) was used to identify CsSWEET proteins in *C. sinensis* genome (*E*-value $\leq$ 1e−3). The Pfam (version 33.1), SMART (version 9), and InterPro (version 5) were used to confirm the above-mentioned identified sequences with *E*-value $\leq$ 1e−3 (Jones et al., 2014; Letunic and Bork, 2018; Mistry et al., 2020). Finally, we manually discarded these sequences with a lack of complete or core MtN3/saliva domain for further analyses.

Phylogenetic Analyses

The MAFFT (version 7) software was used to execute the multiple alignments of all SWEET proteins with default parameters (Katoh et al., 2005). The best substitution model of these SWEET proteins was determined by ModelFinder software (Kalyaanamoorthy et al., 2017). The maximum likelihood (ML) tree was determined by IQ-tree (version 2.1.2) software (Nguyen et al., 2015) with an SH-aLRT test for 1,000 random addition replicates and a bootstrap test for 10,000 replicates as described by Cao et al. (2020a). The Figtree software was used to visualize this ML tree.

Internal Repeats and Topological Analyses

The ClustalX (version 2) software (Thompson et al., 2003) was used to create the multiple alignments of SWEETs from *C. sinensis* and *A. thaliana*. The HHrepID was used to identify the internal repeats in SWEET proteins (Söding et al., 2006). In addition, we also used the AveHAS to calculate the topological conservation (Zhai and Saier, 2001a,b) as well as create the hydropathy, amphipathicity, and similarity plots, respectively (Zhai and Saier, 2001a). The GFF annotation file was obtained from the TPIA (see text footnote 1) database (Xia et al., 2019). The TBtools (version 1) software was used to generate this gene structure map based on the GFF annotation file (Chen et al., 2020). The MEME online tool was used to predict the conserved motifs (Bailey et al., 2015).

Microsynteny Analysis

The MicroSyn (version 1) software was used to detect the microsynteny of *CsSWEET* genes in *C. sinensis* genome with a threshold *E*-value of <1e−5 (Cai et al., 2011). In this study, we determined a syntenic block, a region which include three

[1]http://tpia.teaplant.org/
[2]https://www.arabidopsis.org/
[3]https://www.ncbi.nlm.nih.gov/
[4]http://pfam.xfam.org

or more conserved homolog genes that were distributed within 15 genes upstream and downstream of *CsSWEET* genes as described by Cao et al. (2020b).

RNA-Seq Expression Analysis

In our study, transcriptome data from various tissues in the public database NCBI contained three biological replicates. The RNA-seq reads, including seed, flower, stem, root, two and a bud, one and a bud, old leaf, mature leaf, the second leaf, the first leaf, lateral bud, apical bud, early stage lateral bud, CP24 (24 h after pollination; CP, cross-pollinated), CP48, CP72, SP24 (SP, self-pollinated), SP48, and SP72, were obtained from NCBI (PRJNA291116 and PRJNA230752). The Trimmomatic (version 0.33) was used to carried out the quality-based trimming (Bolger et al., 2014). The HISAT2 was used to map the paired reads to the *C. sinensis* genome with default parameters (Kim et al., 2019). The StringTie was used to calculate the fragments per kilobase of exon model per million reads mapped values of differently expressed genes (Pertea et al., 2016).

RESULTS AND DISCUSSION

The Phylogenetic Analysis Divides *CsSWEET* Genes Into Four Subfamilies in *C. sinensis*

In the present study, we identified 26 *CsSWEET* genes in *C. sinensis* genome by using HMMER 3.0 software, similarly as in *Manihot esculenta*, *Jatropha curcas*, *Ricinus communis*, *Vernicia fordii*, *Malus* × *domestica*, *Oryza sativa*, *Zea mays*, *Glycine max*, and *Pyrus bretschneideri* (Chen et al., 2010; Yuan and Wang, 2013; Chong et al., 2014; Cao et al., 2019; Yin et al., 2020). Subsequently, we renamed the newly identified *CsSWEET*s based on the nomenclature of the *AtSWEET*s of *A. thaliana*. The detailed information of each *CsSWEET*, including chromosome position, gene identifiers, isoelectric point, molecular weight, and gene name, is shown in **Table 1**.

To gain insight into the evolutionary relationship of *CsSWEET*s in *C. sinensis*, the *SWEET*s from *C. sinensis* and *A. thaliana* were aligned by MAFFT software. Remarkably, *HsSWEET1* from *Homo sapiens* was an outgroup, which has only one copy and could transport glucose in *H. sapiens*. In *V. fordii*, Cao et al. (2019) found that *VfSWEET1* contributes to the balance of sucrose levels, and other *VfSWEET*s may play key roles in the growth and development of plants (Cao et al., 2019). In *A. thaliana*, *AtSWEET*s were identified with versatile functions, such as control fructose content and sucrose efflux (Chen et al., 2010, 2012). In our study, the *SWEET*s from *C. sinensis* and *A. thaliana* were clustered into four subfamilies, including subfamily I, subfamily II, subfamily III, and subfamily IV (**Figure 1**). Subsequently, there was at least one *AtSWEET* from *A. thaliana* in each subfamily. Previously published papers have confirmed that members from different *SWEET* subfamilies may have multiple biological functions, such that the members from subfamily I and subfamily II were efficient hexose transporters (**Figure 1**), the members from subfamily III may be responsible for sucrose transporters, and the members from subfamily IV

TABLE 1 | Detailed information of SWEET family members in *C. sinensis*.

Gene name	Gene identifiers	Chromosome	Forward (+)/reverse (–)	3′ end	5′ end	p*I*	MW
CsSWEET1a	CSS0018035.1	Chr4	–	47,319,514	47,322,176	9.64	27,695.99
CsSWEET1b	CSS0018347.1	Chr9	–	71,038,986	71,045,420	9.68	26,377.41
CsSWEET2a	CSS0040201.1	Contig440	–	7,089	14,988	8.84	21,873.99
CsSWEET2b	CSS0003324.2	Chr5	–	46,529,212	46,534,196	8.81	25,966.73
CsSWEET2c	CSS0042514.1	Chr8	+	68,320,865	68,324,641	9.38	23,055.65
CsSWEET3a	CSS0003069.1	Chr7	+	174,818,978	174,820,449	8.61	25,941.05
CsSWEET3b	CSS0039909.1	Contig86	–	502,517	504,121	9.22	30,808.87
CsSWEET3c	CSS0028613.1	Chr1	+	98,968,341	98,969,945	9.22	30,736.72
CsSWEET5a	CSS0040324.1	Chr15	–	79,558,505	79,562,983	9.21	26,678.06
CsSWEET5b	CSS0037258.1	Chr2	+	6,354,873	6,363,319	8.54	51,845.3
CsSWEET5c	CSS0009453.1	Chr2	–	83,321,970	83,325,289	6.72	27,143.31
CsSWEET7a	CSS0014422.1	Chr1	+	220,693,547	220,710,933	5.25	80,809.52
CsSWEET7b	CSS0001069.1	Chr12	–	158,359,825	158,365,331	8.93	28,302.76
CsSWEET7c	CSS0033641.1	Chr12	–	158,634,675	158,640,280	9.06	28,329.83
CsSWEET9a	CSS0020524.1	Chr11	–	96,099,923	96,104,109	6.81	27,224.46
CsSWEET9b	CSS0016012.1	Chr11	–	96,128,414	96,130,178	8.72	30,791.65
CsSWEET10a	CSS0014835.1	Chr11	+	13,881,045	13,882,601	9.26	31,961.51
CsSWEET10b	CSS0011593.1	Chr15	–	49,947,912	49,948,754	7.66	21,250.28
CsSWEET10c	CSS0026264.1	Chr2	–	97,204,012	97,206,001	9	36,861.14
CsSWEET12	CSS0010858.1	Chr15	–	49,282,933	49,284,143	8.97	28,258.6
CsSWEET15a	CSS0049101.1	Chr6	+	3,037,473	3,039,942	7.75	30,227.9
CsSWEET15b	CSS0007875.1	Chr11	+	96,022,451	96,035,117	9.27	32,181.35
CsSWEET15c	CSS0017308.1	Chr11	+	96,051,665	96,053,111	9.5	26,881.44
CsSWEET17a	CSS0009124.1	Chr3	+	61,755,153	61,761,768	8.73	33,441.32
CsSWEET17b	CSS0021289.1	Chr9	–	8,102,406	8,106,125	6.19	26,228.91
CsSWEET17c	CSS0005451.1	Chr3	+	61,588,229	61,597,620	8.89	22,789.63

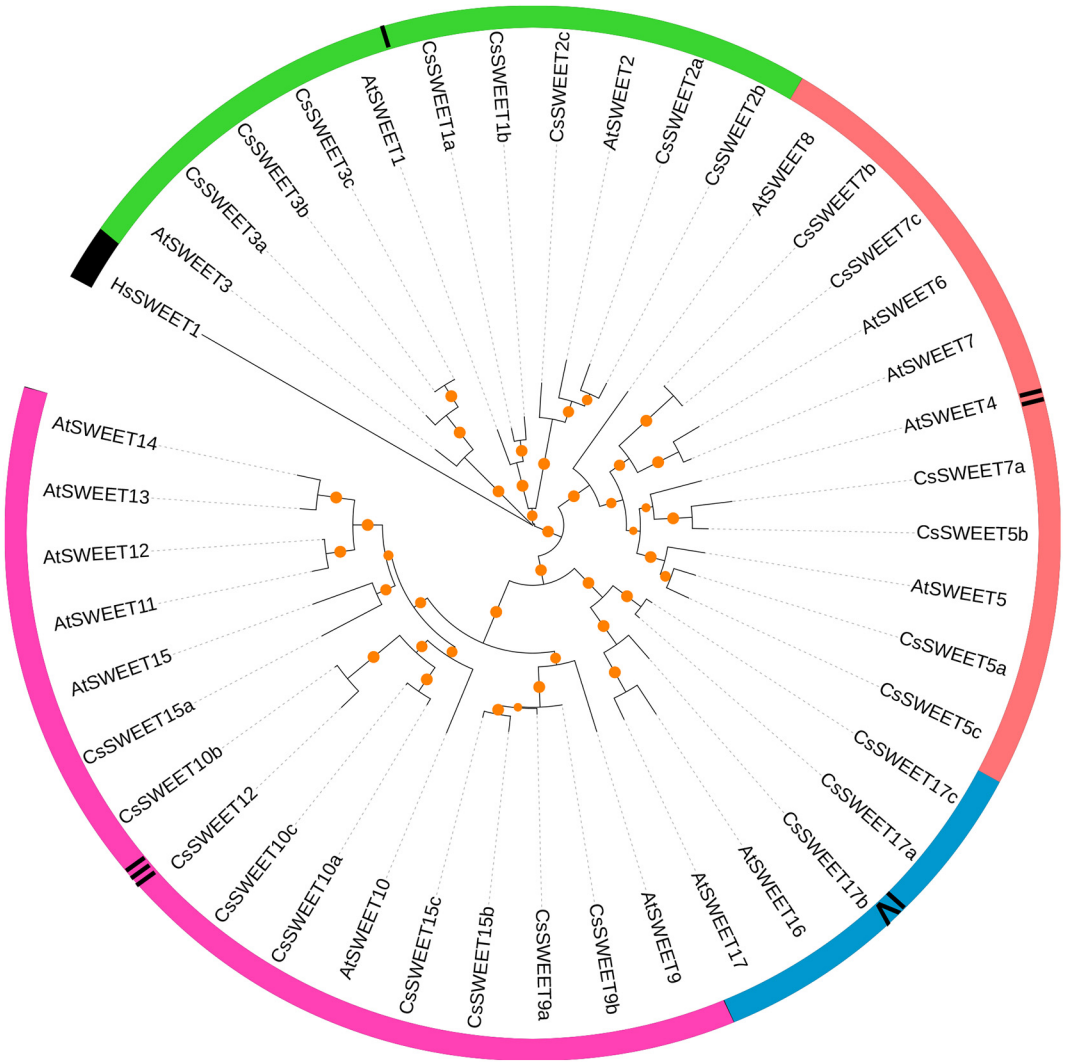


FIGURE 1 | The phylogenetic relationships of *SWEET* genes in *Camellia sinensis*, *Arabidopsis thaliana*, and *Homo sapiens*. The maximum likelihood tree was built by IQ-tree software with bootstrap test (10,000 replicates).

appear to be predominantly fructose transporters (Chen et al., 2010, 2012; Hu et al., 2018; Cao et al., 2019).

The Structure Analyses Reveal the Distribution of Exon–Intron and the Conserved Composition of Motif in *C. sinensis*

According to the predicted sequences and GFF annotation files, we determined the distribution of exon–intron of the 26 *CsSWEET* genes in *C. sinensis*. As shown in **Supplementary Figure 1**, most *CsSWEET* genes contained five introns; however, several of the members from the *CsSWEET* gene family contained two, three, or four introns. Remarkably, there are extreme differences in the number of introns from the *CsSWEET* subfamily II, ranging from 4 to 14, indicating that these *CsSWEET*s may have experienced intron loss or gain some during evolution (**Supplementary Figure 1**). These results were also confirmed by previous studies. For example, Cao et al. (2019) found that the number of introns from *VfSWEET*s ranged from four to 10. Additionally, we found that, although the lengths of members from *CsSWEET* gene family vary, introns stepped into almost the same position of the gene open reading frame.

The MEME website was used to identify the sequence characteristics of CsSWEET genes. Subsequently, we determined 20 distinct motifs for CsSWEETs and annotated these motifs by Pfam and SMART database (Letunic and Bork, 2018; Mistry et al., 2020). The SWEET domain (MtN3/saliva) was encoded by motifs 1, 2, 3, and 4 (**Supplementary Figure 2**). However, the remaining motifs were not scanned for function annotations in Pfam and SMART database. Overall, the CsSWEET proteins within the same subfamily exhibited similar conserved motifs, especially in paralogs gene pairs, suggesting that these sequences might contain similar functions in *C. sinensis*.

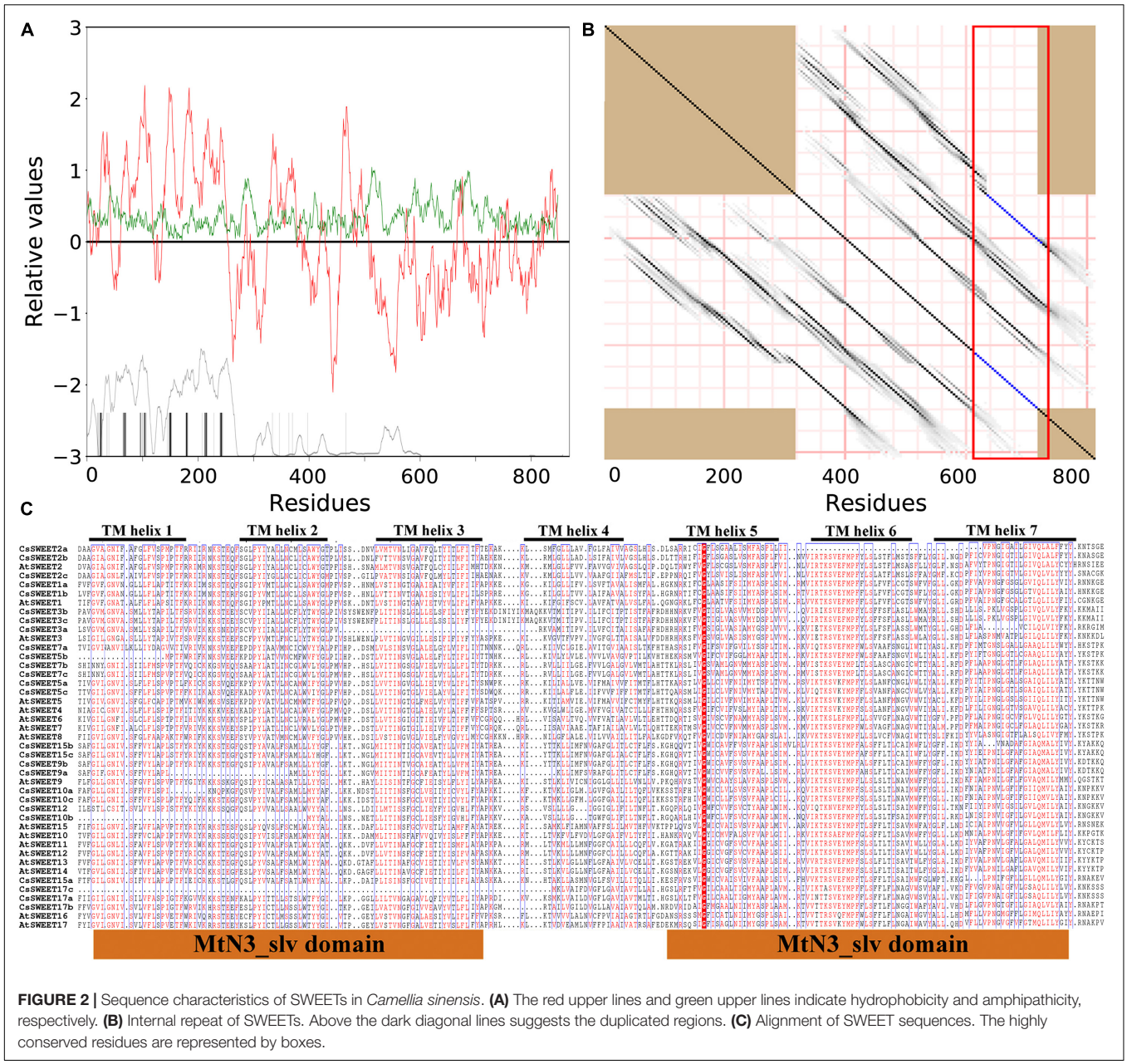


FIGURE 2 | Sequence characteristics of SWEETs in *Camellia sinensis*. **(A)** The red upper lines and green upper lines indicate hydrophobicity and amphipathicity, respectively. **(B)** Internal repeat of SWEETs. Above the dark diagonal lines suggests the duplicated regions. **(C)** Alignment of SWEET sequences. The highly conserved residues are represented by boxes.

Internal Repeat and Transmembrane Domain Analysis of the *CsSWEET*s Reveal Insight Into Key Functional Residues

In general, SWEET proteins have been confirmed to contain seven transmembrane helixes (TMHs) in eukaryotes (Feng and Frommer, 2015; Jia et al., 2017). As expected, the membrane topology of CsSWEETs and AtSWEETs revealed that these proteins contained a moderately amphipathic nature with seven well-conserved hydrophobicity peaks (TMH1–TMH7) as determined by AveHAS plot (**Figures 2A,B**). Compared with TMHs, the loop area between two TMHs is not conservative in this study (**Figure 2C**). Previously published manuscripts have shown that the internal fusion and duplication of small fragments play an important role in the SWEETs evolution, which has been the focus of discussion by researchers (Hu et al., 2018; Cao et al., 2019). Researchers have studied the internal duplication of the SWEET proteins (Li et al., 2017; Hu et al., 2018; Cao et al., 2019), but direct evidence of the CsSWEETs from *C. sinensis* has not been reported. To confirm the presence of internal duplication, a repeat analysis of CsSWEETs and AtSWEETs was performed by the HHrepID program (Söding et al., 2006). The results confirmed that TMH1–TMH3 and TMH5–TMH7 were located in the duplicated regions, implying that SWEETs originate from an ancestral three-TMH unit which

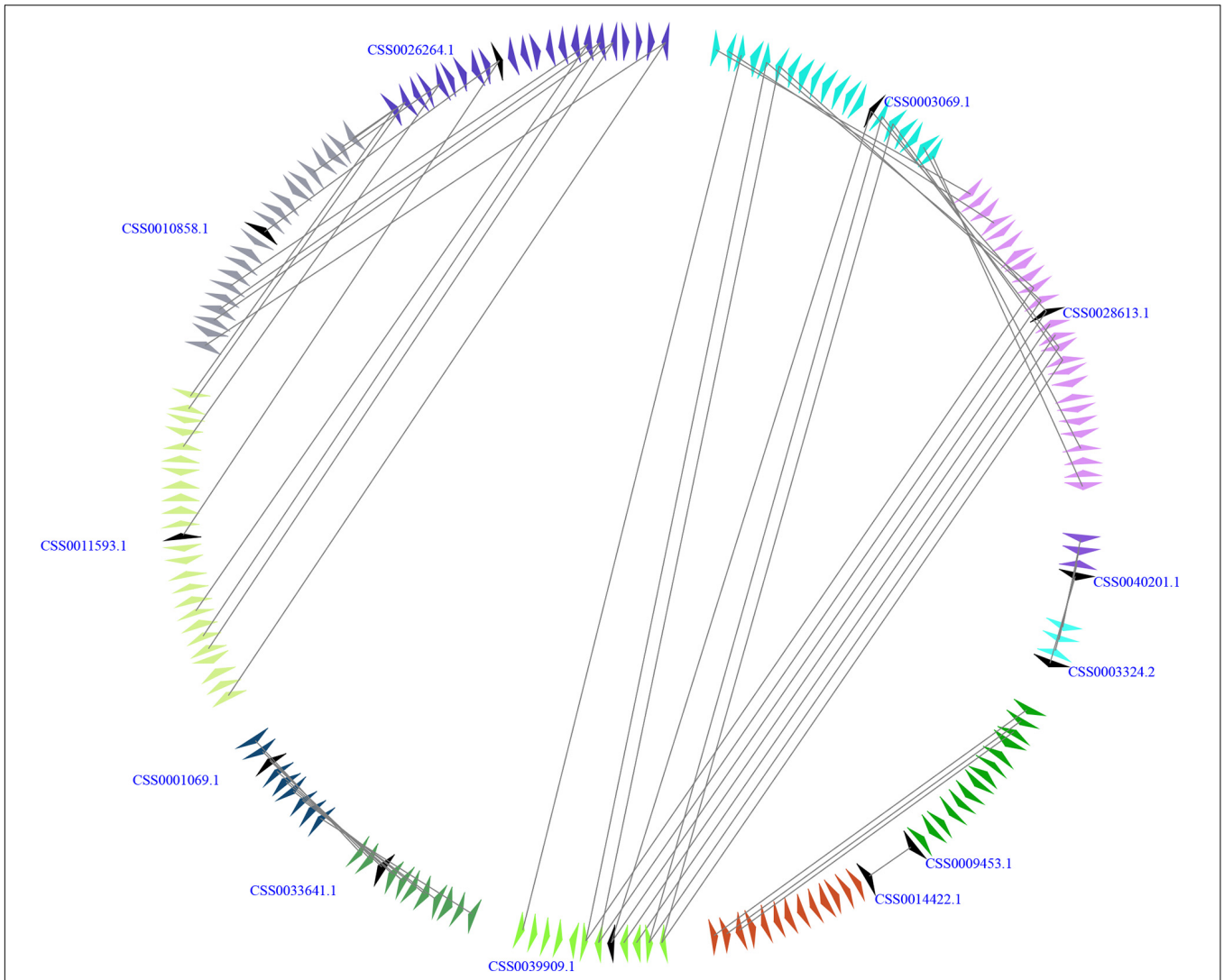


FIGURE 3 | Microsynteny relationships of *SWEET* genes in *Camellia sinensis*. A series of triangles indicate genomic fragments and the genes' orientations. The duplicated gene pairs on the two fragments are linked by gray lines.

experienced an internal repeat duplication, which is in agreement with previously published papers (Li et al., 2017; Hu et al., 2018; Cao et al., 2019).

Li et al. (2017) revealed that Y, P, and D from TMH2, TMH5, and TMH7 were fully conserved residues in pear, respectively (Li et al., 2017). Among these three residues, Y mutation into A will lead to the loss of transport function in *A. thaliana* (Xuan et al., 2013). Proline (P) is a key component of TMHs and allows for dynamic processes during the transport cycle as described by Deber and Therien (2002). In the present study, we suggested that these three residues (Y, P, and D) might contain important functions within the activity of CsSWEETs in *C. sinensis*. In addition, the mutation of 58 amino acid residues G into D can significantly reduce AtSWEET1 activity in *A. thaliana* (Xuan et al., 2013). However, this amino acid is not completely conserved in *C. sinensis*, including the residue G in CsSWEET12 that had changed to N and in CsSWEET15a that had changed to K, but these proteins also have transport activity (**Figure 2C**), which has been verified by RNA-seq. We also noted that the first and the second MtN3/saliva domain included TMH1–TMH3 and TMH5–TMH7, respectively.

Microsynteny Analysis Indicates That Large-Scale Duplication Events Contribute to the Expansion for *CsSWEET* Genes in *C. sinensis*

The genome of *C. sinensis* experienced two rounds of whole-genome duplications (WGDs), including an ancient WGD event (∼90 to 100 Mya) and a recent WGD event (∼30 to 40 Mya) (Wei et al., 2018; Wu et al., 2018; Liu et al., 2020; Xia et al., 2020; Zhang et al., 2020b). To gain insight into the expansion mechanism of the members from *CsSWEET* gene family, we carried out a microsynteny analysis with *E*-value 10−5 as described by

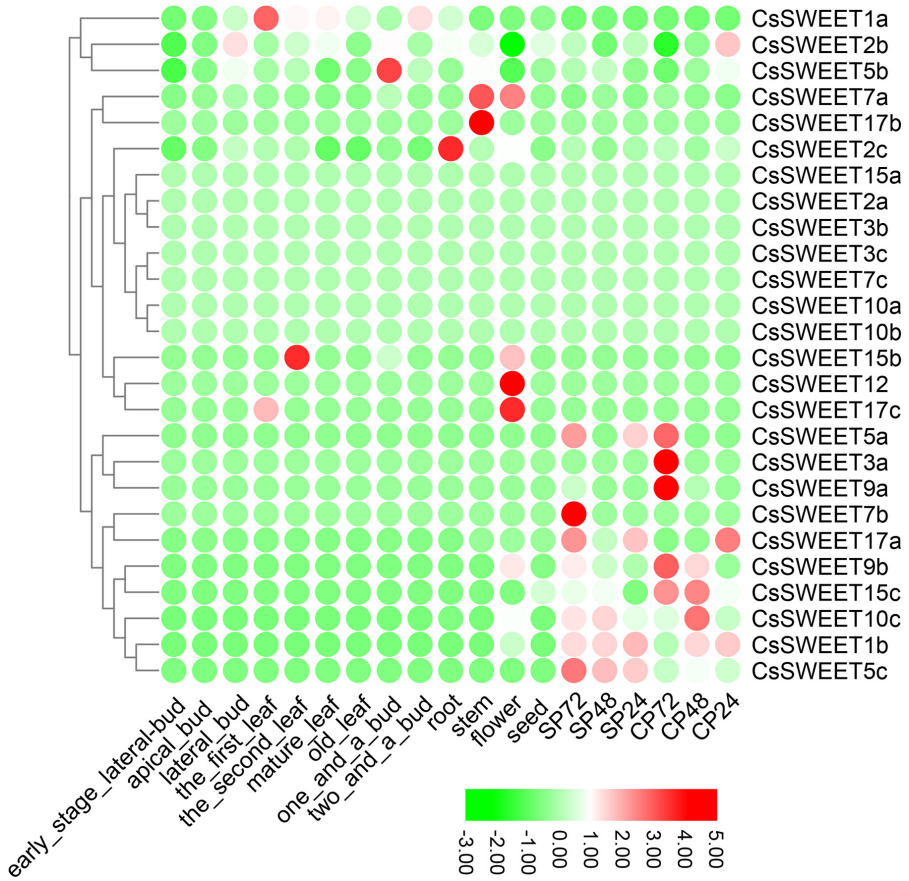


FIGURE 4 | Expression profiles of the *CsSWEET*s in different tissues, including seed, flower, stem, root, two and a bud, one and a bud, old leaf, mature leaf, the second leaf, the first leaf, lateral bud, apical bud, and early stage lateral bud. CP24 = 24 h after pollination; CP, cross-pollinated – CP48, CP72; SP24 (SP, self-pollinated), SP48, and SP72.

Cao et al. (2020b). In our study, *CsSWEET3c/CsSWEET3a*, *CsSWEET10c/CsSWEET12*, *CsSWEET10c/CsSWEET10b*, *CsSWEET7b/CsSWEET7c*, *CsSWEET3c/CsSWEET3b*, *CsSWEET3a/CsSWEET3b*, *CsSWEET7a/CsSWEET5c*, and *CsSWEET2b/CsSWEET2a* contained more than three pairs of conserved flanking genes, suggesting that significant microsynteny might occur in the *CsSWEET* genes (**Figure 3**). These results indicated that the large-scale duplication events contribute to the expansion for *CsSWEET* genes in *C. sinensis*.

Generally, the evolutionary data of WGD events were estimated by *Ks* values (Wang et al., 2011; Tiley et al., 2018; Zwaenepoel and Van De Peer, 2019). As shown in **Table 1**, we found that the *Ks* values of *CsSWEET* duplication pairs ranged from 0.0103 to 2.1158 (**Supplementary Table 1** and **Supplementary Figure 3**). Subsequently, we found that *CsSWEET3c/CsSWEET3a*, *CsSWEET3a/CsSWEET3b*, and *CsSWEET2b/CsSWEET2a* might be generated through a recent WGD event, while *CsSWEET3c/CsSWEET3b* and *CsSWEET10c/CsSWEET12* might be generated through an ancient WGD event, suggesting that these two rounds of WGD events might play key roles in the expansion for *CsSWEET* genes in *C. sinensis*.

The Expression of *CsSWEET*s Reveals Possible Functions in *C. sinensis*

Previously published work provided transcriptome data for *CsSWEET* genes in *C. sinensis* (Xia et al., 2019; Xia et al., 2020). The members from *SWEET* gene family are found to play diverse functional roles in various tissues and contribute to different sugar transport mechanisms in plants (Chen, 2014; Hedrich et al., 2015; Li et al., 2017; Cao et al., 2019). Differential expression analysis of *CsSWEET*s in *C. sinensis* is helpful for us to find out the specialized functions of these *CsSWEET*s in sugar transport from the practical application point of view. In the present study, we collected transcriptome data from 19 different tissues, including seed, flower, stem, root, two and a bud, one and a bud, old leaf, mature leaf, the second leaf, the first leaf, lateral bud, apical bud, early stage lateral bud, CP24 (24 h after pollination; CP, cross-pollinated), CP48, CP72, SP24 (SP, self-pollinated), SP48, and SP72. The published papers indicated that the expression of *SWEET*s in seeds pollen, flower, and embryo sacs in *Sorghum bicolor*, *V. fordii*, *Litchi chinensis*, *O. sativa*, and *A. thaliana* was higher (Yuan and Wang, 2013; Mizuno et al., 2016; Cao et al., 2019; Xie et al., 2019), implying that these genes may

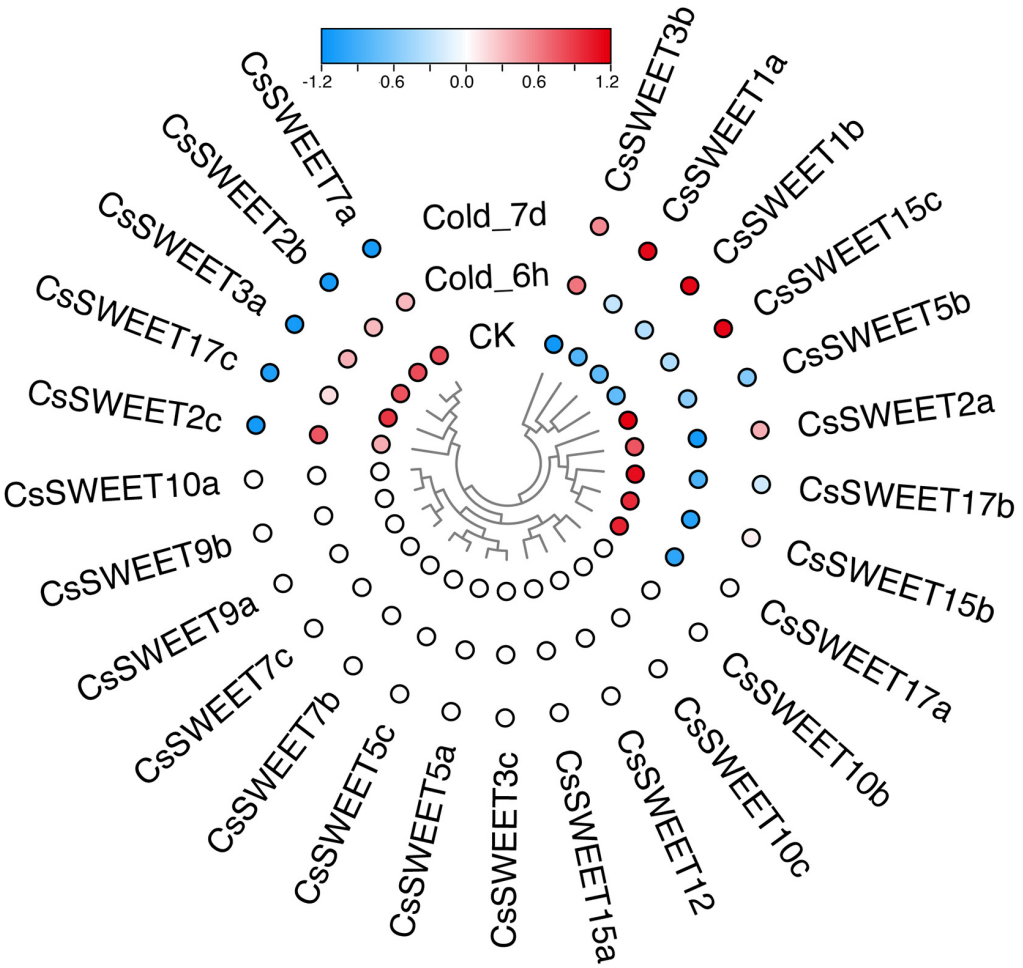


FIGURE 5 | Expression profiles of the *CsSWEET*s in response to cold treatment.

contribute to reproductive development. Here *CsSWEET1a* that was clustered into the same subfamily with *VfSWEET1* from *V. fordii*, *PbSWEET14* from *P. bretschneideri*, and *AtSWEET1* from *A. thaliana* was extremely highly expressed in the leaf of *C. sinensis*. *PbSWEET14* and *VfSWEET1* contribute to the efflux of sucrose in leaves (Li et al., 2017; Cao et al., 2019), while *AtSWEET1* plays an important role as a single glucose transporter in multiple systems. Therefore, the *CsSWEET1a* might play crucial roles in the efflux of sucrose and act as a glucose uniporter in the leaves of *C. sinensis* (**Figure 4**). *AtSWEET16* and *AtSWEET17* were shown to be vacuolar hexose transporters that controlled fructose content (Chen et al., 2010). The *CsSWEET17b*, which is the orthologous gene of *AtSWEET16* and *AtSWEET17*, was extremely highly expressed in the stems of *C. sinensis*, suggesting that this gene could control fructose content as a hexose transporter. The expression patterns of *CsSWEET2a*, *CsSWEET5b*, and *CsSWEET7a* were diverse in several tissues, implying that these *CsSWEET*s might play a role in these tested tissues of *C. sinensis* (**Figure 4**).

In *C. sinensis*, the expression of *CsSWEET*s was examined in pollen development. We found that 10 *CsSWEET*s, including *CsSWEET5a*, *CsSWEET3a*, *CsSWEET9a*, *CsSWEET7b*, *CsSWEET17a*, *CsSWEET9b*, *CsSWEET15c*, *CsSWEET10c*, *CsSWEET1b*, and *CsSWEET5c*, were highly expressed in one or several pollen stages of *C. sinensis* (**Figure 4**). As previously reported, Chen et al. (2012) identified *AtSWEET11* and *AtSWEET12* that contributed to phloem loading and pollen nutrition. Gao et al. (2018) found that the *OsSWEET11* from *O. sativa* can greatly affect the starch pollen content. In our study, *CsSWEET9b*, *CsSWEET9a*, and *CsSWEET10c* are the orthologous genes of *AtSWEET11* and *AtSWEET12*, indicating that these highly expressed genes may share similar roles and functions.

Stress-Induced Expression Profiles of *CsSWEET*s in *C. sinensis*

Plants often suffer from a variety of abiotic stresses in the process of growth and development (Franco et al., 2011; Keunen et al., 2013; Etesami, 2018). Previous studies have shown that *SWEET*s help to control the responses to environmental stresses in plants (Li et al., 2018; Zhang et al., 2020a). Therefore, the

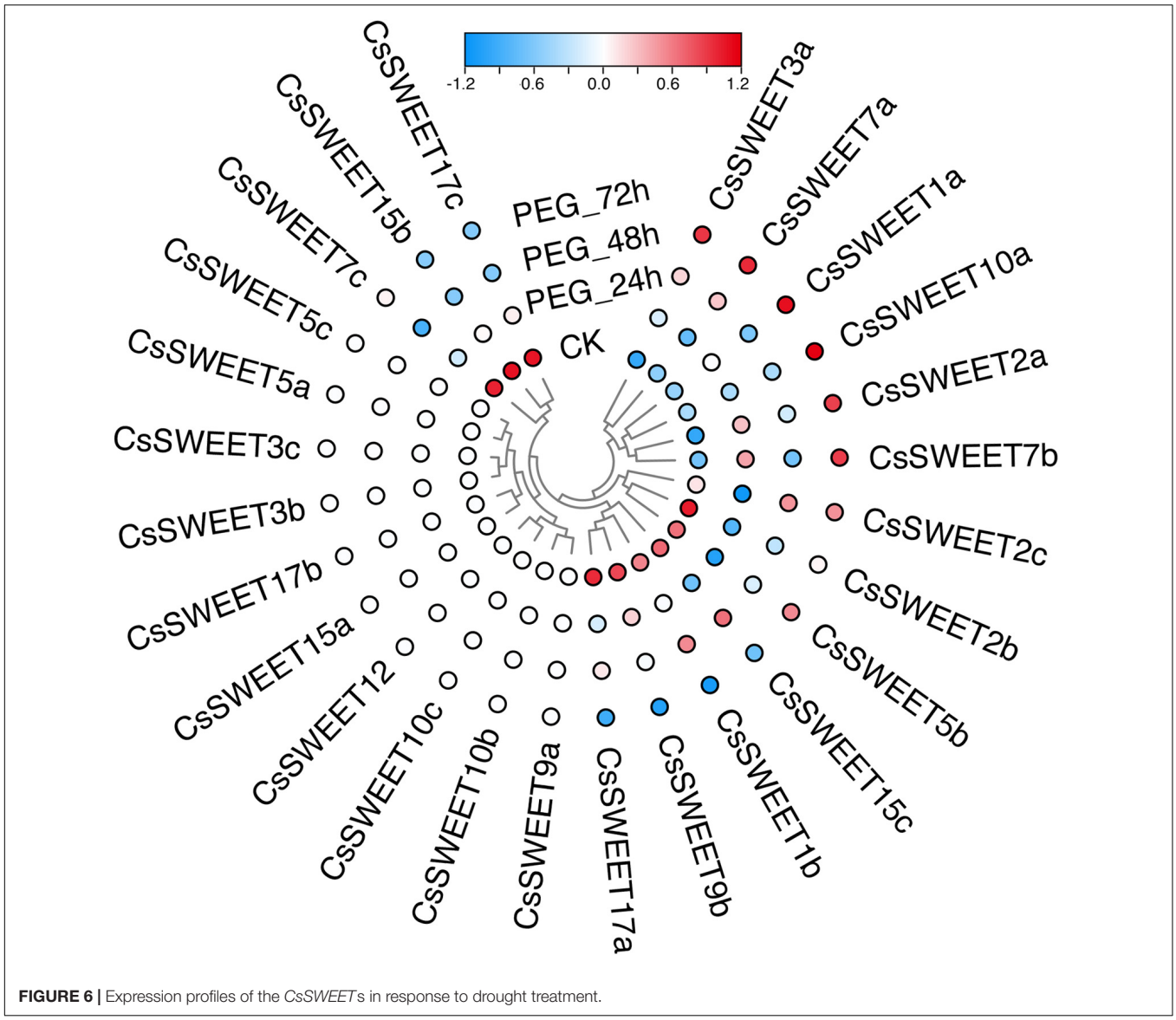


FIGURE 6 | Expression profiles of the *CsSWEET*s in response to drought treatment.

expression patterns of *CsSWEET*s were investigated in response to different stresses, including cold stress and drought stress. In our study, the RNA-seq data for cold stress and drought stress were obtained from the TPIA (see text footnote 1) database (Xia et al., 2019). In response to drought, seven genes (*CsSWEET3a*, *CsSWEET7a*, *CsSWEET1a*, *CsSWEET10a*, *CsSWEET2a*, *CsSWEET7b*, and *CsSWEET2c*) were up-regulated under polyethylene glycol (PEG) for 72 h. Compared with control, nine genes (*CsSWEET2b*, *CsSWEET5b*, *CsSWEET15c*, *CsSWEET1b*, *CsSWEET9b*, *CsSWEET17a*, *CsSWEET17c*, *CsSWEET7c*, and *CsSWEET15b*) were down-regulated under all PEG treatments (**Figure 5**). In response to cold, four genes (*CsSWEET3b*, *CsSWEET1a*, *CsSWEET1b*, and *CsSWEET15c*) were up-regulated under all cold treatments. Compared with control, nine genes (*CsSWEET5b*, *CsSWEET2a*, *CsSWEET17b*, *CsSWEET15b*, *CsSWEET17a*, *CsSWEET7a*, *CsSWEET2b*, *CsSWEET3a*, and *CsSWEET17c*) were down-regulated under all cold treatments (**Figure 6**). Among them, the expression of *CsSWEET1a* was significantly up-regulated in all chosen two treatments, indicating that this gene might help tea resist environmental stresses. This phenomenon was different from previous studies. For example, the banana *MaSWEET1a* was not induced by osmotic, cold, and salt stresses (Miao et al., 2017). The expression level of *GhSWEET1* was only upregulated at 6 h under cold stress, but it was almost not expressed when induced by drought stress (Li et al., 2018). Taken together, we believed that tea plants might have developed specialized regulatory mechanisms for different abiotic stresses.

CONCLUSION

In the present study, we identified 28 *CsSWEET*s in the *C. sinensis* genome. Further analyses for ML tree, intron–exon, and

duplication suggested that the conservation of *CsSWEET*s was accompanied by a certain degree of divergence. The expression profiles of all *CsSWEET*s suggested that several genes were highly expressed in reproductive tissues, indicating that these genes played important roles in sugar transport. Additionally, 28 *CsSWEET*s were differently expressed for RNA-seq data under cold and drought stresses, indicating that tea plants might have developed specialized regulatory mechanisms for different abiotic stresses.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

LJ and XZ performed the experiments, analyzed the data, and wrote the manuscript. LJ, CS, XZ, and JY conceived the research and revised the manuscript. All the authors read and approved the final manuscript.

FUNDING

This work was supported by the Talent Scientific Research Start-up Foundation of Yijishan Hospital, Wannan Medical College (grant no. YR202001), the Opening Foundation of Key Laboratory of Non-coding RNA Transformation Research of Anhui Higher Education Institution (grant no. RNA202004), and the Key Projects of Natural Science Research of Universities in Anhui Province (grant no. KJ2020A0622).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.655843/full#supplementary-material

Supplementary Figure 1 | The gene structure of *CsSWEET*s in *Camellia sinensis*. Green boxes suggest exons, and gray lines indicate introns.

Supplementary Figure 2 | The distribution of conserved motifs in CsSWEETs of *Camellia sinensis*. The MEME was used to identify the motif. The different colors of the boxes indicate different motifs numbered 1–20.

Supplementary Figure 3 | Ka/Ks analysis for duplicated *CsSWEET*s paralogs.

Supplementary Table 1 | Ka/Ks analysis for duplicated *CsSWEET*s paralogs.

REFERENCES

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

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

Cai, B., Yang, X., Tuskan, G. A., and Cheng, Z.-M. (2011). MicroSyn: a user friendly tool for detection of microsynteny in a gene family. *BMC Bioinformatics* 12:79. doi: 10.1186/1471-2105-12-79

Cao, Y., Liu, M., Long, H., Zhao, Q., Jiang, L., and Zhang, L. (2020a). Hidden in plain sight: Systematic investigation of Leucine-rich repeat containing genes unveil the their regulatory network in response to Fusarium wilt in tung tree. *Int. J. Biol. Macromole.* 163, 1759–1767. doi: 10.1016/j.ijbiomac.2020.09.106

Cao, Y., Liu, W., Zhao, Q., Long, H., Li, Z., Liu, M., et al. (2019). Integrative analysis reveals evolutionary patterns and potential functions of SWEET transporters in Euphorbiaceae. *Int. J. Biol. Macromole.* 139, 1–11. doi: 10.1016/j.ijbiomac.2019.07.102

Cao, Y., Xu, X., and Jiang, L. (2020b). Integrative analysis of the RNA interference toolbox in two Salicaceae willow species, and their roles in stress response in poplar (Populus trichocarpa Torr. & Gray). *Int. J. Biol. Macromole.* 162, 1127–1139. doi: 10.1016/j.ijbiomac.2020.06.235

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. *BioRxiv* 2020:289660.

Chen, L. Q. (2014). SWEET sugar transporters for phloem transport and pathogen nutrition. *N. Phytol.* 201, 1150–1155. doi: 10.1111/nph.12445

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X.-Q., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468, 527–532.

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A. R., et al. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* 335, 207–211. doi: 10.1126/science.1213351

Chong, J., Piron, M.-C., Meyer, S., Merdinoglu, D., Bertsch, C., and Mestre, P. (2014). The SWEET family of sugar transporters in grapevine: VvSWEET4 is involved in the interaction with Botrytis cinerea. *J. Exp. Bot.* 65, 6589–6601. doi: 10.1093/jxb/eru375

Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20, 1250–1255. doi: 10.1101/gad.1416306

Deber, C. M., and Therien, A. G. (2002). Putting the β-breaks on membrane protein misfolding. *Nat. Struct. Biol.* 9, 318–319. doi: 10.1038/nsb0502-318

Etesami, H. (2018). Can interaction between silicon and plant growth promoting rhizobacteria benefit in alleviating abiotic and biotic stresses in crop plants? *Agricult. Ecosyst. Env.* 253, 98–112. doi: 10.1016/j.agee.2017.11.007

Feng, L., and Frommer, W. B. (2015). Structure and function of SemiSWEET and SWEET sugar transporters. *Trends Biochem. Sci.* 40, 480–486. doi: 10.1016/j.tibs.2015.05.005

Franco, J. A., Bañón, S., Vicente, M. J., Miralles, J., and Martínez-Sánchez, J. J. (2011). Root development in horticultural plants grown under abiotic stress conditions–a review. *J. Horticult. Sci. Biotechnol.* 86, 543–556. doi: 10.1080/14620316.2011.11512802

Gamas, P., De Carvalho Niebel, F., Lescure, N., and Cullimore, J. V. (1996). Use of a subtractive hybridization approach to identify new Medicago truncatula genes induced during root nodule development. *MPMI-Mole. Plant Microb. Interact.* 9, 233–242. doi: 10.1094/mpmi-9-0233

Gao, Y., Zhang, C., Han, X., Wang, Z. Y., Ma, L., Yuan, D. P., et al. (2018). Inhibition of OsSWEET11 function in mesophyll cells improves resistance of rice to sheath blight disease. *Mole. Plant Pathol.* 19, 2149–2161. doi: 10.1111/mpp.12689

Guan, Y.-F., Huang, X.-Y., Zhu, J., Gao, J.-F., Zhang, H.-X., and Yang, Z.-N. (2008). RUPTURED POLLEN GRAIN1, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in Arabidopsis. *Plant Physiol.* 147, 852–863. doi: 10.1104/pp.108.118026

Hamada, M., Wada, S., Kobayashi, K., and Satoh, N. (2005). Ci-Rga, a gene encoding an MtN3/saliva family transmembrane protein, is essential for tissue differentiation during embryogenesis of the ascidian Ciona intestinalis. *Differentiation* 73, 364–376. doi: 10.1111/j.1432-0436.2005.00037.x

Hedrich, R., Sauer, N., and Neuhaus, H. E. (2015). Sugar transport across the plant vacuolar membrane: nature and regulation of carrier proteins. *Curr. Opin. Plant Biol.* 25, 63–70. doi: 10.1016/j.pbi.2015.04.008

Hu, W., Hua, X., Zhang, Q., Wang, J., Shen, Q., Zhang, X., et al. (2018). New insights into the evolution and functional divergence of the SWEET family in Saccharum based on comparative genomics. *BMC Plant Biol.* 18:270.

Jia, B., Zhu, X. F., Pu, Z. J., Duan, Y. X., Hao, L. J., Zhang, J., et al. (2017). Integrative view of the diversity and evolution of SWEET and SemiSWEET sugar transporters. *Front. Plant Sci.* 8:2178.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., Mcanulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Katoh, K., Kuma, K.-I., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518. doi: 10.1093/nar/gki198

Keunen, E. L. S., Peshev, D., Vangronsveld, J., Van Den Ende, W. I. M., and Cuypers, A. N. N. (2013). Plant sugars are crucial players in the oxidative challenge during abiotic stress: extending the traditional concept. *Plant Cell Env.* 36, 1242–1255. doi: 10.1111/pce.12061

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Lastdrager, J., Hanson, J., and Smeekens, S. (2014). Sugar signals and the control of plant growth and development. *J. Exp. Bot.* 65, 799–807. doi: 10.1093/jxb/ert474

Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496.

Li, J., Qin, M., Qiao, X., Cheng, Y., Li, X., Zhang, H., et al. (2017). A new insight into the evolution and functional divergence of SWEET transporters in Chinese white pear (Pyrus bretschneideri). *Plant Cell Physiol.* 58, 839–850. doi: 10.1093/pcp/pcx025

Li, W., Ren, Z., Wang, Z., Sun, K., Pei, X., Liu, Y., et al. (2018). Evolution and stress responses of Gossypium hirsutum SWEET genes. *Int. J. Mole. Sci.* 19:769. doi: 10.3390/ijms19030769

Liu, Z.-W., Li, H., Liu, J.-X., Wang, Y., and Zhuang, J. (2020). Integrative transcriptome, proteome, and microRNA analysis reveals the effects of nitrogen sufficiency and deficiency conditions on theanine metabolism in the tea plant (Camellia sinensis). *Horticulture Res.* 7, 1–13.

May, M. J., Vernoux, T., Leaver, C., Montagu, M. V., and Inzé, D. (1998). Glutathione homeostasis in plants: implications for environmental sensing and plant development. *J. Exp. Bot.* 49, 649–667. doi: 10.1093/jexbot/49.321.649

Miao, H., Sun, P., Liu, Q., Miao, Y., Liu, J., Zhang, K., et al. (2017). Genome-wide analyses of SWEET family proteins reveal involvement in fruit development and abiotic/biotic stress responses in banana. *Sci. Rep.* 7, 1–15.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 8, D412–D419.

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–e121.

Mizuno, H., Kasuga, S., and Kawahigashi, H. (2016). The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling. *Biotechnol. Biofuels* 9:127.

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

Pan, J., Li, D., Zhu, J., Shu, Z., Ye, X., Xing, A., et al. (2020). Aluminum relieves fluoride stress through stimulation of organic acid production in Camellia sinensis. *Physiol. Mole. Biol. Plants* 26, 1127–1137. doi: 10.1007/s12298-020-00813-2

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protocols* 11:1650. doi: 10.1038/nprot.2016.095

Rahimi, M., Kordrostami, M., and Mortezavi, M. (2019). Evaluation of tea (Camellia sinensis L.) biochemical traits in normal and drought stress conditions to identify drought tolerant clones. *Physiol. Mole. Biol. Plants* 25, 59–69. doi: 10.1007/s12298-018-0564-x

Saddhe, A. A., Manuka, R., and Penna, S. (2020). Plant sugars: homeostasis and transport under abiotic stress in plants. *Physiol. Plantarum* 2020: 13283.

Saier, M. H. Jr., Tran, C. V., and Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34, D181–D186.

Samarina, L. S., Bobrovskikh, A. V., Doroshkov, A. V., Malyukova, L. S., Matskiv, A. O., Rakhmangulov, R. S., et al. (2020). Comparative Expression Analysis of Stress-Inducible Candidate Genes in Response to Cold and Drought in Tea Plant [Camellia sinensis (L.) Kuntze]. *Front. Genet.* 2020:11.

Seo, P. J., Park, J.-M., Kang, S. K., Kim, S.-G., and Park, C.-M. (2011). An Arabidopsis senescence-associated protein SAG29 regulates cell viability under high salinity. *Planta* 233, 189–200. doi: 10.1007/s00425-010-1293-8

Shammai, A., Petreikov, M., Yeselson, Y., Faigenboim, A., Moy-Komemi, M., Cohen, S., et al. (2018). Natural genetic variation for expression of a SWEET transporter among wild species of Solanum lycopersicum (tomato) determines the hexose composition of ripening tomato fruit. *Plant J.* 96, 343–357. doi: 10.1111/tpj.14035

Söding, J., Remmert, M., and Biegert, A. (2006). HHrep: de novo protein repeat detection and the origin of TIM barrels. *Nucleic Acids Res.* 34, W137–W142.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protocols Bioinform.* 2003, 2–3.

Tiley, G. P., Barker, M. S., and Burleigh, J. G. (2018). Assessing the performance of Ks plots for detecting ancient whole genome duplications. *Genome Biol. Evol.* 10, 2882–2898.

Wang, L., Xie, W., Chen, Y., Tang, W., Yang, J., Ye, R., et al. (2010). A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.* 61, 752–766. doi: 10.1111/j.1365-313x.2009.04100.x

Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S. P., Feltus, F. A., et al. (2011). Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *PLoS One* 6:e28150. doi: 10.1371/journal.pone.0028150

Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., et al. (2018). Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* 115, E4151–E4158.

Wipf, D., Pfister, C., Mounier, A., Leborgne-Castel, N., Frommer, W. B., and Courty, P.-E. (2020). Identification of Putative Interactors of Arabidopsis Sugar Transporters. *Trends Plant Sci.* 26:009.

Wu, L.-Y., Fang, Z.-T., Lin, J.-K., Sun, Y., Du, Z.-Z., Guo, Y.-L., et al. (2018). Complementary iTRAQ proteomic and transcriptomic analyses of leaves in tea plant (Camellia sinensis L.) with different maturity and regulatory network of flavonoid biosynthesis. *J. Proteom. Res.* 18, 252–264.

Xia, E. H., Li, F. D., Tong, W., Li, P. H., Wu, Q., Zhao, H. J., et al. (2019). Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol. J.* 17, 1938–1953. doi: 10.1111/pbi.13111

Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., et al. (2020). The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mole. Plant* 13, 1013–1026. doi: 10.1016/j.molp.2020.04.010

Xie, H., Wang, D., Qin, Y., Ma, A., Fu, J., Qin, Y., et al. (2019). Genome-wide identification and expression analysis of SWEET gene family in Litchi chinensis reveal the involvement of LcSWEET2a/3b in early seed development. *BMC Plant Biol.* 19:499.

Xuan, Y. H., Hu, Y. B., Chen, L.-Q., Sosso, D., Ducat, D. C., Hou, B.-H., et al. (2013). Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc. Natl. Acad. Sci.* 110, E3685–E3694.

Yin, Q., Zhu, L., Du, P., Fan, C., Wang, J., Zhang, B., et al. (2020). Comprehensive analysis of SWEET family genes in Eucalyptus (Eucalyptus grandis). *Biotechnol. Biotechnol. Equip.* 34, 595–604. doi: 10.1080/13102818.2020.1790417

Yuan, M., and Wang, S. (2013). Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. *Mole. Plant* 6, 665–674. doi: 10.1093/mp/sst035

Yue, C., Cao, H.-L., Wang, L., Zhou, Y.-H., Huang, Y.-T., Hao, X.-Y., et al. (2015). Effects of cold acclimation on sugar metabolism and sugar-related gene expression in tea plant during the winter season. *Plant Mole. Biol.* 88, 591–608. doi: 10.1007/s11103-015-0345-7

Zhai, Y., and Saier, M. H. (2001a). A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. *J. Mole. Microbiol. Biotechnol.* 3, 501–502.

Zhai, Y., and Saier, M. H. Jr. (2001b). A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. *J. Mole. Microbiol. Biotechnol.* 3, 285–286.

Zhang, R., Niu, K., and Ma, H. (2020a). Identification and Expression Analysis of the SWEET Gene Family from Poa pratensis Under Abiotic Stresses. *DNA Cell Biol.* 39, 1606–1620. doi: 10.1089/dna.2020.5418

Zhang, W., Zhang, Y., Qiu, H., Guo, Y., Wan, H., Zhang, X., et al. (2020b). Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. *Nat. Commun.* 11, 1–12. doi: 10.1017/cbo9781316036198.002

Zhou, Y., Liu, L., Huang, W., Yuan, M., Zhou, F., Li, X., et al. (2014). Overexpression of OsSWEET5 in rice causes growth retardation and precocious senescence. *PLoS One* 9:e94210. doi: 10.1371/journal.pone.0094210

Zhou, Y., Liu, Y., Wang, S., Shi, C., Zhang, R., Rao, J., et al. (2017). Molecular cloning and characterization of galactinol synthases in Camellia sinensis with different responses to biotic and abiotic stressors. *J. Agricult. Food Chem.* 65, 2751–2759. doi: 10.1021/acs.jafc.7b00377

Zwaenepoel, A., and Van De Peer, Y. (2019). wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35, 2153–2155. doi: 10.1093/bioinformatics/bty915

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 April 2021
doi: 10.3389/fgene.2021.661296

Combined Analysis of Volatile Terpenoid Metabolism and Transcriptome Reveals Transcription Factors Related to Terpene Synthase in Two Cultivars of *Dendrobium officinale* Flowers

*Ninghong Li†, Yingxue Dong†, Min Lv, Li Qian, Xu Sun, Lin Liu, Yongping Cai and Honghong Fan**

School of Life Sciences, Anhui Agricultural University, Hefei, China

OPEN ACCESS

Edited by:
Yunpeng Cao,
Central South University Forestry and Technology, China

Reviewed by:
Kaikai Zhu,
Nanjing Forestry University, China
Dongfeng Yang,
Zhejiang Sci-Tech University, China

***Correspondence:**
Honghong Fan
hhfan0551@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 January 2021*
Accepted: *05 March 2021*
Published: *22 April 2021*

Citation:
Li N, Dong Y, Lv M, Qian L, Sun X, Liu L, Cai Y and Fan H (2021) Combined Analysis of Volatile Terpenoid Metabolism and Transcriptome Reveals Transcription Factors Related to Terpene Synthase in Two Cultivars of Dendrobium officinale Flowers. Front. Genet. 12:661296. doi: 10.3389/fgene.2021.661296

Dendrobium officinale is a kind of traditional Chinese herbal medicine. Its flowers could be used as health care tea for its aroma flavor and medicinal value. Most recent studies demonstrated that terpenoids are the main components of the aromatic compounds in the flowers, but the biosynthesis of terpenoids is poorly understood in *D. officinale*. In the experiment, the flowers from two cultivars of *D. officinale* with different smells were collected. The transcriptome analysis and combined volatile terpenoids determination were performed to identify the genes related to the biosynthesis of the terpenoids. The results showed that the different products of volatile terpenoids are α-thujene, linalool, α-terpineol, α-phellandrene, γ-muurolene, α-patchoulene, and δ-elemene in two cultivar flowers. The transcriptome analysis detected 25,484 genes in the flowers. And 18,650 differentially expressed genes were identified between the two cultivars. Of these genes, 253 genes were mapped to the terpenoid metabolism pathway. Among these genes, 13 terpene synthase (TPS) genes may have correlations with AP2/ERF, WRKY, MYB, bHLH, and bZIP transcription factors by weighted gene co-expression network analysis (WGCNA). The transcription factors have regulatory effects on TPS genes. These results may provide ideas for the terpenoid biosynthesis and regulatory network of *D. officinale* flowers.

Keywords: *Dendrobium officinale*, terpenoid, WGCNA, transcription factors, terpene synthase gene

INTRODUCTION

Terpenoids are a class of highly diverse natural products. There are more than 80,000 known terpenoids, and at least half of them are synthesized by plants (Nagegowda and Gupta, 2020). Terpenoids have diverse biological functions in nature. It plays an essential role in the interaction and mediation of antagonism between organisms (Abbas et al., 2017). It can be used to attract pollinators, defend against ground and underground herbivores, and transmit signals between plants (Raguso, 2008; Ali et al., 2012; Abbas et al., 2017). Otherwise, it can be used as natural flavors

and aroma compounds, which have beneficial effects on human health (Wagner and Elmadfa, 2003). Volatile terpenoids found in almost all plant organs are the main part of volatile compounds in flowers (Das et al., 2013). The common volatile terpenes in orchid flowers are limonene, pinene, myrcene, linalool, and ocimene (Gao et al., 2018; Baek et al., 2019; Robustelli Della Cuna et al., 2019).

Terpenoids are synthesized from two common five-carbon precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (David, 2008). In plants, the mevalonate pathway (MVA pathway) and the methyl-D-erythritol-4-phosphate pathway (MEP pathway) are responsible for forming these C5-isoprene building blocks. The MVA pathway produces volatile sesquiterpenes (C_{15}), while the MEP pathway provides precursors for volatile sesquiterpenes (C_5), monoterpenes (C_{10}), and diterpenes (C_{20}) (Karunanithi and Zerbe, 2019). The MVA pathway consists of six enzymatic reactions. Three molecules of acetyl-CoA are gradually condensed into 3-hydroxy-3-methylglutaryl-CoA and then reduced to MVA, followed by two phosphorylation and decarboxylation reactions. Elimination eventually forms IPP (Lange et al., 2000). The MEP pathway involves seven enzymatic steps. It starts with the condensation of D-glycerol 3-phosphate (GAP) and pyruvate (Pyr) to form 1-deoxy-D-xylulose 5-Phosphoric acid, then undergoes isomerization/reduction to generate MEP.

Terpene synthase (TPS) is a key enzyme for terpene synthesis, which has been identified in *Arabidopsis*, *Malus domestica*, *and Solanum lycopersicum* (Aubourg et al., 2002; Nieuwenhuizen et al., 2013; Zhou and Pichersky, 2020). According to the different products formed, it can be classified as monoterpene synthase, sesquiterpene synthase, diterpene synthase, and so on. They are, respectively, geranyl pyrophosphate (GPP) or neryl pyrophosphate (NPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP) as precursor substrates to synthesize corresponding monoterpenes, sesquiterpenes, and diterpenes (Schilmiller et al., 2009). Besides, a variety of transcription factors (TFs) have been involved in the regulation of terpenoid biosynthesis, such as WRKY, bHLH, MYB, bZIP (Zhang et al., 2015), and AP2/ERF (Suttipanta et al., 2011; Lu et al., 2013; Zhang et al., 2015; Alex et al., 2016; Jiang et al., 2016; Paul et al., 2017; Pu et al., 2018; Majid et al., 2019). For example, *PbbHLH4* can induce high expression level of TPS genes in *Phalaenopsis bellina* (Orchidaceae) (Chuang et al., 2018). *Wintersweet* (*Chimonanthus praecox* L.) *CpMYC2* and *CpbHLH13* TFs may be involved in the positive regulation of biosynthesis of monoterpenes (linalool) and sesquiterpenes (β-caryophyllene) in transgenic plants of Wintersweet (*C. praecox* L.) (Ji et al., 2014). Overexpression of this gene can upregulate the transcript level of terpenoid-related genes in transient transformed *Conyza blinii* leaves (Sun et al., 2018).

Dendrobium officinale Kimura et Migo is a valuable Chinese traditional medicine for its beneficial effects, including anti-tumor, anti-angiogenesis, immune enhancement, anti-oxidation, and alleviating diabetes (Tang et al., 2017; Teixeira da Silva and Ng, 2017). Previous studies have identified 34 TPS genes in *D. officinale*. They were classified into four subfamilies (TPS-a, TPS-b, TPS-c, and TPS-e/f) (Yu et al., 2020). In the experiment, the volatile terpenoids in flowers were identified and quantitatively analyzed for two cultivars of Wanhu No.5 and Wanhu No.6 of *D. officinale*. By using transcriptome sequence and weighted gene co-expression network analysis (WGCNA), the terpene synthesis genes were identified. And the correlation between volatile terpenoids and differential gene expression levels was analyzed. The relationship between TFs AP2/ERF, WRKY, MYB, bHLH, and bZIP and terpenoid metabolism was explored.

MATERIALS AND METHODS

Plant Material

Two cultivars of flowers, Wanhu No.5 and Wanhu No.6, came from the laboratory of Professor Cai Yongping of Anhui Agricultural University. The two cultivars of flowers were sampled at the flowering period in the full bloom period. A total of 12 flower samples were collected. There were three replicates for each of the two types of flowers, for terpenoid metabolic profiling and transcriptome analysis. To analyze the natural volatile compounds, the flowers of Wanhu No.5 and Wanhu No.6 were enclosed and sampled in an extraction bottle. Each cultivar of flowers was sealed into solid-phase microextraction (SPME) vials immediately for further analysis (Sun et al., 2016).

To investigate the spatiotemporal correlation between the TFs related to TPS and the emission of volatile terpene compounds, a range of samples, including two cultivars of flowers, were collected for RNA extraction. All samples were immediately frozen in liquid nitrogen and stored at −80°C until required.

Gas Chromatography-Mass Spectrometry Analysis of Volatile Compounds in Flowers of Wanhu No.5 and Wanhu No.6

Headspace SPME was employed to collect the volatile compounds from flower tissues, which were absorbed by a 75 μm CAR/PDMS fiber (Sigma-Aldrich) for 2 h at 25°C. Total trapped volatile compounds were subsequently thermally desorbed and transferred to an Agilent 5975-6890N gas chromatography-mass spectrometry (GC-MS) apparatus (Agilent Technologies) equipped with HP5-MS quartz capillary column (250 μm diameter, 60 m length, and 0.25 μm film thickness). The instrument used for the gas chromatography–mass spectrometry analysis was an Agilent 7890B-7000B triple quadrupole gas-mass spectrometer. The carrier gas was helium with 1 ml/min of flow rate. The temperature of the electron ionization (EI) ion source is 230°C, and the electron energy was 70 eV. The temperature of the quadrupole is 150°C for 280°C of interface temperature, and the mass scanning range was 50–400 amu. The temperature program was isothermal at 60°C for 3 min, then increased at a rate of 5°C min^{-1} to 300°C, and was then mainteined at 300°C for 5 min.

Through the NIST (National Institute of Standards and Technology) 2011 standard library, the volatile compounds detected during the experiment were identified and qualitatively

FIGURE 1 | Two cultivars of *Dendrobium officinale* flowers. **(A)** Wanhu No.5. **(B)** Wanhu No.6.

analyzed. The obtained compounds were compared with the literature to obtain the determined volatile terpenoids.

Transcriptome Data Analysis

The Illumina NEBNext® Ultra™ RNA Library Prep Kit was used in the library construction. AMPure XP beads were used to screen 200 bp cDNA, and polymerase chain reaction (PCR) amplification was performed. AMPure XP beads were used to purify the PCR products again, and the library was finally obtained. RNA quality was evaluated on a NanoPhotometer® spectrophotometer (IMPLEN, München, Germany). RNA-Seq was performed by LC-bio (Hangzhou, China) on the Illumina HiSeq 4000 platform. Raw reads obtained from RNA-Seq were pre-processed; adapters were trimmed, and low-quality and shorter reads were removed. The *Dendrobium* genome was selected as a reference for Wanhu No.5 and Wanhu No.6. Clean reads were aligned to the *Dendrobium* genome, and the GenBank assembly accession was GCA_001605985.2 (latest). The transcriptome data could be obtained on the National Center for Biotechnology Information (NCBI); the BioProject accession number was PRJNA703321. Q20, Q30, and GC contents of the clean data were calculated. Then the fragments per kilobase million (FPKM) of each gene was calculated based on the length of the gene, and the reads mapped to that gene were calculated (Bray et al., 2016).

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding the advanced functions and utilities of biological systems, such as cells, organisms, and ecosystems, from molecular-level information, especially large-scale molecular data sets generated by genome sequencing and other high-throughput databases (Kanehisa et al., 2017). We used clusterProfiler R software to analyze the statistical enrichment of differentially expressed genes (DEGs) in the KEGG pathway. To map the target genes to metabolic pathways, all sequences of DEGs were uploaded to the Mercator v.3.6[1] to generate a root map file, then it was imported to the Mapman software (V3.6.0 RC1) to obtain the map. In detail, the DEGs in the terpernoids biosynthesis pathway were displayed with the value of log2.Fold_change between Wanhu No.5 and Wanhu No.6.

Gene Co-expression Analysis

We performed WGCNA on all DEGs screened out in transcriptome sequencing (Langfelder and Horvath, 2008). We used the WGCNA package to run the FPKM expression of DEGs in the R software. It had a module with default settings, the power was 6, minModuleSize was 30, and the minimum height of the combined module was 0.25. The gene with the highest connectivity within the module was called the intra-mold hub gene. Networks were visualized by Cytoscape software v 3.6.1 (Shannon et al., 2003).

Quantitative Real-Time PCR (qRT-PCR) Assays

RNAprep pure plant kit (Biofit, Chengdu, China) was used to isolate total RNA from fresh Wanhu No.5 and Wanhu No.6

[1]https://www.plabipd.de/portal/web/guest/mercator-sequence-annotation

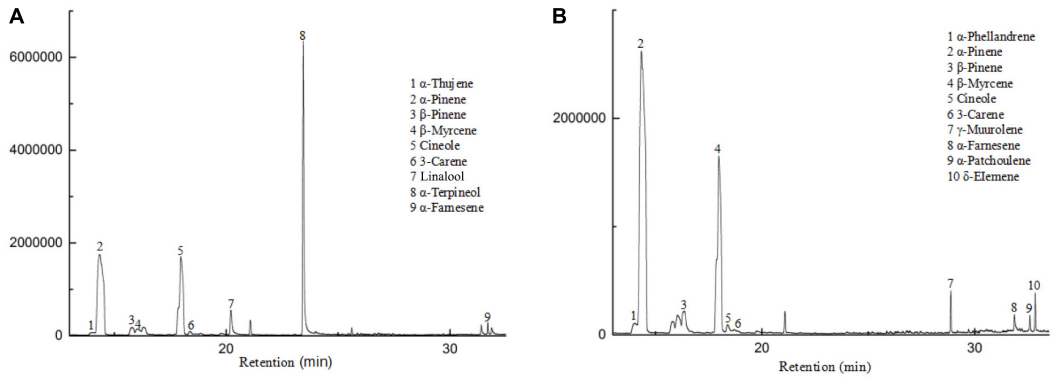


FIGURE 2 | GC-MS analysis of the products formed in two cultivars of *Dendrobium officinale* flowers. **(A)** Wanhu No.5. **(B)** Wanhu No.6.

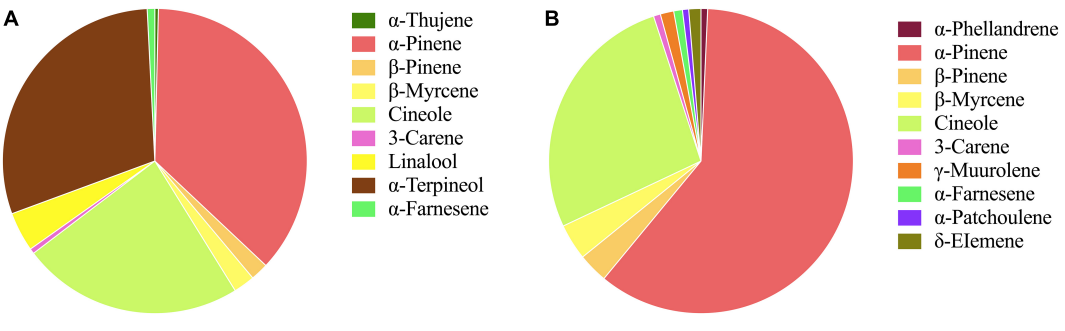


FIGURE 3 | Volatile terpenoids compounds of two cultivars of *Dendrobium officinale* flowers. **(A)** Wanhu No.5. **(B)** Wanhu No.6. The same color represents the same terpenoid.

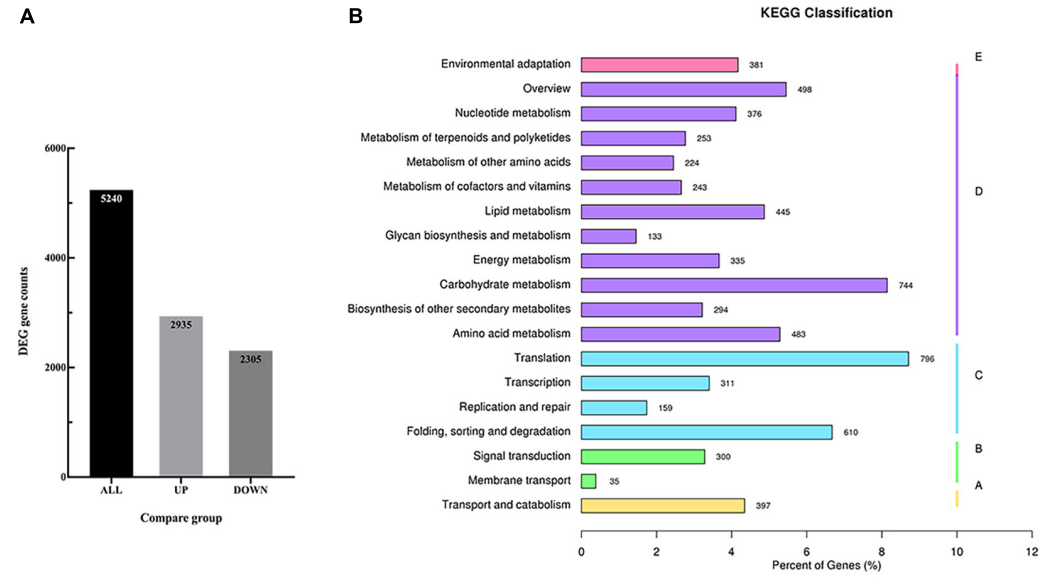


FIGURE 4 | RNA-seq and different expression genes (DEGs) analysis. **(A)** Statistical histogram of the number of different genes in the different comparison combinations. The numbers on the columns indicate the number of differential genes. **(B)** KEGG pathway classification map of DEGs.

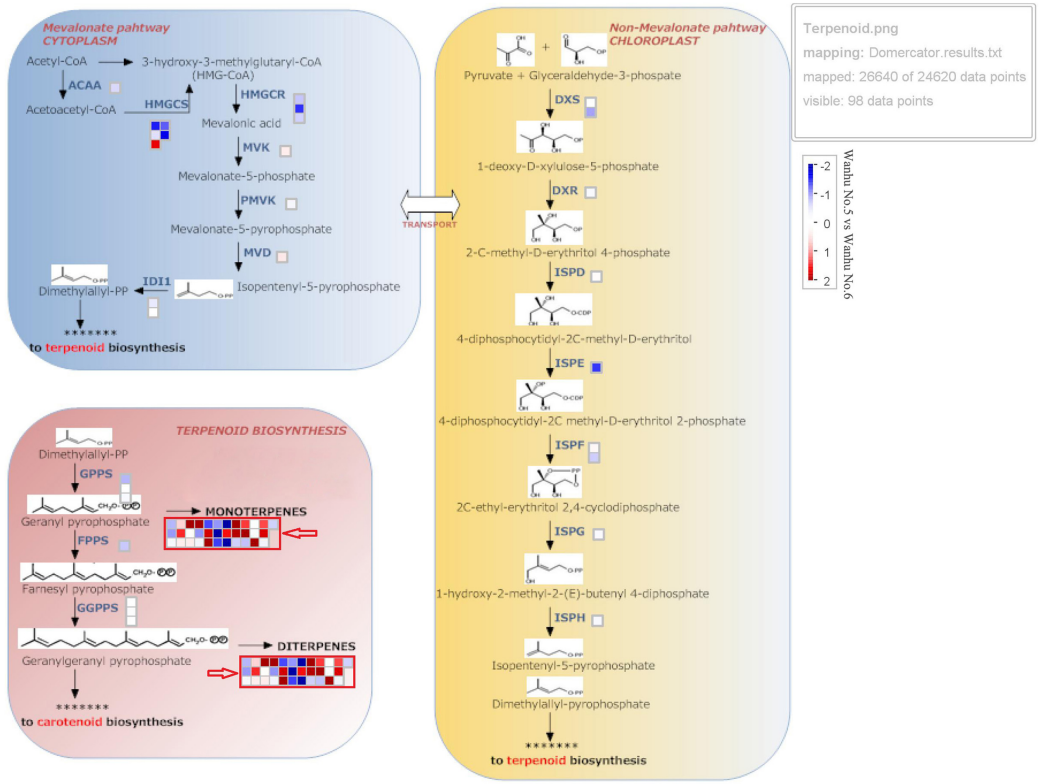


FIGURE 5 | Expression pattern of genes related to terpenoid biosynthesis. Gene expression levels in the fully opened flowers of Wanhu No.5 and Wanhu No.6 are represented by color gradations. AACT, acetyl-CoA acetyltransferase; CMK, 4-(cytidine 59-diphospho)-2-C-methyl-D-erythritol kinase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS, DXP synthase; FPPS, farnesyl diphosphate synthase; GPPS, geranyl diphosphate synthase; HDS, 4-hydroxy-3-methyl-2-butenyl pyrophosphate synthase; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; IDI, isopentenyl diphosphate isomerase; IDS, isoprenyl diphosphate synthase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; ME, ME-CDP synthase; MPDC, mevalonate diphosphate decarboxylase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; TPS, terpene synthase.

flowers (100 mg). According to the manufacturer's instructions, RNA (1 μg) was used to synthesize cDNA using PrimeScript^{TM} RT kit with gDNA eraser (January, Perfect Real Time, Takara, Tokyo, Japan). Using QuantStudio 6 Flex real-time PCR system (Thermo Fisher, Waltham, MA, United States) and SYBR® PremixExTaq^{TM}II (2x) (Japan, Takara), the gene expression level was detected by qRT-PCR. We used NCBI-BLAST online software[2] to design fluorescent quantitative primers for the key genes in the terpene synthesis pathway (**Supplementary Table 1**). The results are attached in **Figure 11**. The reaction steps are 50°C 2 min, 95°C 30 s, 95°C 5 s, 60°C 34 s, 40 cycles, and 72°C for 10 min (Jin et al., 2013). The $2^{-\Delta\Delta CT}$ method was used to calculate the relative gene expression, and the experiment was repeated three times (Livak and Schmittgen, 2001).

Statistical Analysis

Average and standard derivations of chemicals were calculated using Microsoft Excel software. The results of GC-MS were drawn using Origin 6.0 (Origin Lab Corporation, United States). Prism 8 was used for some figures. R software was used to calculate correlation factors. The phylogenetic analysis included the genomes of *D. officinale*. The protein sequences were aligned using the default parameters in MUSCLE[3], and then the neighbor-joining tree was generated by guided analysis (1,000 repeats) using MEGA 7.0 software (Fan et al., 2020).

RESULTS

GC-MS Analysis of Volatile Terpenoids in Flowers of Wanhu No.5 and Wanhu No.6

The flowers from two cultivars, Wanhu No.5 and Wanhu No.6 of *D. officinale*, were collected for GC-MS analysis (**Figure 1**). The results showed that the compositions of volatile compounds in the two cultivars differed from each other. There were 18 volatile compounds detected in flowers of Wanhu No.5, which contained 80% terpenoids and 8.6% alkane compounds. Otherwise, 20 volatile compounds were detected in flowers of Wanhu No.6. The ratios of terpenes and alkanes accounted for 84.57 and 5.51%, respectively. Among these detected volatile terpenoids, a total of 13 volatile terpenoids were detected in the two

[2]https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome

[3]https://www.ebi.ac.uk/Tools/msa/muscle/

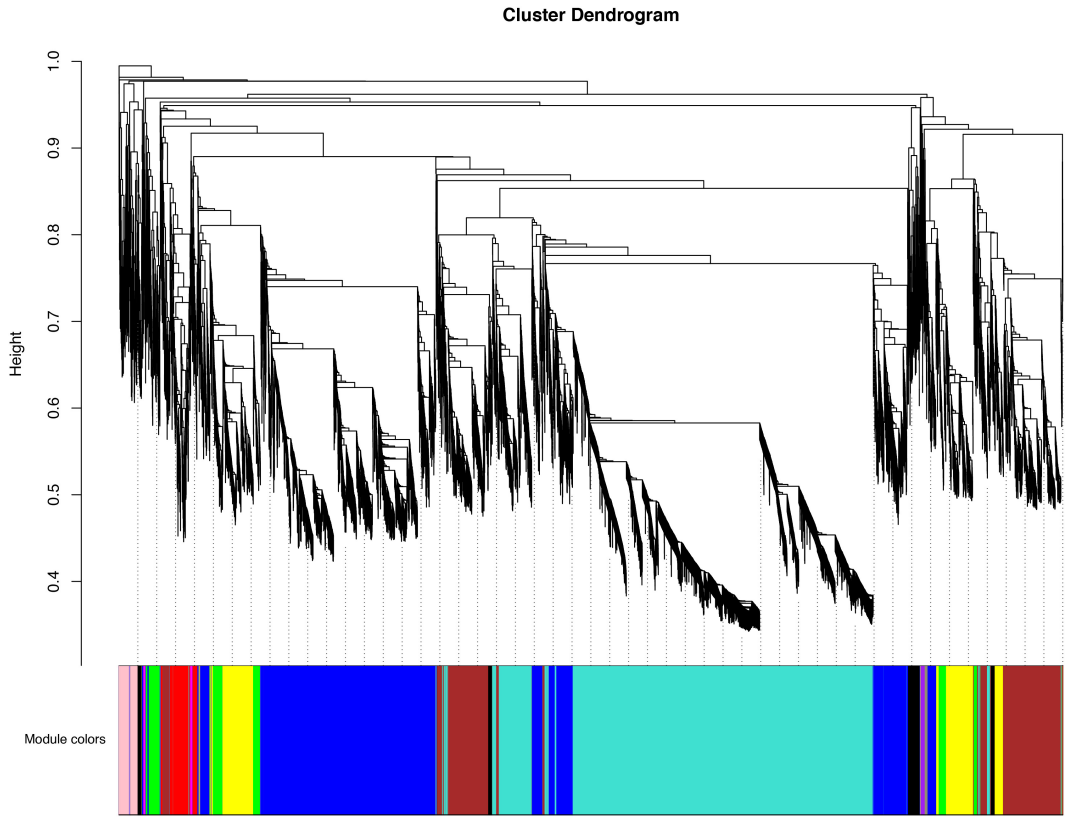


FIGURE 6 | Clustering dendrogram of expressed genes. Gene modules were identified by dynamic hierarchical tree cut and shown in different colors. Height cut = 0.25, minimal module size = 30.

cultivars. Of these components, nine volatile terpenoids were detected in Wanhu No.5, mainly including α-pinene (36.63%), cineole (23.51%), α-terpineol (29.81%), linalool (4.18%), β-myrcene (2.26%), and β-pinene (1.89%) (**Figure 2A**). Ten volatile terpenoids were detected in Wanhu No.6. The main terpenoids were α-pinene (60.29%), cineole (26.96%), β-myrcene (3.83%), β-pinene (3.17%), γ-muurolene (1.43%), and δ-elemene (1.3%) (**Figure 2B**). So volatile terpenoids' content and compositions in the two cultivars, Wanhu No.5 and Wanhu No.6, differed from each other. α-Thujene, linalool, and α-terpineol were only detected in Wanhu No.5, and six volatile terpenoids, such as α-phellandrene, γ-muurolene, α-patchoulene, and δ-elemene, were only detected in Wanhu No.6. The volatile terpenoids content was higher in Wanhu No.6 than in Wanhu No.5 (**Figure 3**).

RNA-Seq and DEGs Analysis

In order to identify the candidate genes related to the volatile terpenoids biosynthesis, RNA-seq was performed using the flowers. As a result, a total of 25,484 genes were detected and mapped to the *D. officinale* genome. Of these genes, 5,240 genes were identified with differentially expressed levels between Wanhu No.5 and Wanhu No.6 (**Figure 4A**). If compared with Wanhu No.5, 2,935 genes were upregulated, and 2,305 genes were downregulated in Wanhu No.6. And KEGG cluster analysis showed 253 unigenes involved in the metabolism of terpenoids and polyketides pathway (**Figure 4B**).

Terpenoids are synthesized through the MVA and methylerythritol phosphate (MEP) pathways, which are independent but complementary to each other. The MEP pathway is mainly responsible for the biosynthesis of monoterpenoids, accounting for about 53% of the total terpenoids in flowers; sesquiterpenes are synthesized through the MVA pathway, accounting for about 28% of the total terpenoids in flowers. **Figure 5** shows that TPS in *D. officinale* Wanhu No.5 was quite different from that in Wanhu No.6, and the number of upregulated genes was greater than that of downregulated genes. The squares in red frames marked with red arrows indicate the differentially expressed TPS genes between Wanhu No.6 and Wanhu No.5. The red squares reveal the upregulated genes in Wanhu No.5 compared with those in Wanhu No.6. The blue frames showed the downregulated genes in Wanhu No.5 in contrast with Wanhu No.6. In the MEP pathway, two genes, HMGR and HMGS, were expressed differently in Wanhu No.5 and Wanhu No.6, which was also the main reason for the different terpenes produced by the two flowers.

Analysis of Gene Correlation Network

In order to obtain the comprehensive transcriptome changes of two cultivars of *D. officinale* flowers, we established a weighted

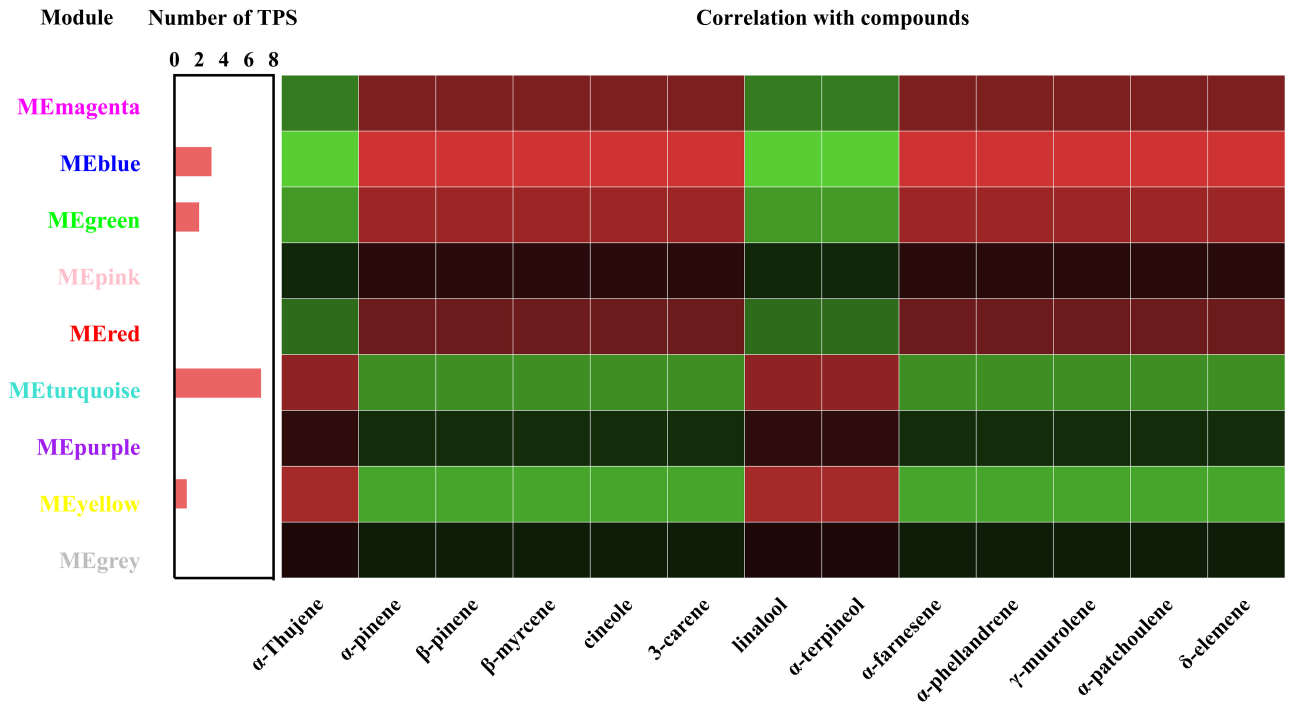


FIGURE 7 | Correlation coefficients between terpene synthase (TPS), module, and volatile compound. Each row represents a module, and each column represents a terpenoid. The color of each block at the row–column intersection indicates the correlation coefficient: red for high positive correlation and green for high negative correlation, with a scale shown on the right of the panel.

gene co-expression network to classify 18650 DEGs. Genes that were more closely related to each other would be gathered in the same module, and finally 10 modules would be obtained (**Figure 6**). The biggest module was the turquoise module, which contains 6,047 genes. However, the smallest module contained 6 genes. The expression patterns of the different modules of the two cultivars of flowers were also quite different from each other. For Wanhu No.5, the turquoise module's gene expression was higher, while for Wanhu No.6 the gene expression of the blue module was higher (**Supplementary Figure 1**).

To understand the relationship between terpenoids and genes, a relationship module diagram between genes and volatile terpenoids was analyzed. As was shown in **Figure 7**, the correlations between different modules and volatile terpenoids were relatively close, except for thujene, linalool, and terpineol, which were the different metabolites in Wanhu No.5 and Wanhu No.6. After screening, we found that the module contains 13 TPS genes, including 7 TPS genes in the turquoise module, 3 TPS genes in the blue module, 2 TPS genes in the green module, and 1 TPS gene in the yellow module. Different TPS genes might be closely related to the production of these volatile terpenoids through the relationship diagram.

To further clarify TPS genes's potential roles in the turquoise module and the blue module, we generated a phylogenetic tree by neighbor-joining method (Yu et al., 2020). **Figure 8** showed that *DoTPSs* proteins were classified into four different clades, 14 in TPS-a, 16 in TPS-b, 3 in TPS-c, and 1 in TPS-e/f. The function of TPS-a family was to synthesize sesquiterpene synthase. TPS-b mainly synthesized monoterpene synthase and isoprene synthase. The function of TPS-c family was to synthesize the bifunctional class I/II (terephthaloyl diphosphate synthase/kaurene) involved in secondary metabolism, and the monofunctional class II included diterpene synthase (terephthaloyl diphosphate synthase enzyme) and diterpene synthase. TPS-e/f was a monofunctional class I diterpene synthase, diterpene synthase, sesquiterpene synthase, and monoterpene synthase involved in secondary metabolism (Alicandri et al., 2020). The phylogenetic tree showed that *DoTPS* in the blue module, including *DoTPS04*, *DoTPS10*, and *DoTPS07*, were located in TPS-a, TPS-b, and TPS-c; *DoTPS* in the turquoise module, including *DoTPS13*, *DoTPS03*, *DoTPS05*, *DoTPS06*, *DoTPS11*, *DoTPS15*, and *DoTPS18*, were located in TPS-a and TPS-b.

TFs Regulating Terpenoid Synthases in Two *D. officinale* Cultivars

As listed in **Supplementary Tables 3–8**, the gene IDs and relative transcript levels of five TF families were obtained from the transcriptomes data of the two cultivars of *D. officinale* flowers. As is shown in **Supplementary Figures 2–6**, the expression patterns of TFs and *DoTPSs* in Wanhu No.5 were different from those in Wanhu No.6. The relative transcript levels were higher in Wanhu No.5. Also, the differences between Wanhu No.5 and Wanhu No.6 were significant in **Supplementary Figures 3–8**. And 10 terpenoid synthases and their read counts were collected in the transcriptome. When compared with Wanhu No.6, six TPS genes

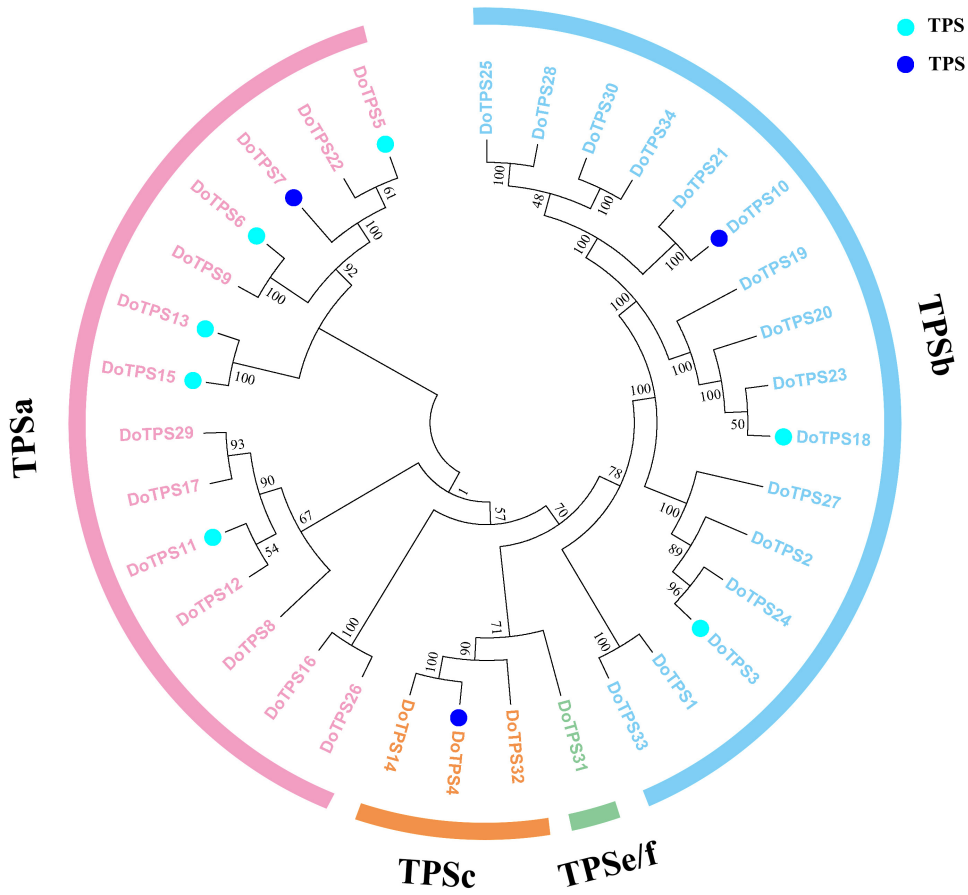


FIGURE 8 | Phylogeny tree analysis of *Dendrobium officinale* TPSs. Triangles represent finger citron TPSs, and filled colors indicate their location in modules shown in **Figure 6**. Sequence alignment was performed by ClustalX. Phylogeny tree was visualized by MEGA7.

were upregulated with more than 2-fold (>2-fold), and eight TPS genes were downregulated (<−2-fold).

In order to understand the relationship between TPS genes and TFs, 12 WRKY, 12 bHLH, 9 MYB, 9 bZIP, and 14 AP2/ERF were selected for the following analysis. Seven TPS genes, *DoTPS13*, *DoTPS03*, *DoTPS05*, *DoTPS06*, *DoTPS11*, *DoTPS15*, and *DoTPS18*, were used based on the WGCNA analysis outcome (**Figure 9**). As a result, WRKY (*DoWRKY02*, DoWRKY01, *DoWRKY36*, *DoWRKY40*, *DoWRKY23*), bHLH (*DobHLH06*, *DobHLH10*, *DobHLH13*, *DobHLH14*, *DobHLH34*, *DobHLH35*), and AP2/ERF (*DoAP2/ERF19*, *DoAP2/ERF19*, *DoAP2/ERF68*, *DoAP2/ERF72*, *DoAP2/ERF86*, *DoAP2/ERF118*) genes were predicted to be closely related to TPS genes. TFs *DoWRKY05*, *DoWRKY61*, *DoMYB07*, *DoMYB18*, *DobHLH42*, *DobZIP39*, *DobZIP56*, *DobZIP21*, and *DoAP2/ERF90* were probably involved in the regulation of *DoTPS04*, *DoTPS10*, and *DoTPS07* (**Figure 10**).

Verification of Gene Expression

In order to verify the transcriptome data, terpenoid synthesis pathway genes and related TFs were selected for real-time PCR analysis (**Figure 11**). The results showed that the expression patterns of the selected genes were consistent with the transcriptome data.

DISCUSSION

The flowers of *D. officinale* not only have a certain ornamental value but are also a kind of Chinese medicine that can be used to make tea with a certain anti-cancer effect (Lai, 2020). In *Dendrobium chrysanthum*, some terpenes including α-phellandrene, α-pinene, α-thujene, L- β-myrcene, α-terpinene, O-cymene, D-limoene, β-ocimene, and carene were detected in the flowers. The volatile components of *Dendrobium lohohense* flowers are mainly esters, and the aroma composition of *Dendrobium densiflorum* is mainly alkanes. The volatile components of *Dendrobium hancockii* and *D. officinale* are mainly terpenes (Li et al., 2015; Lv et al., 2016). In the experiment, 18 and 20 volatile compounds were detected from Wanhu No.5 and Wanhu No.6 flowers, respectively. There were 9 volatile terpenoids compounds detected in Wanhu No.5, and 10 in Wanhu No.6. α-Pinene is the most abundant compound in the flowers of the two cultivars. The volatile terpenoids are quite different in the two cultivars

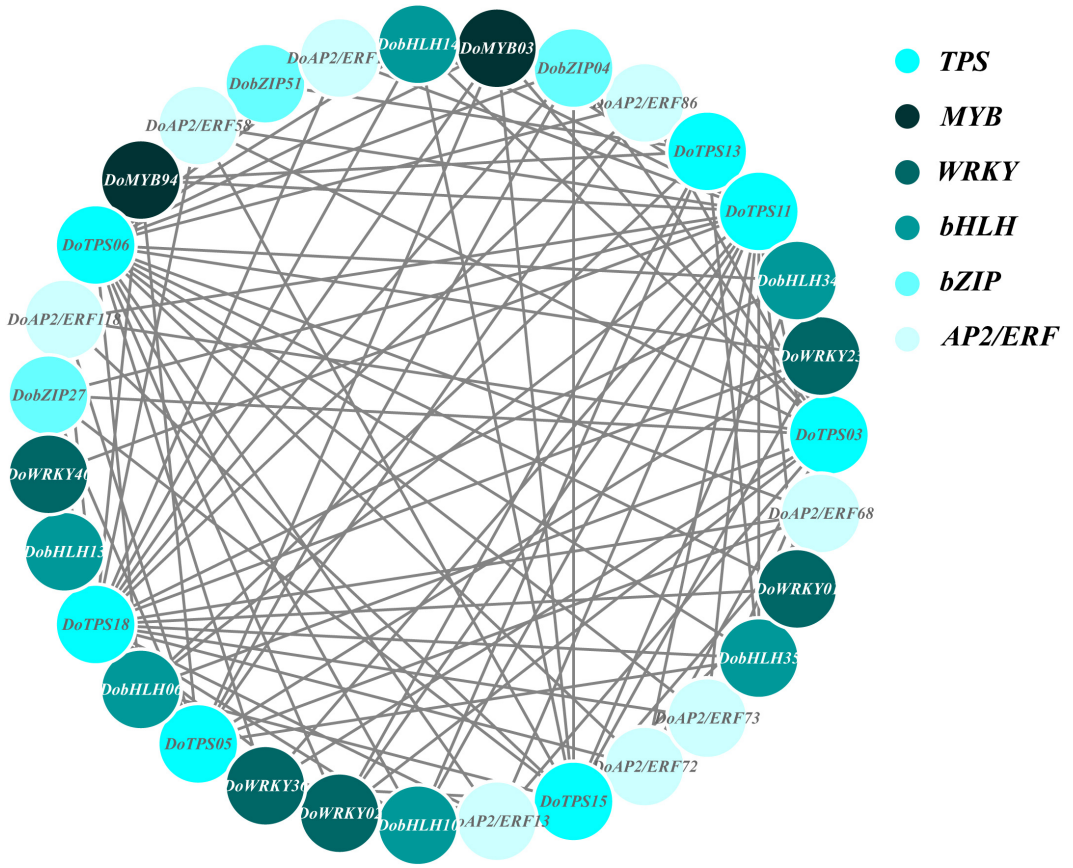


FIGURE 9 | Gene co-expression subnetwork of the turquoise module. Network was reconstructed by edge weight cutoff = 0.25 and visualized by Cytoscape.

of *D. officinale*. For example, α-thujene, linalool, and α-terpineol were only detected in Wanhu No.5, while α-phellandrene, γ-muurolene, α-patchoulene, and δ-elemene were only found in Wanhu No.6.

The comprehensive analysis of gene co-expression and terpenoid accumulation has recently provided new insights into the regulation of terpenoid metabolism (Tai et al., 2018). In order to understand the regulation of terpenoid biosynthesis in *D. officinale*, the expressed genes detected by RNA-Seq were connected by using the method of WGCNA, which provides a network of nodes (genes) and edges (connections). This method is mainly about obtaining connections in the network based on gene co-expression data. This strategy has been used to discover potential target genes and TFs in plants (Ferreira et al., 2016). We obtained 10 different expression modules after WGCNA analysis of the transcriptome data. The results can provide a way to build a network of mining potential target genes and TFs in plants. In *D. officinale*, with 34 TPS genes, 13 TPS genes were obtained in WGCNA analysis. The function of *DoTPS10* has been verified; located in chloroplasts, *DoTPS10* uniquely converted geranyl diphosphate to linalool *in vitro* (Zhao et al., 2020).

In the present study, the top two modules enriched for TPS genes were selected for further analysis. Based on the WGCNA analysis, we obtained 13 TPS genes and 5 kinds of TFs that are related to the synthesis of TPS: AP2/ERF, bHLH, MYB, WRKY, and bZIP (Lu et al., 2013; Zhang et al., 2015; Pu et al., 2018; Majid et al., 2019). Through the correlation network diagram, we found that different TFs regulate different TPS differently. Transient expression of *AaERF1* and *AaERF2* can increase the transcription of amorpha-4,11-diene synthase (ADS) and *CYP71AV1* and increase accumulation of artemisinin and artemisinic acid (Sun et al., 2018). The *Arabidopsis MYC2* TF could bind to the promoter regions of the *TPS21* and *TPS11* that catalyzed sesquiterpenes' formation to activate their expression, thereby increasing the release of sesquiterpenes (Hong et al., 2012). A peltate glandular trichomes (PGT)-specific R2R3-MYB gene, *MsMYB*, was identified in the RNA-Seq comparison data in spearmint. The analysis of the transgenic lines showed increased levels of monoterpenes. In contrast, the levels of *MsMYB* overexpression lines decreased (Reddy et al., 2017). In *Catharanthus roseus*, overexpression of *Cr-WRKY1* could downregulate the expression levels of *ORCA2/3*, *CrMYC2*, and zinc-finger *C. roseus* transcription factors (ZCTs) to regulate the synthesis of monoterpenes (Suttipanta et al., 2011). *AaAPK1* interacted with *AabZIP1* in *Artemisia annua*, and *AaAPK1* enhanced the transactivation activity of *AabZIP1* on artemisinin biosynthesis genes through phosphorylation (Zhang et al., 2018). Regarding the TPS genes identified based on WGCNA, there are eight in the TPS-a family, four in the TPS-b family, and one in the TPS-c family. TPS-a members are

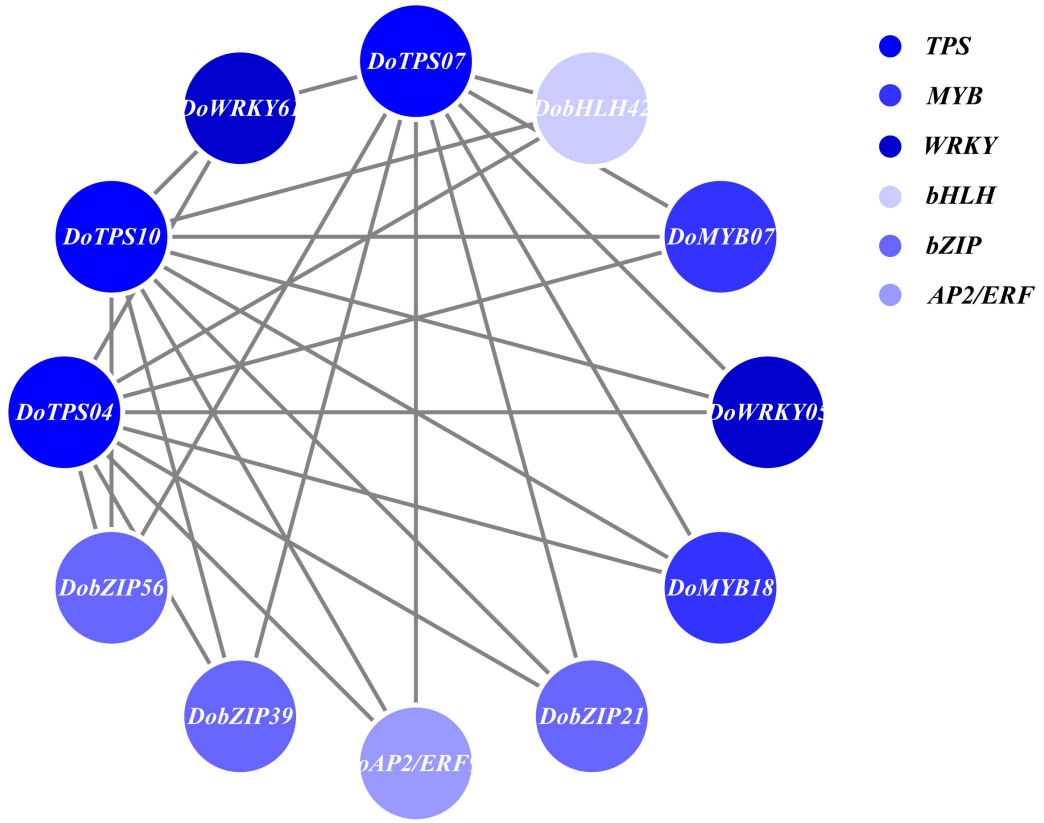


FIGURE 10 | Gene co-expression subnetwork of the blue module. Network was reconstructed by edge weight cutoff = 0.25 and visualized by Cytoscape.

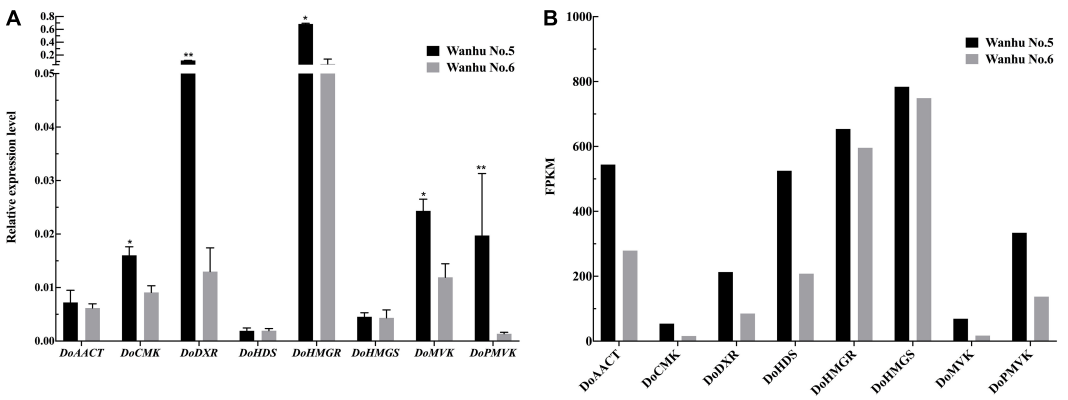


FIGURE 11 | Expression patterns of eight genes as verified by qRT-PCR. **(A)** The relative expression levels of Wanhu No.5 and Wanhu No.6. **(B)** FPKM values from RNA-Seq. Values shown are mean ± SE of three replicates. "*" indicates that the difference is significant, "**" indicates that the difference is very significant.

related to sesquiterpene formation, and TPS-b is related to monoterpene biosynthesis (Alicandri et al., 2020). Therefore, these identified TPSs might play an important role in producing volatile terpenoids in *D. officinale*. As the last enzymatic step of the MVA and MEP pathways, TPS is responsible for the direct synthesis of terpenoids (Schilmiller et al., 2009). However, in most cases, the expression level of these TPS genes has no linear relationship with their product content. There are two main reasons for uncertainty (Degenhardt et al., 2009). Firstly, a considerable amount of TPS was a multi-product enzyme that can produce multiple volatiles from a single substrate. Secondly, the replication and evolution of the TPS family produced isozymes that express different functions in time and space (Takehiko et al., 2014). There were too few studies on the TPS gene of *D. officinale*, so more experiments are needed to test the functions of these 13 *DoTPS*, such as

overexpression in *Escherichia coli* for *in vitro* enzymatic analysis and stable transformation for *in vivo* functional analysis (Zhou and Pichersky, 2020). The regulation between *DoTPS* and TFs will be detected in more experiments.

CONCLUSION

The transcriptome analysis in the two cultivars of *D. officinale* with differences in volatile terpenoid products was performed in order to mine the biosynthetic pathway related genes and regulatory mechanisms of the terpenoid metabolites of *D. officinale*. In the analysis of the two cultivars of *D. officinale* transcriptomes, the expression of upstream genes in the MVA and MEP pathways did not change much, and the TPS genes were quite different. Therefore, the diversity of terpenoids was caused by the differential expression of TPS. We obtained 10 gene modules from WGCNA. From the gene module, we screened 13 TPS genes and AP2/ERF, WRKY, MYB, bHLH, and bZIP TFs, analyzed the correlation between these TFs and TPS expression, and found that these TFs were displayed in the position of the correlation network. They played a role in regulating terpenoid metabolism. Future work should focus on the direct and indirect interactions between TPS and related TFs to clarify the functional network that controls terpene production. These results might provide ideas for the terpenoid biosynthesis and regulatory network of *D. officinale* flowers.

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, https://www.ncbi.nlm.nih.gov/bioproject/, PRJNA703321.

AUTHOR CONTRIBUTIONS

HF designed the study. NL wrote the manuscript. NL, YD, and ML performed the experiments. XS, LL, and LQ helped in data analysis and manuscript preparation. YC revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Anhui Natural Science Foundation (1908085MC59), Anhui Provincial Engineering Technology Research Center for Development and Utilization of Regional Characteristic Plants, The University Synergy Innovation Program of Anhui Province (GXXT-2019-043), and Innovation Project for College Students in Anhui Province (S202010364173).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.661296/full#supplementary-material

Supplementary Figure 1 | Functional annotation of unigenes from *D. officinale* flower transcriptome against databases. **(A)** Venn diagram showing the number of unigenes annotated in two cultivars of *D. officinale* flowers. **(B)** The cluster of transcriptome gene Wanhu No.5 and Wanhu No.6.

Supplementary Figure 2 | Hierarchical cluster tree showing 10 modules of co-expressed genes.

Supplementary Figure 3 | Heatmap about *DoTPS* of two cultivars of *D. officinale* flower.

Supplementary Figure 4 | Heatmap about *DobHLH* of two cultivars of *D. officinale* flower.

Supplementary Figure 5 | Heatmap about *DobZIP* of two cultivars of *D. officinale* flower.

Supplementary Figure 6 | Heatmap about *DoAP2/ERF* of two cultivars of *D. officinale* flower.

Supplementary Figure 7 | Heatmap about *DoWRKY* of two cultivars of *D. officinale* flower.

Supplementary Figure 8 | Heatmap about *DoMYB* of two cultivars of D. officinale flower.

Supplementary Table 1 | Primers used for RT-qPCR analysis.

Supplementary Table 2 | Gene IDs and relative transcript levels of *DobZIP* of two cultivals of *D. officinale* flower.

Supplementary Table 3 | Gene IDs and relative transcript levels of *DoMYB* of two cultivals of *D. officinale* flower.

Supplementary Table 4 | Gene IDs and relative transcript levels of *DoAP2/ERF* of two cultivals of *D. officinale* flower.

Supplementary Table 5 | Gene IDs and relative transcript levels of *DoWRKY* of two cultivals of *D. officinale* flower.

Supplementary Table 6 | Gene IDs and relative transcript levels of *DobHLH* of two cultivals of *D. officinale* flower.

REFERENCES

Abbas, F., Ke, Y., Yu, R., Yue, Y., Amanullah, S., Jahangir, M. M., et al. (2017). Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. *Planta* 246, 803–816.

Alex, V. M., Priscille, S., Ivo, G., Javiera, E., Fabian, S., Karel, M., et al. (2016). The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant J.* 88, 3–12.

Ali, J. G., Alborn, H. T., Campos-Herrera, R., Kaplan, F., Duncan, L. W., Rodriguez-Saona, C., et al. (2012). Subterranean, herbivore-induced plant volatile increases biological control activity of multiple beneficial nematode species in distinct habitats. *PLoS One* 7:e38146. doi: 10.1371/journal.pone.0038146

Alicandri, E., Paolacci, A. R., Osadolor, S., Sorgona, A., Badiani, M., and Ciaffi, M. (2020). On the evolution and functional diversity of terpene synthases in the *Pinus* species: a review. *J. Mol. Evol.* 88, 253–283. doi: 10.1007/s00239-020-09930-8

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267, 730–745. doi: 10.1007/s00438-002-0709-y

Baek, Y. S., Ramya, M., An, H. R., Park, P. M., Lee, S. Y., Baek, N. I., et al. (2019). Volatiles Profile of the floral organs of a new hybrid *Cymbidium*, 'Sunny Bell' using headspace solid-phase microextraction gas chromatography-mass spectrometry analysis. *Plants* 8:251. doi: 10.3390/plants8080251

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Erratum: near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34:888.

Chuang, Y. C., Hung, Y. C., Tsai, W. C., Chen, W. H., and Chen, H. H. (2018). PbbHLH4 regulates floral monoterpene biosynthesis in *Phalaenopsis orchids*. *J. Exp. Bot.* 69, 4363–4377. doi: 10.1093/jxb/ery246

Das, A., Lee, S. H., Hyun, T. K., Kim, S. W., and Kim, J. Y. (2013). Plant volatiles as method of communication. *Plant Biotechnol. Rep.* 7, 9–26.

David, W. C. (2008). Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* 12, 141–150.

Degenhardt, J., Köllner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637.

Fan, H., Cui, M., Li, N., Li, X., Liang, Y., Liu, L., et al. (2020). Genome-wide identification and expression analyses of R2R3-MYB transcription factor genes from two orchid species. *PeerJ* 8:e9781. doi: 10.7717/peerj.9781

Ferreira, S. S., Hotta, C. T., Poelking, V. G., Leite, D. C., Buckeridge, M. S., Loureiro, M. E., et al. (2016). Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant. Mol. Biol.* 91, 15–35. doi: 10.1007/s11103-016-0434-2

Gao, F., Liu, B., Li, M., Gao, X., Fang, Q., Liu, C., et al. (2018). Identification and characterization of terpene synthase genes accounting for volatile terpene emissions in flowers of *Freesia* x hybrida. *J. Exp. Bot.* 69, 4249–4265. doi: 10.1093/jxb/ery224

Hong, G. J., Xue, X. Y., Mao, Y. B., Wang, L. J., and Chen, X. Y. (2012). Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *Plant Cell* 24, 2635–2648. doi: 10.1105/tpc.112.098749

Ji, Y., Xiao, J., Shen, Y., Ma, D., Li, Z., Pu, G., et al. (2014). Cloning and characterization of AabHLH1, a bHLH transcription factor that positively regulates artemisinin biosynthesis in *Artemisia annua*. *Plant Cell Physiol.* 55, 1592–1604. doi: 10.1093/pcp/pcu090

Jiang, W., Fu, X., Pan, Q., Tang, Y., Shen, Q., Lv, Z., et al. (2016). Overexpression of AaWRKY1 leads to an enhanced content of artemisinin in *Artemisia annua*. *Biomed. Res. Int.* 2016:7314971. doi: 10.1155/2016/7314971

Jin, Q., Yao, Y., Cai, Y., and Lin, Y. (2013). Molecular cloning and sequence analysis of a phenylalanine ammonia-lyase gene from *Dendrobium*. *PLoS One* 8:e62352. doi: 10.1371/journal.pone.0062352

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Karunanithi, P. S., and Zerbe, P. (2019). Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front. Plant Sci.* 10:1166. doi: 10.3389/fpls.2019.01166

Lange, B. M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci.* 97, 13172–13177. doi: 10.1073/pnas.240454797

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

Lai, Y. (2020). Study on the efficacy of *Dendrobium* flower. *Technol. Wind* 18:277. doi: 10.19392/j.cnki.1671-7341.202018211

Li, C., Mingzhong, H., Shaohua, H., and Junmei, Y. (2015). Volatile components in flowers of four *Dendrobium* species. *J. Trop. Subtrop. Bot.* 4, 454–462.

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

Lu, X., Zhang, L., Zhang, F., Jiang, W., Shen, Q., Zhang, L., et al. (2013). AaORA, a trichome-specific AP2/ERF transcription factor of *Artemisia annua*, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance toBotrytis cinerea. *New Phytol.* 198, 1191–1202.

Lv, S., Meng, X., Xinfeng, Z., Jinjing, L., and Jinping, S. (2016). Studies on volatile constituents of 11 families of *Dendrobium* officinale flowers. *Chin. J. Exp. Tradit. Med. Formula.* 6, 52–57. doi: 10.13422/j.cnki.syfjx.201606 0052

Majid, I., Kumar, A., and Abbas, N. (2019). A basic helix loop helix transcription factor, AaMYC2-Like positively regulates artemisinin biosynthesis in *Artemisia annua* L. *Ind. Crops Prod.* 128, 115–125.

Nagegowda, D. A., and Gupta, P. (2020). Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Sci.* 294: 110457.

Nieuwenhuizen, N. J., Green, S. A., Chen, X., Bailleul, E. J., Matich, A. J., Wang, M. Y., et al. (2013). Functional genomics reveals that a compact terpene synthase gene family can account for terpene volatile production in apple. *Plant Physiol.* 161, 787–804. doi: 10.1104/pp.112.208249

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L. (2017). A differentially regulated AP 2/ ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytol.* 213, 1107–1123.

Pu, S., Xueqing, F., Qian, S., Meng, L., Qifang, P., Yueli, T., et al. (2018). The roles of AaMIXTA1 in regulating the initiation of glandular trichomes and cuticle biosynthesis in *Artemisia annua*. *New phytol.* 217, 261–276.

Raguso, R. A. (2008). Wake up and smell the roses: the ecology and evolution of floral scent. *Annu. Rev. Ecol. Evol. Syst.* 39, 549–569. doi: 10.1146/annurev.ecolsys.38.091206.095601

Reddy, V. A., Wang, Q., Dhar, N., Kumar, N., Venkatesh, P. N., Rajan, C., et al. (2017). Spearmint R2R3-MYB transcription factor MsMYB negatively regulates monoterpene production and suppresses the expression of geranyl diphosphate synthase large subunit (MsGPPS.LSU). *Plant Biotechnol. J.* 15, 1105–1119. doi: 10.1111/pbi.12701

Robustelli Della Cuna, F. S., Calevo, J., Bari, E., Giovannini, A., Boselli, C., and Tava, A. (2019). Characterization and antioxidant activity of essential oil of four sympatric orchid species. *Molecules* 24:3878. doi: 10.3390/molecules24213878

Schilmiller, A. L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A. L., Schmidt, A., et al. (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10865–10870. doi: 10.1073/pnas.0904113106

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sun, W., Liang, L., Meng, X., Li, Y., Gao, F., Liu, X., et al. (2016). Biochemical and molecular characterization of a flavonoid 3-O-glycosyltransferase responsible for anthocyanins and flavonols biosynthesis in *Freesia* hybrida. *Front. Plant Sci.* 7:410. doi: 10.3389/fpls.2016.00410

Sun, W.-J., Zhan, J.-Y., Zheng, T.-R., Sun, R., Wang, T., Tang, Z.-Z., et al. (2018). The jasmonate-responsive transcription factor CbWRKY24 regulates terpenoid biosynthetic genes to promote saponin biosynthesis in *Conyza blinii* H. *Lév. J. Genet.* 97, 1379–1388. doi: 10.1007/s12041-018-1026-5

Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S. K., and Yuan, L. (2011). The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* 157, 2081–2093. doi: 10.1104/pp.111.181834

Tai, Y., Liu, C., Yu, S., Yang, H., Sun, J., Guo, C., et al. (2018). Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (*Camellia sinensis*). *BMC Genomics* 19:616. doi: 10.1186/s12864-018-4999-9

Takehiko, S., Tomoko, E., Hiroshi, F., Ana, R., Leandro, P., and Mitsuo, O. (2014). Characterization of three linalool synthase genes from *Citrus unshiu* marc. and analysis of linalool-mediated resistance against *Xanthomonas citri* subsp. citri and *Penicilium italicum* in citrus leaves and fruits. *Plant Sci.* 229, 154–166.

Tang, H., Zhao, T., Sheng, Y., Zheng, T., Fu, L., and Zhang, Y. (2017). *Dendrobium officinale* Kimura et Migo: a review on its ethnopharmacology, phytochemistry, pharmacology, and industrialization. *Evid. Based Complement. Alternat. Med.* 2017:7436259. doi: 10.1155/2017/7436259

Teixeira da Silva, J. A., and Ng, T. B. (2017). The medicinal and pharmaceutical importance of *Dendrobium* species. *Appl. Microbiol. Biotechnol.* 101, 2227–2239. doi: 10.1007/s00253-017-8169-9

Wagner, W. H., and Elmadfa, I. (2003). Biological relevance of terpenoids. *Ann. Nutr. Metab.* 47, 95–106.

Yu, Z., Zhao, C., Zhang, G., Teixeira da Silva, J. A., and Duan, J. (2020). Genome-wide identification and expression profile of TPS gene family in *Dendrobium officinale* and the role of DoTPS10 in linalool biosynthesis. *Int. J. Mol. Sci.* 21:5419. doi: 10.3390/ijms21155419

Zhang, F., Fu, X., Lv, Z., Lu, X., Shen, Q., Zhang, L., et al. (2015). A basic leucine zipper transcription factor, AabZIP1, connects abscisic acid signaling with artemisinin biosynthesis in *Artemisia annua*. *Mol. Plant* 8, 163–175. doi: 10.1016/j.molp.2014.12.004

Zhang, F., Xiang, L., Yu, Q., Zhang, H., Zhang, T., Zeng, J., et al. (2018). Artemisinin biosynthesis promoting kinase 1 positively regulates artemisinin biosynthesis through phosphorylating AabZIP1. *J. Exp. Bot.* 69, 1109–1123. doi: 10.1093/jxb/erx444

Zhao, C., Yu, Z., Silva, J., He, C., Wang, H., Si, C., et al. (2020). Functional characterization of a *Dendrobium officinale* geraniol synthase DoGES1 involved in floral scent formation. *Int. J. Mol. Sci.* 21:7005. doi: 10.3390/ijms21197005

Zhou, F., and Pichersky, E. (2020). The complete functional characterisation of the terpene synthase family in tomato. *New Phytol.* 226, 1341–1360. doi: 10.1111/nph.16431

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 April 2021
doi: 10.3389/fgene.2021.670794

Bioinformatics and Expression Analysis of IDA-Like Genes Reveal Their Potential Functions in Flower Abscission and Stress Response in Tobacco (*Nicotiana tabacum* L.)

Cun Guo[1,2†], Qi Wang[1,2†], Zhiyuan Li[1,2], Jinhao Sun[1,2], Zenglin Zhang[1], Xiaoxu Li[1,3] and Yongfeng Guo[1*]*

[1] Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, China, [2] Graduate School of Chinese Academy of Agricultural Sciences, Beijing, China, [3] Technology Center, China Tobacco Hunan Industrial Co., Ltd., Changsha, China

OPEN ACCESS

Edited by:
Yunpeng Cao,
Central South University Forestry and Technology, China

Reviewed by:
Feng Yu,
Hunan University, China
Ethan Andersen,
Francis Marion University, United States
Zhi-yong Ni,
Xinjiang Agricultural University, China

***Correspondence:**
Xiaoxu Li
82101171073@caas.cn
Yongfeng Guo
guoyongfeng@caas.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 22 February 2021
Accepted: 31 March 2021
Published: 27 April 2021

Citation:
Guo C, Wang Q, Li Z, Sun J, Zhang Z, Li X and Guo Y (2021) Bioinformatics and Expression Analysis of IDA-Like Genes Reveal Their Potential Functions in Flower Abscission and Stress Response in Tobacco (Nicotiana tabacum L.). Front. Genet. 12:670794. doi: 10.3389/fgene.2021.670794

The inflorescence deficient in abscission-like (*IDL*) genes have been shown to play critical roles in floral organ abscission, lateral root formation and various stress responses in *Arabidopsis*. The *IDL* gene family has been characterized in a number of plant species, while limited information is available about *IDL* genes of tobacco. In the current study, 15 NtIDL members were identified in the tobacco genome, and were classified into six groups together with IDL members from other species. Evolution analysis suggested that the NtIDL members form group VI might have originated from duplication events. Notably, NtIDL06 shared high similarities with AtIDA in the EPIP sequence, and its encoding gene was highly expressed in the abscission zone of flowers at late developmental stages, implying that NtIDL06 might regulate tobacco flower abscission. In addition, the results from *cis*-elements analysis of promoters and expression after stress treatments suggested that *NtIDL* members might be involved in various stress responses of tobacco. The results from this study provide information for further functional analysis related to flower abscission and stress responses of *NtIDL* genes.

Keywords: flower abscission, IDA peptide, IDL, tobacco, *cis*-element, abiotic stresses

INTRODUCTION

Abscission is a highly coordinated cell separation process in plants. From an evolutionary perspective, active abscission is advantageous in many aspects for plants, such as dispersal, propagation, pollination and defense (Lewis et al., 2006). It allows parent plants to abandon damaged organs no longer needed (Reichardt et al., 2020). A prerequisite for abscission to transpire is the presence of an abscission zone, which is composed of small densely cytoplasmic cells that respond to abscission signals (Patterson, 2001; Liljegren, 2012; Gubert et al., 2014). Also, abscission can be triggered by multiple factors, including seasonal changes, pathogen attack, abiotic stresses, and hormones (Patharkar and Walker, 2019). Abscisic acid (ABA) and Methyl Jasmonate (MeJA) were reported to accelerate organ abscission, while auxin and brassinosteroids were negative regulators of shedding (Hartmond et al., 2000; Chandler, 2011; Marciniak et al., 2018; Wilmowicz et al., 2018; Mesejo et al., 2021).

In *Arabidopsis*, flower organ abscission is dependent on the function of a small peptide that is released from the IDA (inflorescence deficient in abscission) precursor protein (Reichardt et al., 2020). The IDA proprotein is composed of 77 amino acids including an N-terminal signal peptide and a C-terminal EPIP domain (FGYLPKGVPIPPSA PSKRHNSFVNSLPH). The EPIP domain (extended PIP) was confirmed to be the main functional domain of the IDA protein (Stenvik et al., 2008b). The abscission of *ida* mutant flower organs failed to appear, while the flowers fall off prematurely in the plants overexpressing the *IDA* gene (Stenvik et al., 2008b; Kumpf et al., 2013; Liu et al., 2013). It has been shown that the IDA peptide functions as a ligand of the receptor-like kinases HAESA and HAESA-LIKE2 (HAE/HSL2), which dominates flower abscission. The IDA-HAE/HSL2 pathway was shown to activate downstream mitogen-activated protein (MAP) kinase cascades, which regulate the expression of hydrolytic and cell wall-modifying enzymes (Jinn et al., 2000; Stenvik et al., 2008a; Kumpf et al., 2013; Meng et al., 2016). Also, somatic embryogenesis receptor-like kinase (SEKR) was reported to act as a co-receptor for IDA with HAE/HSL2 to transmit the abscission signal (Santiago et al., 2016; Patharkar and Walker, 2018).

Except for being involved in flower abscission, the IDA-HAE/HSL2 signaling module was reported to be important for lateral root emergence (Matsubayashi and Sakagami, 2006; Kumpf et al., 2013; Shi et al., 2018; Zhang et al., 2020). Several *IDL* genes was recently reported to be involved in responding to multiple stresses in *Arabidopsis* (Vie et al., 2017; Wang et al., 2017). *AtIDL6* expression was up-regulated by *Pseudomonas syringae* pv. tomato (*Pst*) DC3000 infection. Overexpression and knockdown lines of *AtIDL6* showed decreased and increased resistance to *Pst* DC3000 in *Arabidopsis*, respectively. Moreover, *AtIDL6* and *AtIDL7* were suggested to be induced rapidly by various stresses as negative modulators of stress-induced reactive oxygen species (ROS) signaling (Vie et al., 2017; Wang et al., 2017).

The regulation of flower abscission by genes encoding IDL peptides seems to be conserved in plants (Tranbarger et al., 2019; Schuster and Van Der Hoorn, 2020). For instance, the *SlIDA1* genes were closely related to drought-induced tomato flower drop (Tucker and Yang, 2012). In *Citrus*, five *CitIDA* genes were identified, and overexpression of *CitIDA3* gene complemented the abscission deficiency of the *ida* mutant in *Arabidopsis* (Estornell et al., 2015). Besides, *LcIDL1* was identified as a homologous gene of *AtIDA* from the litchi genome, and it was reported to play a role in regulating the shedding of floral organs in *Arabidopsis* (Ying et al., 2016). Interestingly, *IDL* genes were also found in root-knot nematodes (*Meloidogyne incognita*), and exogenous treatments of *ida* mutant plants with synthetic MiIDL1 peptides caused petals to abscise in *Arabidopsis* (Kim et al., 2018).

Tobacco (*Nicotiana tabacum* L.) is one of the most important non-food crops and has been widely used as a model plant for analyzing gene function (Li et al., 2021). A recent genome-wide study revealed a range of flower-related genes in tobacco, such as the *MADS-box* gene family (Bai et al., 2019). While study of *IDL* genes in the control of flower abscission is limited in *N. tabacum*. Here, we report the identification of *IDL* and *HAE-Like* genes from the tobacco genome. Expression analysis showed that individual *NtIDL* members and *HAE-Like* genes were highly expressed in the abscission zone of the late stages of flower development. Furthermore, the results from expression analysis also suggested that *NtIDL* genes might be involved in stress responses in tobacco.

MATERIALS AND METHODS

Identification and Sequence Analysis of NtIDL and NtHAE Members

The protein sequences of *Arabidopsis* IDA and IDL1-8 were downloaded from TAIR (Lamesch et al., 2012) and used as probe sequences to search the tobacco genome database (Edwards et al., 2017) with the E-value cutoff of 0.01. Newly identified genes were named according to the information of chromosomes and scaffold numerically. Similarly, the protein sequences of *Arabidopsis* HAE and HSL2 were used as queries to carry out BLASTP searches against the tobacco genome database under the E-value cutoff of 0.001. Newly identified *NtHAE-Like* genes were named according to the evolutionary analysis. Each sequence was submitted to ProtParam[1] to predict isoelectric point and molecular weight.

Multiple Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignment of NtIDL and reported IDLs from other species was performed using MAFFT, with their full-length amino acid sequences under default settings (Katoh and Standley, 2013). Base on the sequence alignment results, MEGA X was used to generate a neighbor-joining (NJ) phylogenetic tree (Kumar et al., 2016). The EPIP sequences of all members were extracted for multiple sequence alignment and visualized together with the results of evolutionary analysis.

Analysis of *Cis*-Elements in the Promoter of *NtIDL* Members

To assess the *cis*-elements of the *NtIDL* promoters, 2000 base pairs of promoter regions upstream of the start codon of the *NtIDL* genes were extracted, according to a previous report (Cao et al., 2016). The PlantCARE database was engaged for *cis*-elements investigation, and the results were visualized by the TBtools (Lescot et al., 2002; Chen et al., 2020).

Plant Growth Conditions

Seeds of tobacco cultivar K326 were germinated and cultured using a floating seedling production system under normal conditions (28°C, 14 h light, 10 h dark). The tissues (root, stem, shoot, leaf and flower) and abscission zones of flowers at different developing stages were collected to analyze *NtIDL* gene expression. For hormones and salt treatments, tobacco seedlings were germinated on MS medium in a light incubator at 25°C

[1]http://au.expasy.org/tools/protparam.html

for 2 weeks and treated with 50 μM ABA, 100 μM MeJA or 150 mM NaCl following a previous report (Li et al., 2019). For low/high temperature and drought treatments, the seedlings were placed in a growth chamber at 4°C/37°C or placed on filter paper. For wounding treatment, a sterile surgical blade was used to mechanically damage the third leaf of tobacco seedlings along the veins. Whole seedlings were collected at 0, 3, and 6 h after treatments, frozen in liquid nitrogen and transferred to −80°C for storage. Triple biological replicates were performed for each sample.

RNA Extraction and qRT-PCR

Total RNA of all samples was extracted following the instructions of Ultrapure RNA Kit (cwbiotech, Beijing, China). The quality and quantity of the isolated total RNA were determined by NanoDrop (Thermo Scientific™) and gel blot analysis. cDNA synthesis was performed using same amount of RNA according to the directions of the kit (R323-01, Vazyme, Nanjing, China). The tobacco ribosomal protein gene *L25* (GenBank No. L18908) was adopted as the control. qRT-PCR was performed on Roche LightCycler® 480 in a 20 μL reaction with SYBR (TaKaRa, Shiga, Japan) 10 μL, 10 mM forward primer 0.4 μL, 10 mM reverse primer 0.4 μL, and diluted cDNA 0.2 μL. Three independent experiments were carried out with three technical replicates, and the average value was taken for analysis based on the $2^{-\Delta\Delta Ct}$ method. The primer pairs used are listed in **Supplementary Table 1**.

RESULTS

Identification of *IDL* Family Genes in Tobacco (*Nicotiana tabacum* L.)

To identify IDL proteins in the tobacco proteome, the *Arabidopsis* IDA and IDL1-8 proteins were employed as queries to search against the local tobacco proteome database using Blastp. After manually removing repeated sequences, a total of 15 *IDL* genes were obtained from tobacco proteome. For consistency, newly identified IDL family members were named NtIDL01-NtIDL15 in the order of chromosome and scaffold. The detailed information of gene localization and protein characteristics were listed in **Table 1**. Amino acid length analysis showed that tobacco IDL family members ranged from 73 aa (NtIDL07) to 126 aa (NtIDL10). Their theoretical isoelectric points were from 5.19 (NtIDL13) to 11.17 (NtIDL15), and the molecular weight ranged from 7,723.83 Da (NtIDL07) to 14,120.9 Da (NtIDL10).

Multiple Sequence Alignment and Evolution Analysis of IDL Family Members

To explore the conservation of tobacco IDLs during evolution, a number of representatives IDL sequences from previous studies (Tucker and Yang, 2012; Estornell et al., 2015; Ying et al., 2016; Kim et al., 2018; Liu et al., 2018) together with the newly identified NtIDL members were subjected to multiple sequence alignments using MAFFT, and a neighbor-joining tree was generated by MEGA X. Thereafter, the EPIP domains of all the IDL members were extracted and displayed together with the results of evolutionary analysis (**Figure 1**).

As a result, all the IDL proteins were classed into six groups, namely I-VI, based on the topology of the phylogenetic tree. Most of the groups contained two or more tobacco IDL members. For example, NtIDL15 and NtIDL11 are in group II, where they were clustered with ZmIDL1, TaIDL1, and AtIDL8. NtIDL06 and NtIDL09 belong to group III together with AtIDA and AtIDL1, and they both share 85.7% similarities with AtIDA in amino acid EPIP sequence (**Figure 1**). In group IV, NtIDL01 and NtIDL14 were clustered together with AtIDL6 and AtIDL7. The remaining NtIDL members are all in group VI, which is unique to tobacco. No tobacco IDL member was grouped in Group I. Group V

TABLE 1 | IDA-Like gene family members in tobacco*.

Genes	Access number	Chr/Scf	5′ End	3′ End	AA	pI	MW	Group
NtIDL01	Nitab4.5_0000788g0070.1	Nt02	82,995,999	82,996,316	105	10.44	11,636.68	IV
NtIDL02	Nitab4.5_0002578g0020.1	Nt03	32,951,224	32,951,508	94	9.86	9,882.36	VI
NtIDL03	Nitab4.5_0001984g0090.1	Nt12	100,853,473	100,853,736	87	8.11	9,236.66	VI
NtIDL04	Nitab4.5_0001027g0160.1	Nt13	22,578,392	22,578,655	87	7.95	9,246.7	VI
NtIDL05	Nitab4.5_0000419g0050.1	Nt14	89,956,424	89,956,687	87	7.96	9,224.67	VI
NtIDL06	Nitab4.5_0000027g0380.1	Nt24	109,387,139	109,387,381	80	9.13	8,945.4	III
NtIDL07	Nitab4.5_0001185g0060.1	Nitab4.5_0001185	437,880	484,907	73	9.52	7,723.83	VI
NtIDL08	Nitab4.5_0003346g0040.1	Nitab4.5_0003346	40,233	49,492	116	7.23	12,812.6	VI
NtIDL09	Nitab4.5_0004688g0060.1	Nitab4.5_0004688	91,864	92,118	84	9.34	9,350.81	III
NtIDL10	Nitab4.5_0004965g0010.1	Nitab4.5_0004965	62,139	67,203	126	6.1	14,120.94	VI
NtIDL11	Nitab4.5_0005426g0020.1	Nitab4.5_0005426	169,912	170,157	81	10.38	8,818.35	II
NtIDL12	Nitab4.5_0005633g0020.1	Nitab4.5_0005633	157,769	159,007	73	5.19	7,972.8	VI
NtIDL13	Nitab4.5_0007980g0010.1	Nitab4.5_0007980	41,307	41,585	92	6.71	9,757.18	VI
NtIDL14	Nitab4.5_0008298g0010.1	Nitab4.5_0008298	85,973	86,302	109	9.51	12,002.92	IV
NtIDL15	Nitab4.5_0012260g0070.1	Nitab4.5_0012260	35,972	36,513	100	11.17	10,920.81	II

**Chr, chromosome; Scf, scaffolds; AA, the number of Amino acids; pI, isoelectric points; MW, molecular weights.*

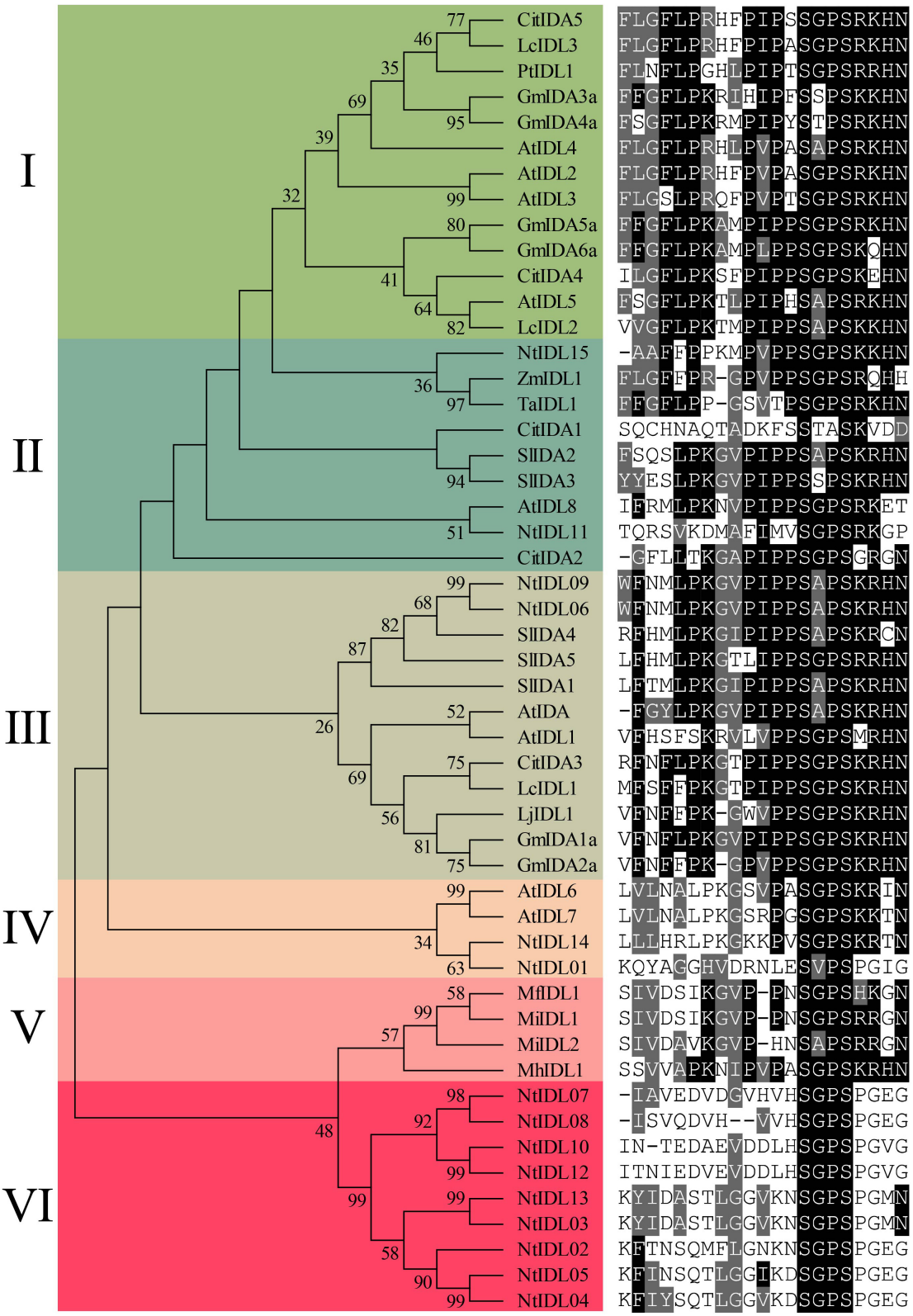


FIGURE 1 | Phylogenetic analysis of NtIDL family members. *Citrus* (Cit), *Litchi chinensis* (Lc), *Populus* (Pt), *Glycine max* (Gm), *Arabidopsis thaliana* (At), *Solanum lycopersicum* (Sl), *Nicotiana tabacum* (Nt), *Zea mays* (Zm), *Triticum aestivum* (Ta), southern root-knot nematode (*Meloidogyne incognita*, Mi), northern root-knot nematode (*Meloidogyne. hapla*, Mh), and peach root-knot nematode (*Meloidogyne. floridensis*, Mf). On the right is the conservative EPIP sequences alignment.

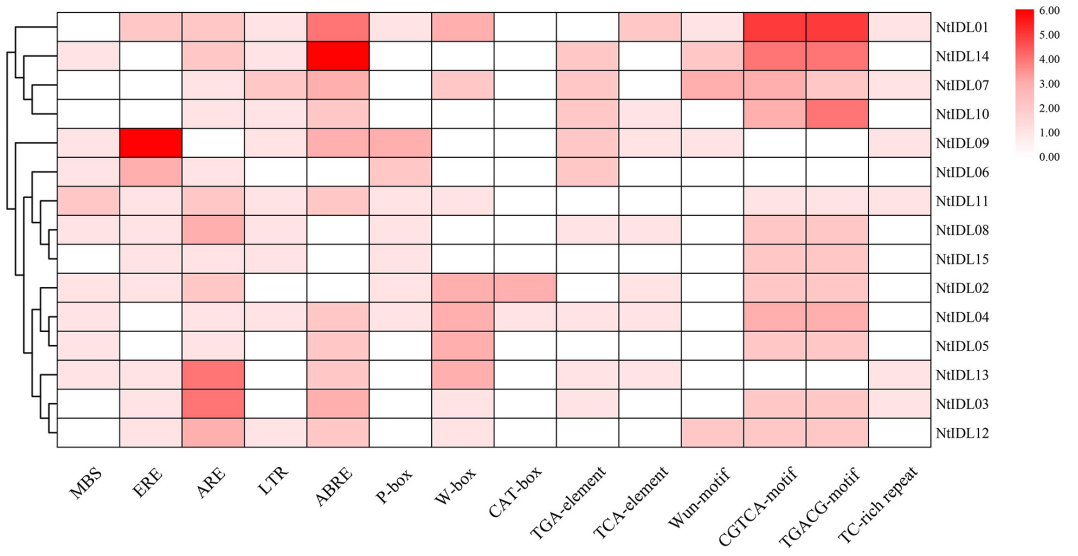


FIGURE 2 | Regulatory elements in the promoter regions of *NtIDL* gene promoters. The color represents the numbers of *cis*-element in *NtIDL* gene promoters. MBS (MYB binding site involved in drought-inducibility), ERE (ethylene-responsive *cis*-acting), ARE (anaerobic induction element), LTR (low-temperature-responsive element), ABRE (*cis*-acting element involved in the abscisic acid responsiveness), P-box (gibberellin-responsive element), W-box (WRKY binding site), CAT-box (*cis*-acting regulatory element related to meristem expression), TGA-element (auxin-responsive element), TCA-element (*cis*-acting element involved in salicylic acid responsiveness), Wun-motif (wound-responsive element), CGTCA-motif (*cis*-acting regulatory element involved in the MeJA-responsiveness), TGACG-motif (*cis*-acting regulatory element involved in the MeJA-responsiveness), and TC-rich repeats (involved in defense and stress responsiveness).

only contains IDL members of root-knot nematodes, implying a unique evolution path shared by the nematode IDLs.

Analysis of *Cis*-Elements in the Promoters of *NtIDLs*

The study of *cis*-elements could provide clues about regulatory pathways of gene expression. Therefore, the promoter regions of 15 *IDL* genes in tobacco were analyzed using the PlantCARE Online toolboxes (Lescot et al., 2002). In general, various *cis*-elements were identified in the tobacco *IDL* gene promoters. 14 *cis*-elements involved in different hormone response, developmental process and stress response were selected for further analysis (**Figure 2**). As a result, 11 *NtIDL* promoters contain ABRE *cis*-acting elements involved in ABA responsiveness. Among them, the *NtIDL14* promoter contains 6 ABRE *cis*-elements. Both CGTCA-motif and TGACG-motif were related to MeJA responsive. 12 *NtIDL* promoters were found to possess these two kinds of elements. Also, ethylene-responsive *cis*-element (ERE), salicylic acid (TCA-element), gibberellin (P-box) and auxin (TGA-element) were identified on *NtIDL* promoters. These results suggest that hormones may play important roles in regulating *NtIDL* expression.

Notably, stress-responsive elements including MBS (MYB binding site involved in drought-inducibility), TC-rich repeats (involved in defense and stress responsiveness), LTR (low-temperature-responsive element), WUN-motif (wound-responsive element), and ARE (anaerobic induction element) were founded to be abundant in the promoter regions of a large number of *NtIDL* genes. Interestingly, nine *NtIDL* genes were predicted to contain W-box *cis*-elements, which act as the WRKY transcription factors' binding site, implying certain WRKY transcription factors might regulate these *NtIDL* genes. Overall, *NtIDL* promoters possess abundant stress-related *cis*-elements, suggesting that tobacco *IDL* genes might be regulated by multiple stresses.

Expression Profiles of *IDL* and *HAE-Like* Genes of Tobacco

To explore the expression patterns of *NtIDL* members, different tissues from tobacco seedlings were collected and analyzed, including roots, stems, leaves, shoots and flowers. The results showed several *NtIDL* genes were detected to be expressed in all these tissues (**Figure 3**), such as *NtIDL01, NtIDL03*, *NtIDL06*, *NtIDL10*, and *NtIDL12*. In comparison, transcripts of some other *NtIDL* genes were presented at high levels in specific tissues. For instance, *NtIDL02* and *NtIDL09* were highly expressed in roots and flowers. Expression of *NtIDL05*, *NtIDL14*, and *NtIDL15* was significantly higher in roots that in the other tissues. It was worth noting that *NtIDL06* and *NtIDL11* were highly expressed in flowers, suggesting that both *NtIDL06* and *NtIDL11* might play significant roles in flower development of tobacco.

In order to explore potential roles of NtIDLs in flower abscission, representative *NtIDL* genes were selected to perform expression analysis in abscission zones during floral organ development. Flower development was divided into five stages as shown in **Figure 4A**. As a result (**Figure 4B**), *NtIDL* genes exhibited various expression patterns in the abscission zone during flower development. The expression of *NtIDL01* was down-regulated during the development of flowers. In contrast, *NtIDL02*, *NtIDL03*, *NtIDL04*, and *NtIDL09* were up-regulated

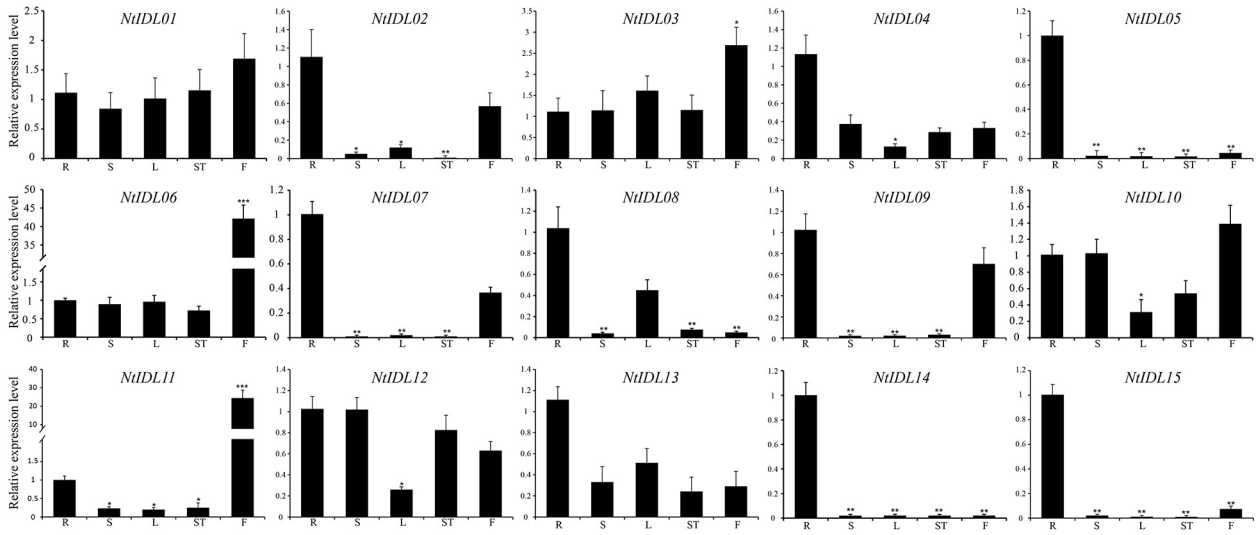


FIGURE 3 | The expression patterns of *NtIDL* genes in selected tissues. "R," "S," "L," "ST," and "F" mean roots, stems, leaves, shoots and flowers, respectively. Their expression was calculated as folds relative to the expression level of the roots to confirm the tissue specificity. All expression levels were calculated through the $2^{-\Delta\Delta Ct}$ method. The data were means ± SD from three independent replications. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ (*t*-test).

during flower development. Especially, *NtIDL06* and *NtIDL07* showed significantly higher expression at the last stage of flower development, implying that they might be closely related to the regulation of flower abscission.

In addition, the expression patterns of putative receptors of the NtIDL peptides in tobacco, *NtHAEa*, *NtHAEb*, *NtHSL2a*, and *NtHSL2b*, were also analyzed (**Supplementary Figure 2**). The results showed that the receptor-encoding genes were expressed in all the tested tissues. All of them showed high-level expression in flowers. Interestingly, the *HAE-Like* genes of tobacco were highly expressed in the abscission zone of at late stages of flower development, which is similar to the expression pattern of some *NtIDL* genes, including *NtIDL06*, *NtIDL07*, and *NtIDL09*.

Expression of *NtIDL* Genes Under Multiple Abiotic Stresses

Promoter regions of *NtIDLs* contain various *cis*-elements that are responsive to hormones and stresses. Therefore, qRT-PCR was performed to study the expression changes of *NtIDLs* under different abiotic stress treatments, including ABA, MeJA, drought, salt, wounding, and low/high temperature. All the 15 *NtIDL* genes were tested and showed complex expression patterns under various abiotic stress treatments (**Figure 5A**). As a result, *NtIDL01*, *NtIDL14*, and *NtIDL04* were up-regulated under ABA treatment. In contrast, the transcription level of *NtIDL02* was down-regulated by ABA. Interestingly, *NtIDL05* was up-regulated after 3 h but not after 6 h of ABA treatment, while *NtIDL08* showed high expression only after 6 h of ABA treatment. For MeJA treatment, nine genes were up-regulated, including *NtIDL04* and *NtIDL14,* while *NtIDL10* and *NtIDL15* were down-regulated by MeJA.

In addition, a number of *NtIDL* genes also responded to abiotic stress treatments. Some of the *NtIDL* genes were induced under multiple stresses (**Figure 5B**). *NtIDL01* and *NtIDL14* were induced by all the seven stress treatments, and *NtIDL04* was induced by all the stresses except for the low/high-temperature treatments. *NtIDL03*, *NtIDL05*, *NtIDL09*, *NtIDL11*, and *NtIDL13* were up-regulated by four different stress treatments. Three of the *NtIDLs*, on the other hand, were only induced by one specific stress treatments: *NtIDL08* was only induced by ABA treatment, *NtIDL02* was only induced by MeJA, and *NtIDL12* was only induced by salt treatment. Notably, *NtIDL07* has not been detected to be induced by any treatment in this study. Among the different stress treatments, salt treatment could induce the most *NtIDL* genes (10), while high-temperature treatment (37°C) only induced three *NtIDL* genes.

DISCUSSION

The IDL peptides have been shown to play critical roles in floral organ abscission, lateral root formation and various stress responses (Jinn et al., 2000; Liljegren, 2012; Wang et al., 2020). Systematic identification and analysis of the *IDL* gene family have been performed in many crops. However, there is less information on the *IDL* genes of tobacco. In the current study, the identification, evolution, classification, and expression profile were performed to study IDL members in tobacco.

A total of 15 NtIDL members were identified form the tobacco genome. These NtIDL members were divided into six groups with IDL members from other plant species (**Figure 1**). Notably, in group VI, all of the IDL members were from tobacco, suggesting that these NtIDL members might be originated from duplication events. Due to the fact that we were not able to map most of the *NtIDL* genes to the tobacco chromosomes (**Table 1**), the related duplication events could not be analyzed yet in the current study. Results from multiple sequence

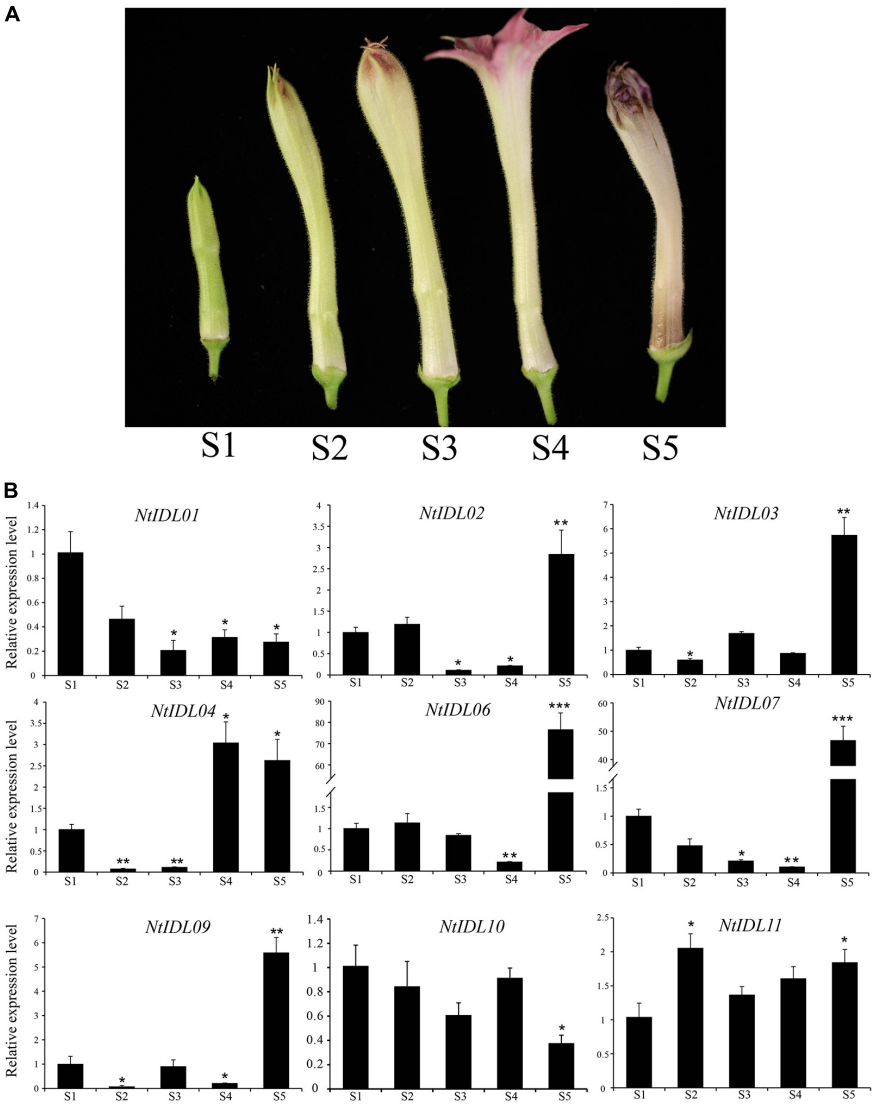


FIGURE 4 | The expression patterns of *NtIDL* genes in abscission zones during flower development. **(A)**. Five stages of tobacco flower development. "S1–S5" means "Stage 1 to Stage 5 of flower development" **(B)**. The expression patterns of selected *NtIDL* genes in the abscission zone during the flower development. All expression levels were calculated through the $2^{-\Delta\Delta Ct}$ method. The data were means ± SD from three independent replications. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ (*t*-test).

alignment analysis indicated that the EPIP sequences of IDL family members have high similarities, suggesting that IDL members might have maintained conserved functions during evolution. Interestingly, IDL proteins were also found in the root-knot nematode (*M. incognita*) genome, and they were shown to be involved in the regulation of plant root development (Kim et al., 2018). In this study, the root-knot nematode IDL members were analyzed and clustered together with NtIDL members of group VI. Root-knot nematode diseases caused by *M. incognita* are one of the most destructive diseases in tobacco production (Li et al., 2018). Most of the *NtIDLs* in group VI, including *NtIDL02*, *NtIDL04*, *NtIDL05*, *NtIDL07*, and *NtIDL13*, were highly expressed in roots (**Figure 3**). Whether the nematode-encoded IDL peptides play a role in the establishment of the infection of root-knot nematodes on tobacco, remains to be elucidated.

Previous studies indicated that some *IDL* genes encode small peptides that mediate in plants' responses to abiotic stresses. In group III, *NtIDL09* was up-regulated under high temperature, salt, and drought treatments (**Figure 5A**). While *NtIDL06*, also in group III, was down-regulated by high-temperature treatment (**Figure 5A**). This result implies that functional divergence might have occurred within this group. In group IV, NtIDL01 and NtIDL14 were clustered with AtIDL6 and AtIDL7 (**Figure 1**). These two *Arabidopsis* members have been reported to be induced rapidly by various stresses (Vie et al., 2017). Interestingly, abundant stress-related *cis*-acting elements were identified in the promoter regions of *NtIDL01* and *NtIDL14* (**Figure 2**), and

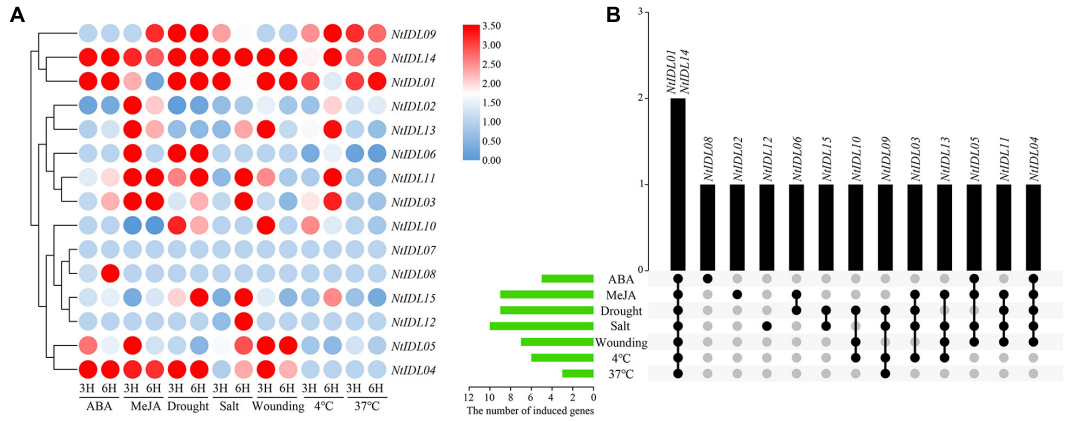


FIGURE 5 | The expression patterns of *NtIDL* genes in tobacco seedlings under multiple abiotic stress treatments. **(A)** The relative expression ratios of abiotic stress treatments, "3H" and "6H" represent 3 and 6 h after stress treatments, respectively. Expression levels were calculated through the $2^{-\Delta\Delta Ct}$ method and normalized relative to gene expression in control plants. Expression levels relative to control plants are given next to the color scale. Red, white and blue represents the induction (values over 1), invariance (values close to 1) and inhibition (values under 1) of genes related to the control treatment, respectively. **(B)** The summarized information of the stress-induced *NtIDL* genes. Black dots indicate *NtIDL* genes responsive to stress treatments and gray dots indicate *NtIDL* genes that did not respond to the stress treatments.

both of them were induced by multiple stresses (**Figure 5A**), suggesting their potential functions in multiple stress responses.

The *cis*-elements analysis showed that *NtIDL* members contained rich response-hormone *cis*-elements on their promoters (**Figure 2**), which suggested that hormones might be involved in the transcriptional regulation of *NtIDL* genes. Phytohormones ABA and MeJA have been reported to accelerate flower abscission in plants (Hartmond et al., 2000; Patharkar and Walker, 2019). Moreover, *NtIDL01*, *NtIDL14*, *NtIDL08*, *NtIDL04*, and *NtIDL05* were found to be up-regulated under ABA treatment (**Figure 5A**). Also, nine *NtIDL* genes, including *NtIDL02* and *NtIDL06*, were induced by MeJA treatment. Taken these results, these NtIDL members may confer flower abscission though the ABA and MeJA signaling pathways.

In *Arabidopsis*, overexpression of the *AtIDA* gene could rescue the deficiency in flower abscission of the *ida* mutant. Notably, NtIDL06 and NtIDL09 were clustered together with AtIDA and AtIDL1 in group III. IDL members from other plant species in this group were also reported to regulate flower abscission, including SlIDA1 (Tucker and Yang, 2012), CitIDA3 (Estornell et al., 2015), LcIDL1 (Ying et al., 2016), and GmIDA1a (Tucker and Yang, 2012; **Figure 1**). Moreover, NtIDL06 shared high similarities with AtIDA in EPIP sequences (**Figure 1**). The qRT-PCR results indicated that *NtIDL06* was highly expressed in the abscission zone at the last stage of flower development (**Figure 4**). Moreover, four *NtHAE-Like* genes identified in this study were detected to show similar expression patterns with *NtIDL06* (**Supplementary Figure 2**). Those genes with high expression levels in the abscission zone at late stages of flower development might be related to cell wall remodeling and abscission of flowers. It is worth mentioning that *NtIDL06* was induced by MeJA treatment (**Figure 5A**), and MeJA was a positive regulator of flower abscission. Combining these results, *NtIDL06* might be involved in tobacco flower abscission.

CONCLUSION

Systematic investigation was adopted to identify 15 *NtIDL* genes in the tobacco genome. The results from expression analysis in different tissues and under various of stress treatments suggested that the tobacco *IDL* genes might play multiple roles in various biological processes. A number of NtIDLs were identified with potential functions in stress responses. Notably, as the closest homolog of AtIDA, *NtIDL06* and its putative receptors were highly expressed in the abscission zone at the last stage of flower development, suggesting that NtIDL06 might be involved in the natural process of corolla abscission. The results from this study provide insights for further exploring the biological functions of tobacco *IDL* genes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YG and XL conceived this research and designed the experiments. CG and QW conducted the research and drafted the manuscript. ZL, JS, and ZZ assisted in data collection and participated in drafting 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 (31571494), the Agricultural Science

and Technology Innovation Program (ASTIP-TRIC02) and the Fundamental Research Funds for Central Non-profit Scientific Institution (Y2017JC27).

SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.670794/full#supplementary-material

Supplementary Figure 1 | Sequence alignment of HAE-Like proteins from tobacco and *Arabidopsis*. The black background indicates that the amino acid similarity is 100%, and the gray indicates that the amino acid similarity is ranged from 60 to 90%.

Supplementary Figure 2 | The expression pattern of *NtHAE-Like* genes in selected tissues **(A)** and in abscission zone during the flower development **(B)**. Panel **(A)**: "R," "S," "L," "ST," and "F" mean roots, stems, leaves, shoots and flowers, respectively. Their expressions were calculated as folds relative to the expression level of the roots to confirm the tissue specificity. Panel **(B)**: "S1–S5" means "Stage 1 to Stage 5 of flower development." All expression levels were calculated through the $2^{-\Delta\Delta Ct}$ method. The data were means ± SD from three independent replications. $^{*}p < 0.05$, $^{**}p < 0.01$, (*t*-test).

Supplementary Table 1 | The qRT-PCR primers used in this study.

Supplementary Table 2 | HAE-Like gene family members in tobacco.

REFERENCES

Bai, G., Yang, D.-H., Cao, P., Yao, H., Zhang, Y., Chen, X., et al. (2019). Genome-wide identification, gene structure and expression analysis of the MADS-Box gene family indicate their function in the development of tobacco (*Nicotiana tabacum L.*). *Int. J. Mol. Sci.* 20:5043. doi: 10.3390/ijms20205043

Cao, Y., Han, Y., Li, D., Lin, Y., and Cai, Y. (2016). MYB transcription factors in chinese pear (Pyrus bretschneideri Rehd.): genome-wide identification, classification, and expression profiling during fruit development. *Front. Plant Sci.* 7:577. doi: 10.3389/fpls.2016.00577

Chandler, J. (2011). The hormonal regulation of flower development. *J. Plant Growth Regul.* 30, 242–254. doi: 10.1007/s00344-010-9180-x

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

Edwards, K., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A., Bombarely, A., et al. (2017). A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18:448. doi: 10.1186/s12864-017-3791-6

Estornell, L. H., Wildhagen, M., Pérez-Amador, M. A., Talón, M., Tadeo, F. R., and Butenko, M. A. (2015). The IDA peptide controls abscission in Arabidopsis and Citrus. *Front. Plant Sci.* 6:1003. doi: 10.3389/fpls.2015.01003

Gubert, C. M., Christy, M. E., Ward, D. L., Groner, W. D., and Liljegren, S. J. (2014). ASYMMETRIC LEAVES1 regulates abscission zone placement in Arabidopsis flowers. *BMC Plant Biol.* 14:195. doi: 10.1186/s12870-014-0195-5

Hartmond, U., Yuan, R., Burns, J. K., Grant, A., and Kender, W. J. (2000). Citrus fruit abscission induced by methyl-jasmonate. *J. Am. Soc. Hortic. Sci.* 125, 547–552. doi: 10.21273/JASHS.125.5.547

Jinn, T.-L., Stone, J. M., and Walker, J. C. (2000). HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev.* 14, 108–117. doi: 10.1101/gad.14.1.108

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, J., Yang, R., Chang, C., Park, Y., and Tucker, M. L. (2018). The root-knot nematode Meloidogyne incognita produces a functional mimic of the Arabidopsis INFLORESCENCE DEFICIENT IN ABSCISSION signaling peptide. *J. Exp. Bot.* 69, 3009–3021. doi: 10.1093/jxb/ery135

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

Kumpf, R. P., Shi, C.-L., Larrieu, A., Stø, I. M., Butenko, M. A., Péret, B., et al. (2013). Floral organ abscission peptide IDA and its HAE/HSL2 receptors control cell separation during lateral root emergence. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5235–5240. doi: 10.1073/pnas.1210835110

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

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

Lewis, M. W., Leslie, M. E., and Liljegren, S. J. (2006). Plant separation: 50 ways to leave your mother. *Curr. Opin. Plant Biol.* 9, 59–65. doi: 10.1016/j.pbi.2005.11.009

Li, X., Guo, C., Ahmad, S., Wang, Q., Yu, J., Liu, C., et al. (2019). Systematic analysis of MYB family genes in potato and their multiple roles in development and stress responses. *Biomolecules* 9:317. doi: 10.3390/biom9080317

Li, X., Xing, X., Tian, P., Zhang, M., Huo, Z., Zhao, K., et al. (2018). Comparative transcriptome profiling reveals defense-related genes against *Meloidogyne incognita* invasion in tobacco. *Molecules* 23:2081. doi: 10.3390/molecules23082081

Li, Z., Chao, J., Li, X., Li, G., Song, D., Guo, Y., et al. (2021). Systematic analysis of the bZIP family in tobacco and functional characterization of NtbZIP62 involvement in salt stress. *Agronomy* 11:148. doi: 10.3390/agronomy11010148

Liljegren, S. J. (2012). Organ abscission: exit strategies require signals and moving traffic. *Curr. Opin. Plant Biol.* 15, 670–676. doi: 10.1016/j.pbi.2012.09.012

Liu, B., Butenko, M. A., Shi, C.-L., Bolivar, J. L., Winge, P., Stenvik, G.-E., et al. (2013). NEVERSHED and INFLORESCENCE DEFICIENT IN ABSCISSION are differentially required for cell expansion and cell separation during floral organ abscission in Arabidopsis thaliana. *J. Exp. Bot.* 64, 5345–5357. doi: 10.1093/jxb/ert232

Liu, C., Zhang, C., Fan, M., Ma, W., Chen, M., Cai, F., et al. (2018). GmIDL2a and GmIDL4a, encoding the inflorescence deficient in abscission-like protein, are involved in soybean cell wall degradation during lateral root emergence. *Int. J. Mol. Sci.* 19:2262. doi: 10.3390/ijms19082262

Marciniak, K., Kućko, A., Wilmowicz, E., Świdziński, M., Przedniczek, K., and Kopcewicz, J. (2018). Gibberellic acid affects the functioning of the flower abscission zone in Lupinus luteus via cooperation with the ethylene precursor independently of abscisic acid. *J. Plant Physiol.* 229, 170–174. doi: 10.1016/j.jplph.2018.07.014

Matsubayashi, Y., and Sakagami, Y. (2006). Peptide hormones in plants. *Annu. Rev. Plant Biol.* 57, 649–674. doi: 10.1146/annurev.arplant.56.032604.144204

Meng, D., Cao, Y., Chen, T., Abdullah, M., Jin, Q., Fan, H., et al. (2019). Evolution and functional divergence of MADS-box genes in Pyrus. *Sci. Rep.* 9:1266. doi: 10.1038/s41598-018-37897-6

Meng, X., Zhou, J., Tang, J., Li, B., De Oliveira, M. V., Chai, J., et al. (2016). Ligand-induced receptor-like kinase complex regulates floral organ abscission in Arabidopsis. *Cell Rep.* 14, 1330–1338. doi: 10.1016/j.celrep.2016.01.023

Mesejo, C., Marzal, A., Martínez-Fuentes, A., Reig, C., and Agustí, M. (2021). On how auxin, ethylene and IDA-peptide relate during mature Citrus fruit abscission. *Sci. Hortic.* 278:109855. doi: 10.1016/j.scienta.2020.109855

Patharkar, O. R., and Walker, J. C. (2018). Advances in abscission signaling. *J. Exp. Bot.* 69, 733–740. doi: 10.1093/jxb/erx256

Patharkar, O. R., and Walker, J. C. (2019). Connections between abscission, dehiscence, pathogen defense, drought tolerance, and senescence. *Plant Sci.* 284, 25–29. doi: 10.1016/j.plantsci.2019.03.016

Patterson, S. E. (2001). Cutting loose. Abscission and dehiscence in Arabidopsis. *Plant Physiol.* 126, 494–500. doi: 10.1104/pp.126.2.494

Reichardt, S., Piepho, H.-P., Stintzi, A., and Schaller, A. (2020). Peptide signaling for drought-induced tomato flower drop. *Science* 367, 1482–1485. doi: 10.1126/science.aaz5641

Santiago, J., Brandt, B., Wildhagen, M., Hohmann, U., Hothorn, L. A., Butenko, M. A., et al. (2016). Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. *Elife* 5:e15075. doi: 10.7554/eLife.15075

Schuster, M., and Van Der Hoorn, R. A. (2020). Plant biology: distinct new players in processing peptide hormones during abscission. *Curr. Biol.* 30, R715–R717. doi: 10.1016/j.cub.2020.04.072

Shi, C.-L., Von Wangenheim, D., Herrmann, U., Wildhagen, M., Kulik, I., Kopf, A., et al. (2018). The dynamics of root cap sloughing in Arabidopsis is regulated by peptide signalling. *Nat. Plants* 4, 596–604. doi: 10.1038/s41477-018-0212-z

Stenvik, G.-E., Butenko, M. A., and Aalen, R. B. (2008a). Identification of a putative receptor-ligand pair controlling cell separation in plants. *Plant Signal. Behav.* 3, 1109–1110. doi: 10.4161/psb.3.12.7009

Stenvik, G.-E., Tandstad, N. M., Guo, Y., Shi, C.-L., Kristiansen, W., Holmgren, A., et al. (2008b). The EPIP peptide of INFLORESCENCE DEFICIENT IN ABSCISSION is sufficient to induce abscission in Arabidopsis through the receptor-like kinases HAESA and HAESA-LIKE2. *Plant Cell* 20, 1805–1817. doi: 10.1105/tpc.108.059139

Tranbarger, T. J., Domonhédo, H., Cazemajor, M., Dubreuil, C., Fischer, U., and Morcillo, F. (2019). The PIP peptide of INFLORESCENCE DEFICIENT IN ABSCISSION enhances Populus leaf and *Elaeis guineensis* fruit abscission. *Plants* 8:143. doi: 10.3390/plants8060143

Tucker, M. L., and Yang, R. (2012). IDA-like gene expression in soybean and tomato leaf abscission and requirement for a diffusible stelar abscission signal. *AoB Plants* 2012:pls035. doi: 10.1093/aobpla/pls035

Vie, A. K., Najafi, J., Winge, P., Cattan, E., Wrzaczek, M., Kangasjärvi, J., et al. (2017). The IDA-LIKE peptides IDL6 and IDL7 are negative modulators of stress responses in Arabidopsis thaliana. *J. Exp. Bot.* 68, 3557–3571. doi: 10.1093/jxb/erx168

Wang, R., Shi, C., Wang, X., Li, R., Meng, Y., Cheng, L., et al. (2020). Tomato SlIDA has a critical role in tomato fertilization by modifying reactive oxygen species homeostasis. *Plant J.* 103, 2100–2118. doi: 10.1111/tpj.14886

Wang, X., Hou, S., Wu, Q., Lin, M., Acharya, B. R., Wu, D., et al. (2017). IDL6-HAE/HSL2 impacts pectin degradation and resistance to *Pseudomonas* syringae pv tomato DC3000 in Arabidopsis leaves. *Plant J.* 89, 250–263. doi: 10.1111/tpj.13380

Wilmowicz, E., Kućko, A., Ostrowski, M., and Panek, K. (2018). INFLORESCENCE DEFICIENT IN ABSCISSION-like is an abscission-associated and phytohormone-regulated gene in flower separation of *Lupinus luteus*. *Plant Growth Regul.* 85, 91–100. doi: 10.1007/s10725-018-0375-7

Ying, P., Li, C., Liu, X., Xia, R., Zhao, M., and Li, J. (2016). Identification and molecular characterization of an IDA-like gene from litchi, LcIDL1, whose ectopic expression promotes floral organ abscission in Arabidopsis. *Sci. Rep.* 6:37135.

Zhang, X., Peng, H., Zhu, S., Xing, J., Li, X., Zhu, Z., et al. (2020). Nematode-encoded RALF peptide mimics facilitate parasitism of plants through the FERONIA receptor kinase. *Mol. Plant* 13, 1434–1454. doi: 10.1016/j.molp.2020.08.014

Conflict of Interest: XL was employed by China Tobacco Hunan Industrial Co., Ltd.

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

The handling editor declared a past co-authorship with one of the authors XL.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 May 2021
doi: 10.3389/fgene.2021.654146

MicroRNAs Involved in Regulatory Cytoplasmic Male Sterility by Analysis RNA-seq and Small RNA-seq in Soybean

Chunbao Zhang[1], Fuyou Fu[2]*, Chunjing Lin[1], Xiaoyang Ding[1], Jingyong Zhang[1], Hao Yan[1], Pengnian Wang[1], Wei Zhang[1], Bao Peng[1] and Limei Zhao[1]**

[1] Soybean Research Institute, The National Engineering Research Center for Soybean, Jilin Academy of Agricultural Sciences, Changchun, China, [2] Saskatoon Research Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:
Lin Zhang,
Central South University Forestry and Technology, China

Reviewed by:
Jianghua Chen,
Xishuangbanna Tropical Botanical Garden (CAS), China
Himanshu Sharma,
National Agri-Food Biotechnology Institute, India
Fangcheng Bi,
Guangdong Academy of Agricultural Sciences, China

***Correspondence:**
Chunbao Zhang
cbzhang@cjaas.com
Fuyou Fu
fuyou.fu@canada.ca
Limei Zhao
lmzhao@cjaas.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 15 January 2021
Accepted: 06 April 2021
Published: 12 May 2021

Citation:
Zhang C, Fu F, Lin C, Ding X, Zhang J, Yan H, Wang P, Zhang W, Peng B and Zhao L (2021) MicroRNAs Involved in Regulatory Cytoplasmic Male Sterility by Analysis RNA-seq and Small RNA-seq in Soybean. Front. Genet. 12:654146. doi: 10.3389/fgene.2021.654146

Cytoplasmic male sterility (CMS) is an important plant characteristic for exploiting heterosis to enhance crop traits during breeding. However, the CMS regulatory network remains unclear in plants, even though researchers have attempted to isolate genes associated with CMS. In this study, we performed high-throughput sequencing and degradome analyses to identify microRNAs (miRNAs) and their targets in a soybean CMS line (JLCMS9A) and its maintainer line (JLCMS9B). Additionally, the differentially expressed genes during reproductive development were identified using RNA-seq data. A total of 280 miRNAs matched soybean miRNA sequences in miRBase, including mature miRNAs and pre-miRNAs. Of the 280 miRNAs, 30, 23, and 21 belonged to the miR166, miR156, and miR171 families, respectively. Moreover, 410 novel low-abundant miRNAs were identified in the JLCMS9A and JLCMS9B flower buds. Furthermore, 303 and 462 target genes unique to JLCMS9A and JLCMS9B, respectively, as well as 782 common targets were predicted based on the degradome analysis. Target genes differentially expressed between the CMS line and the maintainer line were revealed by an RNA-seq analysis. Moreover, all target genes were annotated with diverse functions related to biological processes, cellular components, and molecular functions, including transcriptional regulation, the nucleus, meristem maintenance, meristem initiation, cell differentiation, auxin-activated signaling, plant ovule development, and anther development. Finally, a network was built based on the interactions. Analyses of the miRNA, degradome, and transcriptome datasets generated in this study provided a comprehensive overview of the reproductive development of a CMS soybean line. The data presented herein represent useful information for soybean hybrid breeding. Furthermore, the study results indicate that miRNAs might contribute to the soybean CMS regulatory network by modulating the expression of CMS-related genes. These findings lay the foundation for future studies on the molecular mechanisms underlying soybean CMS.

Keywords: microRNA, degradome, RNA-seq, cytoplasmic male sterility, soybean

Abbreviations: CMS, cytoplasmic male sterility; miRNA, microRNA; sRNA, small RNA; DEG, differentially expressed gene; FDR, false discovery rate.

INTRODUCTION

Soybean is an important crop that is cultivated for its protein and oil content. In some countries, especially China, soybean production is less profitable for farmers than the production of corn, rice, or other crops, resulting in yearly decreases in the arable land area used for cultivating soybean. Crop yields may be increased by exploiting heterosis (Ivanov and Dymshits, 2007; Virmani, 2012). One of the most important methods for exploiting heterosis involves the application of cytoplasmic male sterility (CMS). Specifically, CMS has been broadly used for breeding major crop species (Yuan, 1966; Bosacchi et al., 2015). Heterosis has been exploited to improve specific soybean traits (Zhao et al., 2004). In 1995, Ru Nan Tian E Dan male sterile cytoplasm was developed as the first soybean CMS system (i.e., RN-CMS; Sun et al., 2001). In 2002, HybSoy 1 became the first soybean hybrid bred using the RN–CMS system, with an average yield of more than 20% greater than that of the control line (Zhao et al., 2004). By 2020, 23 soybean hybrids had been bred with the RN–CMS system. However, it remains unclear how RN–CMS works.

The small RNAs (sRNAs) are short [approximately 18–30 nucleotides (nt)] non-coding RNA molecules that can regulate gene expression in the cytoplasm and nucleus via post-transcriptional gene silencing, chromatin-dependent gene silencing, or RNA activation. The three classes of sRNAs are microRNAs (miRNAs), small interfering RNAs, and Piwi-interacting RNAs. The miRNAs are a class of short (20–24 nt) non-coding RNAs that regulate gene expression at the post-transcriptional level by degrading their target mRNAs and/or inhibiting translation (Banisch et al., 2012).

Previous research confirmed that sRNAs, including miRNAs, can modulate anther and pollen development, leading to male sterility (Chambers and Shuai, 2009; Grant-Downton et al., 2009; Wang et al., 2009; Xing et al., 2010; Shen et al., 2011; Wei et al., 2012; Yang J. et al., 2013). For example, miR156, miR167, and miR399 influence pollen development in *Arabidopsis thaliana* and citrus plants (Wu et al., 2006; Wang et al., 2009, 2020; Xing et al., 2010; Wei et al., 2012). An earlier deep-sequencing analysis proved that miRNAs contribute to the flower and pollen development of a soybean CMS line and its maintainer line (Ding et al., 2016). Recent studies revealed the opposite expression patterns of gma-miR156b and its target *GmSPL* genes, including *GmSPL9*, in flowers during the early flower bud development stage of the soybean CMS A-line and its maintainer B-line (Ding et al., 2019, 2020). In this study, RN-CMS line JLCMS9A and its maintainer line JLCMS9B were analyzed by sequencing their sRNAs to identify which sRNAs, especially miRNAs, induce sterility.

RESULTS

sRNA Sequencing and Degradome Profiling of JLCMS9A and JLCMS9B

The high-throughput sequencing of two independent sRNA libraries for the flower buds of the CMS line JLCMS9A and its maintainer line JLCMS9B generated 28,736,898 and 32,977,823 raw reads, respectively (**Supplementary Table 1**). Our small RNA bioinformatics analysis was performed using the development pipeline (**Figure 1**). After removing adapter contaminants, oversized insertions, low-quality reads, poly-A tags, short and long tags (tags < 18 nt and >25 nt), and other non-coding RNAs (rRNA, snoRNA, snRNA, and tRNA), 5,996,889 and 7,397,799 unique clean reads (18–25 nt long), including 3,639,335 and 4,791,157 valid unique sRNA reads, were obtained for JLCMS9A and JLCMS9B, respectively, (**Supplementary Table 1**). Most of the sRNAs in the two libraries were 21–24 nt long, with 24 nt being the most common length, followed by 21 and 22 nt (**Figure 2**). This is consistent with previously reported sRNA length distributions for various soybean tissues, including the roots, nodules, flowers, developing seeds, and cotyledons (Joshi et al., 2010; Song et al., 2011; Goette et al., 2014).

Identification of Known miRNAs

All sRNA sequences were mapped to known soybean miRNAs in the miRBase 21.0 database[1] to identify conserved miRNAs. In the present study, 1,327 miRNAs from 250 families were predicted and corresponded to 511 pre-miRNAs (**Supplementary Table 2**). Additionally, 21 nt was the most common length among the known miRNAs, followed by 22 and 24 nt (**Supplementary Figure 1**). Of the identified miRNA families, miR166 had the most members (Hu et al., 2016), followed by miR171 and miR169, with 48 and 45 members, respectively (**Supplementary Figure 2**). In contrast, most families had fewer than five members, with 73, 98, 13, and 19 families comprising one, two, three, and four members, respectively, (**Supplementary Table 3**). Among the 250 miRNA families, 39 identified families were conserved in other species (47 species in the miRBase 21.0 database). The rest were non-conserved soybean-specific miRNA families, including 13, 8, and 15 that were revealed to be conserved at high (>10 plant species), moderate (five to nine plant species), and low (two to four plant species) levels on the basis of a previous classification (Zhang et al., 2006). Additionally, there were 183 non-conserved miRNA families (**Supplementary Table 4**).

Identification of Known Conserved miRNAs

To assess the accuracy of the miRNA predications, the miRNAs were divided into six groups. A total of 828 conserved miRNAs (820 and eight in gp1a and gp1b, respectively) and 491 non-conserved miRNAs (71, 238, 20, and 162 in gp2a, gp2b, gp3, and gp4, respectively) were identified (**Supplementary Figure 3** and **Supplementary Table 2**). In this study, only conserved miRNAs were analyzed further. A total of 280 miRNAs matched soybean miRNA sequences in the miRbase database, including mature miRNA and pre-miRNA sequences. Moreover, 85 and 170 miRNAs were highly and moderately expressed in the JLCMS9A and JLCMS9B flower buds, respectively. Among the 280 miRNAs, 30, 23, and 21 belonged to the miR166, miR156, and miR171 families, respectively. Accordingly, known conserved

[1] http://www.mirbase.org/

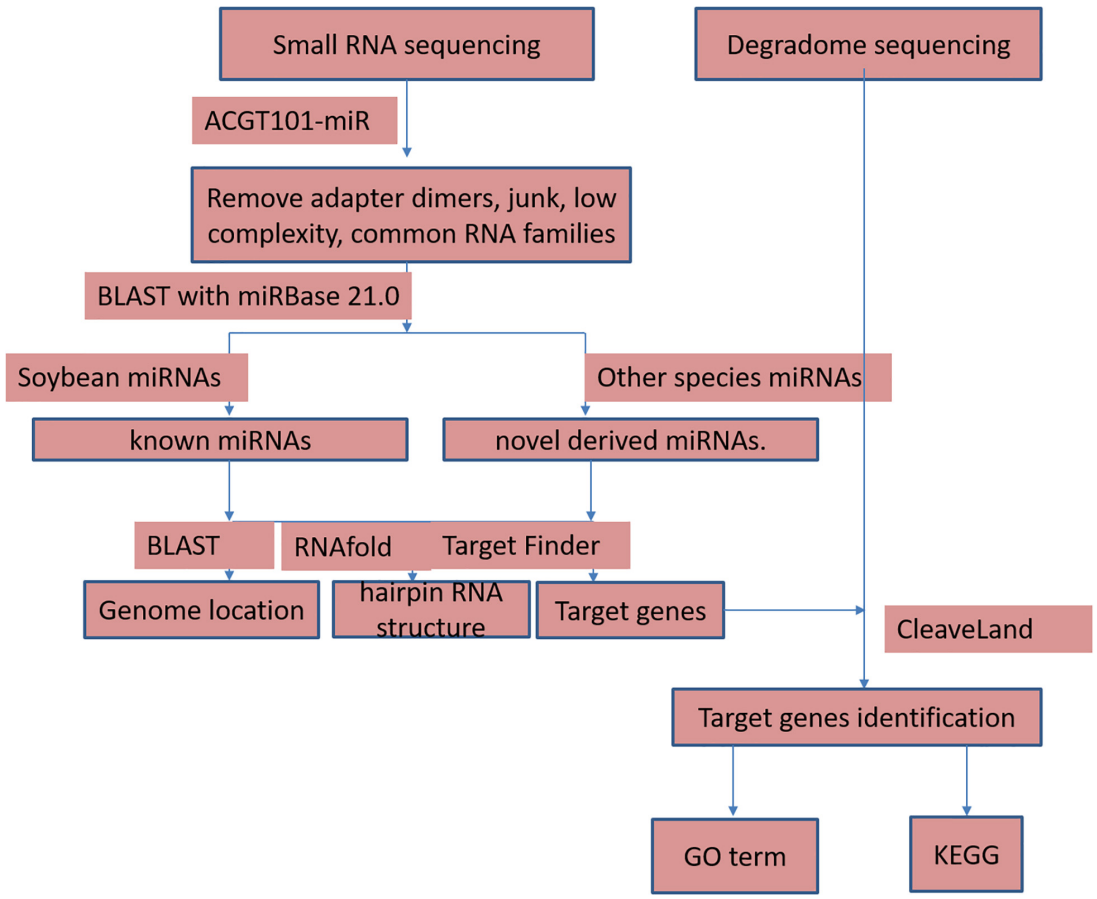


FIGURE 1 | The developmental bioinformatics pipeline of small RNA sequencing analysis in this study.

miRNAs appear to regulate the flower bud development of JLCMS9A and JLCMS9B.

Identification of Known Newly Conserved miRNAs

Known newly conserved miRNAs usually match confirmed miRNA sequences in miRBase, but we detected some sequence diversity in this study. For example, the gma-miR156b_L + 1R-1 sequence (TTGACAGAAGAGAGAGAGCAC) was confirmed using the miRBase database, but it differed from the gma-miR156b sequence (TGACAGAAGAGAGAGAGCACA; **Supplementary Table 2**). A total of 144 known newly conserved miRNAs were confirmed in this study. Additionally, 15 and 92 miRNAs were, respectively, identified as highly and moderately expressed in the JLCMS9A and JLCMS9B flower buds. Among these 144 miRNAs, 10, eight, and five were identified as members of the miR393, miR395, and miR169 families, respectively.

Identification of Novel miRNAs

Novel miRNAs, especially new 5p and 3p sequences, are not included in miRBase. In this study, 410 novel miRNAs were identified, including gma-miR156m-p3, which was a new 3p sequence (**Supplementary Table 2**). A total of 404 novel miRNAs were identified in this study. Among the 144 miRNAs, 10, eight, and five miRNAs belonged to the miR393, miR395, and miR169 families, respectively. However, only the following six were highly expressed in soybean flower buds: gma-miR166e-p5, gma-miR482c-p5, gma-miR1513c-p5, gma-miR4997-p3, gma-miR5368-p5, and gma-miR5368-p3_2ss1AT18CA. Among the 404 novel miRNAs, 26, 15, and 15 miRNAs belonged to the miR1520, miR156, and miR171 families, respectively. Thus, the novel miRNAs were not highly abundant in the JLCMS9A and JLCMS9B flower buds.

Identification of miRNA Target Genes via a Degradome Analysis

Independent degradome libraries for the JLCMS9A and JLCMS9B flower buds were constructed and sequenced, resulting in 13,604,916 and 15,876,060 raw reads, respectively (**Supplementary Table 5**). After eliminating the adapter sequences and/or low-quality reads from the raw reads longer than 15 nt, 13,484,912 (99.12%) and 15,741,865 (99.14%) were perfectly aligned to the soybean reference genome (Gmax_275_v2.0; **Supplementary Table 3**). After CleaveLand4 was used to process and analyze the data

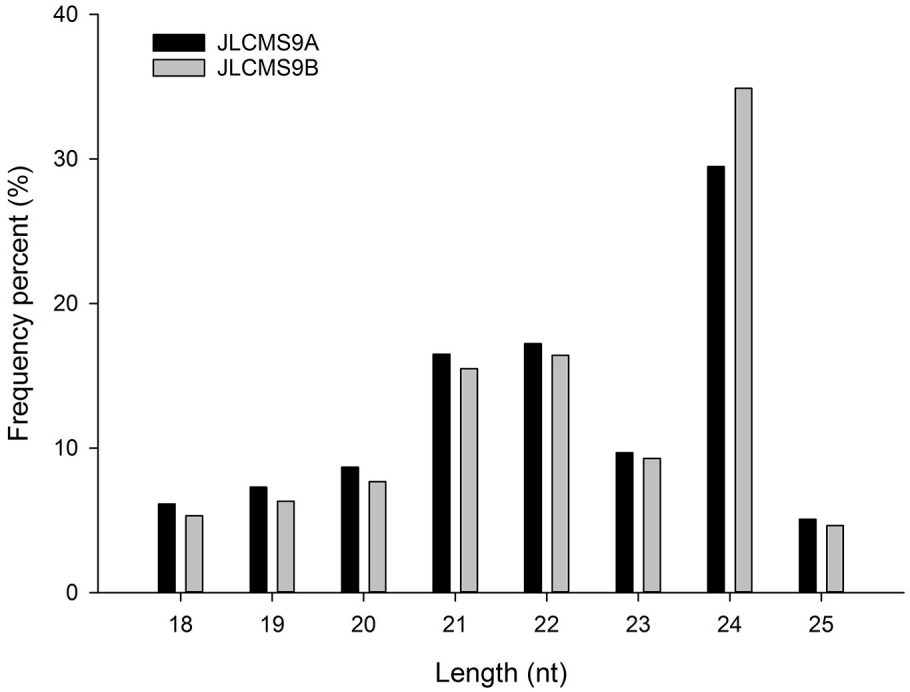


FIGURE 2 | The length distribution of small RNAs in JLCMS9A and JLCMS9B flower buds.

(Addo-Quaye et al., 2009), 1,617 target genes were predicted to be cleaved by 349 miRNAs in JLCMS9A and JLCMS9B.

The sliced-target genes were divided into five categories (0, 1, 2, 3, and 4) on the basis of the relative abundance of tags at the target sites as previously described (Addo-Quaye et al., 2008; Li et al., 2010; Liu et al., 2014). The miRNAs and their targets in the five categories are presented in **Supplementary Table 6** and **Figure 3A**. More gene targets were detected in JLCMS9B (1,095) than in JLCMS9A (1,255). Specifically, the number of gene targets in categories 0 and 2 was substantially higher for JLCMS9B than for JLCMS9A (**Figure 3A**). A total of 303 and 462 unique targets were, respectively, predicted for JLCMS9A and JLCMS9B, in addition to 782 common targets following the degradome analysis (**Figure 3B**).

The target gene analysis revealed that a single miRNA can simultaneously regulate the expression of several target genes, usually from a large gene family. The predicted highly conserved miRNAs, such as miR169, miR156, miR396, miR166, miR172, and miR171 family members, regulated multiple target genes. For example, the miR167 family members were detected as regulators of the expression of 206 target genes, including those encoding mitochondrial carrier proteins, small nuclear ribonucleoprotein component-like proteins, MYB domain proteins, transcription initiation factors, and squamosa promoter-binding-like (SPL) proteins (**Supplementary Table 6**). Moreover, a single mRNA may be targeted by multiple miRNAs, even from different miRNA families. For example, *SPL6* was identified as a target for 15 and three miRNAs from the miR156 and miR157 families, respectively (**Supplementary Table 6**). Meanwhile, the expression of six selected miRNAs were tested using quantitative real-time PCR (qRT-PCR) analysis, which further validated the differential expression data obtained from sequencing on this study. The expression pattern of selected miRNAs was consistent with the sequence reads (**Supplementary Figure 4**).

GO and KEGG Analysis of Target Genes

The identified miRNA target genes were subjected to a Gene Ontology (GO) analysis (**Figure 3C**) and assigned to metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (**Figure 3D**) to functionally characterize the target genes and elucidate the regulatory effects of miRNAs on pollen development. The GO analysis indicated that the target genes affected various biological processes, cellular components, and molecular functions, including transcriptional regulation, the nucleus, meristem maintenance, meristem initiation, cell differentiation, auxin-activated signaling, plant ovule development, and anther development (**Figure 4A**). The enriched KEGG pathways among the target genes were associated with the spliceosome, RNA transport, RNA degradation, the ribosome, and plant hormone signal transduction. These results may reflect the importance of these miRNAs for regulating gene expression during pollen development in soybean plants.

Identification of Differentially Expressed Genes

A total of 103,976,346 and 90,354,809 high-quality RNA-seq reads were generated for JLCMS9A and JLCMS9B, respectively, using a Life Technologies Ion Proton sequencer (**Supplementary Table 2**). After filtering, trimming the adapters,

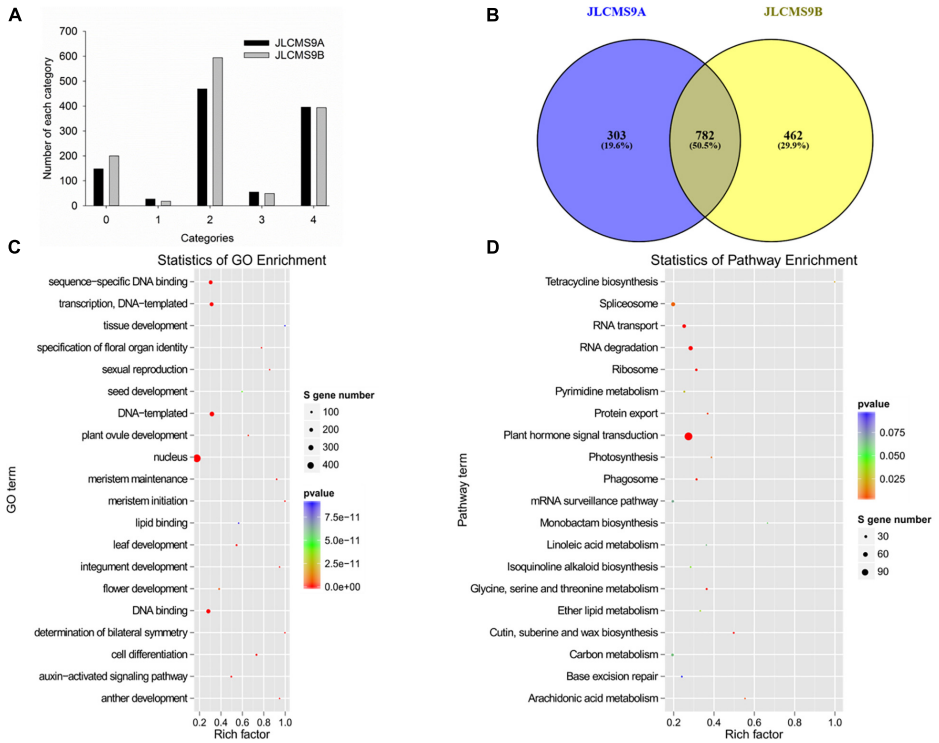


FIGURE 3 | The profiles of small RNAs (sRNAs) and target genes in JLCMS9A and JLCMS9B. **(A)** Number of sRNAs in different categories. **(B)** Number of unique and common sRNAs in JLCMS9A and JLCMS9B. **(C)** Gene Ontology (GO) functional analysis of target genes. **(D)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of target genes.

and removing the low-quality reads, 85,715,721 and 75,773,490 high-quality reads were retained for JLCMS9A and JLCMS9B, respectively. These reads were mapped to the soybean reference genome (Gmax_275_v2.0) using HISAT2, with an average matching rate of 86 and 84% for JLCMS9A and JLCMS9B, respectively (**Supplementary Table 2**; Kim et al., 2015).

The DEGs were detected following the pair-wise comparisons of the two lines using the DEGseq algorithm, with a false discovery rate ≤ 0.05 and $[\log_2(\text{fold-change})] \geq 1$ applied as the threshold (Wang et al., 2010). A total of 440 DEGs were identified between JLCMS9A and JLCMS9B (**Supplementary Table 4**). The expression levels of 428 and 12 of these DEGs were down-regulated and up-regulated, respectively, in JLCMS9A relative to the corresponding expression level in JLCMS9B. Interestingly, 174 of the down-regulated DEGs were only expressed in JLCMS9B, and five of the up-regulated DEGs were uniquely expressed in JLCMS9A. Thus, CMS in soybean may be related to the lack of expression and the induced expression of these 174 and 5 genes, respectively. These DEGs may be closely related to soybean CMS.

Gene Ontology Annotation and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis

On the basis of the annotated GO terms, 440 DEGs were assigned to 25 categories ($P < 10^{-5}$), including 10 biological processes, four cellular components, and 11 molecular function categories (**Figure 4A**). Among the biological process categories, pectin catabolic process was the main annotated GO term, followed by cell wall modification and catalytic activity. Among the cellular component categories, cell wall was the most significantly enriched GO term. Of the molecular function categories, the most common GO term among the DEGs was pectinesterase activity, followed by aspartyl esterase activity and enzyme inhibitor activity.

The enriched metabolic pathways among the DEGs were identified with the KEGG pathway database. Specifically, 37 DEGs were assigned to 15 KEGG pathways (**Figure 4B**), including biosynthesis of secondary metabolites, phenylalanine metabolism, and phenylpropanoid biosynthesis.

miRNA-Gene Regulatory Network Analysis

To investigate the functions of differentially expressed miRNAs and miRNA target genes, a regulatory network was built for the miRNAs and target genes on the basis of the enriched GO terms and KEGG pathways. **Figure 5** presents the regulatory network for 204 target genes, 55 miRNAs, and five GO terms (sexual reproduction, anther development, meristem maintenance, meristem initiation, and nucleus). **Figure 6** presents the regulatory network for 103 target genes, 42 miRNAs, and five KEGG pathways (ribosome, RNA degradation, RNA

transport, spliceosome, and plant hormone signal transduction). The members of the gma-miR169 family were the most common regulators of the miRNA target genes.

DISCUSSION

Identification of miRNAs in JLCMS9A and JLCMS9B Flower Buds

In plants, sRNAs are pivotal regulators of male fertility during anther and pollen development (Chambers and Shuai, 2009; Grant-Downton et al., 2009; Wang et al., 2009; Xing et al., 2010; Shen et al., 2011; Wei et al., 2012; Yang J. et al., 2013). A previous deep-sequencing study revealed the involvement of miRNAs in the flower and pollen development of a soybean CMS line and its maintainer line (Ding et al., 2016). However, the mechanism underlying the relationships among the miRNAs, target genes, and pollen development remained unclear in soybean. In the present study, to clarify the regulation of miRNAs and target genes during pollen development in soybean, CMS line JLCMS9A and its maintainer line JLCMS9B were analyzed by sRNA, RNA, and degradome sequencing. The analyzed sRNAs were mainly 24 nt long, but sRNAs that were 21 and 22 nt long were also relatively common (**Figure 2**). The sRNA lengths were consistent with the results of earlier related investigations of other plants, including *A. thaliana* (Rajagopalan et al., 2006), *Citrus sativus* (Kou et al., 2012; Fang et al., 2016), *Medicago truncatula* (Wang et al., 2011; Chen et al., 2012), *Oryza sativa* (Wei et al., 2011), and *Zea mays* (Li et al., 2013).

Several researchers identified miRNAs in diverse male sterile crops, including maize (Shen et al., 2011), rice (Zhou et al., 2012), cotton (Wei et al., 2013b; Yang X. et al., 2013; Yu et al., 2020), wheat (Bai et al., 2017; Sun et al., 2018), and *Brassica juncea* (Yang J. et al., 2013). In the current study, miRNAs and their target genes were identified following the deep-sequencing analysis of the flower buds from soybean CMS and maintainer lines. Additionally, the DEGs between the CMS and maintainer lines were identified. The 828 conserved, 491 non-conserved, and 410 novel miRNAs revealed in this study exceeded the corresponding number of miRNAs in an earlier study (Ding et al., 2016). The known miRNAs have been classified into four categories in accordance with how conserved they are (i.e., high, moderate, low, and non-conserved). In this study, we identified 828 conserved miRNAs (820 in gp1a and eight in gp1b) and 491 non-conserved miRNAs (71 in gp2a, 238 in gp2b, 20 in gp3, and 162 in gp4; **Supplementary Figure 3** and **Supplementary Table 2**). Furthermore, 410 novel miRNAs were detected (i.e., not in a public soybean miRNA database).

miRNAs May Regulate Pollen Development by Targeting Transcription Factors

In plants, miRNAs mediate gene expression at the post-transcriptional level by cleaving mRNAs at specific sites (Bartel, 2009; Shukla et al., 2011). The miRNA targets have been predicted using bioinformatics-based methods (Pla et al., 2018; Chen and Wang, 2020). Recently, a high-throughput method combining 5' RACE and next-generation sequencing technology was developed to identify miRNA targets (Meng et al., 2010). In the present study, 1,617 genes were predicted to be targeted by 349 miRNAs in JLCMS9A and JLCMS9B. The predicted target genes encoded proteins that mediate a wide range of biological processes. Most of the miRNA targets were transcription factor genes (775 genes), including an auxin response factor, an AP2-like factor, a zinc knuckle (CCHC-type) family protein, a MYB domain family protein, a NAC domain-containing protein, a TCP transcription factor, an F-box family protein, and a basic helix–loop–helix DNA-binding superfamily protein. The target genes also encoded a splicing factor, a nucleosome assembly protein, and heat shock protein 70. Interestingly, a few targets identified in our degradome analysis were previously reported to be involved in plant reproductive development, including MYB proteins controlled by miR159 (Rhoades et al., 2002), AP2-like transcription factors targeted by miRNA172 (Aukerman and Sakai, 2003; Chen, 2004), and ARF6 or ARF8 regulated by miR167 (Wu et al., 2006). In this study, we predicted that the expression of the genes for 29 MYB domain proteins, 21 AP2-like proteins, and 44 ARF6 or ARF8 proteins is regulated by miR159, miR172, and miR167, respectively. Accordingly, miRNAs may regulate pollen development by targeting transcription factor genes. Our findings are consistent with those of a previous study on cotton (Yu et al., 2020).

Abnormal Cell Wall Metabolism May Be an Important Factor Leading to Pollen Abortion in JLCMS9A

Pollen cell wall development is a crucial part of pollen production, and an abnormal pollen cell wall may be associated with male sterility in plants (Li et al., 2006; Ma et al., 2007; Wijeratne et al., 2007; Wei et al., 2013a; Hu et al., 2016; Xu et al., 2017; Zhou et al., 2017; Chen et al., 2018, 2019; Hamid et al., 2019; Nugent et al., 2019; Sutthinon et al., 2019; Mondol et al., 2020). Some genes controlling pollen development have been identified and cloned (Shi et al., 2011), including *DMD1* (Ren et al., 2020), *DPW2* (Xu et al., 2017), and *DPW3* (Mondol et al., 2020). In the current study, we identified 440 DEGs between JLCMS9A and JLCMS9B (**Supplementary Table 4**), with 428 and 12 of these genes expressed at lower and higher levels, respectively, in JLCMS9A than in JLCMS9B. The GO functional analysis assigned the 440 DEGs to 25 categories (10 biological processes, four cellular components, and 11 molecular functions; **Figure 4A**). Among the biological process categories, pectin catabolic process was the main functional term, followed by cell wall modification and catalytic activity. Regarding the cellular component categories, cell wall was the most significantly enriched functional term. Of the molecular function categories, pectinesterase activity was the main enriched GO term, followed by aspartyl esterase activity and enzyme inhibitor activity.

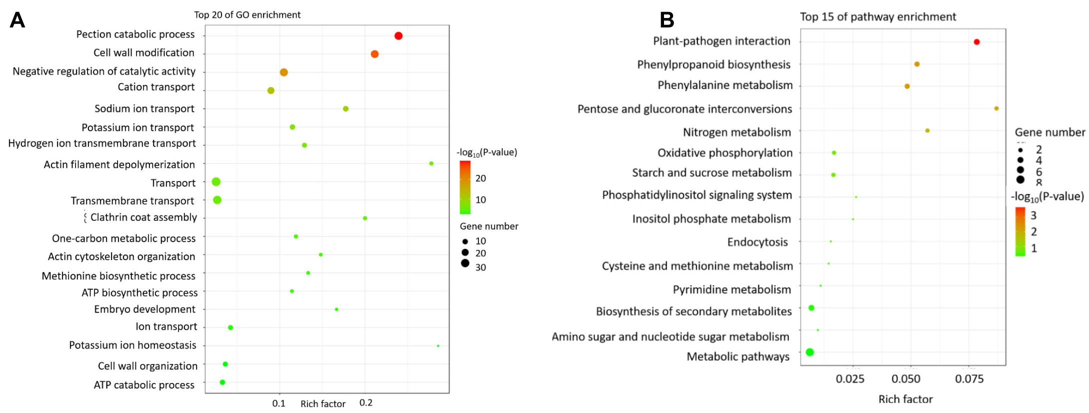


FIGURE 4 | The GO and KEGG pathway analyses of the differentially expressed genes between JLCMS9A and JLCMS9B. **(A)** Top 20 GO terms. The *x*-axis presents the GO terms, whereas the *y*-axis presents the enrichment factors. **(B)** Top 15 KEGG pathways. The *x*-axis presents the KEGG pathways, whereas the *y*-axis presents the enrichment factors.

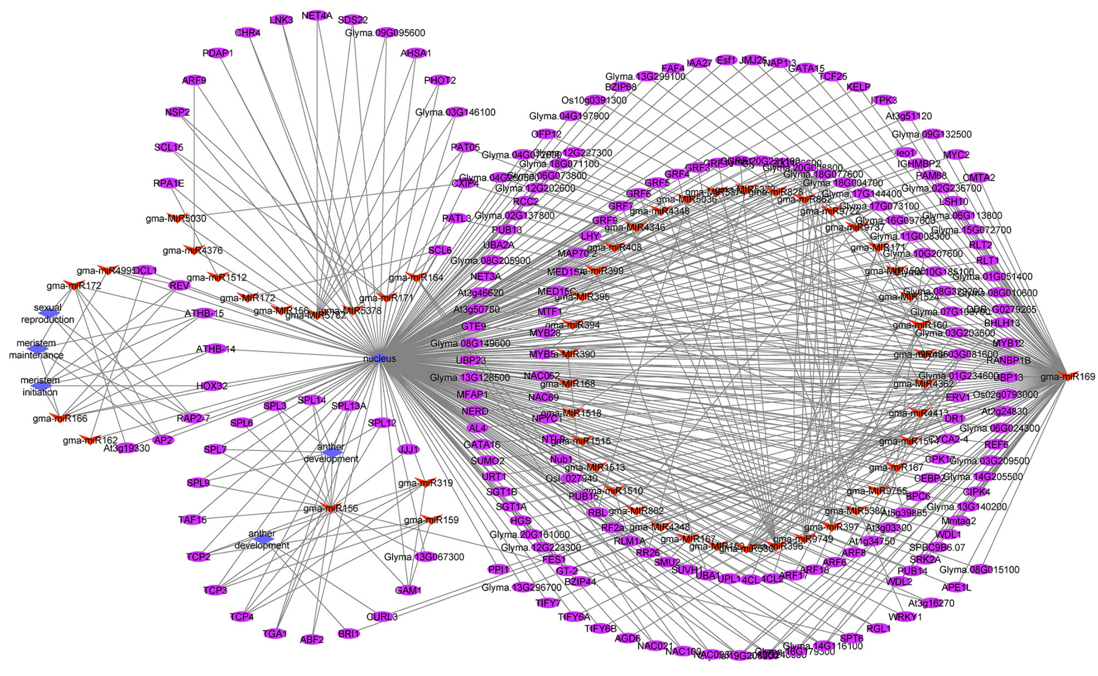


FIGURE 5 | The regulatory network comprising GO terms, microRNAs (miRNAs), and target genes in JLCMS9A and JLCMS9B. Diamonds, ellipses, and arrowheads represent the GO terms, target genes, and miRNAs, respectively.

miR169 May Regulate Pollen Development in Cytoplasmic Male Sterile Soybean

Earlier research confirmed that miR156, miR167, and miR399 contribute to pollen development in *A. thaliana* and a citrus species (Wu et al., 2006; Wang et al., 2009, 2020; Xing et al., 2010; Wei et al., 2012). To functionally characterize the differentially expressed miRNAs and miRNA target genes, we developed a regulatory network that included GO terms and KEGG pathways. The five GO terms for the regulatory network with 204 target genes and 55 miRNAs were sexual reproduction, anther development, meristem maintenance, meristem initiation, and nucleus (**Figure 5**). The five KEGG pathways in the regulatory network for 103 target genes and 42 miRNAs were ribosome, RNA degradation, RNA transport, spliceosome, and plant hormone signal transduction (**Figure 6**). The gma-miR169 family was involved in the regulation of the most miRNA target genes. Many studies proved that miR169 is a ubiquitous regulator of plant responses to various abiotic stresses (heat, cold, dehydration, and salt) and pathogens as well as developmental pathways (Liu et al., 2013; Ni et al., 2013; Potkar et al., 2013; Sorin et al., 2014; Xu et al., 2014, 2016;

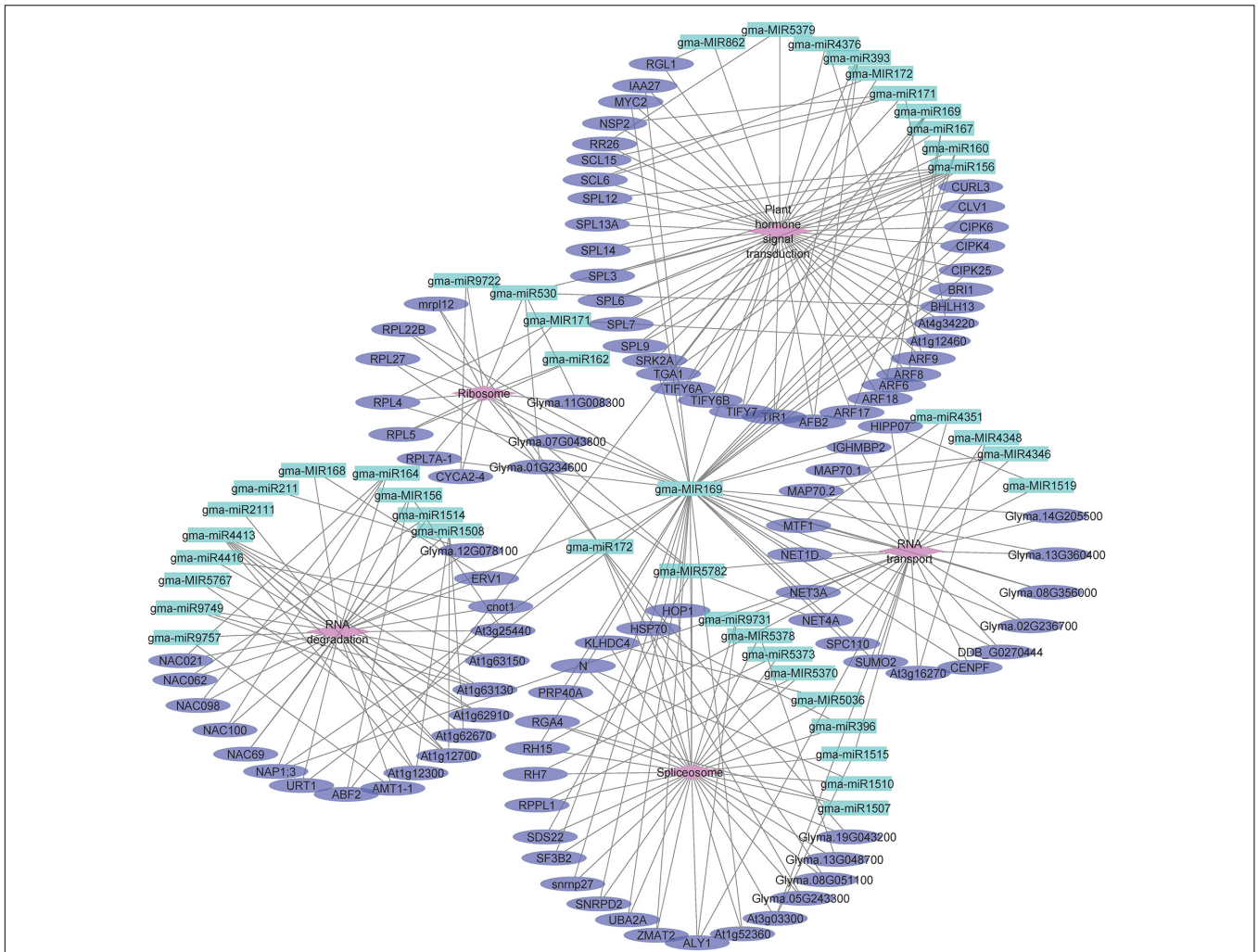


FIGURE 6 | The regulatory network comprising KEGG pathways, microRNAs (miRNAs), and target genes in JLCMS9A and JLCMS9B. Diamonds, ellipses, and arrowheads represent the KEGG pathways, target genes, and miRNAs, respectively.

Luan et al., 2015; Hanemian et al., 2016; Du et al., 2017; Li et al., 2017; Serivichyaswat et al., 2017; Rao et al., 2020). However, to the best of our knowledge, there are no reports describing the involvement of miR169 in the male sterility of soybean or other plant species.

CONCLUSION

In this study, we performed high-throughput sequencing and degradome analyses to identify miRNAs and their targets in a soybean CMS line (JLCMS9A) and its maintainer line (JLCMS9B). Additionally, DEGs during reproductive development were identified using RNA-seq data. The target genes that were revealed as differentially expressed between the CMS line and the maintainer line by an RNA-seq analysis were annotated with diverse functions related to biological processes, cellular components, and molecular functions, including transcriptional regulation, the nucleus, meristem maintenance, meristem initiation, cell differentiation, auxin-activated signaling, plant ovule development, and anther development. Finally, a network was built based on the interactions. Analyses of the miRNA, degradome, and transcriptome datasets generated in this study provided a comprehensive overview of the reproductive development of a CMS soybean line. The data presented herein represent useful information for soybean hybrid breeding. Furthermore, the study results indicate that miRNAs contribute to the soybean CMS regulatory network by modulating the expression of CMS-related genes.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The RN–CMS soybean line JLCMS9A and its maintainer line JLCMS9B were used in this study. All plants were

grown using a randomized block design (three replicates) at the Jilin Academy of Agricultural Sciences, China. More specifically, plants were cultivated in rows (5 m long and 65 cm wide), with 15 cm between plants. Mature flower buds were collected from 12 plants per genotype and stored at -80°C prior to the RNA-seq and sRNA-seq analyses, which were completed using three biological replicates per genotype.

Small RNA Library Construction and Sequencing

Total RNA was extracted using TRK-1001 (LC Sciences, Houston, TX, United States) following the manufacturer's instructions. The RNA quantity and purity were determined using the 2100 Bioanalyzer system and the RNA 6000 Nano LabChip Kit (Agilent Technologies, Santa Clara, CA, United States). High-quality RNA samples were those with an RNA integrity number greater than 7.0. Total RNA was ligated to the RNA 30 and RNA 50 adapters, then reverse transcribed and amplified by PCR to produce cDNA constructs of the sRNAs. The small cDNA fractions (22–30 nt long) were then isolated via 6% denaturing polyacrylamide gel electrophoresis. Finally, the cDNA constructs were purified, and the library was validated. We then performed single-end sequencing (50 bp) on an Illumina HiSeq 2500 system at LC-BIO (Hangzhou, China) following the vendor's recommended protocol.

Identification of Known and Potential Novel miRNAs

Raw reads were analyzed using ACGT101-miR (LC Sciences, Houston, TX, United States) to remove adapter dimers, junk reads, reads with low complexity, reads for common RNA families (rRNA, tRNA, snRNA, and snoRNA), and repeats. Unique sequences (18–25 nt long) were mapped to precursors in specific species in miRBase 21.0 on the basis of a BLAST search to identify known miRNAs and novel 3p- and 5p-derived miRNAs. Length variations at the 3' and 5' ends and one mismatch within the sequence were allowed during the alignment. The unique sequences mapped to the hairpin arm corresponding to a mature miRNA were identified as known miRNAs. The unique sequences mapped to the other hairpin arm were considered to be novel 5p- or 3p-derived miRNA candidates. The remaining sequences were mapped to precursors in other selected species in miRBase 21.0 on the basis of a BLAST search. The mapped pre-miRNAs were used as queries for a BLAST search of genomes from specific species to determine their genomic locations. The above two were designated as known miRNAs. The unmapped sequences served as queries for a BLAST search of specific genomes, and the hairpin RNA structures containing these sequences were predicted according to the 120-nt flanking sequences using the RNAfold program[2]. The following criteria were used for predicting secondary structures: (Virmani, 2012) number of nucleotides in one stem bulge $\leq$ 12 (Ivanov and Dymshits, 2007) number of base pairs in the stem region of the predicted hairpin $\geq$ 16 (Yuan, 1966) free energy (kCal/mol) cut-off $\leq$-15, (Bosacchi et al., 2015) hairpin length (upstream and downstream of stems and the terminal loop) $\geq$50 (Zhao et al., 2004) hairpin loop length $\leq$ 200 (Sun et al., 2001) number of nucleotides in one bulge in the mature region $\leq$ 4 (Banisch et al., 2012) number of biased errors in one bulge in the mature region $\leq$ 2 (Chambers and Shuai, 2009) number of biased bulges in the mature region $\leq$ 2 (Grant-Downton et al., 2009) number of errors in the mature region $\leq$ 4 (Shen et al., 2011) number of base pairs in the mature region of the predicted hairpin $\geq$ 12, and (Wang et al., 2009) proportion of the mature region in the stem $\geq$ 80%.

[2]http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Analysis of Differentially Expressed miRNAs

Differentially expressed miRNAs revealed by the normalized deep-sequencing read counts were analyzed by Student's *t*-test. The following criteria were used to identify significantly up-regulated and down-regulated miRNAs: $\log_2$ expression level fold-change $\geq$ 1 and $P < 0.05$.

The expression of six selected miRNAs was assayed in JLCMS9A and JLCMS9B using Platinum SYBR Green-based q-RT-PCR (Invitrogen, United States) with analytikjena-qTOWER2.2 (Analytik Jena, Germany). The primers of six selected miRNAs and internal control gene (U6 snRNA) are available in **Supplementary Table 7**.

Prediction of miRNA Target Genes

To predict the genes targeted by the most abundant miRNAs, computational target prediction algorithms (Target Finder) were used to identify miRNA binding sites. The predicted miRNA target genes were annotated with GO terms and assigned to KEGG pathways.

Degradome Sequencing and Target Identification

Two degradome libraries were constructed as previously described (Ma et al., 2010). Polyadenylated RNAs were obtained and ligated to a 5p adapter, after which cDNA was generated by PCR. The cDNA was purified and then sequenced using an Illumina HiSeq 2500 system (LC Sciences, Hangzhou, China). By eliminating the low-quality data, the raw reads were obtained using Illumina Pipeline (version 1.5). After removing ADTs and reads shorter than 15 nt, the remaining reads were compared with the sequences in a cDNA library in the soybean genome database. The mapped reads were then aligned with the identified miRNAs using CleaveLand 3.0 (alignment score $\leq$ 4). Furthermore, on the basis of the number of degradome sequences and their abundance values, the miRNA targets were classified into five categories (0, 1, 2, 3, and 4) as previously described. To further elucidate their potential functions, the miRNA target genes were annotated using the GO and KEGG databases.

Total RNA Extraction, cDNA Library Construction, and Ion Proton Deep Sequencing

Total RNA was extracted from each sample using TRIzol Reagent (Life Technologies, United States) according to the manufacturer's protocol. The concentration of each sample was determined using the NanoDrop 2000 spectrophotometer (Thermo Scientific, United States), whereas the quality was assessed using the Agilent 2200 TapeStation system (Agilent). A sequencing library for each RNA sample was prepared using the Ion Total RNA-Seq Kit (version 2) according to the manufacturer's protocol (Life Technologies). Briefly, polyadenylated mRNA was purified from 5 μg of total RNA using Dynabeads (Life Technologies). The mRNA was fragmented using RNase III and purified; after which, it was hybridized and ligated with an ion adapter. The RNA fragments were reverse transcribed and amplified to produce double-stranded cDNA, which was then purified using magnetic beads. After determining the molar concentration of each cDNA library, an emulsion PCR amplification was performed using the cDNA library as a template. Template-positive Ion PITM Ion Sphere^{TM} Particles were enriched and loaded onto the ion PITM chip for sequencing.

Analysis of RNA-seq Data

Raw data (raw reads) in the FASTQ format were first processed using in-house Perl scripts. During this step, clean data (clean reads) were obtained by removing reads containing adapters or poly-N sequences as well as low-quality reads. Additionally, Q20 and Q30 values and the GC content of the clean data were calculated. All downstream analyses were completed using the high-quality clean data. The reference genome and gene model annotation files available online were downloaded[3]. A reference genome index was built using Bowtie (version 2.2.3; Langmead et al., 2009), after which paired-end clean reads were aligned to the reference genome using TopHat (version 2.0.12; Trapnell et al., 2009). The TopHat program was used for mapping because it can generate databases of splice junctions from the gene model annotation files, resulting in better mapping than that produced by other non-splice mapping tools. The HTSeq program (version 0.6.1) was used to determine the number of reads mapped to each gene (Anders et al., 2015). The fragments per kilobase of transcript per million base pairs sequenced (FPKM) value was then calculated for each gene based on the gene length and the number of reads mapped to the gene. Expression analyses involving the FPKM value are very common because they simultaneously consider the effects of sequencing depth and gene length on read counts. Pearson's correlation coefficient among samples was used to evaluate the quality of the RNA-seq data.

Analysis of Differential Expression

The differential expression between two conditions was analyzed using the DEGSeq R package (version 1.20.0; Anders and Huber, 2012). The P values were adjusted according to the Benjamini–Hochberg method. A corrected P value < 0.005 and a $\log_2$ (fold-change) of 1 were set as the threshold for identifying significant DEGs. The DEGs were functionally annotated with GO terms using the GOseq R package (Young et al., 2010), which corrects the gene length bias. Specifically, DEGs were annotated with GO terms on the basis of a corrected P value < 0.05.

The KEGG database comprises molecular information, including large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. It is useful for elucidating high-level functions and biological system activities (Kanehisa et al., 2008) in a cell, organism, and ecosystem[4]. We used the KOBAS program (Mao et al., 2005) to identify the enriched KEGG pathways among the DEGs.

DATA AVAILABILITY STATEMENT

The sequencing data has been deposited into the National Genomics Data Center (accession: CRA003993; link: https://bigd.big.ac.cn/search/?dbId=gsa&q=CRA003993&page=1).

AUTHOR CONTRIBUTIONS

CZ, FF, and LZ conceived and designed the study. CZ and FF performed the experiments and wrote the manuscript. JZ and BP collected the plant materials. FF and XD analyzed and modified the data. CL, HY, PW, and WZ provided advice and assistance. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Agricultural Science and Technology Innovation Project (CXGC2017JQ004) and the Science and Technology Development Program (20190101007JH) of Jilin Province and the National Natural Science Foundation of China (31971969).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.654146/full#supplementary-material

Supplementary Figure 1 | Distribution of known microRNAs (miRNAs) in JLCMS9A and JLCMS9B.

Supplementary Figure 2 | Number of identified microRNAs (miRNAs) in each family.

[3]http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax

[4]http://www.genome.jp/kegg/

Supplementary Figure 3 | Number of identified conserved microRNAs (miRNAs).

Supplementary Figure 4 | Detection of selected miRNAs expression in JLCMS9A and JLCMS9B using q-RT-PCR. U6 was chosen as an endogenous control. The reulsts were obtained from three biological replicates with three technical replicates and the error bars indicated the standard error of the mean.

Supplementary Table 1 | Overview of small RNA sequencing reads from raw data to clean reads.

Supplementary Table 2 | Summary of known and predicted microRNAs (miRNAs) in this study.

Supplementary Table 3 | Distribution of microRNA family members.

Supplementary Table 4 | Summary of the identified microRNAs (miRNAs) and their families in JLCMS9A and JLCMS9B flower buds.

Supplementary Table 5 | Overview of the degradome sequences from JLCMS9A and JLCMS9B.

Supplementary Table 6 | microRNA (miRNA) target genes in JLCMS9A and JLCMS9B.

Supplementary Table 7 | The primers sequence of in this study.

REFERENCES

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* 18, 758–762. doi: 10.1016/j.cub.2008.04.042

Addo-Quaye, C., Miller, W., and Axtell, M. J. (2009). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* 25, 130–131. doi: 10.1093/bioinformatics/btn604

Anders, S., and Huber, W. (2012). *Differential Expression of RNA-Seq Data at the Gene Level–the DESeq Package*. Heidelberg: European Molecular Biology Laboratory (EMBL).

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Aukerman, M. J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell.* 15, 2730–2741. doi: 10.1105/tpc.016238

Bai, J. F., Wang, Y. K., Wang, P., Duan, W. J., Yuan, S. H., Sun, H., et al. (2017). Uncovering male fertility transition responsive miRNA in a wheat photo-thermosensitive genic male sterile line by deep sequencing and degradome analysis. *Front. Plant Sci.* 8:1370. doi: 10.3389/fpls.2017.01370

Banisch, T. U., Goudarzi, M., and Raz, E. (2012). Small RNAs in germ cell development. *Curr. Top. Dev. Biol.* 99, 79–113. doi: 10.1016/b978-0-12-387038-4.00004-5

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Bosacchi, M., Gurdon, C., and Maliga, P. (2015). Plastid genotyping reveals the uniformity of cytoplasmic male sterile-T maize cytoplasms. *Plant Physiol.* 169, 2129–2137. doi: 10.1104/pp.15.01147

Chambers, C., and Shuai, B. (2009). Profiling microRNA expression in *Arabidopsis* pollen using microRNA array and real-time PCR. *BMC Plant Biol.* 9:87. doi: 10.1186/1471-2229-9-87

Chen, G., Ye, X., Zhang, S., Zhu, S., Yuan, L., Hou, J., et al. (2018). Comparative transcriptome analysis between fertile and CMS flower buds in wucai (*Brassica campestris* L.). *BMC Genomics* 19:908. doi: 10.1186/s12864-018-5331-4

Chen, H., Jin, J., Zhang, H., Wang, Y., Li, Q., Zou, Y., et al. (2019). Comparative analysis of proteomics and transcriptomics during fertility transition in a two-line hybrid rice line wuxiang S. *Int. J. Mol. Sci.* 20:4542. doi: 10.3390/ijms20184542

Chen, L., Wang, T., Zhao, M., Tian, Q., and Zhang, W. H. (2012). Identification of aluminum-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *Planta* 235, 375–386. doi: 10.1007/s00425-011-1514-9

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* 303, 2022–2025. doi: 10.1126/science.1088060

Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48, D127–D131. doi: 10.1093/nar/gkz757

Ding, X., Li, J., Zhang, H., He, T., Han, S., Li, Y., et al. (2016). Identification of miRNAs and their targets by high-throughput sequencing and degradome analysis in cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B of soybean. *BMC Genomics* 17:24. doi: 10.1186/s12864-015-2352-0

Ding, X., Ruan, H., Yu, L., Li, Q., Song, Q., Yang, S., et al. (2020). miR156b from soybean CMS line modulates floral organ development. *J. Plant Biol.* 63, 141–153. doi: 10.1007/s12374-020-09237-7

Ding, X., Zhang, H., Ruan, H., Li, Y., Chen, L., Wang, T., et al. (2019). Exploration of miRNA-mediated fertility regulation network of cytoplasmic male sterility during flower bud development in soybean. *3 Biotech* 9:22. doi: 10.1007/s13205-018-1543-1

Du, Q., Zhao, M., Gao, W., Sun, S., and Li, W. X. (2017). microRNA/microRNA* complementarity is important for the regulation pattern of NFYA5 by miR169 under dehydration shock in *Arabidopsis*. *Plant J.* 91, 22–33. doi: 10.1111/tpj.13540

Fang, Y. N., Zheng, B. B., Wang, L., Yang, W., Wu, X. M., Xu, Q., et al. (2016). High-throughput sequencing and degradome analysis reveal altered expression of miRNAs and their targets in a male-sterile cybrid pummelo (Citrus grandis). *BMC Genomics* 17:591. doi: 10.1186/s12864-016-2882-0

Goette, W., Liu, Z. R., Xia, J., Zhang, W. X., and Zhao, P. X. (2014). Systems and evolutionary characterization of microRNAs and their underlying regulatory networks in soybean cotyledons. *PLoS One* 9:e86153. doi: 10.1371/journal.pone.0086153

Grant-Downton, R., Le Trionnaire, G., Schmid, R., Rodriguez-Enriquez, J., Hafidh, S., Mehdi, S., et al. (2009). MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*. *BMC Genomics* 10:643. doi: 10.1186/1471-2164-10-643

Hamid, R., Marashi, H., Tomar, R. S., Malekzadeh Shafaroudi, S., and Sabara, P. H. (2019). Transcriptome analysis identified aberrant gene expression in pollen developmental pathways leading to CGMS in cotton (*Gossypium hirsutum* L.). *PLoS One* 14:e0218381. doi: 10.1371/journal.pone.0218381

Hanemian, M., Barlet, X., Sorin, C., Yadeta, K. A., Keller, H., Favery, B., et al. (2016). *Arabidopsis* CLAVATA1 and CLAVATA2 receptors contribute to *Ralstonia solanacearum* pathogenicity through a miR169-dependent pathway. *New Phytol.* 211, 502–515. doi: 10.1111/nph.13913

Hu, J., Chen, G., Zhang, H., Qian, Q., and Ding, Y. (2016). Comparative transcript profiling of alloplasmic male-sterile lines revealed altered gene expression related to pollen development in rice (*Oryza sativa* L.). *BMC Plant Biol.* 16:175. doi: 10.1186/s12870-016-0864-7

Ivanov, M. K., and Dymshits, G. M. (2007). Dymshits: cytoplasmic male sterility and restoration of pollen fertility in higher plants. *Russ. J. Genet.* 43, 354–368. doi: 10.1134/S1022795407040023

Joshi, T., Yan, Z., Libault, M., Jeong, D. H., Park, S., and Green, P. J. (2010). Prediction of novel miRNAs and associated target genes in Glycine max. *BMC Bioinformatics* 11:S14. doi: 10.1186/1471-2105-11-s1-s14

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

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

Kou, S. J., Wu, X. M., Liu, Z., Liu, Y. L., Xu, Q., and Guo, W. W. (2012). Selection and validation of suitable reference genes for miRNA expression normalization by quantitative RT-PCR in citrus somatic embryogenic and adult tissues. *Plant Cell Rep.* 31, 2151–2163. doi: 10.1007/s00299-012-1325-x

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Li, D., Wang, L., Liu, X., Cui, D., Chen, T., Zhang, H., et al. (2013). Deep sequencing of maize small RNAs reveals a diverse set of microRNA in dry and imbibed seeds. *PLoS One* 8:e55107. doi: 10.1371/journal.pone.0055107

Li, N., Zhang, D. S., Liu, H. S., Yin, C. S., Li, X. X., Liang, W. Q., et al. (2006). The rice tapetum degeneration retardation gene is required for tapetum degradation and anther development. *Plant Cell* 18, 2999–3014. doi: 10.1105/tpc.106.044107

Li, Y., Zhao, S. L., Li, J. L., Hu, X. H., Wang, H., Cao, X. L., et al. (2017). Osa-miR169 negatively regulates rice immunity against the blast fungus *Magnaporthe oryzae*. *Front. Plant Sci.* 8:2. doi: 10.3389/fpls.2017.00002

Li, Y. F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., et al. (2010). Transcriptome-wide identification of microRNA targets in rice. *Plant J.* 62, 742–759. doi: 10.1111/j.1365-313X.2010.04187.x

Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., et al. (2014). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC Genomics* 15:25. doi: 10.1186/1471-2164-15-25

Liu, Z. X., Zeng, C. Z., and Tan, X. F. (2013). [Molecular evolution of the poplar MIR169 gene family]. *Yi Chuan* 35, 1307–1316. doi: 10.3724/sp.j.1005.2013.01307

Luan, M., Xu, M., Lu, Y., Zhang, L., Fan, Y., and Wang, L. (2015). Expression of zma-miR169 miRNAs and their target ZmNF-YA genes in response to abiotic stress in maize leaves. *Gene* 555, 178–185. doi: 10.1016/j.gene.2014.11.001

Ma, X., Xing, C., Guo, L., Gong, Y., Wang, H., Zhao, Y., et al. (2007). Analysis of differentially expressed genes in genic male sterility cotton (*Gossypium hirsutum* L.) using cDNA-AFLP. *J. Genet. Genomics* 34, 536–543. doi: 10.1016/S1673-8527(07)60059-9

Ma, Z., Coruh, C., and Axtell, M. J. (2010). *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* 22, 1090–1103. doi: 10.1105/tpc.110.073882

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Meng, Y., Gou, L., Chen, D., Wu, P., and Chen, M. (2010). High-throughput degradome sequencing can be used to gain insights into microRNA precursor metabolism. *J. Exp. Bot.* 61, 3833–3837. doi: 10.1093/jxb/erq209

Mondol, P. C., Xu, D., Duan, L., Shi, J., Wang, C., Chen, X., et al. (2020). Defective Pollen Wall 3 (DPW3), a novel alpha integrin-like protein, is required for pollen wall formation in rice. *New Phytol.* 225, 807–822. doi: 10.1111/nph.16161

Ni, Z., Hu, Z., Jiang, Q., and Zhang, H. (2013). GmNFYA3, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. *Plant Mol. Biol.* 82, 113–129. doi: 10.1007/s11103-013-0040-5

Nugent, J. M., Byrne, T., McCormack, G., Quiwa, M., and Stafford, E. (2019). Progressive programmed cell death inwards across the anther wall in male sterile flowers of the gynodioecious plant Plantago lanceolata. *Planta* 249, 913–923. doi: 10.1007/s00425-018-3055-y

Pla, A., Zhong, X., and Rayner, S. (2018). miRAW: a deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. *PLoS Comput. Biol.* 14:e1006185. doi: 10.1371/journal.pcbi.1006185

Potkar, R., Recla, J., and Busov, V. (2013). ptr-MIR169 is a posttranscriptional repressor of PtrHAP2 during vegetative bud dormancy period of aspen (Populus tremuloides) trees. *Biochem. Biophys. Res. Commun.* 431, 512–518. doi: 10.1016/j.bbrc.2013.01.027

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

Rao, S., Balyan, S., Jha, S., and Mathur, S. (2020). Novel insights into expansion and functional diversification of MIR169 family in tomato. *Planta* 251:55. doi: 10.1007/s00425-020-03346-w

Ren, L., Zhao, T., Zhang, L., Du, G., Shen, Y., Tang, D., et al. (2020). Defective microspore development 1 is required for microspore cell integrity and pollen wall formation in rice. *Plant J.* 103, 1446–1459. doi: 10.1111/tpj.14811

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520. doi: 10.1016/S0092-8674(02)00863-2

Serivichyaswat, P. T., Susila, H., and Ahn, J. H. (2017). Elongated Hypocotyl 5-Homolog (HYH) negatively regulates expression of the ambient temperature-responsive MicroRNA gene MIR169. *Front Plant Sci.* 8:2087. doi: 10.3389/fpls.2017.02087

Shen, Y., Zhang, Z., Lin, H., Liu, H., Chen, J., Peng, H., et al. (2011). Cytoplasmic male sterility-regulated novel microRNAs from maize. *Funct. Integr. Genomics* 11, 179–191. doi: 10.1007/s10142-010-0202-3

Shi, J., Tan, H., Yu, X. H., Liu, Y., Liang, W., Ranathunge, K., et al. (2011). Defective pollen wall is required for anther and microspore development in rice and encodes a fatty acyl carrier protein reductase. *Plant Cell* 23, 2225–2246. doi: 10.1105/tpc.111.087528

Shukla, G. C., Singh, J., and Barik, S. (2011). MicroRNAs: processing, maturation, target recognition and regulatory functions. *Mol. Cell Pharmacol.* 3, 83–92. doi: 10.4255/mcpharmacol.11.13

Song, Q. X., Liu, Y. F., Hu, X. Y., Zhang, W. K., Ma, B., and Chen, S. Y. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol.* 11:5. doi: 10.1186/1471-2229-11-5

Sorin, C., Declerck, M., Christ, A., Blein, T., Ma, L., Lelandais-Briere, C., et al. (2014). A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. *New Phytol.* 202, 1197–1211. doi: 10.1111/nph.12735

Sun, H., Zhao, L. M., and Huang, M. (2001). *Cytoplasmic-genetic Male Sterile Soybean and Method for Producing Hybrid Soybean*. United State Patent. Application No. US 6,320,098B1, 2001-11-20.

Sun, L., Sun, G., Shi, C., and Sun, D. (2018). Transcriptome analysis reveals new microRNAs-mediated pathway involved in anther development in male sterile wheat. *BMC Genomics* 19:333. doi: 10.1186/s12864-018-4727-5

Sutthinon, P., Samuels, L., and Meesawat, U. (2019). Pollen development in male sterile mangosteen (*Garcinia mangostana* L.) and male fertile seashore mangosteen (*Garcinia celebica* L.). *Protoplasma* 256, 1545–1556. doi: 10.1007/s00709-019-01397-9

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Virmani, S. S. (2012). *Heterosis and Hybrid Rice Breeding*. Berlin: Springer Science & Business Media.

Wang, J. W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Wang, R., Fang, Y. N., Wu, X. M., Qing, M., Li, C. C., Xie, K. D., et al. (2020). The miR399- CsUBC24 module regulates reproductive development and male fertility in citrus. *Plant Physiol.* 183, 1681–1695. doi: 10.1104/pp.20.00129

Wang, T., Chen, L., Zhao, M., Tian, Q., and Zhang, W. H. (2011). Identification of drought-responsive microRNAs in Medicago truncatula by genome-wide high-throughput sequencing. *BMC Genomics* 12:367. doi: 10.1186/1471-2164-12-367

Wei, L. Q., Yan, L. F., and Wang, T. (2011). Deep sequencing on genome-wide scale reveals the unique composition and expression patterns of microRNAs in developing pollen of *Oryza sativa*. *Genome Biol.* 12:R53. doi: 10.1186/gb-2011-12-6-r53

Wei, M., Song, M., Fan, S., and Yu, S. (2013a). Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. *BMC Genomics* 14:97. doi: 10.1186/1471-2164-14-97

Wei, M., Wei, H., Wu, M., Song, M., Zhang, J., Yu, J., et al. (2013b). Comparative expression profiling of miRNA during anther development in genetic male sterile and wild type cotton. *BMC Plant Biol.* 13:66. doi: 10.1186/1471-2229-13-66

Wei, S., Gruber, M. Y., Yu, B., Gao, M. J., Khachatourians, G. G., Hegedus, D. D., et al. (2012). *Arabidopsis mutant* sk156 reveals complex regulation of SPL15 in a miR156-controlled gene network. *BMC Plant Biol.* 12:169. doi: 10.1186/1471-2229-12-169

Wijeratne, A. J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., et al. (2007). Differential gene expression in *Arabidopsis* wild-type and mutant anthers:

insights into anther cell differentiation and regulatory networks. *Plant J.* 52, 14–29. doi: 10.1111/j.1365-313X.2007.03217.x

Wu, M. F., Tian, Q., and Reed, J. W. (2006). *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* 133, 4211–4218. doi: 10.1242/dev.02602

Xing, S., Salinas, M., Hohmann, S., Berndtgen, R., and Huijser, P. (2010). miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in *Arabidopsis*. *Plant Cell* 22, 3935–3950. doi: 10.1105/tpc.110.079343

Xu, D., Shi, J., Rautengarten, C., Yang, L., Qian, X., Uzair, M., et al. (2017). Defective pollen wall 2 (DPW2) encodes an Acyl transferase required for rice pollen development. *Plant Physiol.* 173, 240–255. doi: 10.1104/pp.16.00095

Xu, M. Y., Zhang, L., Li, W. W., Hu, X. L., Wang, M. B., Fan, Y. L., et al. (2014). Stress-induced early flowering is mediated by miR169 in *Arabidopsis thaliana*. *J. Exp. Bot.* 65, 89–101. doi: 10.1093/jxb/ert353

Xu, M. Y., Zhu, J. X., Zhang, M., and Wang, L. (2016). Advances on plant miR169/NF-YA regulation modules. *Yi Chuan* 38, 700–706. doi: 10.16288/j.yczz.15-526

Yang, J., Liu, X., Xu, B., Zhao, N., Yang, X., and Zhang, M. (2013). Identification of miRNAs and their targets using high-throughput sequencing and degradome analysis in cytoplasmic male-sterile and its maintainer fertile lines of *Brassica juncea*. *BMC Genomics* 14:9. doi: 10.1186/1471-2164-14-9

Yang, X., Wang, L., Yuan, D., Lindsey, K., and Zhang, X. (2013). Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis. *J. Exp. Bot.* 64, 1521–1536. doi: 10.1093/jxb/ert013

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

Yu, D., Li, L., Wei, H., and Yu, S. (2020). Identification and profiling of microRNAs and differentially expressed genes during anther development between a genetic male-sterile mutant and its wildtype cotton via high-throughput RNA sequencing. *Mol. Genet. Genomics* 295, 645–660. doi: 10.1007/s00438-020-01656-y

Yuan, L. (1966). A preliminary report on male sterility in rice. *Oryza sativa* L. *Chin Sci. Bull.* 7:322.

Zhang, B. H., Pan, X. P., Cannon, C. H., Cobb, G. P., and Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant J.* 46, 243–259. doi: 10.1111/j.1365-313X.2006.02697.x

Zhao, L., Sun, H., Wang, S., Wang, Y., Huang, M., and Li, J. (2004). A breeding report of hybrid soybean 'HybSoy 1'. *Chinese J. Soil Crop Sci.* 26, 15–17. doi: 10.3321/j.issn:1007-9084.2004.03.004

Zhou, H., Liu, Q., Li, J., Jiang, D., Zhou, L., Wu, P., et al. (2012). Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. *Cell Res.* 22, 649–660. doi: 10.1038/cr.2012.28

Zhou, X., Liu, Z., Ji, R., and Feng, H. (2017). Comparative transcript profiling of fertile and sterile flower buds from multiple-allele-inherited male sterility in Chinese cabbage (*Brassica campestris* L. ssp. pekinensis). *Mol. Genet. Genomics* 292, 967–990. doi: 10.1007/s00438-017-1324-2

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fgene.2021.669959

Comparative Genomic Analysis of *TCP* Genes in Six Rosaceae Species and Expression Pattern Analysis in *Pyrus bretschneideri*

*Yu Zhao[1†], Xueqiang Su[2†], Xinya Wang[1], Mengna Wang[1], Xujing Chi[1], Muhammad Aamir Manzoor[1], Guohui Li[1] and Yongping Cai[1]**

[1] School of Life Sciences, Anhui Agricultural University, Hefei, China, [2] Institute of Sericulture, Anhui Academy of Agricultural Sciences, Hefei, China

OPEN ACCESS

Edited by:
Yunpeng Cao,
Central South University Forestry and Technology, China

Reviewed by:
Awais Shakoor,
Universitat de Lleida, Spain
Wu Yuejin,
Institute of Intelligent Machines, Hefei Institutes of Physical Science (CAS), China
Ling Xu,
Zhejiang Sci-Tech University, China
Hui Song,
Qingdao Agricultural University, China

***Correspondence:**
Yongping Cai
ypcaiah@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *19 February 2021*
Accepted: *19 April 2021*
Published: *17 May 2021*

Citation:
Zhao Y, Su X, Wang X, Wang M, Chi X, Aamir Manzoor M, Li G and Cai Y (2021) Comparative Genomic Analysis of TCP Genes in Six Rosaceae Species and Expression Pattern Analysis in Pyrus bretschneideri. Front. Genet. 12:669959. doi: 10.3389/fgene.2021.669959

TCP is a plant-specific transcription factor that plays an important role in flowering, leaf development and other physiological processes. In this study, we identified a total of 155 *TCP* genes: 34 in *Pyrus bretschneideri*, 19 in *Fragaria vesca*, 52 in *Malus domestica*, 19 in *Prunus mume*, 17 in *Rubus occidentalis* and 14 in *Prunus avium*. The evolutionary relationship of the *TCP* gene family was examined by constructing a phylogenetic tree, tracking gene duplication events, performing a sliding window analysis. The expression profile analysis and qRT-PCR results of different tissues showed that *PbTCP10* were highly expressed in the flowers. These results indicated that *PbTCP10* might participated in flowering induction in pear. Expression pattern analysis of different developmental stages showed that *PbTCP14* and *PbTCP15* were similar to the accumulation pattern of fruit lignin and the stone cell content. These two genes might participate in the thickening of the secondary wall during the formation of stone cells in pear. Subcellular localization showed that *PbTCPs* worked in the nucleus. This study explored the evolution of *TCP* genes in six Rosaceae species, and the expression pattern of *TCP* genes in different tissues of "Dangshan Su" pear. Candidate genes related to flower induction and stone cell formation were identified. In summary, our research provided an important theoretical basis for improving pear fruit quality and increasing fruit yield by molecular breeding.

Keywords: *TCP* genes, flowers, development of fruit, expression patterns, genome-wide

INTRODUCTION

TCP (TEOSINTE BRANCHED I, CYCLOIDEA, PROLIFERATING CELL FACTOR I) transcription factors are unique to plants and play an important role in all aspects of plant growth and development (Uberti-Manassero et al., 2016; Lucero et al., 2017). The amino acid sequences encoded by members of the *TCP* family generally have a basic helix loop helix structure. The second helical region has a specific LXXLL motif, which can interact with DNA or protein. Based on their structures, the *TCP* family can be divided into two subfamilies. Class I, the TCP-P subfamily, is also called PCF subfamily. Class II, the TCP-C subfamily, includes CYC/TB1, and CIN. The most significant difference between the two subfamilies is that PCF subfamily lacks four amino acids in

the basic region, and the members of CYC/TB1 subfamily specifically contain a hydrophilic α helix (R domain) rich in polar amino acids which does not exist in other members (Cubas et al., 1999).

The *TCP* gene was first identified in maize (*Zea mays*) (teosinte branched 1, *TB1*), snapdragon (*Antirrhinum majus*) (cycloidea, *CYC*) and rice (*Oryza sativa*) (proliferating cell factors 1 and 2, *PCF1/PCF2*) (Luo et al., 1996; Doebley et al., 1997; Kosugi and Ohashi, 1997; Cubas et al., 1999). Class I transcription factors can promote cell differentiation and plant growth (Aguilar-Martinez and Sinha, 2013). For example, *TCP14* and *TCP15* can regulate *Arabidopsis* seed germination by activating the gibberellin-dependent cell cycle (Resentini et al., 2015). At the same time, it has been reported that *TCP14* and *TCP15* regulate cell proliferation in leaves and flowers, thus affecting the length between nodes and leaf traits (Kieffer et al., 2011). Overexpression of *TCP16* can regulate the process of plant differentiation, resulting in the formation of ectopic meristems (Uberti-Manassero et al., 2016). Class II, compared with Class I, mainly inhibit cell differentiation and plant growth (Manassero et al., 2013; Huang and Irish, 2015). The *CYC* gene affects the symmetry of flowers in many plants, such as *Antirrhinum majus* (Luo et al., 1996, 1999), *Lotus corniculatus* (Feng et al., 2006), and *Gerbera happipot* (Broholm et al., 2008). It inhibits the formation of floral organs by inhibiting the differentiation of cells, and ultimately affects floral symmetry. The transcription factors of the CIN subfamily can regulate the development of plant leaves. Compared with the wild type, the leaf area of snapdragon mutant (*cin*) and *Arabidopsis* mutant (*cin*) increased, and the leaves were curled and wrinkled (Nath et al., 2003; Palatnik et al., 2003). There were many leaflets on the compound leaves of the tomato mutant (*cin*), and the excessive growth of the leaf edge caused bending deformation (Ori et al., 2007). The *TCP* gene of maize (*TB1*) and *Arabidopsis* (*BRC1*) can inhibit the growth of axillary buds and reduce the number of branches above ground (Hubbard et al., 2002; Aguilar-Martinez and Sinha, 2013).

Flowering is an important life activity in plants and is a key step in the transformation from vegetative growth to reproductive growth. The *TCP* transcription factor plays an important role in flower induction (Zhao et al., 2018; Li et al., 2019). Previous studies found that *TCP15* can regulate flowering by binding to the promoter of SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (*SOC1*) (Lucero et al., 2017). In contrast to *TCP15*, *TCP20,* and *TCP22* delay flowering by CIRCADIAN CLOCK ASSOCIATED 1 (*CCA1*) (Wu et al., 2016). CIN-TCP subfamily, represented by *TCP4*, can interact with FLOWERING BHLH (*FBH*) and its co-promoter to regulate the flowering process (Liu et al., 2017). *TCP5* can regulate petal growth by ethylene (Van Es et al., 2018). In conclusion, two subfamily members of the *TCP* are involved in regulating flower growth and development (Li et al., 2019). Among the Rosaceae species, fruit trees make up for the majority. Flowering is the starting point of the reproductive stage of fruit trees. The quantity and quality of flowering is an important factor directly affecting the yield of fruit.

With a long history of cultivation, "Dangshan Su" pear (*Pyrus bretschneideri* cv. Dangshan Su) is one of the most important pear resources in China and occupies an important position in the fruit market (Su et al. 2019). The stone cell content and the size in pear are important factors affecting the quality of fruit (Zhang et al., 2017). Thickening of the secondary wall is an important step in the formation of stone cells (Cheng et al., 2019). Therefore, the thickening of the secondary wall and the deposition of lignin have a great influence on the quality of pear. In the previous study, *TCP4* can activate the promoter of *VND7* to increase the formation of the secondary wall and up-regulate genes related to lignin and cellulose synthesis (Nag et al., 2009). *TCP24* negatively regulates the secondary wall thickness of anther endothecium, resulting in anther dehiscence and pollen release and eventually male sterility (Wang et al., 2015). *GbTCP5* is involved in the formation of secondary wall (Wang et al., 2020). Up-regulation of *GhTCP4* expression in cotton can activate the synthesis of secondary walls in fibrocyte, thus obtaining fiber with thicker cell walls (Cao et al., 2020). In conclusion, we speculate that the *TCP* family members may be involved in flower induction and stone cell formation during fruit development of "Dangshan Su" pear. Systematic study of the *TCP* family is of great importance for improving pear fruit quality.

Although the identification and functions of *TCP* genes have been studied in *Arabidopsis*, snapdragon, the *TCP* genes in pear remain unstudied. In this study, 155 *TCP* genes were identified in pear (*Pyrus bretschneideri*), strawberry (*Fragaria vesca*), plum (*Prunus mume*), raspberry (*Rubus occidentalis*), cherry (*Prunus avium*), apple (*Malus domestica*). The phylogenetic relationships of *TCP* genes in six Rosaceae species were elucidated by constructing phylogenetic tree, tracking gene duplication events, performing a sliding window analysis. Candidate genes related to flowering regulation (*PbTCP10*) were identified by qRT-PCR and expression profile analysis. In addition, based on the analysis of expression patterns in pear and bioinformatics analysis results, we predicted that *PbTCP14* and *PbTCP15* were the key factors of pear fruit stone cell development. This study provided important theoretical basis and gene resources for improving pear fruit quality.

MATERIALS AND METHODS

Identification of *TCP* Genes in Rosaceae

In this study, the *Pyrus bretschneideri* genome was downloaded from GIGADB datasets[1]. In addition, Rosaceae genomes (*Fragaria vesca*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, *Prunus mume*) were obtained from the following website (see text footnote 1)[2]. Bioedit software was used to construct the local protein database. The conserved domain of *TCP* was used as the query sequence for Blastp search ($E = 0.001$) from the local protein database (**Supplementary Table 1**). The SMART online software was used to search and analyze *TCP* conserved region (Letunic et al., 2012). ExPASY online website was used to predict the molecular weight and basic information of *TCP* genes (Artimo et al., 2012). Wolf PSORT was

[1]http://gigadb.org/dataset/10008

[2]https://www.rosaceae.org/

used to the predicted subcellular localization of all *TCP* genes[3] (Horton et al., 2007). Blast2GO sofware was used to implement Gene Ontology (GO) annotation analysis. Visualization of GO classifcations was used the WEGO online tool (Ye et al., 2006). The data of different tissues of Chinese white pear were downloaded from NCBI under the following accession numbers SRR8119889, SRR8119890, SRR8119891, SRR8119892, SRR8119893, SRR8119894, SRR8119895, SRR8119896, SRR8119897, SRR8119898, SRR8119899, SRR8119900, and SRR8119901 (Cao et al., 2019).

Phylogenetic Construction and Conserved Structure Analysis of *TCP* Genes

All TCP proteins sequenced were analyzed by ClustalW tool in MEGA7.0 software. The phylogenetic tree was constructed by MEGA7.0 software with the Neighbor-Joining method and other default parameters (Kumar et al., 2016). The *TCP* genes of *Arabidopsis* were obtained from previous study (Yao et al., 2007). Subsequently, the TCP protein sequence was used to obtain the conserved motif region by MEME online software (Bailey et al., 2015). In the conservative region prediction, we chose the interval range of 6–200, and the number of conservative regions was generally not less than 20.

Chromosomal Localization and Gene Duplication Events

The chromosome information of six Rosaceae species was obtained from the public genomic database, and MapInspect software was used to display the members of *TCP* gene family on their respective chromosomes (Ma et al., 2015; Zhu et al., 2015). The determination of gene duplication events mainly depended on the following principles: (1) Two genes were located in the same branch of the evolutionary tree, and the similarity of amino acid sequence was more than 80% (2) Two genes were located on the same chromosome and the distance between them was at least 200 kb, we considered these two genes tandem duplicated events (3) Two genes located on different chromosomes were defined as fragment duplication events (4) The non-synonymous substitution (Ka) and synonymous substitution (Ks) values of a replicated gene pair were calculated by DnaSP v5.0 software (Ka/Ks $>$ 1 was positive selection, Ka/Ks $<$ 1 was purification selection, Ka/Ks = 1 was neutral selection). Finally, DnaSP v5.0 software was used to analyze the gene duplication events by sliding window to determine the selection modest each amino acid site (Librado and Rozas, 2009). The specific parameters were as follows: the window size was 150 bp, and each step moved 9 bp.

Chinese White Pear *TCP* Gene Promoter *cis*-Acting Element Analysis

We obtained the promoter sequence of *TCP* genes from the pear genome database. In the database, we found promoter about 1,500–2,000 bp upstream of the initiation codon (ATG) of each *TCP* genes. The online Plantcare database was used to analyze *cis*-acting elements[4] (Rombauts et al., 1999).

RNA Extraction and qRT-PCR Analysis

The plant material was collected from the "Dangshan Su" pear, which grown in the Dangshan County (Anhui Province, China). The fruits samples were taken on 15, 39, 47, 55, 63, 79, and 102 DAP (days after pollination), as well as the other tissue samples such as flowers, stems, and leaves were also collected on the same year. The 102 DAP fruit was used for expression analysis in different tissues. The buds of "Dangshan Su" pear were treated with gibberellin (GA) (700 mg·L^{-1}). Then, the samples of 0, 2, 4, 6, 8, and 12 HPT (h post-treatment) were collected and stored at −80°C. Finally, the RNA was extracted using a plant RNA extraction kit from Tiangen (Beijing, China). Reverse transcribed by PrimeScript^{TM} RT reagent kit (Takara, Kusatsu, Japan) and each reaction consisted of 1 μg of RNA. The qRT-PCR primers of *TCP* genes were designed by Beacon Designer 7 (**Supplementary Table 2**). The qRT-PCR system consisted of 10 μL SYBR Premix Ex TaqTM II, 2 μL cDNA, 6.4 μL water, and 0.8 μL forward primer and reverse primer. The pear *Tubulin* gene (AB239680.1) used as an internal reference (Su et al. 2019). Introduction manual used for the procedure and repeat 3 times for each sample. The relative expression levels were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Subcellular Localization of *PbTCP6, 13,* and *17*

Full length sequence specific primers and primers with restriction sites were designed using Primer Premier 5.0 software based on the full-length sequences of *PbTCP6*, *13,* and *17* (*PbTCP6*, *PbTCP13,* and *PbTCP17* both used *Ncob* I and *Spe*I restriction endonuclease sites) (**Supplementary Table 3**). "Dangshan Su" pear fruit cDNA as template was used. Finally, each gene fragment was ligated into the pCAMBIA1304 (GenBank: AF234300.1) vector used T_4 DNA ligase (Takara, China) at 16°C for 3 h to obtain complete pCAMBIA1304-*PbTCP6*, *13,* and *17* recombinant plasmids.

The pCAMBIA1304-*PbTCP6*, *13,* and *17* recombinant plasmid, and pCAMBIA1304 empty plasmid *Agrobacterium tumefaciens* were cultured. Then mixed the infection liquid (10 mM MES, 10 mM MgCl2, 0.1 mM AS). Finally, the OD_{600}

[3]http://www.genscript.com/wolf-psort.html

[4]http://bioinformatics.psb.ugent.be/webtools/plantcarere/html/

TABLE 1 | Number of genes in each subfamily of 7 species.

	PCF	CIN	CYC	Total
Fragaria vesca	10	6	3	19
Malus domestica	22	26	4	52
Prunus mume	10	6	3	19
Arabidopsis thaliana	13	8	3	24
Pyrus bretschneideri	14	12	8	34
Rubus occidentalis	10	4	3	17
Prunus avium	5	6	3	14

value of bacteria solution was adjusted between 0.6 and 0.8. The growing well and flat tobacco leaves were selected for injection. The infection solution was injected into the lower epidermis of tobacco leaf and cultured in the dark for 48 h (Sufficient water should be kept during dark culture). After dark culture, the tobacco leaf tissue near the injection hole was selected and placed on the glass slide. The fluorescence of GFP protein was observed under confocal laser scanning microscopy.

RESULTS

Identification, Characterization, and Phylogenetic Analysis of *TCP* Genes

Firstly, we used the HMM for obtaining PF03634 of the conservative domain as the search criteria to compare six Rosaceae species in the protein database (**Supplementary Table 1**). Thirty-four *TCP* genes were identified in pear and 121 genes were identified in the other five Rosaceae species, including strawberry (19), apple (52), plum (19), raspberry (17), and cherry (14) (**Table 1**). Finally, we constructed phylogenetic trees with six Rosaceae species and *Arabidopsis* using the Neighbor-Joining method. The phylogenetic tree was divided into two subgroups: PCF was in Class I, CIN, and CYC were in Class II (**Figure 1**). Among them, CYC members had the least members, 3 in strawberry, 4 in apple, 3 in plum, 3 in *Arabidopsis*, 8 in pear, 3 in raspberry, 3 in cherry. Compared to CYC, PCF had more members, apple had 22 members, followed by pear (14), *Arabidopsis* (13), strawberry (10), raspberry (10), plum (10) and cherry (5) (**Table 1**). Additionally, we calculated the physicochemical parameters of *TCP* genes in six Rosaceae species. Among these six Rosaceae species, the pI value was 4.62–10.65 and the molecular weight ranges from 12.82 to 69.01. The GRAVY values of all TCP proteins were negative. 99% of *TCP* genes were located in the nucleus (**Table 2** and **Supplementary Table 4**).

To further understanding about the potential function of *TCP* family in six Rosaceae species, we analyzed 155 *TCP* genes by GO analysis. The results showed that *TCP* genes could be divided into three categories: cellular component, biological process and molecular function. In molecular function, most genes were enriched in transcriptional regulatory activity. In biological process, *TCP* genes of six Rosaceae species were found in three GO terms (regulation of biological process, biological regulation, metabolic process). Among cellular components, organelle part and membarane enclosed activity were only found in a few genes of apple (**Supplementary Figure 1** and **Supplementary Table 5**).

Conserved Structure Analysis of *TCP* Genes

In order to study the evolutionary relationship of *TCP* genes, we analyzed the conservative structure in six Rosaceae species. In this study, we used MEME to predict 20 motifs of *TCP* genes, and the results showed that members of *TCP* genes were highly conservative (motif 1, 2, 3) (**Figure 2**). We used ClustalX 2.0 to align the protein sequences. After alignment, TCP proteins were divided into two subgroups. Most of the *TCP* domains in each species were composed of 55–60 amino acids, which

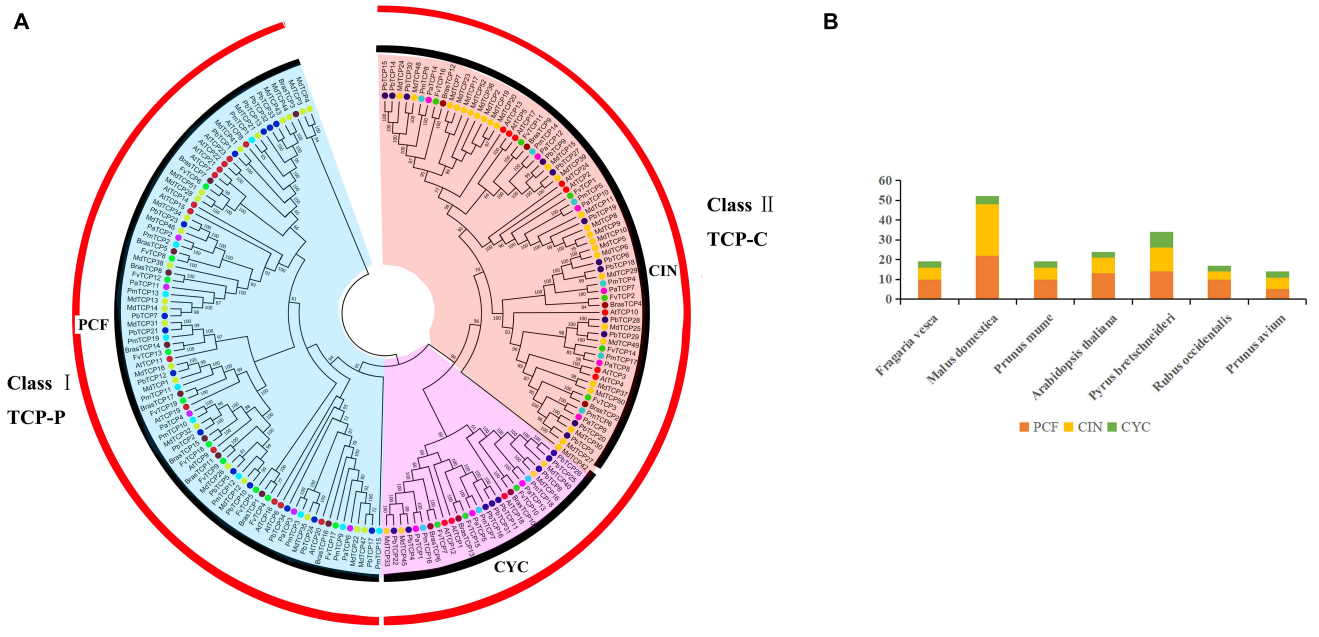


FIGURE 1 | Phylogenetic relationships and subfamily designations in TCP proteins from *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, and *Arabidopsis thaliana*. **(A)** Interspecific phylogenetic tree of TCP protein sequences from 7 species. **(B)** Number of genes in each subfamily of 7 species.

TABLE 2 | Basic information of *TCP* genes in *Pyrus bretschneideri*.

Gene name	Gene ID	Chromosome	AA	KD	pI	GRAVY	Preditced subcellular localization
PbTCP1	Pbr018420.1	Chr1	250	26.93	9.76	–0.410	nucl
PbTCP2	Pbr018814.1	Chr2	307	33.03	5.64	–0.473	nucl
PbTCP3	Pbr025856.1	Chr3	440	47.80	6.53	–0.767	nucl
PbTCP4	Pbr013244.1	Chr3	461	51.33	9.12	–0.866	nucl
PbTCP5	Pbr021770.1	Chr4	380	39.69	5.68	–0.474	nucl
PbTCP6	Pbr000450.1	Chr5	217	24.57	6.75	–0.841	nucl
PbTCP7	Pbr011454.1	Chr6	411	43.73	7.13	–0.668	nucl
PbTCP8	Pbr020246.1	Chr6	462	52.90	6.97	–1.014	nucl
PbTCP9	Pbr020171.1	Chr6	377	42.02	8.00	–0.761	nucl
PbTCP10	Pbr001559.1	Chr6	377	39.36	5.63	–0.411	nucl
PbTCP11	Pbr013717.1	Chr6	376	42.17	8.99	–0.707	nucl
PbTCP12	Pbr013906.1	Chr7	366	37.88	6.25	–0.328	nucl
PbTCP13	Pbr026562.3	Chr8	601	62.82	7.58	–0.705	nucl
PbTCP14	Pbr006457.1	Chr9	383	42.07	7.32	–0.791	nucl
PbTCP15	Pbr006477.1	Chr9	383	42.10	7.97	–0.790	nucl
PbTCP16	Pbr030633.1	Chr9	390	43.73	9.17	–0.758	nucl
PbTCP17	Pbr041545.1	Chr9	323	34.51	9.02	–0.802	nucl
PbTCP18	Pbr016172.1	Chr10	217	24.71	8.71	–0.909	nucl
PbTCP19	Pbr039609.1	Chr10	483	52.57	8.36	–0.828	nucl
PbTCP20	Pbr038238.1	Chr11	430	46.70	6.44	–0.664	nucl
PbTCP21	Pbr020546.1	Chr12	220	24.01	6.60	–0.441	nucl
PbTCP22	Pbr027488.1	Chr13	477	52.91	9.06	–0.823	nucl
PbTCP23	Pbr039105.1	Chr13	401	42.85	6.73	–0.624	nucl
PbTCP24	Pbr035636.1	Chr13	249	26.82	9.67	–0.575	nucl
PbTCP25	Pbr007075.1	Chr14	497	56.49	7.44	–0.978	nucl
PbTCP26	Pbr007125.1	Chr14	471	53.46	6.47	–0.955	nucl
PbTCP27	Pbr007197.1	Chr14	373	41.40	7.30	–0.739	nucl
PbTCP28	Pbr031206.1	Chr15	345	37.83	6.40	–0.725	nucl
PbTCP29	Pbr022498.1	Chr17	351	38.30	5.98	–0.752	nucl
PbTCP30	Pbr006641.1	Chr17	380	41.88	7.03	–0.832	nucl
PbTCP31	Pbr003924.1	scaffold1180.0	307	34.67	9.22	–0.747	Chlo
PbTCP32	Pbr039926.1	scaffold868.0	603	63.39	8.74	–0.722	Nucl
PbTCP33	Pbr039901.1	scaffold868.0	603	63.39	8.74	–0.722	Nucl
PbTCP34	Pbr037196.1	scaffold751.0	247	27.76	10.42	–0.592	Nucl

The TCP genes of Pyrus bretschneideri identified in this study are listed.

conform to the basic HLH structure (**Supplementary Figures 2–7**). In the basic region of *TCP*, several specific amino acids could bind to DNA, which was relatively conservative. In the region of helix 1 and 2, the amino acid sequences of TCP-P and TCP-C were different. In the TCP-C subfamily, most of the *CYC* genes contained an R domain (**Supplementary Figures 2–7**).

Chromosomal Location and Duplication Events of *TCP* Genes in Six Rosaceae Species

According to the whole genome data of strawberry, apple, plum, pear, raspberry, cherry, the exact chromosome physical location information of all *TCP* genes were determined (**Figure 3**). In the pear, 4 out of 34 *TCP* genes were not located on any chromosome, and 30 *TCP* genes were located on 16 chromosomes (except chromosome 16). In strawberry, *TCP* genes were located on chromosome 3, 4, 5, 6, and 7. Five genes in plum were not located on any chromosome, and other genes were located on chromosome 2, 3, 4, 5, and 7. In raspberry, *TCP* genes were mainly distributed on chromosomes 3 and 5, and other genes were distributed on chromosomes 4, 6, and 7. In cherry, four genes were distributed on chromosome 4, three genes on chromosome 1 and 5, and the remaining three genes on chromosome 2 and 3. In apple, there are no genes on chromosome 3 and three genes are not located on any chromosome.

Among the six Rosaceae species, only 11 gene duplication events were identified in pear and apple (**Supplementary Table 6**). Seven duplication events were identified in pear and 4 in apple. In order to study the effect of duplication events on gene evolution, we counted the values of Ka, Ks, and Ka/Ks of 11 duplicated gene pairs were analyzed them. Among these 11 gene duplication events, Ka/Ks values were <1, with the maximum value of 0.907 (*MdTCP24-MdTCP48*) and the minimum value

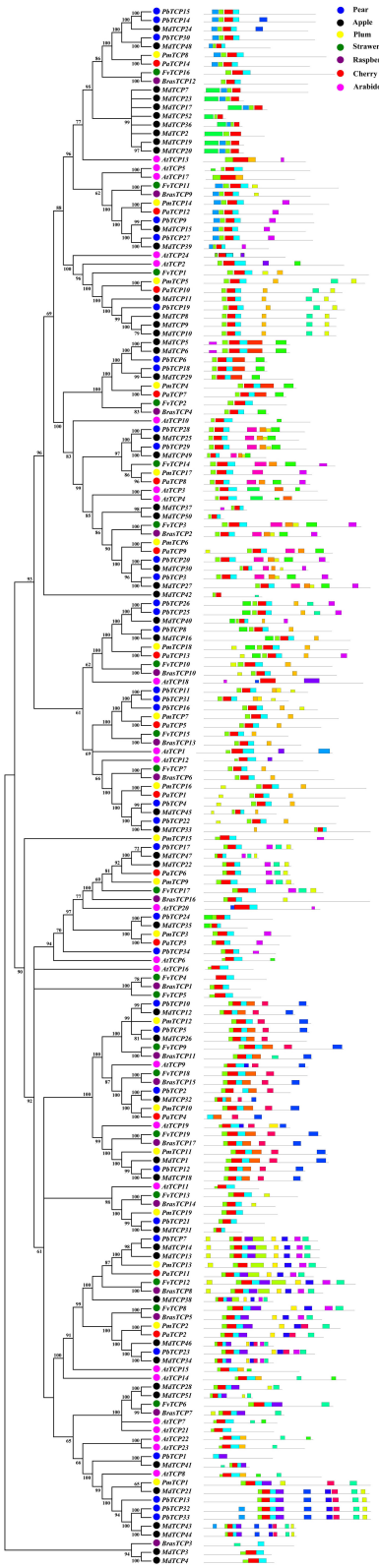


FIGURE 2 | Predicted *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Malus domestica*, *Prunus avium,* and *Arabidopsis thaliana* TCP protein conserved motifs.

of 0.097 (*MdTCP28-MdTCP51*). These results indicated that *TCP* family genes were mainly affected by purifying selection during evolution.

Among the 11 gene duplication events, 9 pairs were fragment duplication events and two pairs were not located any chromosome. These results indicated that the expansion of *TCP* genes were mainly driven by fragment duplication. To understand the selection pressure of *TCP* family in the evolution process, we performed sliding window analysis (**Figure 4**). Sliding window analysis, results implied that the Ka/Ks values of *TCP* conservative domains were <1. Most coding site Ka/Ks ratios were <1, with exceptions for one or several distinct peaks (Ka/Ks > 1).

Analysis of *cis*-Acting Elements in *TCP* Gene Promoter

In plant growth and development, gene-specific expression was mainly related to *cis*-acting elements of upstream promoter. In this experiment, we had been analyzed the *cis*-acting elements of 34 members of *TCP* gene promoter in pear. We divided the functional elements into three types: plant growth and development, biological and abiotic stress responses, and phytohormone responses (**Figure 5** and **Supplementary Table 7**). In phytohormone responses, there were many *cis*-acting elements related to the responses to hormones, including responses to methyl jasmonate (CGTCA-motif, TGACG-motif), gibberellin (TATC-box, GARE-motif), auxin (TGA-element, AuxRR-core), abscisic acid (ABRE), and salicylic acid (TCA-element). In 34 members of *TCP* family, the *cis*-acting element related to the responses to abscisic acid appeared 75 times. In plant growth and development, including the light response elements (MRE, Box 4, G-Box), cell cycle regulation (MSA-like), zein metabolism regulation (O_2-site), day and night control (circadian), in which the proportion of light response elements was more, Box4, G-box each appeared 61 times. In biological and abiotic stress responses mainly included drought (MBS), defense and stress (TC rich repeats), hypoxia specific inducible enhancer like elements (GC motif), anaerobic (ARE), and low temperature (LTR).

Expression Profile Analysis of *PbTCP*s in Different Tissues of Chinese White Pear

In order to further study the function of *PbTCP*s in flower, we analyzed the expression patterns of 34 *TCP* genes in petal, sepal, ovary, bud, stem, leaf according to the RNA-seq database. As shown in **Figure 6**, three genes (*PbTCP1*, *25*, *27*) were not expressed in all tissues. *PbTCP2*, *3*, *12*, *14*, *15,* and *30* were highly expressed in petals. The expression levels of *PbTCP16*, *18*, *31*, *32,* and *33* were higher in ovary, which might affect the growth and development of fruits in the later stage. Comparing with other tissues, the expression of almost all genes in mature fruit was relatively low. Four genes (*PbTCP10*, *20*, *22*, *29*) were highly expressed in sepal. About 30% genes were highly expressed in buds and stems.

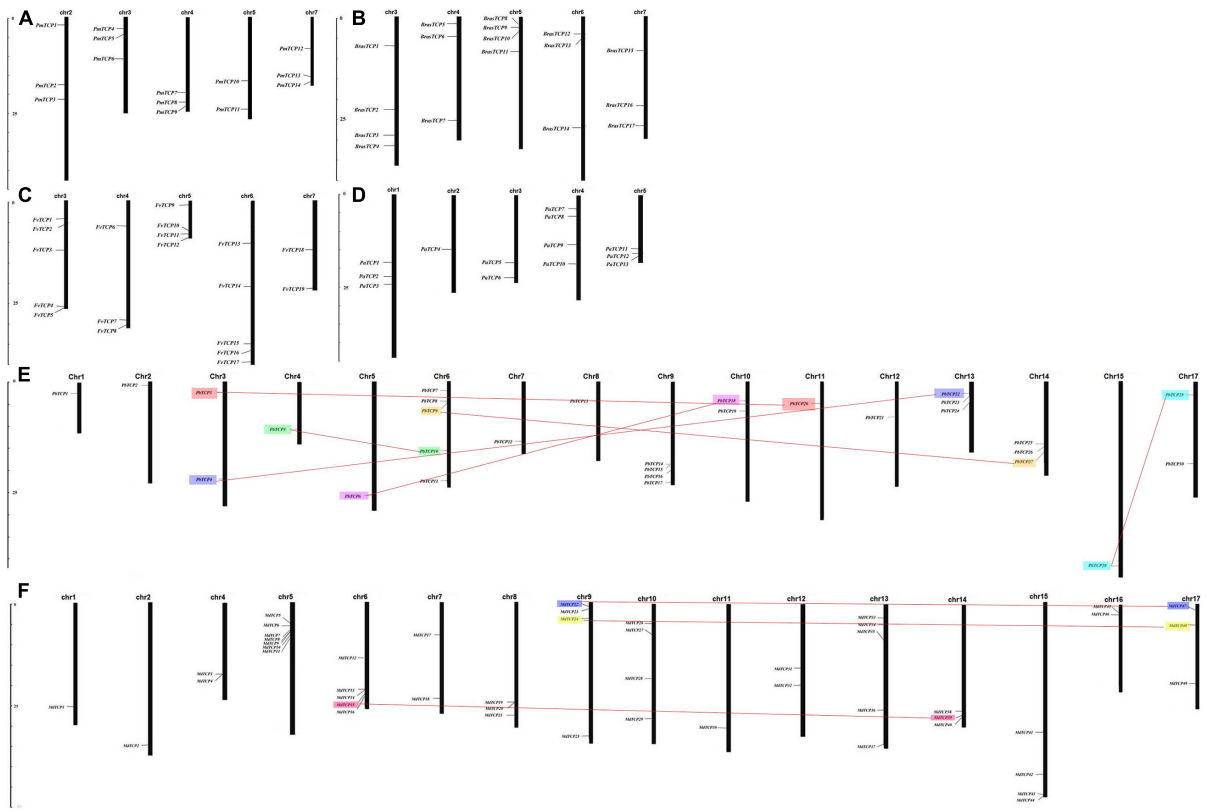


FIGURE 3 | Chromosomal locations of Six Rosaceae species. Chromosomal locations of *TCP* genes in **(A)** *Prunus mume*, **(B)** *Rubus occidentalis*, **(C)** *Fragaria vesca*, **(D)** *Prunus avium*, **(E)** *Pyrus bretschneideri*, and **(F)** *Malus domestica.* Duplicated gene pairs are connected with colored lines.

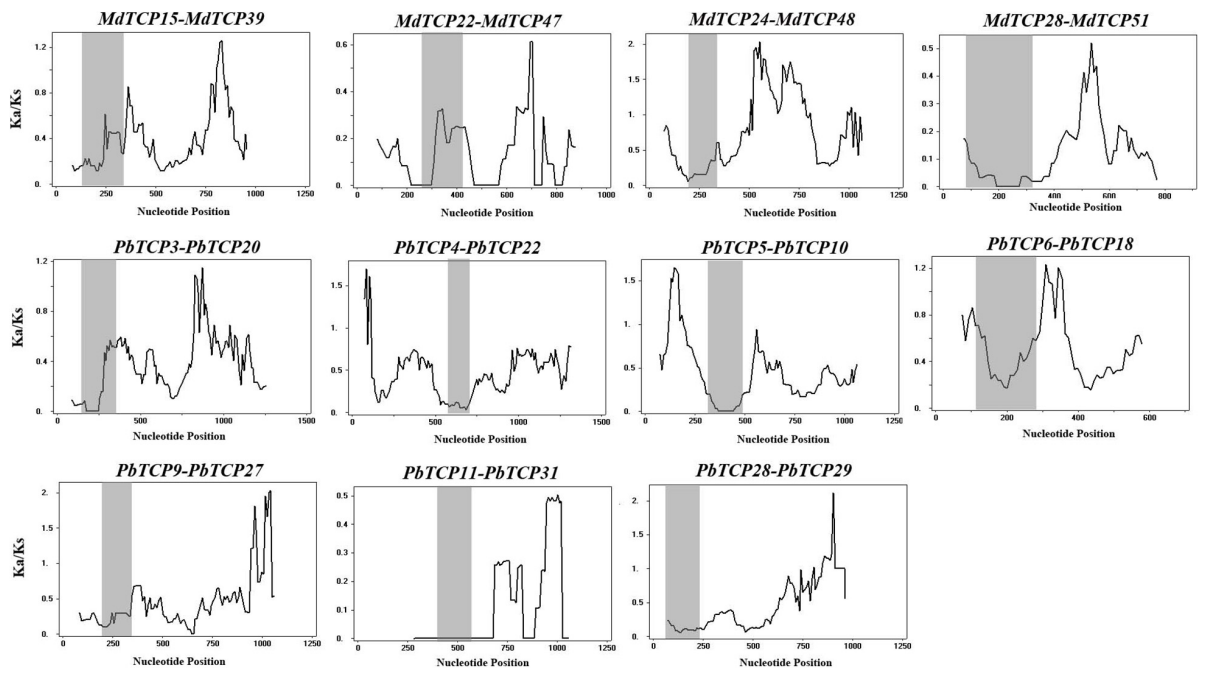


FIGURE 4 | Sliding window plots of duplicated *TCP* genes in *Pyrus bretschneideri* and *Malus domestica.* The gray shaded portion indicates conserved *TCP* domain. The *X*-axis indicates the synonymous distance within each gene.

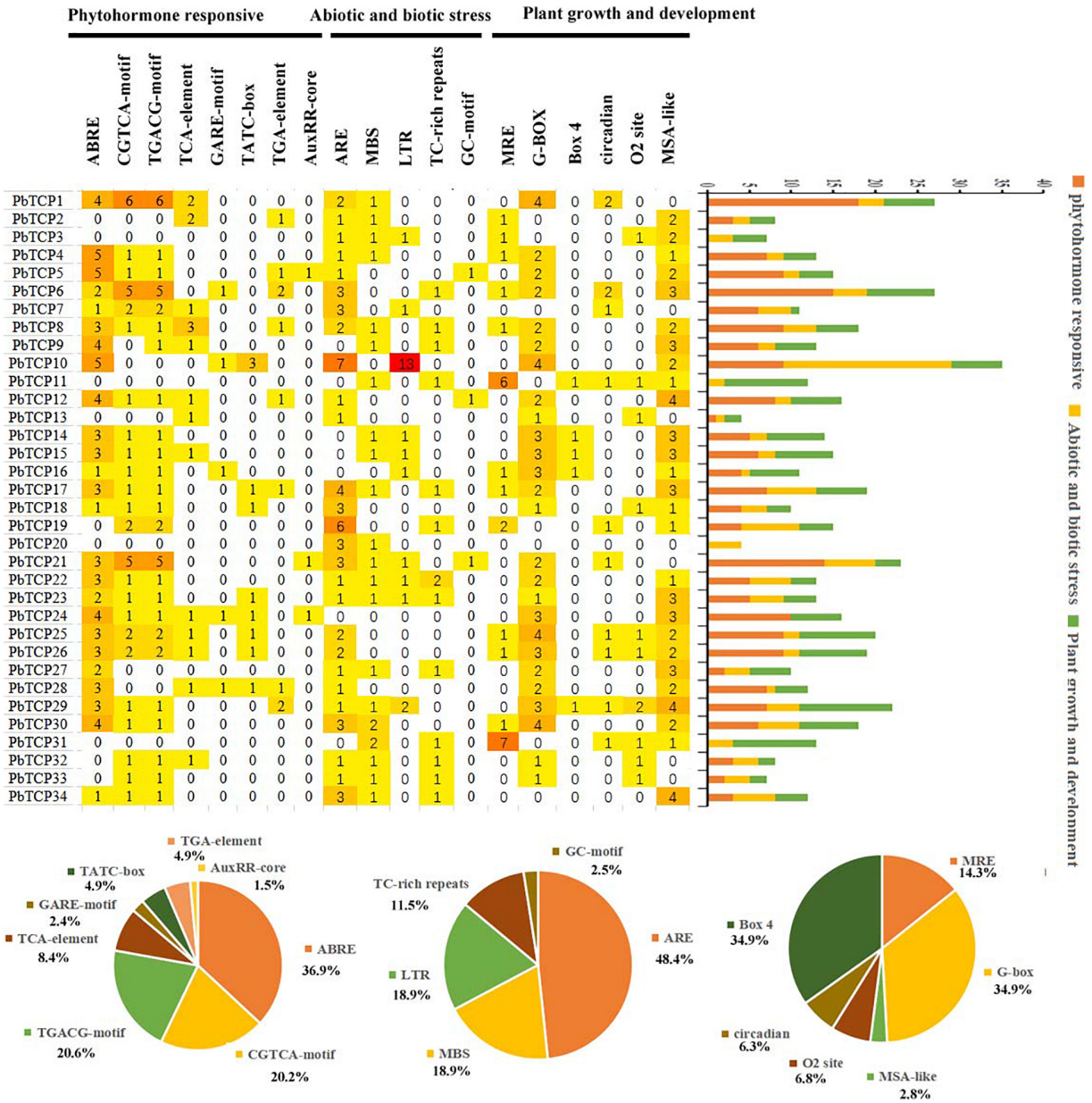


FIGURE 5 | Putative *cis*-acting regulatory elements in the *PbTCP* promoters.

Expression Characteristics of Chinese White Pear *TCP* Genes

In order to further study the function of *TCP* genes in pear, we studied the expression of *TCP* genes in different tissues. As shown in **Figure 7**, *PbTCP26*, *31*, *32,* and *33* were not expressed in any tissues. *PbTCP10*, *19,* and *22* were highly expressed in flowers, and *PbTCP1* was highly expressed in leaves. Other genes were highly expressed in fruits.

We analyzed the expression level in seven development stages of Chinese white pear (**Figure 8**). These results showed that the expression pattern of *PbTCP7* reached a peak at 63 DAP. The expression of *PbTCP9* and *PbTCP19* reached a peak only at 55 DAP, but the expression level was very low at other developmental stages. Firstly, the expression level of *PbTCP14*, *15*, *23,* and *24* were increased, then the expression level decreased during fruit development. The expression of *PbTCP6* and *PbTCP18* reached a peak at the 15 DAP.

Gibberellin Response Pattern Analysis of *PbTCPs*

The results of expression profile analysis and qRT-PCR showed that *PbTCP10*, *19,* and *22* were highly expressed in flowers (**Supplementary Figure 8**). In this experiment, the buds of "Dangshan Su" pear were treated with exogenous GA, and the expression patterns of *PbTCP10*, *19,* and *22* were analyzed. After GA treatment, these three genes expressed two response

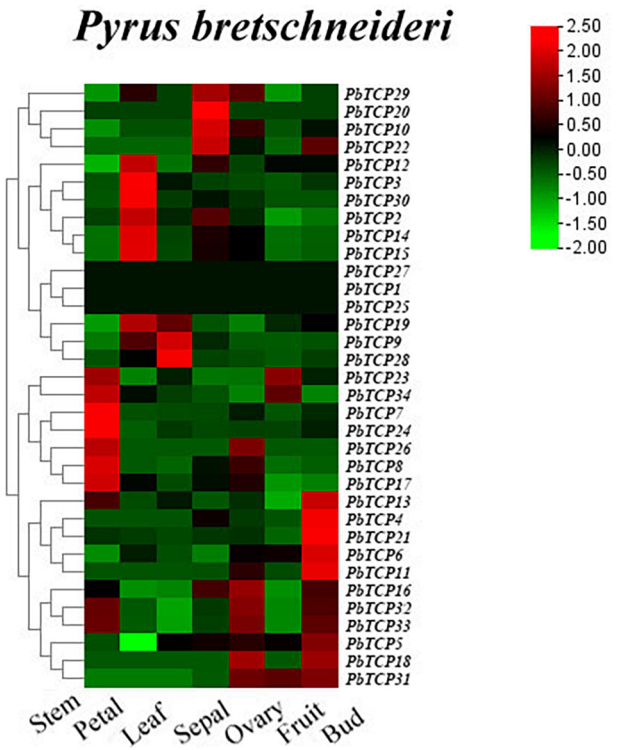


FIGURE 6 | Expression profiles of *PbTCP* genes in different tissues of *Pyrus bretschneideri*. The heatmap was generated by TBtools software according to the RNA-seq database. Log2-based-fold changes were used to create a heatmap. As shown in the bar at the upper right corner, the gene transcription level is expressed in different colors on the map.

patterns. The expression of *PbTCP10* increased significantly at 2 HPT, maintained at a high level at 4–8 HPT, and returned to the initial level at 12 HPT. There was no significant change in the transcriptional level of *PbTCP19* and *PbTCP22* under exogenous GA treatment.

Subcellular Localization of *PbTCP6, 13,* and *17*

The main function of transcription factors is to connect with *cis*-acting elements of gene promoter in the nucleus. In order to study the subcellular localization of *TCP* genes in pear, three *TCP* genes were connected with 35S promoter containing green fluorescent protein (GFP). These three genes and empty vector were transiently expressed in tobacco. As shown in **Figure 9**, these three genes were located in the nucleus, and the empty vector was located in the nucleus and cell membrane, which was consistent with the predicted results.

DISCUSSION

TCP proteins are transcription factors that are unique in plants and are involved in leaf development, flower symmetry, stem branching, and other biological processes. TCP proteins can also regulate the flowering process and secondary wall formation and ultimately affect plant growth and development (Nag et al., 2009; Wang et al., 2015). In this study, 155 genes were identified in six Rosaceae plants. All genes contained a *TCP* conserved domain, and their proteins were hydrophilic with a negative GRAVY value (**Table 2** and **Supplementary Table 4**). In the six Rosaceae species, all genes were divided into two subgroups, and the number of Class I (TCP-P) members was generally greater than that of Class II (TCP-C) members. However, in apple, pear and cherry, there were more TCP-C members than TCP-P members (**Table 1**). According to the number of *TCP* genes in each species, there are the most *TCP* genes in apple (52), followed by pear (34). The number of *TCP* family members of apple and pear were more than other species (**Table 1**). These differences might be related to the evolution of the *TCP* family.

Whole-genome duplication (WGD) or polyploidy is an important driving force shaping plant evolution (Tang et al., 2008). Previous studies indicated that pear, strawberry, apple and other dicotyledons had a whole-genome duplication event before 140 million years ago (Mya). However, apple and pear experienced a whole-genome duplication event 30–40 Mya (Shulaev et al., 2011; Wu et al., 2013). After that, the chromosome number of pear and apple changed to 17, strawberry changed to 7, plum changed to 8 and raspberry changed to 7, and cherry changed to 8. These results indicated that the second WGD, the 9 chromosomes in the common ancestor of Rosaceae underwent doubling, breaking, hybridization and fusion. The conserved domains were closely related to the diversity of gene functions. The structures within a subfamily were similar, which indicated that these genes might have similar functions. In the HLH domain, the second helix region had a specific LXXLL motif, and members of the CYC/TB1 subfamily specifically contained a hydrophilic α helix (R domain) rich in polar amino acids, which did not exist in other members (**Supplementary Figures 2–7**). The difference in gene number and the retention of conserved structures might be due to the loss of *TCP* family genes, chromosome doubling, and selection pressure in the process of WGDs.

To understand the evolutionary patterns of *TCP* family genes in six Rosaceae species, we calculated the values of Ka and Ks (**Figure 4** and **Supplementary Table 6**). These results showed that collateral gene pairs only existed in apple and pear, and the Ka/Ks values of all gene pairs were <1, which indicated that the *TCP* family had undergone obvious purifying selection in the evolutionary process. Interestingly, there were two gene pairs (*PbTCP28-PbTCP29*, *MdTCP24-MdTCP48*) with relatively high Ka/Ks values (>0.5), which might be due to the rapid evolution and diversification of these two genes after the duplication event.

Plant flowering is an important life activity in the process of plants transitioning from vegetative growth to reproductive growth. Many genes are involved in the flowering process of plants, such as *FLS* (Park et al., 2020), *MADS* (Tang et al., 2020), and *CDF* (Corrales et al., 2014). Recent studies have shown that *TCP* genes also play a regulatory role in plant flowering (Lucero et al., 2017; Li et al., 2019). We used public transcriptome data and qRT-PCR to obtain the expression pattern of *TCP* genes

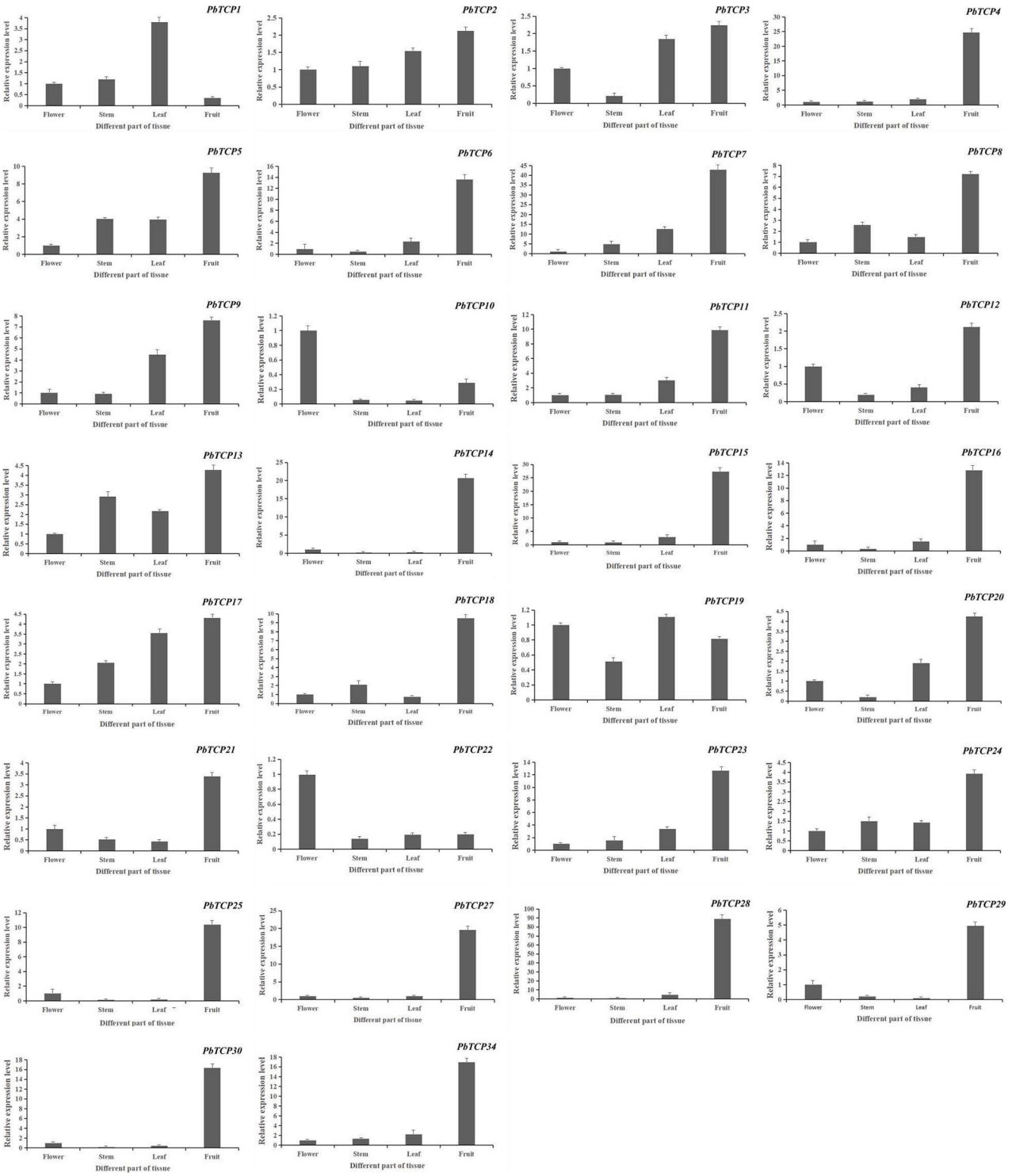


FIGURE 7 | Expression pattern of *PbTCPs* across different tissues (fruit, stems, flowers, and leaves). Error bars show the standard error between three replicates.

in "Dangshan Su" pear. These results showed that there was expression in all tissues results of *TCP* genes, which indicated that *TCP* genes played an important role during growth and development in pear. The qRT-PCR results in different tissues showed that *PbTCP10*, *19*, and *22* were highly expressed in flowers (**Figure 7**). According to expression profile analysis, *PbTCP10* and *PbTCP22* were highly expressed in the sepal. *PbTCP19* was highly expressed in the petal (**Figure 6**). Previous

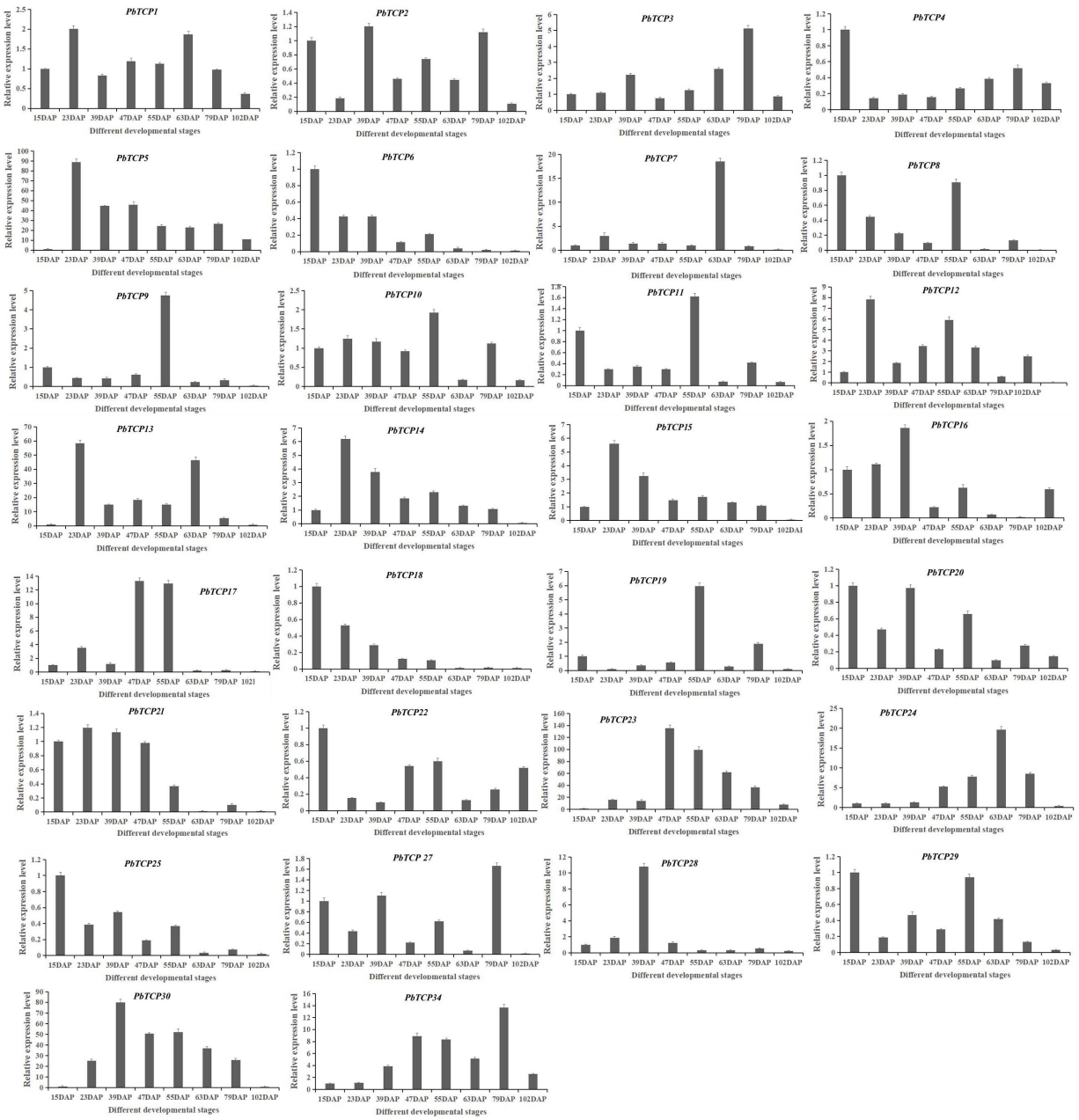


FIGURE 8 | Expression pattern of *PbTCPs* across different developmental stages (15DAP, 23DAP, 39DAP, 47DAP, 55DAP, 63DAP, 79DAP, and 102DAP). Error bars show the standard error between three replicates.

studies showed that hormones (especially GA), sugar and light also play an important role in flowering regulation (Srikanth and Schmid, 2011; Osnato et al., 2012; Cao et al., 2018). In the analysis of *cis*-acting elements, it could be seen that the promoter regions of *PbTCP10* contained GA-responsive elements (TATC-box, GARE-motif). After exogenous GA treatment, the expression patterns of *PbTCP10*, *19,* and *22* showed that the expression of *PbTCP19* and *PbTCP22* were almost not induced by GA, and the expression of *PbTCP10* increased significantly at 2 HPT (**Supplementary Figure 8**). These phenomenons might be due to the absence of GA response element in the promoter of *PbTCP19* and *PbTCP22*. In addition, we found that the *cis*-acting elements of *PbTCP10* promoter contained light response elements, which indicated that *PbTCP10* might be involved in photoperiodic signal (**Figure 5**). In conclusion, *PbTCP10* might be involved in GA regulated flowering induction pathway and regulate photoperiod.

Previous studies found that the formation of stone cells in "Dangshan Su" pear mainly occurred in the early stage of fruit formation (15–47 DAP) (Su et al. 2019). Therefore, *TCP*

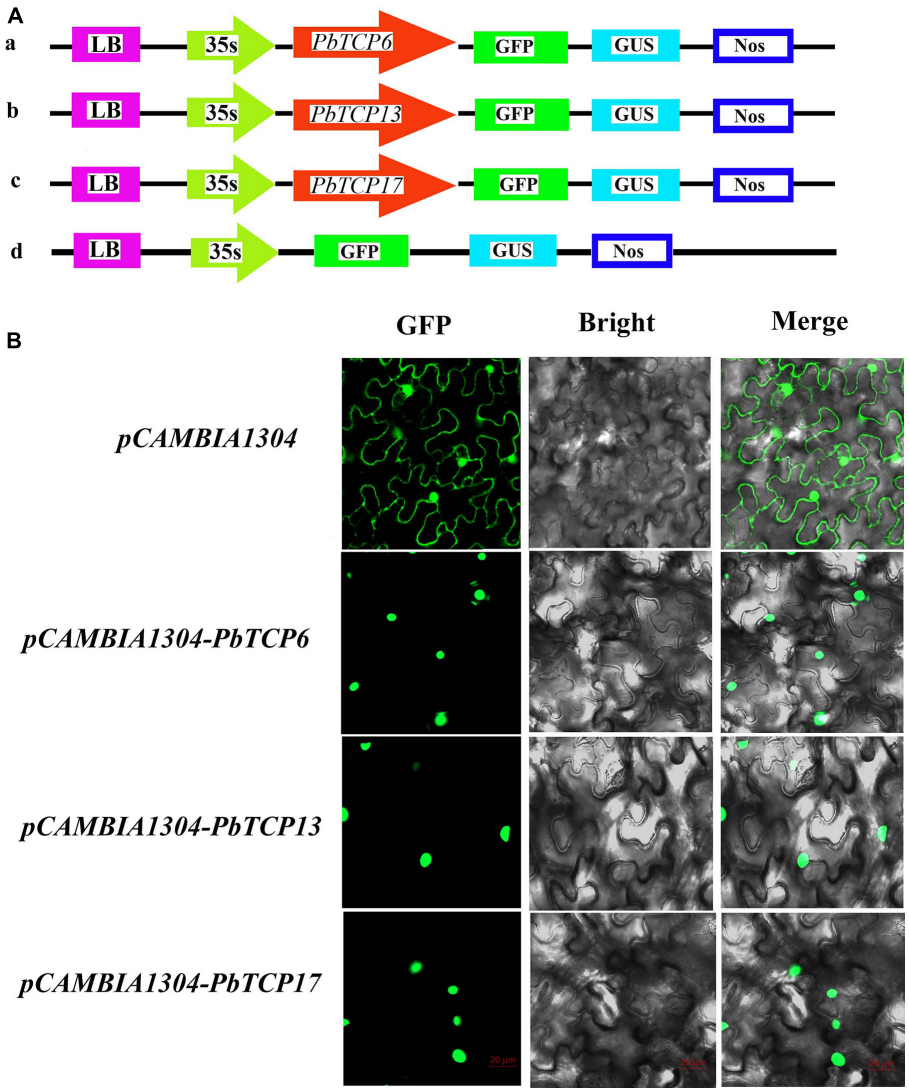


FIGURE 9 | Subcellular localization of three *PbTCP* genes. **(A)** Schematic illustration of vectors pCAMBIA1304 and *PbTCPs*. **(B)** The three pCAMBIA1304-*PbTCPs* fusion proteins (pCAMBIA1304-*PbTCP6*, pCAMBIA1304-*PbTCP13*, pCAMBIA1304-*PbTCP17*), and pCAMBIA1304 as a control were transiently expressed in tobacco leaf and observed under fluorescence microscope.

genes with high expression level in the early stage of fruit development might be involved in the formation of stone cells. The genes with high expression in late stage might be involved in the accumulation of sugar and the response of hormone during fruit ripening. In order to determine the effect of *TCP* genes on secondary wall formation during fruit development of "Dangshan Su" pear, qRT-PCR analysis was conducted at different stages of fruit development (**Figure 8**). The results showed that *PbTCP14*, *15*, *23*, and *24* increased firstly and then decreased during fruit development, which was consistent with the trend of stone cell formation, but only *PbTCP14* and *PbTCP15* were highly expressed in the early stage of fruit development. Therefore, *PbTCP14* and *PbTCP15* might be involved in the stone cell formation during fruit development of "Dangshan Su" pear.

Through comparative genomics analysis, we identified the evolution of *TCP* genes in six Rosaceae species, and screened candidate regulatory genes related to flowering (*PbTCP10*) and stone cell formation (*PbTCP14* and *PbTCP15*). In the following study, we will analysis the biological functions of these genes and provide an important theoretical basis for improving pear quality.

CONCLUSION

In this work, 155 *TCP* genes were identified in six Rosaceae species. According to bioinformatics analysis, we explained the possible evolutionary patterns of *TCP* genes in six Rosaceae species. By qRT-PCR analysis of 34 *TCP* genes in different development stages and tissues of pear, we found that *PbTCP14*

and *PbTCP15* might be involved in the formation of secondary wall during pear fruit development, and *PbTCP*10 might be involved in the process of flowering induction by GA. In general, these results provided a theoretical basis for improving the quality of pear.

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

YZ and XS performed the experiments and wrote the manuscript. XW, MW, and XC analyzed the data. MA and GL helped to polish the language. YC conceived and designed the experiments. All authors read and approved the final manuscript.

FUNDING

The research was supported by the National Natural Science Foundation of China (Grant no. 31640068). The funding bodies were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.669959/full#supplementary-material

Supplementary Figure 1 | Gene Ontology (GO) analysis of TCP genes in six Rosaceae species (*Prunus mume*, *Rubus occidentalis*, *Fragaria vesca*, *Prunus avium*, *Malus domestica*, *Pyrus bretschneideri*).

Supplementary Figure 2 | Multiple sequence alignment of pear *TCP* transcription factors. Alignment of the *TCP* domain for the predicted pear TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 3 | Multiple sequence alignment of apple *TCP* transcription factors. Alignment of the *TCP* domain for the predicted apple TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 4 | Multiple sequence alignment of raspberry *TCP* transcription factors. Alignment of the *TCP* domain for the predicted raspberry TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 5 | Multiple sequence alignment of strawberry *TCP* transcription factors. Alignment of the *TCP* domain for the predicted strawberry TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 6 | Multiple sequence alignment of cherry *TCP* transcription factors. Alignment of the *TCP* domain for the predicted cherry TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 7 | Multiple sequence alignment of plum *TCP* transcription factors. Alignment of the *TCP* domain for the predicted plum TCP proteins. The basic, helix I, loop, and helix II regions are indicated. Alignment of the R domain of Class II subfamily members. The sequences were aligned with ClustalW.

Supplementary Figure 8 | Expression modes of candidate *PbTCP10*, *19,* and *22* in Chinese white Pear buds treated with gibberellin. Error bars show the standard error between three replicates.

Supplementary Table 1 | All TCP protein sequences list.

Supplementary Table 2 | Primer sequences used in qRT-PCR.

Supplementary Table 3 | Primers for vector construction.

Supplementary Table 4 | Basic information of *TCP* genes in five Rosaceae species (*Prunus mume*, *Rubus occidentalis*, *Fragaria vesca*, *Prunus avium*, and *Malus domestica*).

Supplementary Table 5 | Blast2GO annotation details of TCP protein sequences of six Rosaceae species (*Prunus mume*, *Rubus occidentalis*, *Fragaria vesca*, *Prunus avium*, *Malus domestica*, *Pyrus bretschneideri*).

Supplementary Table 6 | Ka/Ks analysis of the *TCP* homologous gene pairs from *Pyrus bretschneideri* and *Malus domestica*.

Supplementary Table 7 | All *TCP* gene promoter sequences list.

REFERENCES

Aguilar-Martinez, J. A., and Sinha, N. (2013). Analysis of the role of *Arabidopsis* class I *TCP* genes *AtTCP7*, *AtTCP8*, *AtTCP22*, and *AtTCP23* in leaf development. *Front. Plant Sci.* 16:406. doi: 10.3389/fpls.2013.00406

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E., et al. (2012). ExPASy:SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). *Nucleic Acids Res.* 43, W39–W49.

Broholm, S. K., Tähtiharju, S., Laitinen, R. A. E., Albert, V. A., Teeri, T. H., and Elomaa, P. (2008). A *TCP* domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proc. Natl. Acad. Sci.* 105, 9117–9122. doi: 10.1073/pnas.0801359105

Cao, J. F., Zhao, B., Huang, C. C., Chen, Z. W., Zhao, T., Liu, H., et al. (2020). The miR319-targeted *GhTCP4* promotes the transition from cell elongation to wall thickening in cotton fiber. *Mol. Plant* 13, 1063–1077. doi: 10.1016/j.molp.2020.05.006

Cao, K., Yan, F., Xu, D. W., Ai, K. Q., Yu, J., Bao, E. C., et al. (2018). Phytochrome B1-dependent control of *SP5G* transcription is the basis of the night break and red to far-red light ratio effects in tomato flowering. *BioMed Central* 18:158.

Cao, Y. P., Li, X. X., and Jiang, L. (2019). Integrative analysis of the core fruit lignification toolbox in pear reveals targets for fruit quality bioengineering. *Biomolecules* 9:504. doi: 10.3390/biom9090504

Cheng, X., Li, M. L., Abdullah, M., Li, G. H., Zhang, J. Y., Manzoor, M. A., et al. (2019). In silico genome-wide analysis of the pear (*Pyrus bretschneideri*) *KNOX* family and the functional characterization of *PbKNOX1*, an *Arabidopsis BREVIPEDICELLUS* orthologue gene, involved in cell wall

and lignin biosynthesis. *Front. Genet.* 10:632. doi: 10.3389/fgene.2019.00632

Corrales, A. R., Nebauer, S. G., Carrillo, L., Fernandez-Nohales, P., Marques, J., Renau-Morata, B., et al. (2014). Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. *J. Exp. Bot.* 65, 995–1012. doi: 10.1093/jxb/ert451

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The *TCP* domain: a motif found in proteins regulating plant growth and development. *Plant J.* 18, 215–222. doi: 10.1046/j.1365-313x.1999.00444.x

Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature* 386, 485–488. doi: 10.1038/386485a0

Feng, X. Z., Zhao, Z., Tian, Z. X., Xu, S. L., Luo, Y. H., Cai, Z. G., et al. (2006). Control of petal shape and floral zygomorphy in lotus japonicus. *Proc. Natl. Acad. Sci. U S A.* 103, 4970–4975. doi: 10.1073/pnas.0600681103

Huang, T. B., and Irish, V. F. (2015). Temporal control of plant organ growth by *TCP* transcription factors. *Curr. Biol.* 25, 1765–1770. doi: 10.1016/j.cub.2015.05.024

Hubbard, L., McSteen, P., Doebley, J., and Hake, S. (2002). Expression patterns and mutant phenotype of teosinte branched1 correlate with growth suppression in maize and teosinte. *Genetics* 162, 1927–1935. doi: 10.1093/genetics/162.4.1927

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587.

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

Kieffer, M., Master, V., Waites, R., Davies, and Brendan. (2011). TCP14 and *TCP15* affect internode length and leaf shape in *Arabidopsis*. *Plant J.* 68, 147–158. doi: 10.1111/j.1365-313x.2011.04674.x

Kosugi, S., and Ohashi, Y. (1997). *PCF1* and *PCF2* specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* 9, 1607–1619. doi: 10.2307/3870447

Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40, D302–D305.

Luo, D., Carpenter, R., Vincent, C., Copsey, L., and Coen, E. (1996). Origin of floral asymmetry in *Antirrhinum*. *Nature* 383, 794–799. doi: 10.1038/383794a0

Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J., and Coen, E. (1999). Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99, 367–376. doi: 10.1016/s0092-8674(00)81523-8

Li, D. B., Zhang, H. Y., Mou, M. H., Chen, Y. L., Xiang, S. Y., Chen, L. G., et al. (2019). *Arabidopsis* class II TCP transcription factors integrate with the FT-FD module to control flowering. *Plant Physiol.* 181, 97–111. doi: 10.1104/pp.19.00252

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Liu, J., Cheng, X. L., Liu, P., Li, D. Y., Chen, T., Gu, X. F., et al. (2017). MicroRNA319-regulated *TCP*s interact with FBHs and PFT1 to activate CO transcription and control flowering time in *Arabidopsis*. *PLoS Genet.* 13:e1006833. doi: 10.1371/journal.pgen.1006833

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

Lucero, L. E., Manavella, P. A., Gras, D. E., Ariel, F. D., and Gonzalez, D. H. (2017). Class I and Class II *TCP* transcription factors modulate *SOC1*-Dependent flowering at multiple levels. *Mol. Plant* 10, 1571–1574. doi: 10.1016/j.molp.2017.09.001

Ma, Y., Zhang, F. Y., Bade, R., Daxibater, A., Men, Z. H., and Hasi, A. (2015). Genome-Wide Identification and Phylogenetic Analysis of the *ERF* gene family in melon. *J. Plant Growth Regulat.* 34, 66–77. doi: 10.1007/s00344-014-9443-z

Manassero, N. G., Uberti, V., Ivana, L., Welchen, E., and Gonzalez, D. H. (2013). *TCP* transcription factors: architectures of plant form. *BioMol. Concepts* 4, 111–127. doi: 10.1515/bmc-2012-0051

Nag, A., King, S., and Jack, T. (2009). MiR319a targeting of *TCP4* is critical for petal growth and development in *Arabidopsis*. *Proc. Natl. Acad. Sci.* 106, 22534–22539. doi: 10.1073/pnas.0908718106

Nath, U., Crawford, B. C. W., Carpenter, R., and Coen, E. (2003). Genetic control of surface curvature. *Science* 299, 1404–1407. doi: 10.1126/science.1079354

Ori, N., Cohen, A. R., Etzioni, A., Brand, A., Yanai, O., Shleizer, S., et al. (2007). Regulation of *LANCEOLATE* by miR319 is required for compound-leaf development in tomato. *Nat. Genet.* 39, 787–791. doi: 10.1038/ng2036

Osnato, M., Castillejo, C., Matias-Hernandez, L., and Pelaz, S. (2012). TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in *Arabidopsis*. *Nat. Communicat.* 3:808.

Palatnik, J. F., Allen, E., Wu, X. L., Schommer, C., Schwab, R., Carrington, J. C., et al. (2003). Control of leaf morphogenesis by microRNAs. *Nature* 425, 257–263. doi: 10.1038/nature01958

Park, S., Kim, D. H., Yang, J. H., Lee, J. Y., and Lim, S. H. (2020). Increased flavonol levels in tobacco expressing *AcFLS* affect flower color and root growth. *Int. J. Mol. Sci.* 21:1011. doi: 10.3390/ijms21031011

Resentini, F. A., Lucia, F. B., Colombo, A., David, B. M., and Masiero, A. S. (2015). TCP14 and *TCP15* mediate the promotion of seed germination by gibberellins in *Arabidopsis thaliana*. *Mol. Plant* 8, 482–485. doi: 10.1016/j.molp.2014.11.018

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a plant *Cis*-acting regulatory element database. *Nucleic Acids Res.* 27, 295–296. doi: 10.1093/nar/27.1.295

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of the woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116.

Su, X. Q., Meng, T. K., Zhao, Y., Li, G. H., Cheng, X., Muhammad, A., et al. (2019). Comparative genomic analysis of the *IDD* genes in five Rosaceae species and expression analysis in Chinese white pear (*Pyrus bretschneideri*). *PeerJ* 26:e6628. doi: 10.7717/peerj.6628

Su, X. Q., Zhao, Y., Wang, H., Li, G. H., Cheng, X., Jin, Q., et al. (2019). Transcriptomic analysis of early fruit development in Chinese white pear (*Pyrus bretschneideri Rehd.*) and functional identification of *PbCCR1* in lignin biosynthesis. *BMC Plant Biol.* 19:417. doi: 10.1186/s12870-019-2046-x

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell Mol. Life Sci.* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y

Tang, H. B., Bowers, J. E., Wang, X. Y., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Tang, X. L., Liang, M. X., Han, J. J., Cheng, J. S., Zhang, H. X., and Liu, X. H. (2020). Ectopic expression of *LoSVP*, a MADS-domain transcription factor from lily, leads to delayed flowering in transgenic *Arabidopsis*. *Plant Cell Rep.* 39, 289–298. doi: 10.1007/s00299-019-02491-1

Uberti-Manassero, N. G., Coscueta, E. R., and Gonzalez, D. H. (2016). Expression of a repressor form of the *Arabidopsis thaliana* transcription factor *TCP16* induces the formation of ectopic meristems. *Plant Physiol. Biochem.* 108, 57–62. doi: 10.1016/j.plaphy.2016.06.031

Van Es, S. W., Silveira, S. R., Rocha, D. I., Bimbo, A., Martinelli, A. P., Dornelas, M. C., et al. (2018). Novel functions of the *Arabidopsis* transcription factor *TCP5* in petal development and ethylene biosynthesis. *Plant J.* 94, 867–879. doi: 10.1111/tpj.13904

Wang, H., Mao, Y. F., Yang, J., and He, Y. K. (2015). *TCP24* modulates secondary cell wall thickening and anther endothecium development. *Front. Plant Sci.* 6:436. doi: 10.3389/fpls.2015.00436

Wang, Y., Yu, Y. H., Wang, J. D., Chen, Q. J., and Ni, Z. Y. (2020). Heterologous overexpression of the *GbTCP5* gene increased root hair length, root hair and stem trichome density, and lignin content in transgenic *Arabidopsis*. *Gene* 758:144954. doi: 10.1016/j.gene.2020.144954

Wu, J., Wang, Z. W., Shi, Z. B., Zhang, S., Ming, R., Zhu, S. L., et al. (2013). The genome of the pear (*Pyrus bretschneideri Rehd.*). *Genome Res.* 23, 396–408.

Wu, J. F., Tsai, H. L., Joanito, I., Wu, Y. C., Chang, C. W., Li, Y. H., et al. (2016). LWD-TCP complex activates the morning gene *CCA1* in *Arabidopsis*. *Nat. Communicat.* 7:13181.

Yao, X. M., Hong, W. J., and Zhang, D. B. (2007). Genome-wide comparative analysis and expression pattern of *TCP* gene families in *Arabidopsis thaliana* and *Oryza sativa*. *J. Integrat. Plant Biol.* 49, 885–897. doi: 10.1111/j.1744-7909.2007.00509.x

Ye, J., Fang, L., Zheng, H. K., Zhang, Y., Chen, J., Zhang, Z. J., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297.

Zhang, J. Y., Cheng, X., Jin, Q., Su, X. Q., Li, M. L., Yan, C. C., et al. (2017). Comparison of the transcriptomic analysis between two Chinese white pear (*Pyrus bretschneideri Rehd*.) genotypes of different stone cells contents. *PLoS One* 12:e0187114. doi: 10.1371/journal.pone.0187114

Zhao, Y. F., Pfannebecker, K., Dommes, A. B., Hidalgo, O., Becker, A., and Elomaa, P. (2018). Evolutionary diversification of *CYC/TB1*-like *TCP* homologs and their recruitment for the control of branching and floral morphology in *Papaveraceae* (basal eudicots). *N. Phytol.* 220, 317–331. doi: 10.1111/nph.15289

Zhu, G. R., Chen, G. X., Zhu, J. T., Zhu, Y., Lu, X. B., Li, X. H., et al. (2015). Molecular characterization and expression profiling of *NAC* transcription factors in *Brachypodium distachyon L. PLoS One* 10:e0139794. doi: 10.1371/journal.pone.0139794

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 handling editor declared a past co-authorship with one of the authors YC.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fgene.2021.683037

De novo Transcriptome Sequencing Coupled With Co-expression Analysis Reveal the Transcriptional Regulation of Key Genes Involved in the Formation of Active Ingredients in *Peucedanum praeruptorum* Dunn Under Bolting Period

Cheng Song[1,2], Xiaoli Li[2,3], Bin Jia[2,3], Li Liu[2,3], Jinmei Ou[3] and Bangxing Han[1,2,3]**

[1] College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an, China, [2] Anhui Engineering Laboratory for Conservation and Sustainable Utilization of Traditional Chinese Medicine Resources, West Anhui University, Lu'an, China, [3] College of Pharmacy, Anhui University of Chinese Medicine, Hefei, China

OPEN ACCESS

Edited by:
Yunpeng Cao,
Central South University Forestry and Technology, China

Reviewed by:
Yunpeng Zhang,
East China Normal University, China
Yichao Li,
Sun Yat-sen University, China
Teng Zhang,
Xiamen University, China

***Correspondence:**
Jinmei Ou
too9319@163.com
Bangxing Han
hanbx1978@sina.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 19 March 2021
Accepted: 26 April 2021
Published: 14 June 2021

Citation:
Song C, Li X, Jia B, Liu L, Ou J and Han B (2021) De novo Transcriptome Sequencing Coupled With Co-expression Analysis Reveal the Transcriptional Regulation of Key Genes Involved in the Formation of Active Ingredients in Peucedanum praeruptorum Dunn Under Bolting Period. Front. Genet. 12:683037. doi: 10.3389/fgene.2021.683037

Peucedanum praeruptorum Dunn is a perennial and one-off flowering plant of the *Peucedanum* genus in Umbelliferae. The cultivated *P. praeruptorum* Dunn usually grows nutritionally in the first year and then moves into the reproductive growth in the second year. The lignification of the roots caused by bolting leads to the quality decline of crude materials. Since most of the previous studies have dealt with coumarin biosynthesis and identification of functional genes in *P. praeruptorum*, the scientific connotation of the inability that the bolted *P. praeruptorum* cannot be used medically is still unclear. Here, we employed a transcriptome sequencing combined with coexpression analysis to unearth the regulation mechanism of key genes related to coumarin synthesis in pre- and postbolting period, and to explore the mechanisms underlying the effects of bolting on the formation and transport of coumarins between the annual and biennial plants. Six cDNA libraries were constructed, and the transcripts were sequenced and assembled by Illumina Hiseq platform. A total of 336,505 unigenes were obtained from 824,129 non-redundant spliced transcripts. Unigenes (114,488) were annotated to the NCBI nr database, 119,017 and 10,475 unigenes were aligned to Gene Ontology (GO) functional groups and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, respectively. Differential expression analysis screened out a series of upregulated and downregulated genes related to the phenylpropanoid pathway. The heatmap clustering showed that the similar expression patterns were both observed in groups C vs. D and groups C vs. F. The WGCNA-based coexpression was performed to elucidate the module and trait relationship to unearth important genes related to the bolting process. Seven pivotal modules on the KEGG functional annotations suggested these genes were mainly enriched in the process of plant–pathogen interaction, plant hormone signal transduction, MAPK signaling pathway, α-linolenic acid metabolism, circadian

rhythm, and phenylpropanoid pathway. Further analysis provided clues that the key genes of the phenylpropanoid pathway, the ABC transporters, the apoptosis-related and circadian rhythm regulatory genes may play pivotal roles in regulating bolting signaling, biosynthesis, and transportation of coumarins.

Keywords: *Peucedanum praeruptorum*, transcriptional regulation, coexpression analysis, key gene, coumarins, bolting period

INTRODUCTION

Peucedanum praeruptorum Dunn, as a traditional Chinese medicine, is well known for the dried root of the *Peucedanum* genus of the Umbelliferae family. It has multiple effects on dispersing wind and heat, and resolving phlegm (Zhao et al., 2015). Dihydropyran-type coumarin compounds are the main medicinal components of *P. praeruptorum*, including praeruptorin A, praeruptorin B, and praeruptorin C. (Wang et al., 2015). In addition to coumarins, the main chemical components of *P. praeruptorum* also include volatile oils, phenanthrenequinones, organic acids, and sterols (Chu et al., 2020). Studies have proven that the coumarin compounds from *P. praeruptorum* have wide applications in the prevention and treatment of cardiovascular and cerebrovascular diseases, anti-inflammatory, reversing multidrug resistance, anticancer, and neuroprotection (Lee et al., 2015). The methanol extract of the *P. praeruptorum* root could reduce the allergic pneumonia symptoms, and lessen the secretion of mucus and histamine in the airway epithelium, as well as the infiltration of eosinophils (Lee et al., 2017). Some other studies had demonstrated that praeruptorin A could inhibit the migration and invasion of liver cancer cells, and inhibit the expression of the MMP1 gene by activating the extracellular signal-regulated kinase (ERK) signaling pathway, thereby inhibiting the movement of liver cancer cells (Yu et al., 2020). Praeruptorin C had good pharmacological effects in anti-inflammatory, antihypertensive, and antiplatelet aggregation (Liu et al., 2020). Studies had shown that praeruptorin C could significantly inhibit the proliferation, colony formation, wound healing, and migration of the non-small cell lung cancer cells. By inhibiting the phosphorylation of the ERK1/2 signaling pathway, it apparently reduced the expression of cathepsin D and thereby inhibited the invasion activity of the non-small cell lung cancer cells (Liu et al., 2020). In addition, praeruptorin C has a good therapeutic role in improving neuroprotection such as motor and cognitive impairment in Huntington's disease (Wang et al., 2017).

Recently, the functional genomics has been widely used for the analysis of biosynthetic pathways of specific metabolites of medicinal plants and the mining of functional genes in crucially synthetic pathways (Zhang et al., 2015; Bains et al., 2019; Yu et al., 2019). In particular, the combination of two or more technologies such as genome, transcriptome, proteomics, and metabolomics helps to clarify the formation and molecular regulation mechanism on the TCM efficacy factors (Devi et al., 2016). Coumarin compounds are derived from the phenylpropanoid metabolic pathway. By using high-throughput sequencing and metabonomics technology, some studies had identified some key genes for the synthesis and transport of coumarin compounds, and speculated that the cytochrome P450 family genes and MDR transporters are probably involved in the synthesis and transport of coumarins (Zhao et al., 2015; Song et al., 2020). Phenylalanine ammonia-lyase (*PAL*), as the first rate-limiting enzyme of the phenylalanine pathway, plays an important role in regulating the synthesis of flavonoids, phenols and coumarin compounds. *PpPAL* could respond to abiotic stresses such as jasmonic acid, UV-B, and cold treatment, leading to a rapid increase in the expression level in *P. praeruptorum* (Sui et al., 2019). 4-Coumarate: CoA ligase (4CL) is an important enzyme in the phenylpropanoid-branching pathway, which is responsible for catalyzing the synthesis of cinnamoyl CoA, p-coumaryl CoA, caffeoyl CoA, and ferulic acid CoA. Among the three 4CL genes identified in *P. praeruptorum*, it was found that *Pp4CL1* mainly used coumaric acid and ferulic acid as substrates for the synthesis of downstream metabolites, and could also catalyze precursors such as caffeic acid, cinnamic acid, and o-coumaric acid. However, its paralogous genes *Pp4CL7* and *Pp4CL10* did not show catalytic activity for hydroxycinnamic acid compounds (Liu et al., 2017). The ortho-hydroxylation of hydroxycinnamate is a key step in the synthesis of coumarins and is mainly used for the cyclization of coumarin lactones. p-Coumaroyl CoA 2'-hydroxylase (*C2'H*) is a rate-limiting enzyme in the upstream-branching pathway of coumarins, which is used for the synthesis of umbelliferone—the precursor of praeruptorin A. The expression level of *PpC2'H* was higher in the roots of *P. praeruptorum*, while the expression was upregulated after methyl jasmonate and UV-B treatments (Yao et al., 2017). The biosynthesis of pectoralum A and B is also required for the participation of some postmodifying enzymes, which are mainly used for methylation, oxy-methylation, prenylation, and redox reactions (Jian et al., 2020). The bergaptol *O*-methyltransferase (*BMT*) participated in the oxygen-methylation reaction of the coumarin bergaptol and had a high substrate specificity (Zhao et al., 2016a,b). Studies had identified a caffeic acid *O*-methyltransferase-similar (*COMT-S*) that from *P. praeruptorum* for the oxygen-methylation reaction of hydroxycoumarin (Zhao et al., 2019b). Further analysis showed that caffeic acid *O*-methyltransferase (*CAOMT*) was obtained from its paralog *BMT* through gene duplication about 37 to 1 million years ago. Due to the increasing need for coumarin compounds in domestic markets, traditional natural extraction can no longer meet the demand. With the rise of synthetic biology, mining specific functional genes will help in the heterologous expression and large-scale production of the medicinal ingredients from *P. praeruptorum*. Using high-throughput sequencing technology, some have identified

PAL, *4CL*, and *C2'H*, three enzymes from *Peucedanum purpurea*, combined with tyrosine ammonia lyase (*TAL*) obtained from *Rhodotorula glutinis*, which removed prephenate dehydratase (*pheA*) and anthranilate synthase (*trpE*), a transcriptional regulatory protein (*tyrR*), to construct a microbial cell production route for the synthesis of the coumarin precursor umbelliferone (Zhao et al., 2019a).

At present, *P. praeruptorum* is still dominated by wild species. With the increase in market demand, wild resources are gradually depleted. The raw materials of *P. praeruptorum* have been unable to meet the needs of the market. As a perennial and one-time flowering plant, the wild *P. praeruptorum* is usually grown for more than 3 years until bolting and flowering. The vegetative growth of *P. praeruptorum* is in the first year of cultivation, and bolting and flowering in the second year (Zhou et al., 2014). Plants such as *Peucedanum*, *Angelica*, *Saposhnikovia*, *Notopterygium*, and *Glehnia* genus of the Umbelliferae family have long been unable to have their roots collected for medicinal use after bolting and flowering. Since the roots of *P. praeruptorum* starts reproductive growth, the roots begin to lignify, and the content of medicinal ingredients is greatly reduced, thus, the cultivated products produced as medicinal materials can only be harvested in the same year (Liang et al., 2018). Early bolting seriously affects the accumulation of secondary metabolites of TCM materials, and has a great impact on the yield and quality of *P. praeruptorum* medicinal materials (Yrjönen et al., 2016). The regulatory mechanism on the inability of the bolted Umbelliferae plants for medicinal use is still unclear. One possible reason is that a large number of nutrients in *P. praeruptorum* are used for bolting in the reproductive growth period, and the roots cannot obtain sufficient nutrition to undertake the normal metabolic activities, which results in increasing the area of secondary xylem and decreasing the content of coumarins (Chen et al., 2019). Another possible reason is that the growth center of *P. praeruptorum* changed upon bolting, and most of the nutrients produced by photosynthesis are transferred from the roots to the apical parts for the development of leaves and floral organs of *P. praeruptorum*. This may be the main reason for the phenomenon of early bolting (Yu et al., 2012). In this study, a high-throughput transcriptome sequencing technology and coexpression analysis was used to analyze the expression patterns of key genes in coumarin biosynthesis before and after bolting, and explored the internal mechanism of bolting on the biosynthesis and transport of coumarins in different *P. praeruptorum*. The experimental samples with different harvest periods, different bolting period and different orientations were collected, and about 330,000 unigenes were identified through transcriptome sequencing. The GO function annotation and KEGG annotation results have shown that pathways such as posttranscriptional modification, signal transduction, and secondary metabolism are highly enriched in these differential genes. Further analysis showed that the key genes of the phenylpropane pathway and ABC transporter, apoptosis-related genes and circadian rhythm regulation genes may both play an important role in regulating bolting, coumarin biosynthesis, and transportation.

MATERIALS AND METHODS

Plant Sample Collection and Pretreatment

The samples used in this experiment were collected from the Donghekou Cultivation Base in Jin'an District, Lu'an City of Anhui Province. The geographical coordinates are 116.6567° east longitude and 31.4044° north latitude. The original plant was identified by Professor Han Bangxing of Wanxi University as *P. praeruptorum* of the *Peucedanum* genus of Umbelliferae. The samples of *P. praeruptorum* were collected between March 2019 and November 2019. First, the enzyme-free tube was precooled with liquid nitrogen. The disinfection and de-RNase treatment of the sampling equipment were performed. Fresh *Peucedanum* root samples were collected, then the surface was quickly cleaned with RNase-free water. The samples were put into an enzyme-free tube for quick freezing in liquid nitrogen. After completely frozen, it was transferred to a refrigerator at −80°C for storage. One hundred milligrams of each *P. praeruptorum* sample was taken and grinded in liquid nitrogen. The RNAprep Pure Plant Kit (Tiangen Biotech (Beijing) Co., Ltd., Beijing, China) was used to extract total RNA from the samples.

The Transcriptome Sequencing and Assembly

First of all, it is necessary to evaluate and control the quality of raw data obtained. The FastQC method[1] was used for quality control of raw data. In order to obtain raw reads, the raw data files obtained by the Illumina Hiseq platform are analyzed by base calling and converted to the sequenced reads. The FASTQ file contains sequence read information corresponding with sequencing quality information. The pieces of information, such as the quality value of the original data, were calculated, and FastQC was used to evaluate the quality of the sequencing data of the sample. The Trimmomatic method (Bolger et al., 2014) was used to remove linkers and low-quality sequences in reads to obtain clean data. Trinity (Haas et al., 2013) was used to *de novo* assemble the clean data into a transcript and set the parameter to min_kmer_cov 2. The transcripts assembled by Trinity are non-redundant, and the longest transcript in each transcript cluster is taken as Unigene, which is used as the reference sequence for subsequent analysis. The original data have been uploaded to the SRA database, and the accession of the BioProject was PRJNA714368.

The Annotations and Gene Structure Analysis

The NCBI blast+ method was used to compare the transcripts with CDD, KOG, COG, NR, NT, PFAM, Swissprot, and TrEMBL databases to obtain functional annotation information (Altschul et al., 1997). According to the annotation of the transcript from Swissprot and TrEMBL, the GO function annotation information is obtained. The KEGG Automatic Annotation Server (KAAS) was used to obtain KEGG annotation information of the

[1] http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

TABLE 1 | The detailed information of the spliced transcripts and unigenes.

	No.	≥500 bp	≥1,000 bp	N50	N90	Max Len	Min Len	Total Len	Average Len
Transcript	824129	363,437	173,390	1,082	294	14,871	201	587,521,132	712.9
Unigene	336505	109,099	45,711	799	247	14,871	201	191,946,933	570.41

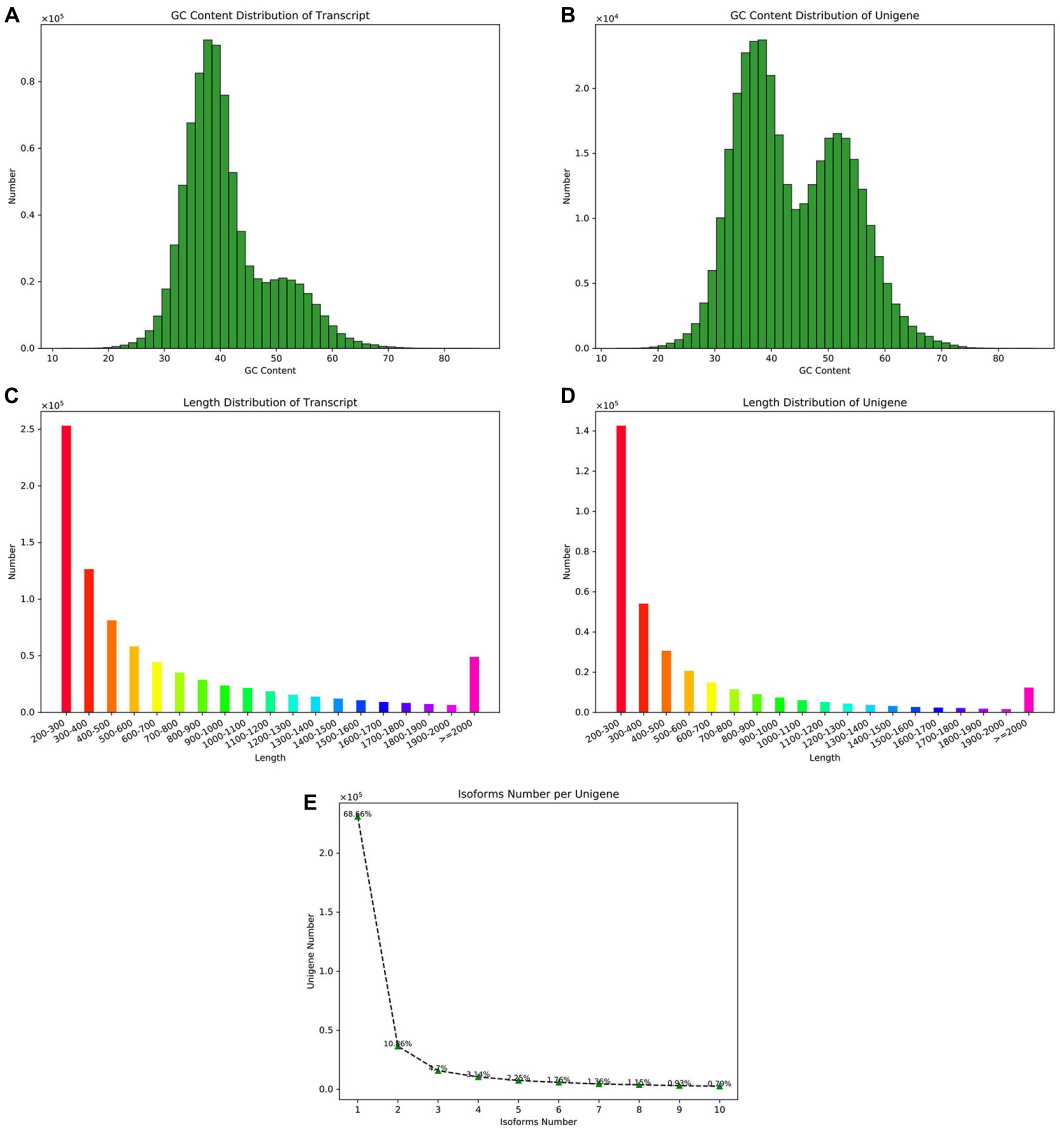


FIGURE 1 | The assembly and sequence length distribution of transcripts and unigenes. **(A)** The GC content distribution of the transcripts. **(B)** The GC content distribution of the unigenes. **(C)** The sequence length distribution of the transcripts. **(D)** The sequence length distribution of unigenes. **(E)** The isoform number of unigenes upon alternative splicing.

transcript (Moriya et al., 2007). By blasting the transcript and the database, the transdecoder software[2] was used for the CDS prediction. In order to better evaluate the quality of the RNA-seq data, the Bowtie2 software was used to compare the effective data of the samples to the spliced transcripts, and the mapping information was counted (Langmead and Salzberg, 2012). The RSeQC software was used to perform the redundant sequences and insert distribution analysis based on the comparison results (Wang et al., 2012). The BEDTools software was used to make the distribution check and the statistical analysis of gene coverage uniform (Quinlan and Hall, 2010). According to the mapping results, we used the BCFtools to perform the SNP analysis, and filtration was done based on the principle that the quality value is greater than 20, and the coverage is greater

[2]http://transdecoder.github.io/

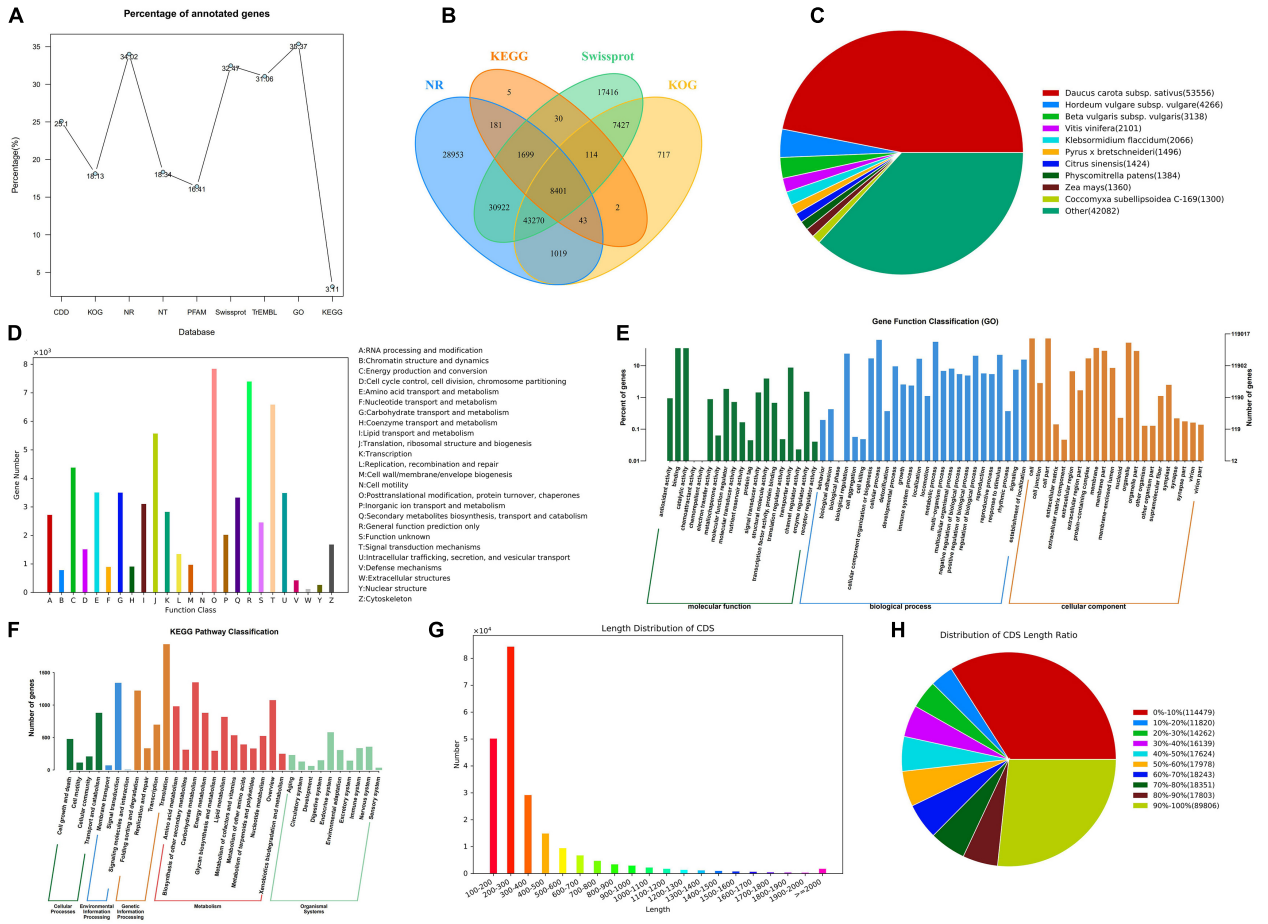


FIGURE 2 | The gene function annotation and the CDS prediction. **(A)** The ratio line chart of the main database annotation. **(B)** The Venn map of the annotated genes based on NR, Kyoto Encyclopedia of Genes and Genomes (KEGG), Swissprot, and KOG. **(C)** The distribution of homologous species based on NR. **(D)** The classification of annotation function based on KOG. **(E)** The Gene Ontology (GO) annotation distribution of different categories. **(F)** The KEGG pathway classification of the annotated genes. **(G)** The length distribution of CDS. **(H)** Distribution of CDS length ratio.

than 8 (Li, 2011). The MISA software[3] was used to perform the SSR analysis based on the sequence information of the spliced transcripts.

The Analysis of Gene Expression Level

Transcript abundance directly reflects the expression level of a certain gene. In this experiment, we used the TPM value to measure the gene expression level between different treatment groups and Salmon to calculate the gene expression level (Patro et al., 2017). For the repeated samples in the same group, the final expression level is the average of all repeated data. The VennDiagram in the R package was used to construct a common and unique Venn diagram of the expressed genes for all samples. The vegan installation package was used for PCA analysis and PCoA analysis. Also, the vegan package was used to calculate the evolution distance, and then the hclust was used to construct the number of clusters or box plots. The gplots installation package was used to construct a heatmap cluster among samples. The software version and R installation package used in the experiment are shown in **Supplementary Table 1**.

The Differential Expression and Co-Expression Analysis

For the samples with biological replicates, DESeq was used for the differential analysis. In order to obtain significantly different genes, we set the selection criteria as: qValue < 0.05 and the multiple of difference| FoldChange| > 2. The scatter plot and volcano plot were used to construct the differentially expressed gene distribution. The VennDiagram was used to construct a Venn diagram of the differential genes. The gplot package was used to construct the heatmap clustering of the differential gene and the expression trend maps of modules. The WGCNA script was to construct a gene set matrix for the correlation analysis of gene coexpression. The differential gene expression profile obtained by transcriptome analysis was used in the WGCNA data set. The soft threshold was further

[3]https://webblast.ipk-gatersleben.de/misa/

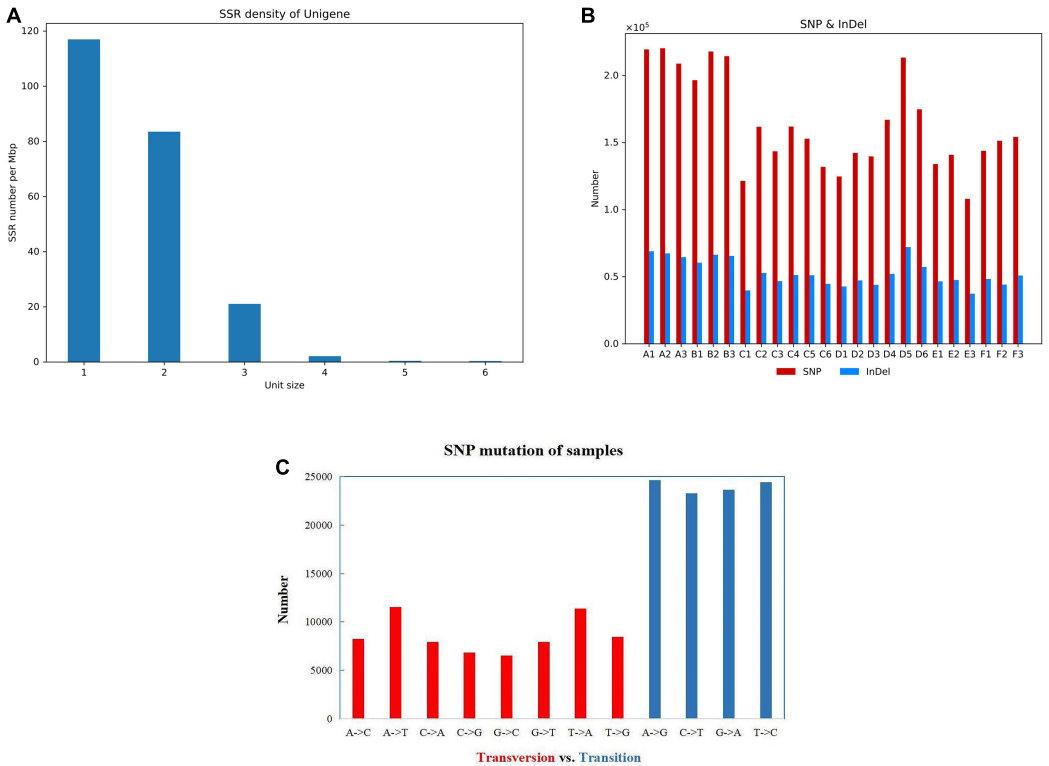


FIGURE 3 | The gene structure and the SNP analysis. **(A)** The density distribution of the SSR marker. The number on the *x*-axis represents the SSR repeat types. **(B)** The amount of SNP and InDel in different groups. **(C)** The SNP mutation of sample A1.

determined by constructing a gene matrix. After selecting a suitable soft threshold, the gene coexpression modules were performed to determine the number of genes in each module. First, the coexpression correlation of coefficient between genes was calculated based on the measured gene expression levels, and then Euclidean distance to cluster the genes by drawing a gene tree was used. The constructed gene tree was pruned by dynamic shearing. The pruned gene tree was fused to obtain gene modules. The differential genes of all groups were selected to visualize the correlation of the genes in the modules by clustering according to the expression amount between genes. A weighted analysis was performed on the phenotypic traits, and the correlation and credibility of all genes in each gene module was calculated with the phenotypic traits. The most relevant and significant modules were chosen as the core module. Finally, the correlation map of module membership and the difference weight of genes were obtained.

The Enrichment Analysis of the Key Genes

The clusterProfiler in the R package was used for the functional enrichment analysis. The topGO was used to visually analyze GO terms generated by GO enrichment. When the corrected *P*-value is less than 0.05, the function is considered to be significantly enriched. The igraph was used for the correlation analysis of functional enrichment.

RESULTS AND DISCUSSION

The cDNA Library Construction and Transcript Assembly

The sample information used for sequencing is shown in **Supplementary Table 2**. A total of 24 samples of *P. praeruptorum* were collected and constructed: A (biennially grown at undrawn phase), B (biennially grown at drawn phase), C (annually grown at undrawn phase), D (annually grown at drawn phase), E (annually grown in north slope), and F (annually grown in south slope) cDNA libraries; each group contains three or six biological replicates. The Illumina Hiseq platform was used to sequence and assemble transcripts of all samples. By removing the linker and the low-quality sequence, the quality control data of all samples is obtained (**Supplementary Table 3**). The number of spliced transcripts were 824,129, and the number of unigenes were 336,505 (**Table 1**). By comparing the distribution of GC content and sequence length in the spliced transcript and unigene, it is indicated that the GC content is basically distributed in the range of 40% to 60%, and the number of sequences with a length of 200 to 300 bp is the largest (**Figures 1A–D**). The variable shear analysis showed that the number of unigenes containing one isoform accounted for 68.6%, and those containing two isoforms accounted for 10.8% (**Figure 1E**). The results indicated that the obtained unigene could be used for further annotation and the differentially expressed gene analysis.

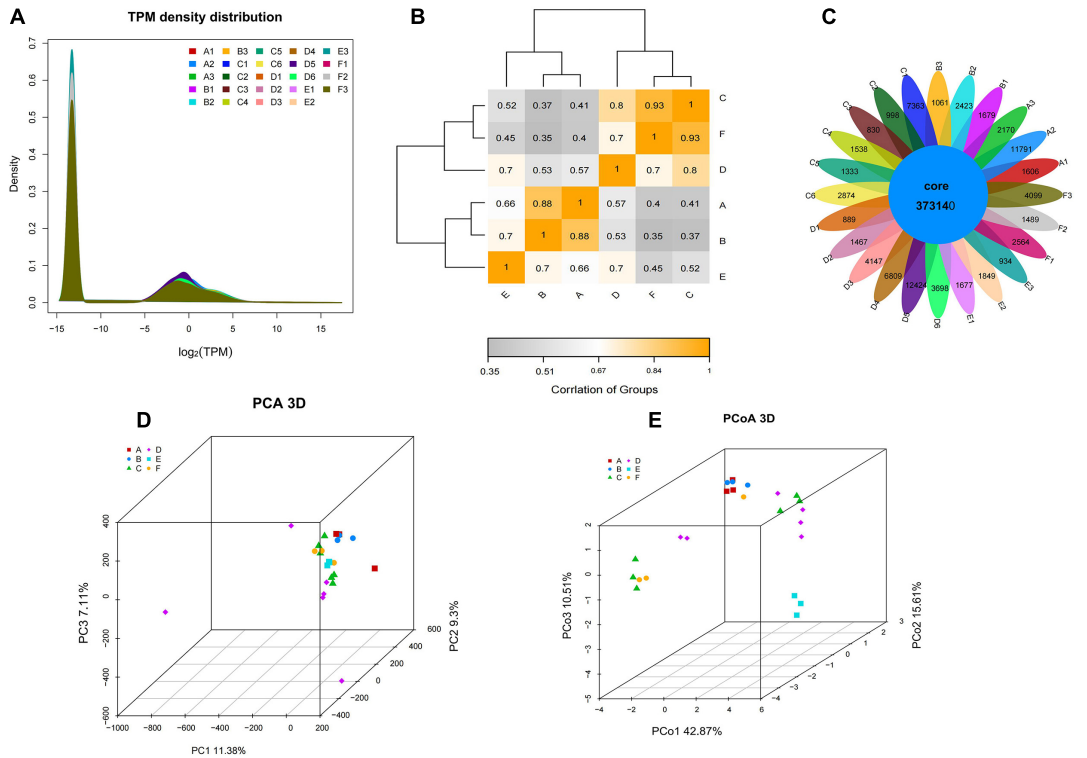


FIGURE 4 | The gene expression level and the intergroup correlation analysis. **(A)** The TPM density distribution between groups. **(B)** Heatmap cluster analysis of the correlation among samples. **(C)** The Venn of coexpression and specific genes. **(D)** PCA visualization of the different groups. **(E)** PCoA visualization of the different groups.

The Analysis of the Gene Annotation

In order to obtain the annotation information of almost all genes, the KOG, COG, NR, NT, PFAM, Swissprot, TrEMBL, and other databases were used for the functional gene annotation. The results indicated that 336,505 genes were annotated under varying databases, of which 114,488 unigenes were annotated to the NR database, and 119,017 were annotated to the GO database, accounting for 34.02% and 35.57% of the total annotated genes, respectively. Swiss-Prot and TrEMBL accounted for 32.47% and 31.06% of the annotated genes, respectively. KEGG had the least functionally annotated genes, accounting for only 3.11% (**Figure 2A**). The NR, KEGG, Swissprot, and KOG jointly annotated 8,401 genes (**Figure 2B**). Through the comparison with the NR library, we analyzed the homologous sequence of the transcript of *P. praeruptorum* and its similar species, which found that the transcript of *P. praeruptorum* has the highest similarity with *Daucus carota* subsp. *sativus* (**Figure 2C**). The annotated genes were classified into 26 groups in KOG. The categories of posttranslational modification, protein turnover, chaperones, signal transduction mechanisms, and general function prediction contained more genes (**Figure 2D**). The GO function annotations in biological process, cellular component, and molecular function indicated that a majority of the genes were involved in molecular functions such as protein binding, enzyme catalysis, and transport. Some were involved in biological processes such as cell processing, metabolic processes, biological regulation, and environmental induction. A few of the genes were involved in the process, such as composition of cells and organelles (**Figure 2E**). The classification annotation of KEGG metabolic pathway indicated that more genes were involved in transport and catabolism, signal transduction, folding, sorting, and degradation, transcription, carbohydrate metabolism, amino acid metabolism, and other pathways (**Figure 2F**). Most CDS lengths are concentrated in 200–300 bp, followed by 100–200 bp (**Figure 2G**). The higher proportion of the CDS region in the total sequence length is less than 10% or greater than 90% of the sequence (**Figure 2H**).

The Gene Structure Analysis

The RSeQC package and BEDTools software were used to analyze the redundant series and gene coverage of the samples. Using MISA performing the SSR detection on the gene transcripts, more than 38,000 SSR markers were identified (**Supplementary Table 4**). According to the combination and number of bases, it mainly included three types of SSR markers such as single-base repeats, double-base repeats, and three-base repeats (**Figure 3A**). SNP is a genetic marker formed by a single nucleotide variation, which reflects the polymorphism of a gene. Using BCFtools, SNP/InDel analysis showed that the number of SNP mutant genes in all samples was higher than the number of InDel mutations. SNP mutation sites include two types: transversion and transformation, and the number

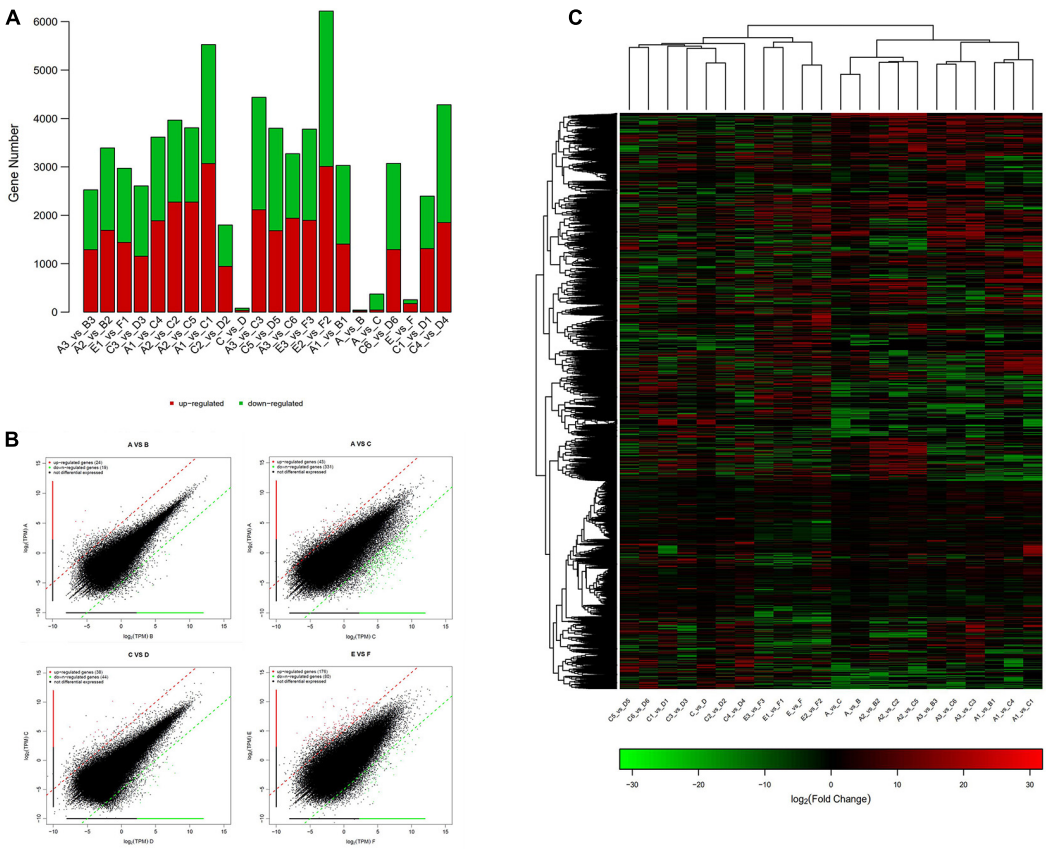


FIGURE 5 | The differential gene expression analysis of *P. praeruptorum*. **(A)** The amount of up- and downregulated DEGs in different groups. **(B)** Scatter plots of the DEGs in A vs. B, A vs. C, C vs. D, and E vs. F. **(C)** Heatmap of the differential genes between groups based on fold change values. The red denotes upregulated expression, and the green denotes downregulated expression.

of transformation mutations is higher than the number of transversion mutation genes (**Figures 3B,C**).

The Analysis of Differentially Expressed Genes

In RNA-seq analysis, the expression level of a gene could be estimated by counting the sequencing sequence (reads) located in the genomic region or gene exon region. In this study, the TPM value was used to measure the expression level of a certain gene (**Figure 4A**). By comparing the gene expression density curves of the six groups of samples, it was found that the relative density of most non-expressed genes was higher, and the expression interval of some normally expressed genes was set between $|\log2(TPM)| < 5$. The correlation analysis between groups showed that the correlation between group C and group F was higher, as well as group A and group B, which indicated that the gene expression of these samples was similar (**Figure 4B**). The Venn diagram denoted the number of shared and unique expressed genes ($TPM > 0$) in the sample. The coexpression Venn showed that 37,314 expressed genes shared in all samples, and some had a certain number of specific expressions (**Figure 4C**). PCA analysis reflects the distance and aggregation between samples. It was found that the three principal components were unable to separate the six groups (**Figure 4D**). To observe the differences between individuals and groups, further PCoA analysis showed that group E was separated from other groups, but group A vs. B, group C vs. F, and group C vs. D could not be entirely separated (**Figure 4E**). It was suggested that these three pairs of groups were collected in the same year (either in the first year or in the second year). Little difference in genetic level was observed in the bolting period, and bolting did not affect the expression level of most genes. The hierarchical cluster analysis and distance analysis between groups also showed that the difference between group A and group B is small, as well as group C vs. D, and group C vs. F. The overall difference between groups is not significant (**Supplementary Figure 1**).

The statistics of the expressed differential genes showed that the number of differential genes in group A vs. B and group C vs. D was less, and the number of differential genes in group A vs. C and E vs. F was more. Among them, there are 97 and 92 upregulated and downregulated genes in groups A and B, respectively. There are 158 and 225 upregulated and downregulated genes in groups C and D, respectively. There were 685 and 2,248 upregulated and downregulated genes in groups A and C, respectively. There are 2,118 and 1,384 genes that

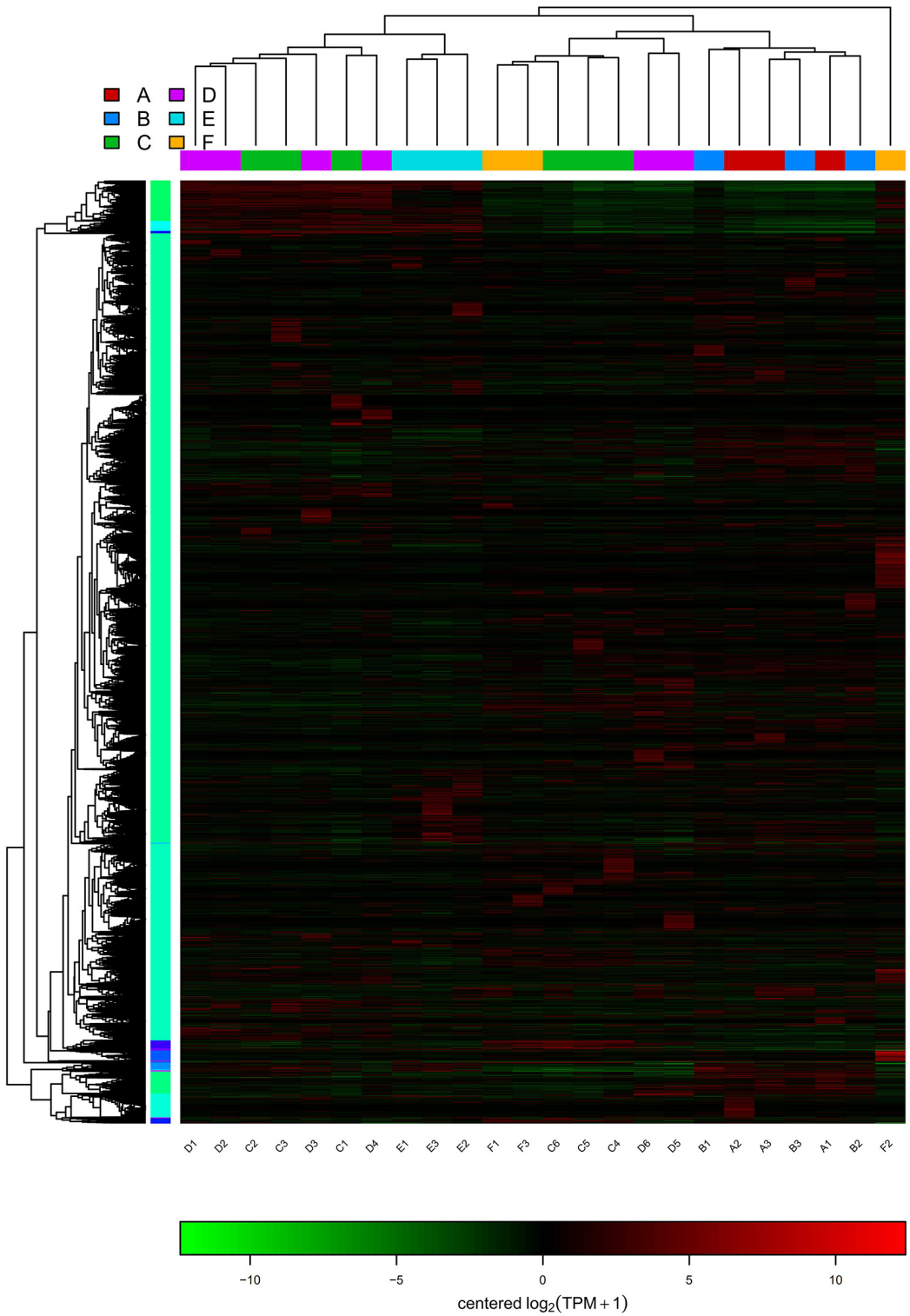


FIGURE 6 | The heatmap clustering of the differential genes between groups.

are significantly different between groups E and F (**Figure 5A**). **Figure 5B** showed the extremely significant differences between upregulated and downregulated genes (q-value < 0.01). The heatmap drawn by fold change indicated that the differential genes of group A vs. B and group A vs. C had similar clustering patterns, and the differential genes of group C vs. D and group

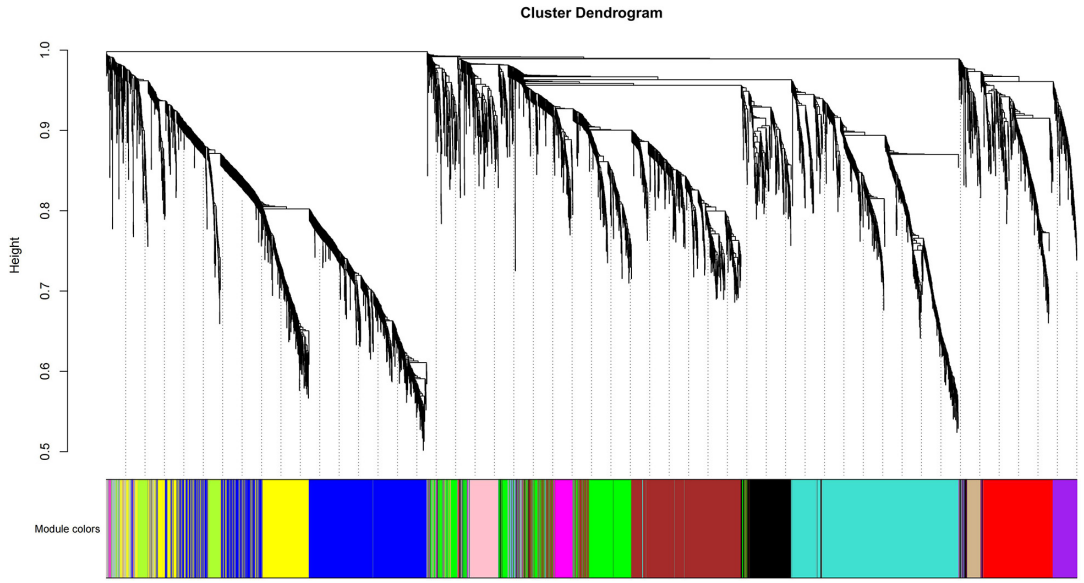


FIGURE 7 | The cluster dendrogram of coexpressed genes.

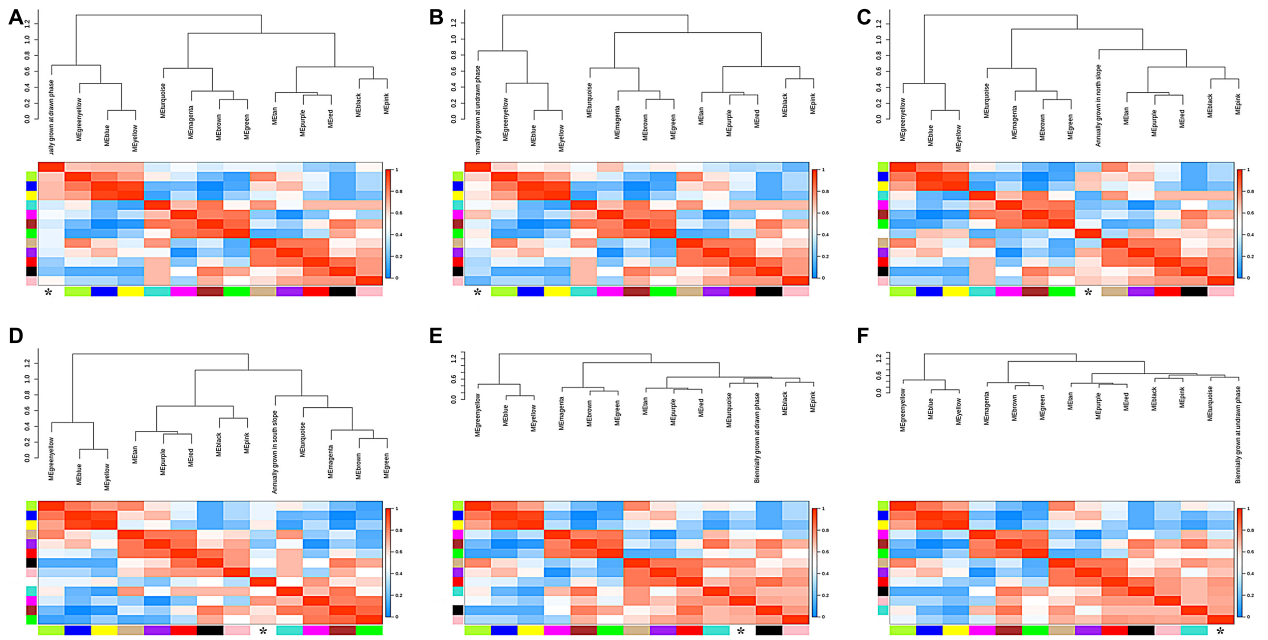


FIGURE 8 | The eigengene adjacency trees of different conditions. **(A)** Annually grown at drawn phase. **(B)** Annually grown at undrawn phase. **(C)** Annually grown in north slope. **(D)** Annually grown in south slope. **(E)** Biennially grown at drawn phase. **(F)** Biennially grown at undrawn phase.

E vs. F had similar clusters (**Figure 5C**). The clustering analysis of DEG clusters showed that the differential genes from groups A and B were clustered into one category, and the differential genes from groups C and D were clustered into one category. However, the distance between two branches of group A and C was relatively long, and group E and F were obviously clustered into two types. The above results suggested that the overall level of gene expression difference of *P. praeruptorum* is relatively small in the bolting phase, and different years and different slope (photoperiod) conditions had a greater contribution to the difference in gene expression (**Figure 6**). In addition, the expression trend of differential gene clusters showed that the expression levels of most gene sets (subcluster 3 and subcluster 4) tended to be stable, and the expression levels of other gene sets (subcluster 1, subcluster 8, and subcluster 15) appeared uneven pattern (**Supplementary Figure 2**).

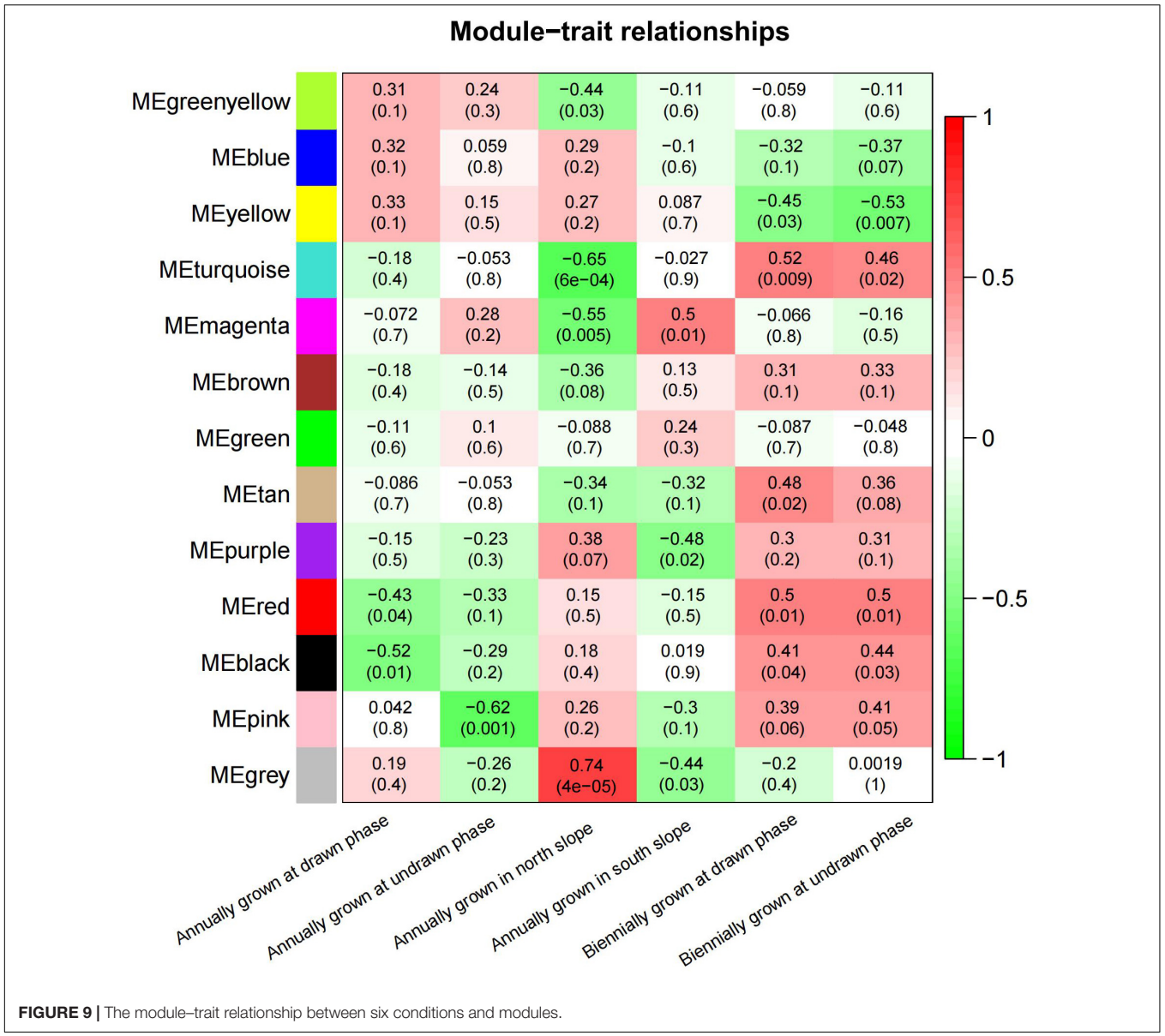


FIGURE 9 | The module–trait relationship between six conditions and modules.

In order to examine the changes in differential gene expression of *P. praeruptorum* before and after bolting, we analyzed the differentially expressed genes of group A vs. B. The results showed that *MAPK1*, *PP2A*, sugar efflux transporter, *DUF4413*, *DUF639*, *PPR*, etc., that related to signal transduction were significantly downregulated after bolting, while *DFR*, *ABCC3*, *ERD6*, protochlorophyllide reductase, seed storage proteins, phloem protein 2 (*PP2*), GDSL lipase, magnesium transporter, and other genes related to growth and development, synthesis, and transportation of secondary metabolites were upregulated (**Supplementary Table 5**). By comparing the differentially expressed genes of group C vs. D, *ABCB29*, serine hydroxymethyltransferase, *AGL80*, GA-regulated protein, *NRT2.5*, polygalacturonase, *bZIP2*, *bHLH93*, glucose and ribitol dehydrogenase, *CYP81D11*, aquaporin, SBP family genes, anthocyanin acyltransferase, *TCP15*, *RGLG2*, and *WRKY16* were downregulated after bolting. *MYC2*, *ABCC15*, auxin response factor, *ABCC3*, *HCT*, ACC oxidase, carotene 3-hydroxylase, *P5CS*, carotenoid oxygenase, *bZIP35*, *TPS*, *CYP82C3*, *UGT90*, *NAC056*, *NAC043*, *MYB113*, and *NAC025* genes are upregulated after bolting (**Supplementary Table 7**). The above results indicated that bolting event initiated the transformation of vegetative growth, and the carbon source and some secondary metabolism began to transport from the compound-synthesizing part to the apical parts. The differentially expressed genes of the 1- and 2-year-old unbolted *P. praeruptorum* mainly involved the interaction between plants and pathogenic bacteria, MAPK signaling pathway, hormone signal transduction, starch and sugar metabolism, linolenic acid synthesis, phenylpropane synthesis, and other processes (**Supplementary Table 6**). *P. praeruptorum* grown in the different directions received different regulation on the photoperiod and intensity. Differential

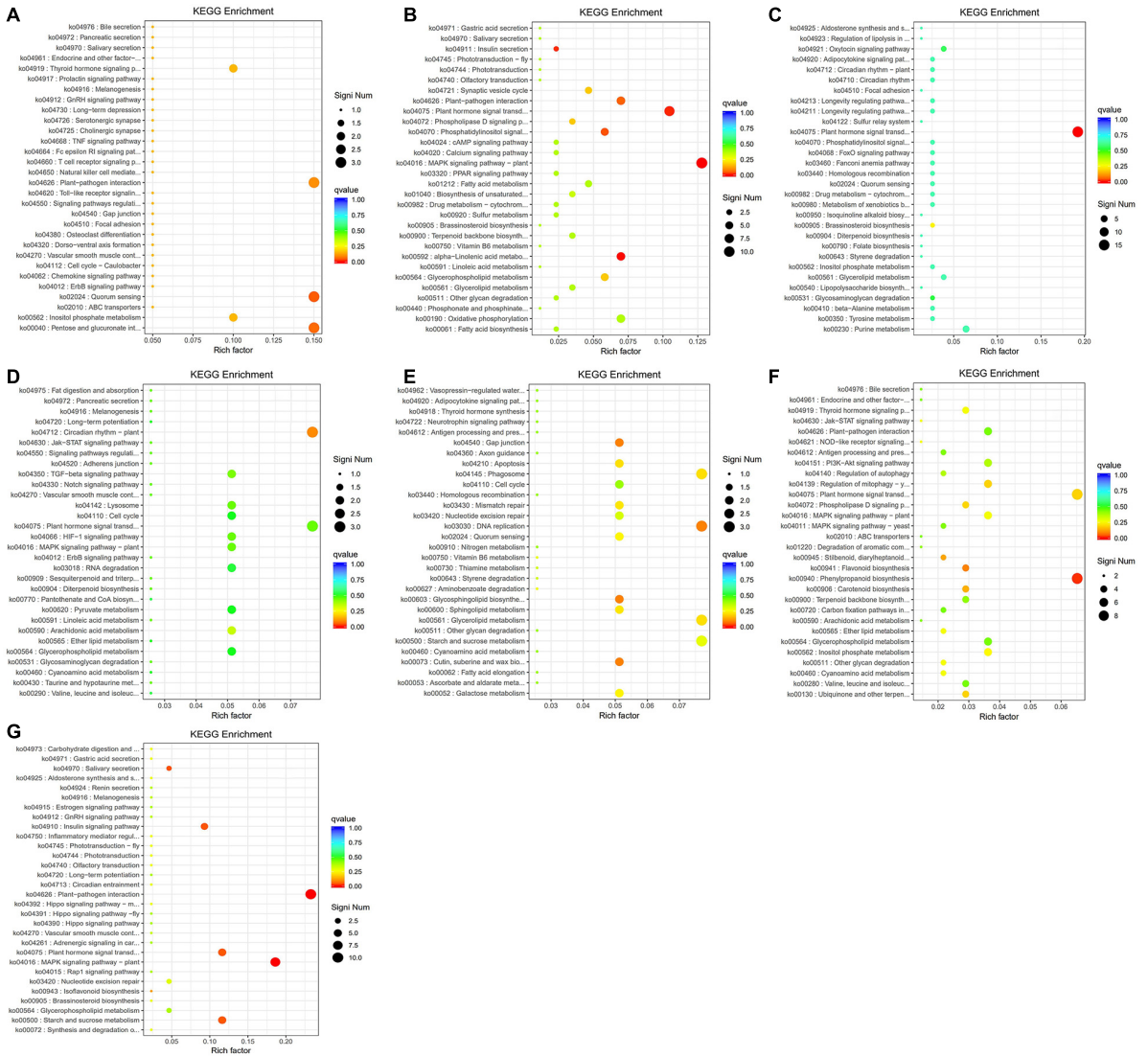


FIGURE 10 | The scatter plots of the significant KEGG enrichment of functional pathway. **(A)** The highly enriched pathways of the black module. **(B)** The highly enriched pathways of the blue module. **(C)** The highly enriched pathways of the brown module. **(D)** The highly enriched pathways of the green module. **(E)** The highly enriched pathways of the red module. **(F)** The highly enriched pathways of the turquoise module. **(G)** The highly enriched pathways of the yellow module. The vertical axis is the function annotation information, and the horizontal axis is rich factor (equal to the number of differential genes annotated to the function divided by the number of genes annotated to the function). *Q* value is shown in different colors, and the number of different genes is shown in the size of dots.

expression analysis showed that starch and sugar metabolism, circadian rhythm, oxidative phosphorylation, hormone signal transduction, carbon fixation in photosynthesis, MAPK signal pathway, fatty acid metabolism, and carbon metabolism are mainly involved in photosynthesis and development processes (**Supplementary Table 8**).

Coexpression Analysis and Functional Enrichment of the Differential Gene

To study the correlation of gene expression, the total expression matrix and all genes were used for the WGCNA analysis. All of 24 samples from six groups were used for the analysis of co-expression. According to the expression profiles of these differential genes, we obtained the WGCNA data matrix to determine the soft threshold. The results showed that if signed R^2 was set to 0.9, the soft threshold was 8, which suggested that it was more appropriate to construct a coexpression matrix under the power value of 8 (**Supplementary Figure 3**). Euclidean distance was used to cluster genes and draw a cluster dendrogram, which displayed each module through a hierarchical clustering tree. It was indicated that there were 13 gene sets that could form representative modules (**Figure 7**). By gene coexpression of correlation coefficient, further differentially expressed genes between groups were selected to construct a visualized TOMplot (**Supplementary Figure 4**). The correlation

TABLE 2 | The significantly differential genes of the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment in seven modules.

Modules	Unigene ID	Description	KEGG term
Black	TRINITY_DN91982_c0_g1,TRINITY_DN97897_c1_g1,TRINITY_DN94663_c1_g5	Pectate lyase, pectate lyase, nonspecific phospholipase	Quorum sensing
	TRINITY_DN91982_c0_g1,TRINITY_DN97897_c1_g1,TRINITY_DN80952_c0_g1	Pectate lyase, pectate lyase, pectinesterase	Pentose and glucuronate interconversions
	TRINITY_DN92836_c0_g1,TRINITY_DN74599_c2_g1,TRINITY_DN78227_c0_g4	CDK30, MKK6, EDS1	Plant–pathogen interaction
Blue	TRINITY_DN69364_c0_g2,TRINITY_DN93921_c0_g2,TRINITY_DN93921_c0_g1, TRINITY_DN87065_c0_g1,TRINITY_DN78780_c1_g1,TRINITY_DN77043_c0_g3, TRINITY_DN90039_c0_g1,TRINITY_DN69537_c0_g1,TRINITY_DN80918_c1_g1, TRINITY_DN96204_c0_g1,TRINITY_DN70380_c1_g4	MKS1, serine/threonine–protein kinase OXI1, OXI1, bHLH14, PP2Cc, MYC2, MKKA, MKKK4, WRKY22, MPK9, MKS1	MAPK signaling pathway—plant
	TRINITY_DN98338_c2_g1,TRINITY_DN91314_c0_g2,TRINITY_DN91314_c0_g1, TRINITY_DN88300_c0_g1,TRINITY_DN98156_c2_g2,TRINITY_DN93109_c0_g3	LOX2, AOC4, AOC4, OPR2, AOC3, ACX1	Alpha-linolenic acid metabolism
	TRINITY_DN82132_c0_g1,TRINITY_DN90894_c0_g2,TRINITY_DN73073_c1_g1, TRINITY_DN87479_c2_g5,TRINITY_DN87065_c0_g1,TRINITY_DN78780_c1_g1, TRINITY_DN77043_c0_g3,TRINITY_DN84473_c4_g1,TRINITY_DN94676_c0_g3	TIFY10A, XTH23, TIFY10A, TIFY10A, bHLH14, PP2Cc, MYC2, GID1B, BZR1	Plant hormone signal transduction
	TRINITY_DN83723_c0_g2,TRINITY_DN92948_c0_g1,TRINITY_DN95105_c1_g1, TRINITY_DN98530_c3_g1,TRINITY_DN93972_c0_g1	CML45, PLC4, DGK1, DGK5, DGK2	Phosphatidylinositol signaling system
	TRINITY_DN79382_c1_g1,TRINITY_DN83723_c0_g2,TRINITY_DN69924_c1_g1, TRINITY_DN66938_c0_g1,TRINITY_DN80918_c1_g1,TRINITY_DN96169_c1_g1	Calcium-binding allergen 8, CML45, CPK7, CML23, WRKY22, respiratory burst oxidase homolog protein C	Plant–pathogen interaction
Brown	TRINITY_DN86120_c2_g3,TRINITY_DN72157_c2_g1,TRINITY_DN89061_c0_g2, TRINITY_DN79989_c1_g2,TRINITY_DN73550_c1_g1,TRINITY_DN88273_c1_g1, TRINITY_DN89338_c2_g1,TRINITY_DN87704_c0_g2,TRINITY_DN75691_c0_g1, TRINITY_DN81580_c1_g1,TRINITY_DN91987_c1_g4,TRINITY_DN76934_c0_g1, TRINITY_DN80774_c0_g1,TRINITY_DN83614_c0_g1,TRINITY_DN87173_c0_g1	SAUR72, ARF18, PP2Cc, TGA-2.1, SAUR32, bZIP7, ARR9, IAA13, ABF2, IAA11, PYL9, bZIP8, GBF4, ARR3, DELLA protein GAI	Plant hormone signal transduction
	TRINITY_DN95742_c1_g1,TRINITY_DN98250_c0_g1	CYP90D1, CYP749A22	Brassinosteroid biosynthesis
	TRINITY_DN79675_c0_g3,TRINITY_DN79269_c0_g2	Heparan-α-glucosaminide *N*-acetyltransferase, heparanase-like protein 3	Glycosaminoglycan degradation
	TRINITY_DN98033_c2_g1,TRINITY_DN94990_c1_g3	GIGANTEA, TCP7	Circadian rhythm—plant
Green	TRINITY_DN96996_c3_g1,TRINITY_DN89650_c0_g2,TRINITY_DN78434_c1_g1	CRY1, APRR5, CCA1	Circadian rhythm—plant
	TRINITY_DN92859_c0_g1,TRINITY_DN94016_c0_g3	GGT1, phospholipase A2-alpha	Arachidonic acid metabolism
	TRINITY_DN94016_c0_g3	Phospholipase A2-alpha	Linoleic acid metabolism
Red	TRINITY_DN93413_c1_g1,TRINITY_DN93413_c0_g1	Alpha-galactosidase, alpha-galactosidase	Glycosphingolipid biosynthesis
	TRINITY_DN92960_c0_g1,TRINITY_DN71353_c4_g2	Fatty acid omega-hydroxylase, fatty aldehyde decarbonylase	Cutin, suberin, and wax biosynthesis
	TRINITY_DN87965_c0_g2,TRINITY_DN93413_c1_g1,TRINITY_DN93413_c0_g1	PDAT2, alpha-galactosidase, alpha-galactosidase	Glycerolipid metabolism
Turquoise	TRINITY_DN77989_c3_g2,TRINITY_DN86350_c0_g1,TRINITY_DN75021_c0_g3, TRINITY_DN95188_c0_g1,TRINITY_DN95807_c3_g1,TRINITY_DN81440_c1_g1, TRINITY_DN95494_c1_g1,TRINITY_DN74684_c0_g2,TRINITY_DN81857_c0_g3	CCoAOMT, PER21, *O*-acetyltransferase, CAOMT, beta-glucosidase, CAOMT, C4H, PER51, CCR1	Phenylpropanoid biosynthesis
	TRINITY_DN77989_c3_g2,TRINITY_DN75021_c0_g3,TRINITY_DN96835_c0_g1, TRINITY_DN95494_c1_g1	CCoAOMT, *O*-acetyltransferase, chalcone isomerase, C4H	Flavonoid biosynthesis
	TRINITY_DN77989_c3_g2,TRINITY_DN75021_c0_g3,TRINITY_DN95494_c1_g1	CCoAOMT, *O*-acetyltransferase, C4H	Stilbenoid, diarylheptanoid and gingerol biosynthesis
	TRINITY_DN92852_c0_g3,TRINITY_DN74872_c1_g1,TRINITY_DN70885_c1_g1, TRINITY_DN89214_c0_g1	NCED1, CCD8B, CYP97A3, VDE	Carotenoid biosynthesis
Yellow	TRINITY_DN75582_c1_g3,TRINITY_DN87724_c1_g3,TRINITY_DN83018_c0_g4, TRINITY_DN95142_c1_g2,TRINITY_DN88604_c0_g1,TRINITY_DN88820_c1_g3, TRINITY_DN85142_c0_g4,TRINITY_DN76663_c0_g1 ,TRINITY_DN76663_c0_g3,TRINITY_DN69751_c0_g2	CML19, WRKY33, CML25, WRKY33, WRKY33, WRKY22, CML27, RIN4, RIN4, CML45	Plant–pathogen interaction
	TRINITY_DN87724_c1_g3,TRINITY_DN95142_c1_g2,TRINITY_DN88604_c0_g1, TRINITY_DN86043_c0_g1,TRINITY_DN88820_c1_g3,TRINITY_DN93901_c2_g1, TRINITY_DN73035_c0_g1,TRINITY_DN86457_c0_g1	WRKY33, WRKY33, WRKY33, ABCB22, WRKY22, PYL4, PYL5, MKK9	MAPK signaling pathway—plant
	TRINITY_DN88900_c1_g4,TRINITY_DN98074_c1_g1,TRINITY_DN95611_c1_g2, TRINITY_DN89504_c0_g1,TRINITY_DN79580_c1_g1	*O*-Glycosyl hydrolase, alpha- amylase 2, beta-amylase 3, UDP-glucuronate 4-epimerase, beta-amylase 3	Starch and sucrose metabolism

TABLE 3 | The significantly different genes between groups.

Groups	Unigene ID	Description	Log_2(fold change)	KEGG term
A vs. C	TRINITY_DN78443_c0_g1	26-*O*-β-glucosidase	−1.69	Phenylpropanoid biosynthesis
	TRINITY_DN90447_c0_g1	Cationic peroxidase 2	−4.51	
	TRINITY_DN88245_c0_g2	PER64	6.33	
	TRINITY_DN88245_c0_g3	PER64	7.17	
	TRINITY_DN88958_c1_g6	β-Glucosidase 12	−1.08	
	TRINITY_DN78489_c3_g1	Glutathione peroxidase	−4.16	
	TRINITY_DN88179_c1_g2	HCT	−3.87	
	TRINITY_DN79471_c0_g4	HCT	−10.53	
	TRINITY_DN92248_c0_g1	PER44	−10.71	
	TRINITY_DN92351_c2_g2	4CL	−1.38	
	TRINITY_DN79471_c0_g4	HCT	−10.54	
	TRINITY_DN95807_c3_g1	β-Glucosidase 11	2.66	
	TRINITY_DN74096_c0_g1	ABCB1	3.02	ABC transporters
	TRINITY_DN96811_c2_g1	ABCB9	2.68	
	TRINITY_DN74096_c1_g1	ABCB1	3.36	
	TRINITY_DN88179_c1_g2	HCT	−3.87	Flavonoid biosynthesis
	TRINITY_DN91517_c0_g1	CHS1	3.86	
	TRINITY_DN79471_c0_g4	HCT	−10.53	
C vs. D	TRINITY_DN88080_c0_g1	β-Glucosidase 44	−3.70	Phenylpropanoid biosynthesis
	TRINITY_DN93179_c0_g3	IPK2b	6.14	Phosphatidylinositol signaling system
	TRINITY_DN68773_c0_g2	IPK2a	4.47	
	TRINITY_DN86197_c2_g3	PLC2	4.89	
	TRINITY_DN86197_c2_g1	PLC2	4.83	
	TRINITY_DN69779_c1_g2	Cytochrome c	−4.95	Apoptosis—multiple species
	TRINITY_DN70114_c2_g1	Cytochrome c	−3.34	
E vs. F	TRINITY_DN78443_c0_g1	β-Glucosidase	−3.01	Phenylpropanoid biosynthesis
	TRINITY_DN93468_c0_g1	PER63	1.88	
	TRINITY_DN91057_c0_g1	4CL	−3.63	
	TRINITY_DN69279_c0_g1	PER52	−4.59	
	TRINITY_DN92351_c2_g2	4CL	−3.83	
	TRINITY_DN97609_c1_g1	DFR	−2.96	Flavonoid biosynthesis
	TRINITY_DN80821_c0_g2	CHI	−7.89	
	TRINITY_DN92803_c1_g2	ABCC2	−2.69	ABC transporters
	TRINITY_DN95608_c1_g1	ABCC2	−3.98	
	TRINITY_DN82168_c0_g4	CRY	−3.10	Circadian rhythm—plant
	TRINITY_DN87544_c1_g2	GI	−4.72	
	TRINITY_DN96996_c3_g1	CRY	−4.12	
	TRINITY_DN77328_c0_g1	PHYA	2.71	

diagram between these modules obtained by clustering suggested that the module eigengene of different treatment groups were clustered with the modules of close correlation (**Figure 8**). The asterisks represented the specific position of the six different traits in the modules. For example, the MEgreenyellow module was combined with annually grown at drawn phase and annually grown at undrawn phase. MEpink and MEblack are grouped together with annually grown in the north slope. MEturquiose combined with annually grown in the south slope, biennially grown at drawn and undrawn phase. Moreover, the weighted analysis of traits was performed to obtain the correlation and confidence of all genes and different traits in each module. The results showed that, except for MEgray, MEred and MEgreen were significantly negatively correlated with annually grown at undrawn phase. MEpink had an extremely significantly negative correlation with annually grown at the drawn phase. MEturquoise and MEmagenta are extremely significantly correlated with annually grown in the north slope negatively, while MEturquoise and biennially grown at drawn phase were significantly positively correlated. MEyellow was significantly negatively correlated with biennially grown at undrawn phase (**Figure 9**). The specific correlation and confidence of module membership and gene significance are shown in **Supplementary Figure 5**.

Based on the WGCNA results above, a total of 13 modules with high correlation were obtained. Among them, seven modules (yellow, turquoise, red, green, brown, blue, and black) contain plentifully coexpressed genes. The heatmap

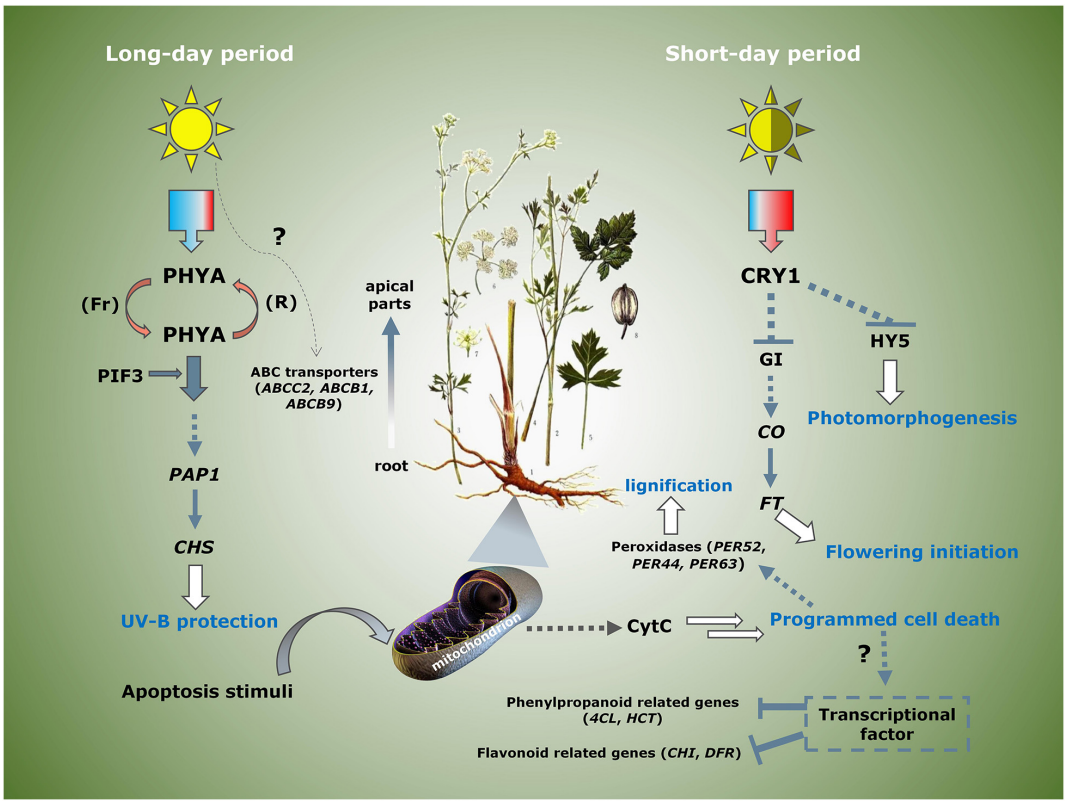


FIGURE 11 | A plausible model on the molecular mechanism of photoperiod regulating bolting and transformation of coumarins in *P. praeruptorum*. *PHYA* (TRINITY_DN77328_c0_g1), *PIF3* (TRINITY_DN83313_c0_g2), *PAP1* (TRINITY_DN63777_c0_g1, TRINITY_DN63757_c0_g2), *CRY1* (TRINITY_DN82168_c0_g4, TRINITY_DN96996_c3_g1), *GI* (TRINITY_DN87544_c1_g2), *CO* (TRINITY_DN75433_c0_g3, TRINITY_DN94065_c2_g1), *FT* (TRINITY_DN110183_c0_g1, TRINITY_DN61953_c0_g1), *HY5* (TRINITY_DN89278_c0_g1, TRINITY_DN5919_c0_g1), *CHS* (TRINITY_DN70261_c1_g1, TRINITY_DN91517_c0_g1, TRINITY_DN80066_c0_g1), *CytC* (TRINITY_DN69779_c1_g2, TRINITY_DN70114_c2_g1).

clustering matrix indicated that there were more coexpressed genes among blue–yellow, blue–brown, blue–green, blue–blue, brown–yellow, brown–brown, and turquoise–blue modules (**Figure 10**). Functional enrichment analysis and KEGG pathway analysis could obtain the expression pattern of one typical gene in a specific metabolic pathway. Further GO-enriched annotation on these seven modules indicated that the GO terms annotated in the black module mainly included protein kinase activity, seed development, developmental process involved in reproduction, response to endogenous stimulus and hormone, meristem development, cell wall organization, intrinsic component of membrane, etc. The terms annotated in the black module included serine/threonine kinase activity, kinase activity, response to stress, transferase activity, plasma membrane, oxidoreductase activity, cellulose biosynthetic process, UDP-glucosyltransferase activity, signal transduction, etc. The terms annotated in the Brown module are mainly responses to endogenous stimulus, hormone-mediated signaling pathway, DNA-binding transcription factor activity, etc. The terms annotated in the green module mainly contained transcription regulator activity, circadian rhythm, response to abiotic stimulus, response to oxygen-containing compound, flower development, reproductive shoot system development, etc. The items annotated in the red module mainly included plant-type cell wall, anchored component of plasma membrane, microtubule, polymeric cytoskeletal fiber, hydrolase activity, cytoskeleton, etc. The items annotated in the yellow module mainly included response to endogenous stimulus, response to hormone, DNA-binding transcription factor activity, defense response, response to abiotic and biotic stress, etc. (**Supplementary Table 9**).

The results of KEGG pathway enrichment analysis are shown in **Supplementary Table 10**. The enriched genes in the black module included quorum sensing, pentose and glucuronate interversion, plant–pathogen interaction, and signal transduction, translation, carbohydrate metabolism, and other biological processes (**Figure 10A**). The enrichment in the blue module contained MAPK signaling pathway, plant hormone signal transduction, α-linolenic acid metabolism, transport, and catabolism, and lipid metabolism (**Figure 10B**). The enriched genes in the brown module are mainly involved in plant hormone signal transduction (**Figure 10C**). The enriched genes in the green module are mainly involved in the circadian rhythm (**Figure 10D**). The red module is mainly involved in DNA replication, glycerolipid metabolism, and starch and sucrose metabolism (**Figure 10E**). The enriched genes in the turquoise

module are mainly involved in the phenylpropanoid biosynthesis and plant hormone signal transduction (**Figure 10F**). The yellow module is mainly involved in plant-pathogen interaction, MAPK signaling pathway, starch and sucrose metabolism, plant hormone signal transduction, and environmental adaptation (**Figure 10G**). Finally, the differential genes from the seven modules based on KEGG enrichment are described in **Table 2**.

To explore the key genes that regulated the phenylpropane pathway further, we performed KEGG enrichment analysis on the differential genes between the groups (**Table 3**). By comparing the differentially expressed genes of group A vs. C, it indicated that most of the peroxidases related to the formation of lignin monomers were apparently downregulated, and the *CHS* related to flavonoid biosynthesis was upregulated. *4CL* and *HCT*, which are involved in the synthesis of phenylpropanoid, were downregulated, while the ABC transporters involved in coumarin transport were upregulated. Group C vs. D involved the differentially expressed genes on bolting. It was indicated that phosphatidylinositol signaling and apoptosis genes were the main differential genes, suggesting that they may be involved in the process of programmed cell necrosis (PCD) of root cells. Compared with group F, the genes involved in the synthesis and transport of phenylpropanoids and flavonoids of group E were almost downregulated, indicating that long-day period was beneficial to induce the expression of these genes. Interestingly, we have also identified several genes related to the circadian rhythm. Phytochrome A was upregulated in short-day period, and *CRY* and *GI* genes are upregulated in long-day period, indicating that the two types of genes that jointly regulated the photoperiod flowering and development of *P. praeruptorum* are in different proportions of red and blue light. In summary, we proposed a photoperiod-dependent working model for the growth and regulation of secondary metabolites of *P. praeruptorum* (**Figure 11**). The photoperiod and apoptosis signaling jointly participated in the regulation of the bolting period and later growth. Under long-day conditions, PHYA and PIF3 were coordinated to regulate the expression of downstream CHS related to UV-B light stress. Under short-day conditions, CRY participated in the expression of CONSTANS, which, in turn, enhanced the upregulated expression of FT to initiate flowering, followed by regulating HY5 to control its photomorphogenesis. The apoptosis process mainly involved the programmed death of root cells, which further affected the lignification process and the expression levels of related biosynthesis genes. This study will provide a scientific reference for the bolting and flowering mechanism of *P. praeruptorum* and the regulation of key genes in coumarin biosynthesis.

CONCLUSION

Bolting is often the turning point of plants from vegetative growth to reproductive growth. However, the scientific connotation for which the crude material of bolted *P. praeruptorum*, hardly used for medicinal purposes, need to be addressed. By transcriptome sequencing technology, the gene expression profiles of *P. praeruptorum* was employed under different conditions, by which a series of key genes related to phenylpropanoid metabolism and transportation was filtered. Differential gene analysis and coexpression further revealed the connection between growth and secondary metabolism. We inferred that the physiological bolting may not be the direct cause of the downregulation of genes involved in the coumarin synthesis pathway. The programmed cell death and the photoperiod regulation could probably be the reason for leading the lignification of roots and caused the chemical components to migrate to the apical parts after bolting. Some evidence proved that numerous materials were supplemented in the reproductive growth period. The underground part like taproot may not obtain sufficient nutrition to undertake the normal metabolism, which leads to thickening of the secondary xylem and the reduction of coumarins. The results will provide theoretical support for the early flowering intervention, coumarin biosynthesis, and transport of *P. praeruptorum*.

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

BH, CS, and XL designed the research. CS, XL, BJ, and LL conducted the experiments. CS, XL, and JO analyzed the data. CS wrote the manuscript. BH, CS, JO, and XL revised the manuscript and improved the English. BH and CS acquired the funding. All authors have read, reviewed, and approved the submitted version.

FUNDING

This work was supported by the National Industry Technology System of Traditional Chinese Medicine (CARS-21), Anhui Outstanding Youth Fund (1808085J17), and High-Level Talents Research Initiation Funding Project (WGKQ202001011).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.683037/full#supplementary-material

Supplementary Figure 1 | The distance similarity analysis between samples. **(A)** Hierarchical clustering tree. **(B)** Heatmap of distance between samples. **(C)** Analysis of anosim group similarity. If *R*-value is close to 1, difference between groups become greater. If *R*-value is close to 0, suggesting there is no significant difference between and within groups. The *p*-value denote the significance of this statistical analysis.

Supplementary Figure 2 | The expression trend of different gene modules.

Supplementary Figure 3 | The soft threshold plot of scale free topology model. The dot line denoted the signed R^2 value at 0.9.

Supplementary Figure 4 | The TOM plot with cluster dendrogram. The yellow matrix across diagonal represents the correlation between modules, and the density represents the number of co-expressed genes.

Supplementary Figure 5 | The module membership of different conditions with gene significant weight. **(A)** Annually grown at drawn phase. **(B)** Annually grown at undrawn phase. **(C)** Annually grown in north slope. **(D)** Annually grown in south slope. **(E)** Biennially grown at drawn phase. **(F)** Biennially grown at undrawn phase.

Supplementary Table 1 | The software version and R packages used in the experiment.

Supplementary Table 2 | The detailed information of the samples in the experiment.

Supplementary Table 3 | The QC data statistics of all samples.

Supplementary Table 4 | The SSR markers of unigenes.

Supplementary Table 5 | The significantly differential expressed genes between group A and B.

Supplementary Table 6 | The significantly differential expressed genes between group A and C.

Supplementary Table 7 | The significantly differential expressed genes between group C and D.

Supplementary Table 8 | The significantly differential expressed genes between group E and F.

Supplementary Table 9 | The functional GO enrichment of seven modules obtained by co-expression analysis.

Supplementary Table 10 | The KEGG enrichment of seven modules obtained by co-expression analysis.

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Bains, S., Thakur, V., Kaur, J., Singh, K., and Kaur, R. (2019). Elucidating genes involved in sesquiterpenoid and flavonoid biosynthetic pathways in saussurea lappa by de novo leaf transcriptome analysis. *Genomics* 111, 1474–1482. doi: 10.1016/j.ygeno.2018.09.022

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

Chen, L. L., Chu, S. S., Zhang, L., Xie, J., Dai, M., Wu, X., et al. (2019). Tissue-specific metabolite profiling on the different parts of bolting and unbolting peucedanum praeruptorum Dunn (Qianhu) by laser microdissection combined with UPLC-Q/TOF-MS and HPLC-DAD. *Molecules* 24:1439. doi: 10.3390/molecules24071439

Chu, S., Chen, L., Xie, H., Xie, J., Zhao, Y. J., Tong, Z. Z., et al. (2020). Comparative analysis and chemical profiling of different forms of peucedani radix. *J. Pharm. Biomed. Anal.* 189:113410. doi: 10.1016/j.jpba.2020.113410

Devi, K., Mishra, S. K., Sahu, J., Panda, D., Modi, M. K., and Sen, P. (2016). Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus. *Sci. Rep.* 6:21026. doi: 10.1038/srep21026

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Jian, X., Zhao, Y., Wang, Z., Li, S., Li, L., Luo, J., et al. (2020). Two CYP71AJ enzymes function as psoralen synthase and angelicin synthase in the biosynthesis of furanocoumarins in peucedanum praeruptorum Dunn. *Plant Mol. Biol.* 104, 327–337. doi: 10.1007/s11103-020-01045-4

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lee, A. R., Chun, J. M., Lee, A. Y., Kim, H. S., Gu, G. L., and Kwon, B. I. (2017). Reduced allergic lung inflammation by root extracts from two species of peucedanum through inhibition of Th2 cell activation. *J. Ethnopharmacol.* 196, 75–83. doi: 10.1016/j.jep.2016.12.015

Lee, J., Lee, Y. J., Kim, J., and Bang, O. S. (2015). Pyranocoumarins from root extracts of Peucedanum praeruptorum Dunn with multidrug resistance reversal and anti-inflammatory activities. *Molecules* 20, 20967–20978. doi: 10.3390/molecules201219738

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Liang, W. H., Chang, T. W., and Charng, Y. C. (2018). Influence of harvest stage on the pharmacological effect of Angelica dahurica. *Bot. Stud.* 59:14. doi: 10.1186/s40529-018-0230-1

Liu, C. M., Shen, H. T., Lin, Y. A., Yu, Y. L., Chen, Y. S., Liu, C. J., et al. (2020). Antiproliferative and antimetastatic effects of praeruptorin C on human non-small cell lung cancer through inactivating ERK/CTSD signalling pathways. *Molecules* 25:1625. doi: 10.3390/molecules25071625

Liu, T., Yao, R., Zhao, Y., Xu, S., Huang, C., Luo, J., et al. (2017). Cloning, functional characterization and site-directed mutagenesis of 4-Coumarate: coenzyme a ligase (4CL) involved in coumarin biosynthesis in peucedanum praeruptorum Dunn. *Front. Plant Sci.* 8:4. doi: 10.3389/fpls.2017.00004

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, 182–185.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

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

Song, J., Luo, H., Xu, Z., Zhang, Y., Xin, H., Zhu, D., et al. (2020). Mining genes associated with furanocoumarin biosynthesis in an endangered medicinal plant, Glehnia littoralis. *J. Genet.* 99:11. doi: 10.1007/s12041-019-1170-6

Sui, Z., Luo, J., Yao, R., Huang, C., Zhao, Y., and Kong, L. (2019). Functional characterization and correlation analysis of phenylalanine ammonia-lyase (PAL) in coumarin biosynthesis from Peucedanum praeruptorum Dunn. *Phytochemistry* 158, 35–45. doi: 10.1016/j.phytochem.2018.11.006

Wang, L., Wang, J., Yang, L., Zhou, S. M., Guan, S. Y., Yang, L. K., et al. (2017). Effect of Praeruptorin C on 3-nitropropionic acid induced Huntington's disease-like symptoms in mice. *Biomed. Pharmacother.* 86, 81–87. doi: 10.1016/j.biopha.2016.11.111

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28:2184. doi: 10.1093/bioinformatics/bts356

Wang, X. Y., Li, J. F., Jian, Y. M., Wu, Z., Fang, M. J., and Qiu, Y. K. (2015). On-line comprehensive two-dimensional normal-phase liquid chromatography× reversed-phase liquid chromatography for preparative isolation of Peucedanum praeruptorum. *J. Chromatogr. A* 1387, 60–68. doi: 10.1016/j.chroma.2015.02.003

Yao, R., Zhao, Y., Liu, T., Huang, C., Xu, S., Sui, Z., et al. (2017). Identification and functional characterization of a p-coumaroyl CoA 2'-hydroxylase involved in the biosynthesis of coumarin skeleton from peucedanum praeruptorum Dunn. *Plant Mol. Biol.* 95, 199–213. doi: 10.1007/s11103-017-0650-4

Yrjönen, T., Eeva, M., Kauppila, T. J., Martiskainen, O., Summanen, J., Vuorela, P., et al. (2016). Profiling of coumarins in peucedanum palustre (L.) Moench populations growing in Finland. *Chem. Biodivers.* 13, 700–709. doi: 10.1002/cbdv.201500198

Yu, C. L., Yu, Y. L., Yang, S. F., Hsu, C. E., Lin, C. L., Hsieh, Y. H., et al. (2020). Praeruptorin A reduces metastasis of human hepatocellular carcinoma cells

by targeting ERK/MMP1 signaling pathway. *Environ. Toxicol.* 36, 540–549. doi: 10.1002/tox.23059

Yu, G., Ma, Y. X., Duan, J. A., Song, B. S., and He, Z. Q. (2012). Identification of differentially expressed genes involved in early bolting of Angelica sinensis (Apiaceae). *Genet. Mol. Res.* 11, 494–502. doi: 10.4238/2012.March.6.2

Yu, G., Zhou, Y., Yu, J., Hu, X., Tang, Y., Yan, H., et al. (2019). Transcriptome and digital gene expression analysis unravels the novel mechanism of early flowering in Angelica sinensis. *Sci. Rep.* 9:10035. doi: 10.1038/s41598-019-46414-2

Zhang, X., Allan, A. C., Li, C., Wang, Y., and Yao, Q. (2015). De Novo assembly and characterization of the transcriptome of the Chinese medicinal herb, Gentiana rigescens. *Int. J. Mol. Sci.* 16, 11550–11573. doi: 10.3390/ijms160511550

Zhao, Y., Jian, X., Wu, J., Huang, W., Huang, C., Luo, J., et al. (2019a). Elucidation of the biosynthesis pathway and heterologous construction of a sustainable route for producing umbelliferone. *J. Biol. Eng.* 13:44. doi: 10.1186/s13036-019-0174-3

Zhao, Y., Liu, T., Luo, J., Zhang, Q., Xu, S., Han, C., et al. (2015). Integration of a decrescent transcriptome and metabolomics dataset of Peucedanum praeruptorum to investigate the CYP450 and MDR genes involved in coumarins biosynthesis and transport. *Front. Plant Sci.* 6:996. doi: 10.3389/fpls.2015.00996

Zhao, Y., Luo, J., Xu, S., Wang, W., Liu, T., Han, C., et al. (2016a). Selection of reference genes for gene expression normalization in peucedanum praeruptorum Dunn under abiotic stresses, hormone treatments and different tissues. *PLoS One* 11:e0152356. doi: 10.1371/journal.pone.0152356

Zhao, Y., Wang, N., Sui, Z., Huang, C., Zeng, Z., and Kong, L. (2019b). The molecular and structural basis of o-methylation reaction in coumarin biosynthesis in peucedanum praeruptorum dunn. *Int. J. Mol. Sci.* 20:1533. doi: 10.3390/ijms20071533

Zhao, Y., Wang, N., Zeng, Z., Xu, S., Huang, C., Wang, W., et al. (2016b). Cloning, functional characterization, and catalytic mechanism of a Bergaptol O-Methyltransferase from Peucedanum praeruptorum dunn. *Front. Plant Sci.* 7:722. doi: 10.3389/fpls.2016.00722

Zhou, J., Wang, W., Liu, M., and Liu, Z. (2014). Molecular authentication of the traditional medicinal plant Peucedanum praeruptorum and its substitutes and adulterants by DNA-Barcoding technique. *Pharmacogn. Mag.* 10, 385–390. doi: 10.4103/0973-1296.141754

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 handling editor declared a past collaboration with one of the authors CS.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fgene.2021.687689

Genome-Wide Analysis of *PEBP* Genes in *Dendrobium huoshanense:* Unveiling the Antagonistic Functions of *FT/TFL1* in Flowering Time

*Cheng Song[1,2], Guohui Li[3], Jun Dai[1,2] and Hui Deng[1,2]**

[1] College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an, China, [2] Anhui Engineering Laboratory for Conservation and Sustainable Utilization of Traditional Chinese Medicine Resources, West Anhui University, Lu'an, China, [3] College of Life Science, Anhui Agricultural University, Hefei, China

OPEN ACCESS

Edited by:
Yunpeng Cao, Central South University of Forestry and Technology, China

Reviewed by:
Quan Gan, Anhui Academy of Agricultural Sciences, Chinese Academy of Agricultural Sciences (CAAS), China
Irfan Ali Sabir, Shanghai Jiao Tong University, China
Cheng Qin, Zunyi Vocational and Technical College, China

***Correspondence:**
Hui Deng
dhup@qq.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *29 March 2021*
Accepted: *18 May 2021*
Published: *09 July 2021*

Citation:
Song C, Li G, Dai J and Deng H (2021) Genome-Wide Analysis of PEBP Genes in Dendrobium huoshanense: Unveiling the Antagonistic Functions of FT/TFL1 in Flowering Time. Front. Genet. 12:687689. doi: 10.3389/fgene.2021.687689

Dendrobium is a semi-shade epiphytic Orchidaceae herb with important ornamental and medicinal value. Parts of the cultivation of Dendrobium germplasm resources, as well as the identification of medicinal components, are more studied, but the functional characterization of the flowering regulation in Dendrobium plants is less reported. Here, six PEBP family genes (*DhFT3, DhFT1, DhMFT, DhTFL1b, DhFT2,* and *DhTFL1a*) were identified from the *Dendrobium huoshanense* genome. The chromosome-level mapping showed that these genes were sequentially distributed on chromosomes 6, 9, 15, and 17. The paralogous gene *DhTFL1b* corresponded to *DhTFL1a*, which was determined through tandem duplication. The gene structure and conserved motif of *DhPEBP* indicated five PEBP genes apart from *DhMFT* contained four exons and three introns entirely. The phylogeny analysis showed that the PEBP gene family in *A. thaliana, O. sativa, Z. mays, S. lycopersicum,* and *P. equestris* were classified into three subclades, FT, TFL, and MFT, which maintained a high homology with *D. huoshanense*. The conserved domain of the amino acid demonstrated that two highly conserved short motifs (DPDXP and GXHR) embed in DhPEBPs, which may contribute to the conformation of the ligand binding bag. The 86th position of DhFTs was tyrosine (Y), while the 83th and 87th of DhTFL1s belonged to histidine (H), suggesting they should have distinct functions in flowering regulation. The promoter of six *DhPEBPs* contained several *cis*-elements related to hormone induction, light response, and abiotic stress, which indicated they could be regulated by the environmental stress and endogenous signaling pathways. The qRT-PCR analysis of *DhPEBPs* in short-term days induced by GA indicated the gene expressions of all *DhFTs* were gradually increased, whereas the expression of *DhTFL1* was decreased. The results implied that DhPEBPs have various regulatory functions in modulating flowering, which will provide a scientific reference for the flowering regulation of Dendrobium plants.

Keywords: *Dendrobium huoshanense*, flowering regulation, flowering locus T, homology, collinearity analysis, PEBP genes

INTRODUCTION

The traditional wild *Dendrobium huoshanense* experiences issues such as difficulty in natural pollination, fewer capsules, and low seed germination. Although asexual reproduction has been widely developed in the tissue culture of *D. huoshanense*, the degradation of germplasm is much more serious than sexual reproduction. Flower organs are the sexual reproduction organs of angiosperms, and the entire process of flower formation is essentially the basis for the procreation of offspring (Chen and Penfield, 2018). The flower-forming transition of plants from vegetative growth to reproductive growth is a key step to adapt to the external environment and ensure the reproduction of offspring (Liu et al., 2019). The timing of these transitions is precisely regulated by endogenous signals and the external environment, among which photoperiod signals can prompt plants to bloom at the most appropriate time (Plackett et al., 2018).

The protein encoded by phosphatidylethanolamine-binding protein (PEBP) is widely present in dicotyledonous monocots, and its family consists of three subfamilies, the FT-like, MFT-like, and TFL1-like (Auge et al., 2018). *FLOWERING LOCUS T (FT)*, as a florigen-inducing gene, has been identified in model plants such as Arabidopsis, rice, and maize (Ahluwalia and Hatsukami, 2015; Jin et al., 2019). FT is an important integration factor in the flowering regulation pathway of plants and has been one of the key genes regulating flowering (Adeyemo et al., 2019; Bi et al., 2019). It transmits signals to the downstream flower development-related *CONSTANS (CO)* by sensing the vernalization pathway, gibberellin pathway, photoperiod pathway, and autonomous regulation pathways (Melzer, 2017; Xiao et al., 2018). Under long-day conditions, *CO* induces the expression of *FT*, and the FT protein binds to *FLOWERING LOCUS D (FD)* protein in the stem end meristem (SAM) to promote the expression of *APETALA1* (*AP1*) (Li et al., 2015). *TWIN SISTER OF FT (TSF)*, the homologous gene of *FT*, regulates the early flowering of *A. thaliana* and exhibits a redundant function similar to that of overexpression of *FT* (Wang et al., 2019). Another branch of the PEBP family is the *TERMINAL FLOWER 1* (*TFL1*) subfamily, whose main function is to maintain vegetative growth and infinite growth of inflorescences. In Arabidopsis, *TFL1* controls the morphological structure of the plant by regulating the meristem genes, *LEAFY (LFY)* and *AP1,* present in SAM (Wang et al., 2017). In addition, *TFL1* plays a role in inhibiting flowering during flowering, and it exhibits completely different characteristics from FT (Jin et al., 2021). *ARABIDOPSIS THALIANA CENTRORADIALIS HOMOLOGU (ATC)* is classified into the same subclass as *TFL1* and is homologous to the snapdragon *CEN* gene. *ATC* overexpression can complement the late flowering phenotype of *tfl1*, but the *atc* mutant has no obvious early flowering phenomenon (González-Suárez et al., 2020). *Brother of FT and TFL1 (BFT)* is another member of the *TFL1* subfamily. Overexpression of *BFT* in *A. thaliana* shows delayed flowering, but the *bft* mutant does not show a similar phenotype to the *tfl1* mutant (Zhang et al., 2016). *MOTHER OF FT AND TFL1 (MFT)* is the ancestral gene of *FT* and *TFL1*. Overexpression of *AtMFT* can lead to early flowering, but it is not as significant as *FT*. In addition, MFT is specifically expressed in seeds, mainly through ABA and GA signaling pathways to participate in seed germination regulation (Yu et al., 2019). In dicotyledonous plants like *A. thaliana*, tomato, grape, and poplar, 6–9 genes are contained in this PEBP family (Luo et al., 2018). However, the number of PEBP family members in monocotyledonous plants is about three times that of dicotyledonous plants. *O. sativa* and *T. aestivum* have 17 and 30 PEBP genes, respectively. Through large fragments and genome-wide replication, plants have produced a large number of repetitive genes in the process of evolution. Some functions are redundant, some genes are silent as non-function, and some neofunctionalization genes have novel functions due to mutations (Schiessl et al., 2017). Six conserved ligand-binding sites in PEBP proteins of different species form a pocket-like structure, in which two amino acids are the key sites that determine the function of *FT/TFL1*. *TFL1* not only affects flowering time but also affects inflorescence morphology. In some plants, *TFL1* homologous genes may have different regulatory functions. For example, the DET and LF genes in pea control the flowering time and the developmental state of the apical meristem, respectively (Johnson et al., 2019).

To solve the bottleneck behind the sexual reproduction of *D. huoshanense*, the genetic background of *D. huoshanense,* the regulation mechanism of key genes related to flowering need to be clarified. Based on the previous genome sequencing, six PEBP family genes were screened from the whole genome of *D. huoshanense*. Among them, the two subfamilies FT and TFL1 have unique regulatory effects in the flowering initiation process. Phylogeny analysis showed that *MFT, FT,* and *TFL1* of *D. huoshanense* could be clustered with PEBP family of other species. By calculating the ka/ks of *DhPEBP* orthologous genes, all genes had suffered purified selections. Collinearity analysis revealed that the *FT/TFL1* of *D. huoshanense* and some PEBP family genes in *O. sativa*, *Z. mays,* and *G. max* have collinearity. Amino acid alignment analysis showed that the 86th amino acid of *DhFT2* and 87th of *DhTFL1a* were different. The promoter analysis indicated that signal elements related to stress and gibberellin binding possessed in *DhFT* and *DhTFL1*. Quantitative fluorescence analysis verified the flower-promoting effect of *DhFT1* and *DhFT3* as well as the flower-inhibiting effect of *DhTFL1a*. These studies would provide scientific reference for elucidating the different roles of PEBP family members in the flowering regulation of *D. huoshanense*.

MATERIALS AND METHODS

The tissue culture material of *D. huoshanense* used in the experiment came from the Anhui Engineering Technology Research Center of Plant Cell Engineering of West Anhui University (Lu'an City, Anhui Province). In total, 0.5 mM GA3 solution was sprayed once a day on the leaves of *D. huoshanense*. The samples were taken every 5 days, and each sample was biologically repeated three times. The control group was sprayed with clean water in the same way. After removing the leaves with scissors, we put them in liquid nitrogen and immediately froze

them before transferring them to a 5mL screw tube to store them in a refrigerator at −80 degrees for later use.

Identification of PEBP Family Genes in *D. huoshanense*

Firstly, the consensus conserved seed file (PF01161) of the hidden Markov model (HMM) was downloaded from the Pfam website[1]. Then, the HMM profile was performed as a query to identify all PBP-containing domain in *D. huoshanense* by retrieving against the genome with a threshold of e-value of $< e^{-3}$. All candidate DhPEBPs were verified by Pfam[2], SMART[3], and InterPro[4] for further confirmation. The same method was used to screen out PEBP genes in *A. thaliana, O. sativa subsp. japonica, Z. mays, S. lycopersicum,* and *P. equestris*. The genome sequence and CDS files of *A. thaliana, O. sativa subsp. japonica, Z. mays,* and *S. lycopersicum* were obtained from Ensembl Plants database[5]. The genome sequence and CDS file of *P. equestris* were got from the NCBI database[6]. Based on the sequence alignments generated by the Muscle method in MEGA software (v.6.06), all putative redundant PEBP sequences were discarded.

The Chromosome Location and Gene Duplication Events

By using TBtools (v.1.089), the chromosome location information of the PEBP gene was obtained (Chen et al., 2020). Firstly, the gene density information from the GFF3 file was extracted, and the gene location visualization from the GTF/GFF tool was used to obtain the chromosome location map. The Genome gene dot plot tool was used to draw a dot plot of the gene duplication events. Using MEME suite[7] for conservative motif analysis, the conservative motif of PEBP gene was obtained by searching the protein sequence. The number of motifs was set at 20, and the width of motifs was in 6 to 200, and select the default parameters to get the meme.xml file. TBtools was used to visualize the motif pattern and the gene structure analysis.

The Phylogeny Analysis of PEBP Genes

Neighbor-joining (NJ) and maximum-likelihood (ML) methods were used to construct the phylogenetic tree of PEBP genes in *A. thaliana, O. sativa subsp. japonica, Z. mays, S. lycopersicum,* and *P. equestris*. MEGA 6 was used to align the target protein sequence (Tamura et al., 2013). The aligned file was used to construct the phylogenetic tree and bootstrap consensus tree by using the NJ method. Bootstrap replications value is set to 1,000, and the substitution model was the Poisson model. Gaps and missing dates were treated in pairwise deletion. Using IQ-TREE (v.1.6.6), a phylogenetic tree was construct based on the ML method.

[1]http://pfam.sanger.ac.uk/family/PF01161

[2]http://pfam.xfam.org/search#tabview=tab1

[3]http://smart.embl-heidelberg.de/

[4]http://www.ebi.ac.uk/interpro/search/sequence/

[5]http://plants.ensembl.org/index.html

[6]https://www.ncbi.nlm.nih.gov/genome?LinkName=bioproject_genome&from_uid~$=$~192198

[7]https://meme-suite.org/

Identification of Orthologous and Paralogous Genes

Using Orthovenn2[8], the orthologous and paralogous genes in *A. thaliana, O. sativa subsp. japonica, Z. mays, S. lycopersicum,* and *G. max* were identified and compared to obtain the phylogenetic trees to identify the PEBP homologous genes of *D. huoshanens*e. Through Orthovenn2 analysis, the common and unique clusters information of the five species is obtained. Using six DhPEBPs family genes as templates, the clusters associated with them are retrieved.

The Pressure Analysis of Evolutionary Selection

The Ka and Ks values of the obtained orthologous gene pairs were used to calculate the ka/ks ratio of all the homologous gene pairs. Meanwhile, DnaSP (v.5.10.01) was used to recalculate each gene pair to remove the gene pair with high disproportionation value, and, finally, we obtained the homologous gene pairs with normal ka/ks ratio (Ge et al., 2017).

The Collinearity Analysis of PEBP Genes

Based on the genome sequence and gene annotation files, we used the one-step MCScanX plug-in in TBtools to obtain the collinearity files between every two species, including collinearity, gene linkage, and basic gene replication.

Amino Acid Alignment and the Conserved Domain Analysis

MEGA 6 was used to compare the PEBP protein sequences of *A. thaliana, O. sativa subsp. japonica, Z. mays, S. lycopersicum,* and *P. equestris*. According to the level of amino acid homology, Genedoc[9] was used for amino acid coloring.

The *cis*-Acting Element Analysis of *DhPEBPs*

To find out the responsive elements in the promoter, TBtools was used to extract the upstream 2,000 bp of gene sequence from the genome sequence of *D. huoshanense* (Carvalho et al., 2015). First, the GFF/GTF sequence extractor tool in the TBtools software was used to obtain the promoter regions of all genes. Then, the quick fasta extractor or filter tool was used to extract the promoter sequence of all PEBP family genes. PlantCARE[10] was used to analyze the *cis*-regulatory elements in the sequences. The BioSequence Viewer tool was used to visualize the promoter elements of *DhPEBPs*.

The qRT-PCR Analysis of *DhPEBPs*

RNAprep Pure Plant Kit (Takara, Japan) was used to extract total RNA from Dendrobium leaves of CK group and GA-treated groups for 5 and 10 days. An ultra-micro spectrophotometer was used to detect RNA concentration and quality. The 7,500 series real-time fluorescent quantitative PCR (Bio-RAD, America) was

[8]https://orthovenn2.bioinfotoolkits.net/

[9]https://www.psc.edu/biomed/genedoc

[10]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

used for the quantitative analysis. The CDS sequences of the six genes *DhFT3, DhFT1, DhMFT, DhTFL1b, DhFT2*, and *DhTFL1a* were obtained from the genome of *D. huoshanense*, and the primers for the fluorescent qPCR were designed respectively (**Supplementary Table 1**). The total RNA of the above samples was reverse transcribed into cDNA with PrimeScript RT reagent Kit (Takara, Japan). A total of 20 μl reaction system was used for qPCR: 10 μl SYBR Premix Ex Taq II (2×), 2 μl cDNA, 0.8 μl DhFT-RT-F, and DhFT-RT-R. The PCR reaction program was as follows: 50°C 2 min, 95°C 30 s, 95°C 5 s, 60°C 34 s, 40 cycles; and 72°C 10 min. With β-actin as the internal reference gene, the $2^{-\Delta\Delta Ct}$ method was used to calculate the relative gene expression, and the experiment was repeated three times.

RESULTS AND DISCUSSION

Identification of PEBP Family in *D. huoshanense*

Many studies had confirmed the PEBP genes were involved in flowering regulation (Ahluwalia and Hatsukami, 2015). We used the established hidden Markov model of PEBP protein (PF01161) to conduct preliminary screening of PEBP genes in the whole genome of *D. huoshanense* (Han et al., 2020). A total of six PEBP family genes were identified, and we further verified the specific PBP-containing domain through pfam, SMART, and the InterPro database, which confirmed these six genes were PEBP genes.

Gene Duplication of *D. huoshanense* and the Chromosomal Location of the PEBP Genes

Large-scale gene duplication and recombination are one of the important driving forces of species evolution (Jiang et al., 2020). However, few reports have focused on the gene duplication event in Dendrobium plants. In this study, we analyzed the gene duplication events in the evolution of *D. huoshanense*. The dot plot shows that *D. huoshanense* has experienced at least two large-scale WGD events since ancient times. There are a large number of WGD events that occurred in chromosome 1, chromosome 4, chromosome 16, chromosome 17, and chromosome 19. A large number of WGT events in chromosome 5, chromosome 6, chromosome 7, chromosome 14, and chromosome 18 were observed. In addition, the ancient γ-WGD events also appeared

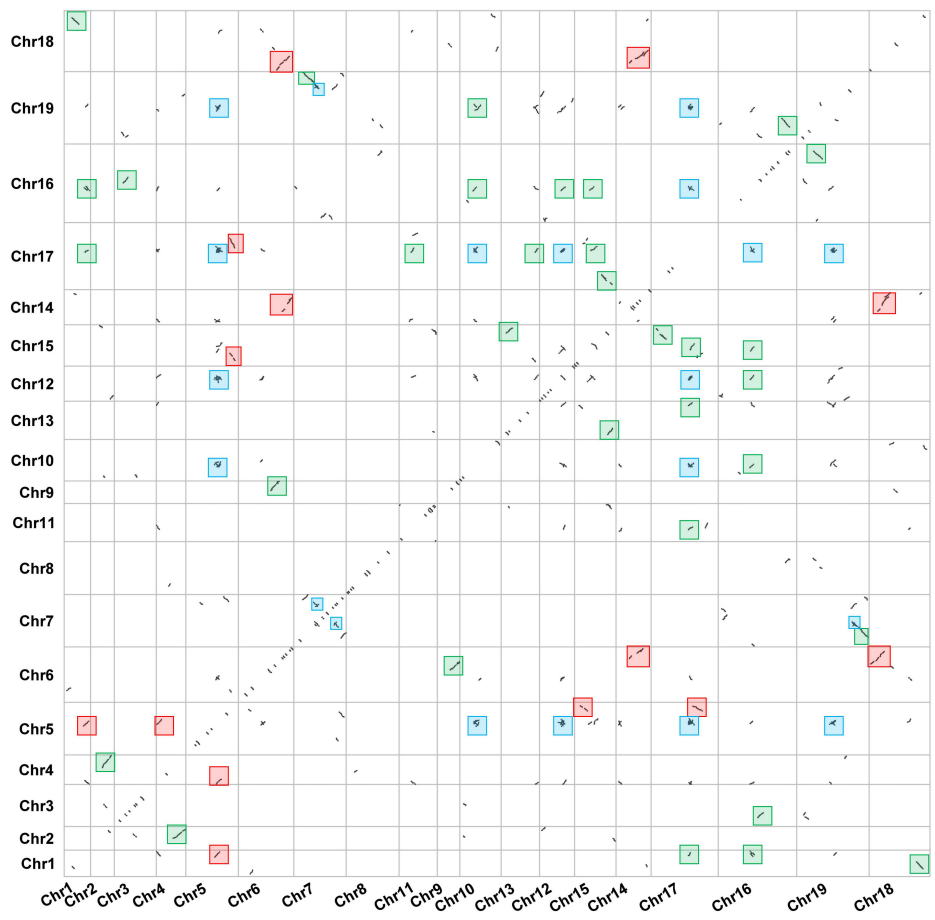


FIGURE 1 | The genome dotplot of *D. huoshanense*. The dot lines denote the duplication events that occur between two chromosomes. The blue boxes represented the ancient WGD events. The red boxes represented the large-scale WGT events. The green boxes represented the recent WGD events.

in great numbers, including chromosome 5, chromosome 10, chromosome 12, chromosome 19, and chromosome 17. The tandem duplication were scattered on the diagonal (**Figure 1**). These results indicated that *D. huoshanense* has undergone such ancient polyploidization events to adapt to changes in the external environment, which are consistent with previous results (Ospina-Zapata et al., 2020). The phylogeny tree indicated that two ancient PEBP duplication events in the lineage leading to the common ancestor of angiosperms after its split with gymnosperms. The first duplication gave rise to the MFT-like subfamily and the ancient lineage of the TFL1-like and FT-like subfamilies, which experienced a second duplication to create the two subfamilies. The *TFL1* ancestor underwent two separate duplication events in the common ancestor of angiosperms, which created three daughter lineages corresponding to *BFT*, *TFL1*, and *ATC* in *A. thaliana* (Wang et al., 2015). The chromosome mapping showed that the PEBP family genes were distributed on chromosomes 6, 9, 17, and 15, among which *DhFT1*, *DhFT2*, and *DhFT3* were located on chromosomes 9, 15, and 17, respectively. The two tandem duplicated genes of the DhTFL1 family are located on chromosome 6 (**Figure 2**).

Gene Structure Analysis and Conserved Motifs of *DhPEBPs*

The gene structure of *DhPEBP* showed that, except for *DhMFT*, the other five genes contained four exons and three introns (**Figure 3A**). The conserved motifs in the six DhPEBP proteins were identified using the MEME suite (**Figure 3B**). In total, 20 motifs were identified in DhPEBP proteins, named motifs 1 to 20, and the motifs identified were 6 to 121 amino acids in length. Motif 1, motif 2, and motif 3 are the three main conserved domains in the PEBP gene (**Figure 3C**). The number of mainly conserved motifs in each PEBP varied between three and four, indicating that the same subgroup of PEBP protein members shared one or more identical motifs. All of the PEBP genes contained motif 2. Except for the *DhMFT*, the other five PEBP genes had similar motif composition, which suggested their similar functions. However, some motifs were only presented in *DhMFT*, indicating that they may perform its particular functions (Zhao et al., 2020).

The Phylogenetic Analysis of PEBP Genes

In this experiment, MEGA 6 was used to perform homology alignment and phylogeny analysis of 81 PEBP-like protein sequences from six species. The evolutionary tree constructed by the NJ method showed that these genes were clearly divided into three clades. According to the classification of the PEBP family in *A. thaliana*, we divided these genes into FT subclade, TFL1 subclade and MFT subclade (**Supplementary Figure 1**). *D. huoshanense* PEBP protein maintained high homology with genes in the same clade of other species. To improve the reliability and accuracy of the established phylogenetic tree, we used the IQ-TREE tools to build an ML tree based on the optimal model (Nguyen et al., 2015). The optimal model is determined by calculation to be JTT + G4. Using this mode, further phylogeny analysis showed that *DhTFL1* is the closest to *OsTFL1*, *DhMFT* is the closest to *PeMFT*, and *DhFT1*, *DhFT2*, and *DhFT3* are the closest to *PeFT6*, *PeFT9*, and *PeFT4*, respectively. These results are also consistent with the adjacent evolutionary relationship between Dendrobium and Phalaenopsis genus (**Figure 4**).

Identification of Homologous Genes With Selection Pressure Analysis

To further clarify the homologous duplication of *DhPEBPs* in the evolutionary process, the plausible functions of homologous genes were speculated. The Orthovenn 2 was used to analyze the homology of five species of *A. thaliana, D. huoshanense, Z. mays, S. lycopersicum,* and *G. max*. The results showed that these species formed 23,615 clusters, including 23,049 orthologous clusters (at least contains two species) and 566 single-copy gene clusters. Among them, the genes from *D. huoshanense* were distributed

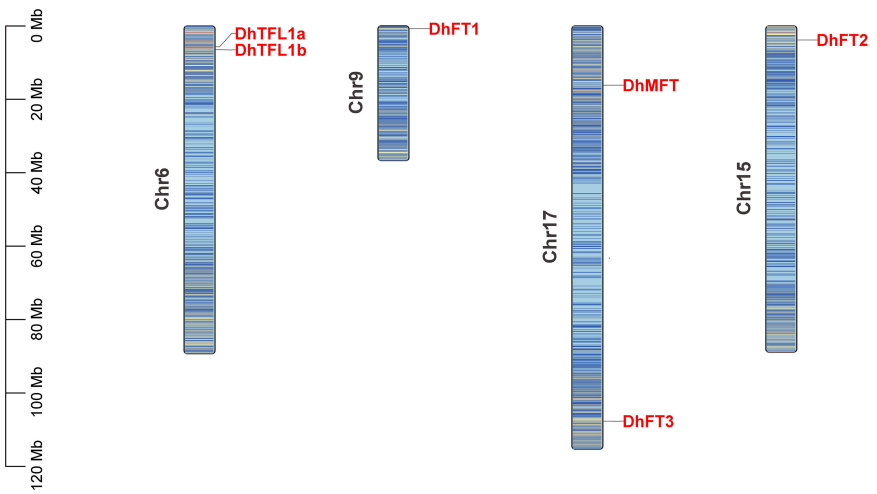


FIGURE 2 | The chromosome location of PEBP genes of *D. huoshanense*.

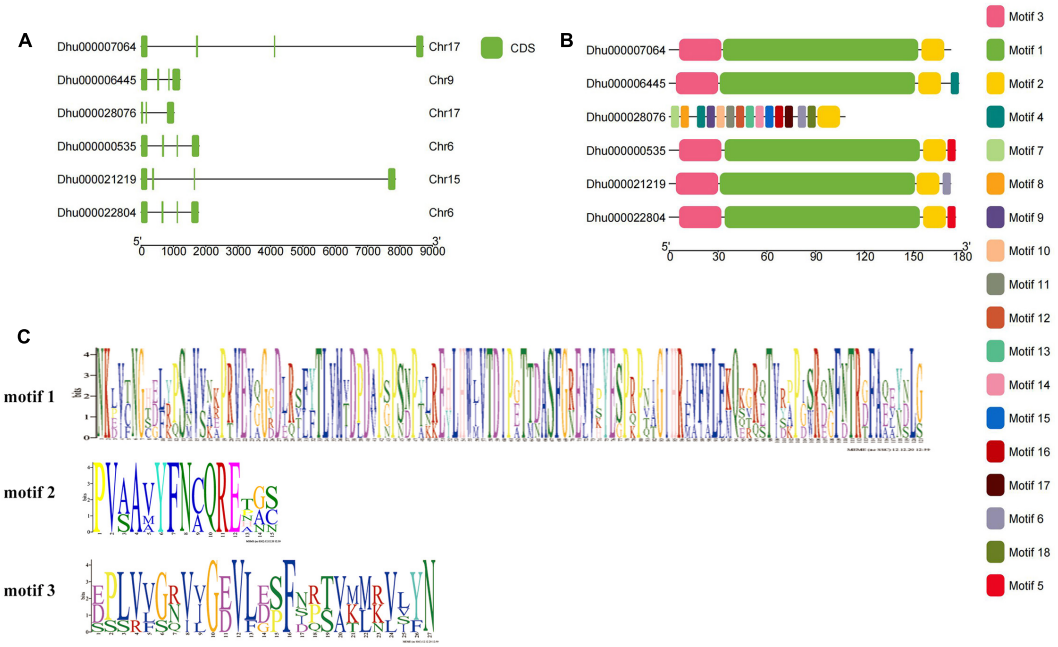


FIGURE 3 | The gene structure and the conserved motif of PEBP genes in *D. huoshanense.* **(A)** The gene structure of exon and intron. **(B)** The conserved motif distributed in PEBP genes. **(C)** The composition of three main motifs.

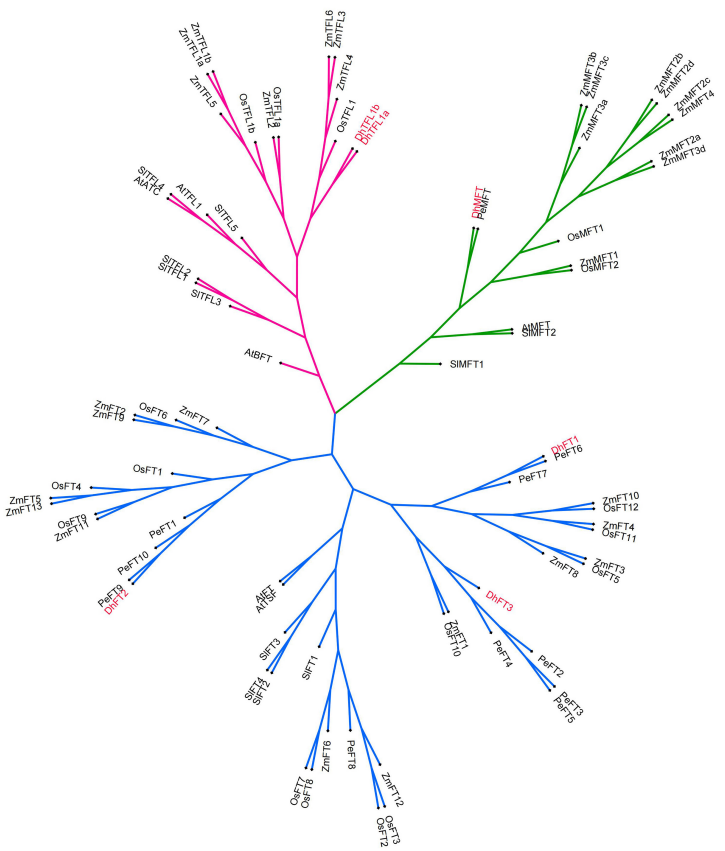


FIGURE 4 | The phylogenic tree of five species by ML method.

TABLE 1 | The orthologous and paralogous genes with Ka/Ks ratio among five species.

Seq_1	Seq_2	Ka	Ks	Ka/Ks
Dendrobium_huoshanense| Dhu000000535	Zea_mays| Zm00001d050649	0.1259	2.8550	0.0441
Dendrobium_huoshanense| Dhu000000535	Zea_mays| Zm00001d052537	0.1302	3.1977	0.0407
Dendrobium_huoshanense| Dhu000000535	Glycine_max| GLYMA_03G194700	0.1653	2.1969	0.0753
Dendrobium_huoshanense| Dhu000000535	Glycine_max| GLYMA_10G071400	0.1927	6.3124	0.0305
Dendrobium_huoshanense| Dhu000000535	Glycine_max| GLYMA_11G209500	0.2028	2.7745	0.0731
Dendrobium_huoshanense| Dhu000000535	Arabidopsis_thaliana| AT5G03840	0.2000	3.0282	0.0661
Dendrobium_huoshanense| Dhu000000535	Glycine_max| GLYMA_19G194300	0.1676	2.2117	0.0758
Dendrobium_huoshanense| Dhu000022804	Zea_mays| Zm00001d052537	0.1303	3.1638	0.0412
Dendrobium_huoshanense| Dhu000022804	Zea_mays| Zm00001d050649	0.1260	2.5464	0.0495
Dendrobium_huoshanense| Dhu000022804	Glycine_max| GLYMA_11G209500	0.2028	2.2936	0.0884
Dendrobium_huoshanense| Dhu000022804	Glycine_max| GLYMA_10G071400	0.1927	2.8590	0.0674
Dendrobium_huoshanense| Dhu000022804	Glycine_max| GLYMA_03G194700	0.1654	1.9347	0.0855
Dendrobium_huoshanense| Dhu000022804	Arabidopsis_thaliana| AT5G03840	0.2001	2.6514	0.0755
Dendrobium_huoshanense| Dhu000022804	Glycine_max| GLYMA_19G194300	0.1677	1.9457	0.0862
Dendrobium_huoshanense| Dhu000007064	Zea_mays| Zm00001d006116	0.1734	1.2652	0.1371
Dendrobium_huoshanense| Dhu000006445	Zea_mays| Zm00001d017134	0.1505	1.8798	0.0800
Dendrobium_huoshanense| Dhu000006445	Glycine_max| GLYMA_19G108100	0.2645	3.7647	0.0703
Dendrobium_huoshanense| Dhu000006445	Arabidopsis_thaliana| AT4G20370	0.2314	3.5402	0.0654
Dendrobium_huoshanense| Dhu000006445	Arabidopsis_thaliana| AT1G65480	0.2392	3.0163	0.0793
Dendrobium_huoshanense| Dhu000006445	Solanum_lycopersicum| Solyc03g063100.1	0.2407	2.8471	0.0845
Dendrobium_huoshanense| Dhu000000535	Dendrobium_huoshanense| Dhu000022804	0	0.0236	0

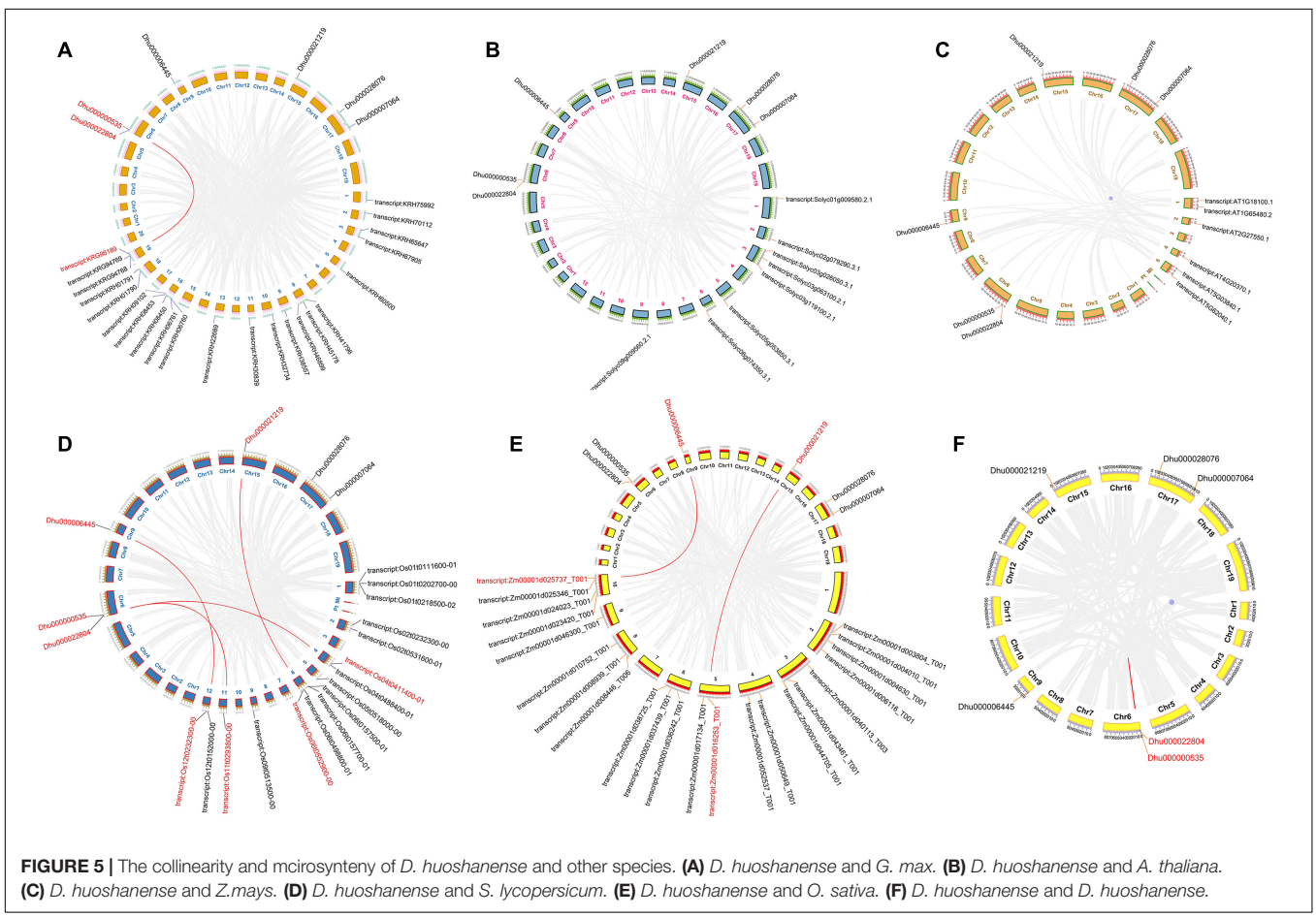


FIGURE 5 | The collinearity and mcirosynteny of *D. huoshanense* and other species. **(A)** *D. huoshanense* and *G. max*. **(B)** *D. huoshanense* and *A. thaliana*. **(C)** *D. huoshanense* and *Z.mays*. **(D)** *D. huoshanense* and *S. lycopersicum*. **(E)** *D. huoshanense* and *O. sativa*. **(F)** *D. huoshanense* and *D. huoshanense*.

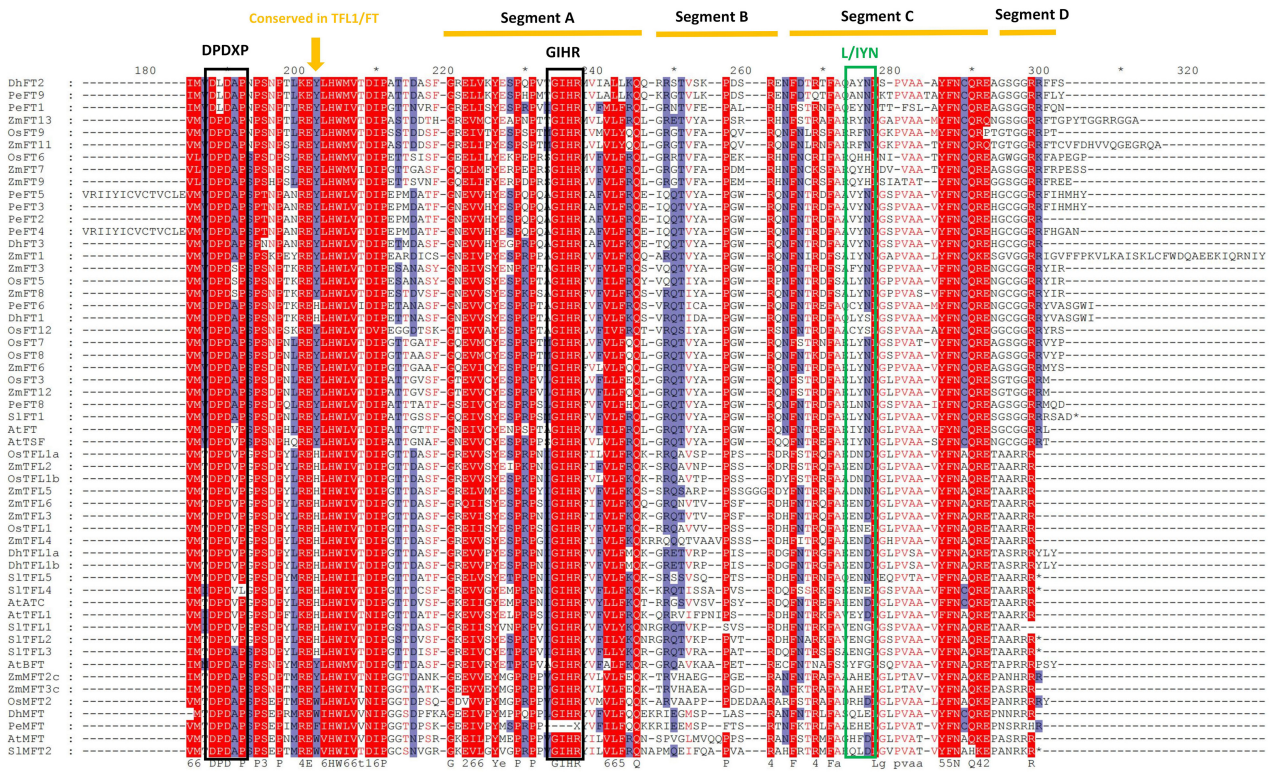


FIGURE 6 | The amino acid comparison and conserved domains of PEBP proteins. The brown arrow indicates a key amino acid residue that determines FT-like and TFL1-like functions. Black boxes denote the conserved DPDxP, GIHR motif, and L/IYN motif. Underlines represent segments A–D.

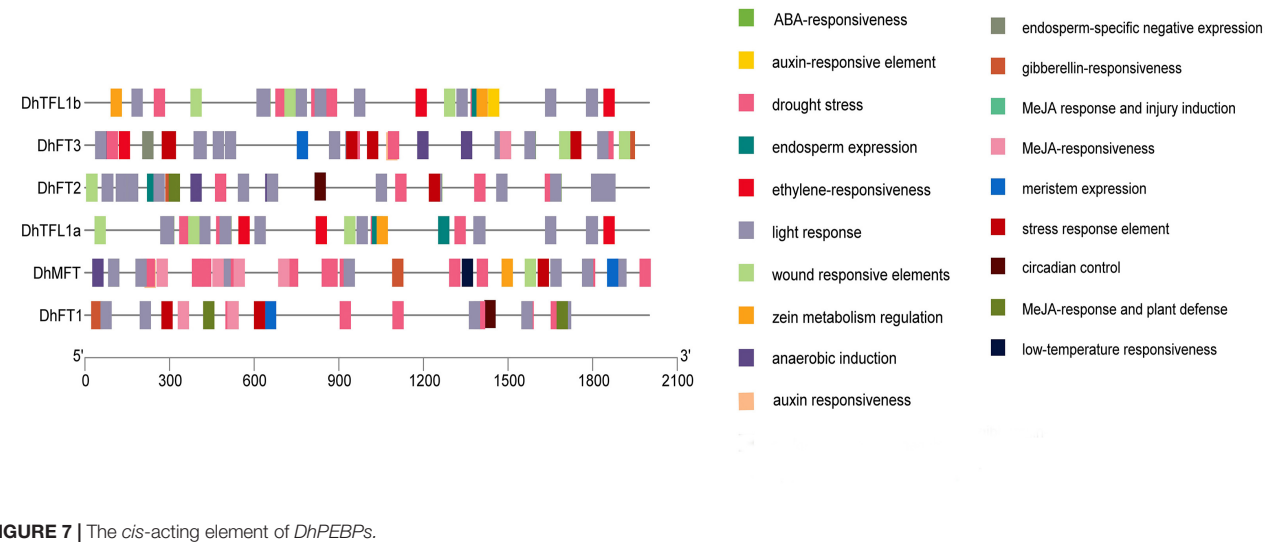


FIGURE 7 | The *cis*-acting element of *DhPEBPs*.

in 8,714 clusters (**Supplementary Figure 2**). We further analyzed the clusters related to *DhPEBP* and found that *DhPEBPs* scattered in cluster 734, cluster 2,452, and cluster 19,662, indicating these genes are orthologous genes (**Supplementary Figure 3**). In order to further reveal the evolutionary selection pressure of these homologous genes, we used DnaSP 5.0 to calculate the ka and ks values of each gene pair and found that the ka/ks ratio of all gene pairs was far less than 1, indicating that they were all subjected to a strongly purified selection (**Table 1**).

Colinearity and Microsynteny Analysis

Collinearity analysis is often used to reveal the genetic relationship between the same species or several different species. In this experiment, we performed microsynteny analysis on the

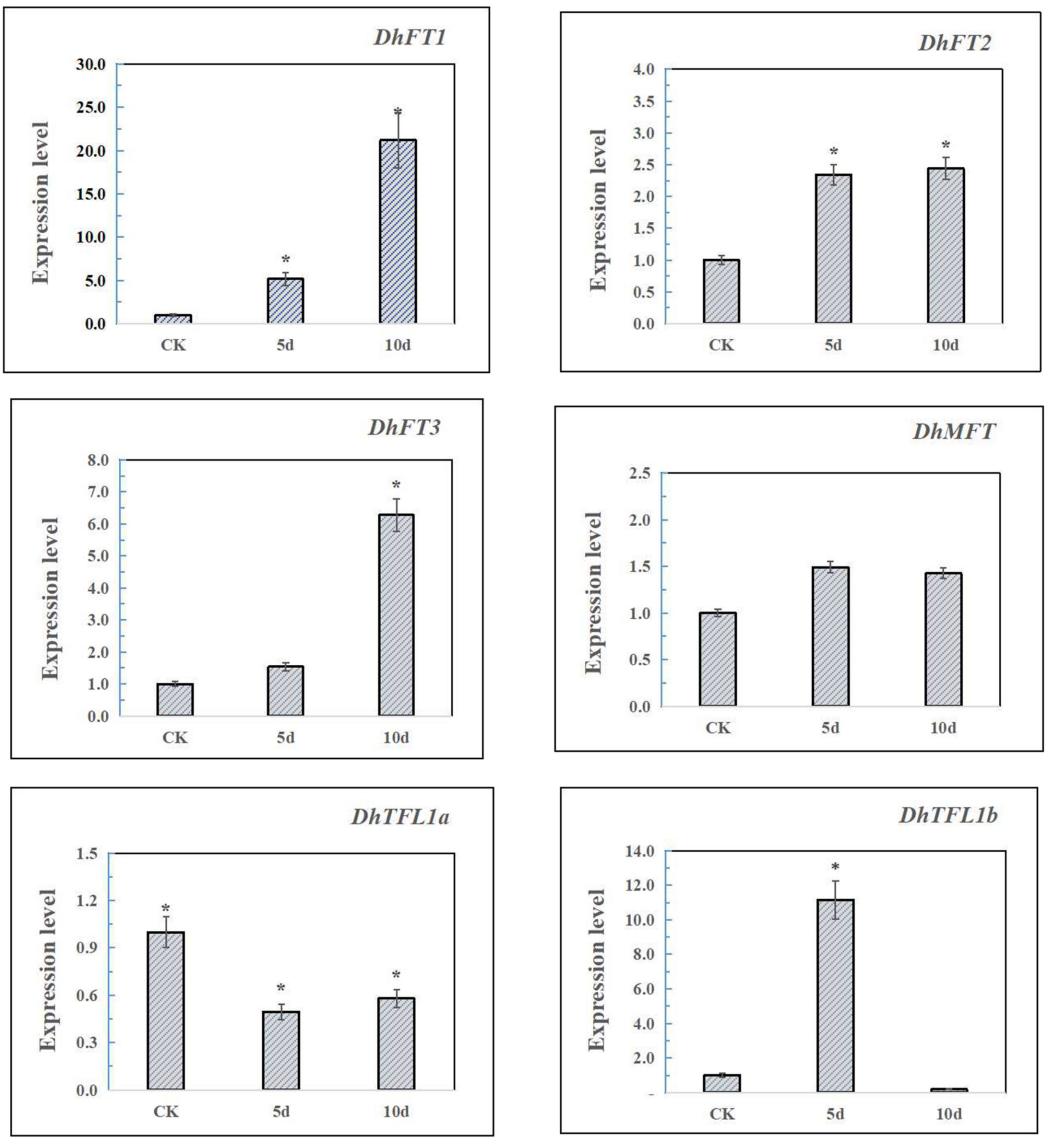


FIGURE 8 | The qRT-PCR analysis of *DhPEBPs*. Asterisks mean that the expression levels have significant differences compared with CK.

genomes of *A. thaliana, D. huoshanense, Z. mays, S. lycopersicum,* and *G. max*. It was indicated that *DhTFL1/TFL2* and *GmTFL1* (KRG96189) have collinearity, *DhFT2* and *OsFT8*, *DhFT1,* and *OsFT2*, and *DhTFL1/TFL2* and *OsTFL1b* have collinearity. *DhFT1* and *ZmFT3*, *DhFT2,* and *ZmFT13* have collinearity (**Figure 5**). In addition, a pair of tandem duplicated genes *DhTFL1* and *DhTFL2* are paralogs. These results indicated there was a homologous evolutionary relationship on PEBP genes between *D. huoshanense* and other species.

Amino Acid Alignment and the Conserved Domain Analysis

Within amino acid comparison, the conserved structural domains of some proteins could be obtained, accompanied by the conserved catalytic sites and DNA-binding sites for the subsequent research (Li et al., 2014). In this experiment, we performed the amino acid alignment on the aligned sequences of 81 proteins from six species. It was indicated that the 86th position of DhFTs was tyrosine (Y), while 83th and 87th positions of of *DhTFL1s* were both histidine (H), suggesting they should have distinct functions in flowering regulation (**Figure 6**). PEBP proteins are characterized by the presence of two highly conserved short motifs, DPDxP and GxHR, which presumably contribute to the conformation of the ligand binding pocket (Guo et al., 2015; Mackenzie et al., 2019). It was reported that substitution of the single amino acid, Tyr85 to His, in FT partially converts FT function to TFL1 function probably through discrimination of structurally related interactors (Yang et al., 2019). In addition, the amino acid sequence encoded by the

fourth exon plays a critical role to determine FT/TFL1 protein functions, which are divided into four segments (A–D). Segment B and segment C containing the LYN/IYN triplet conserved motif are especially important for the determination of functional specifcity between FT and TFL1.

Analysis of *cis*-Acting Elements

The PLANTCARE service was performed to analyze the upstream promoter sequences of six *DhPEBP* genes to discover hormone-responsive elements related to flowering regulation, such as GARE elements (**Figure 7**). The results showed that the promoter region of *DhFT1, DhFT2,* and *DhMFT* all contained a GARE-motif (TCTGTTG), while that of *DhFT1* and *DhFT3* both contained a P-box motif (CCTTTTG), implying that they may be involved in flowering regulation under the GA signaling pathway (**Supplementary Table 1**).

The Expression Profile of *DhPEBPs* Induced by GA

In the experiment, qRT-PCR was performed on Dendrobium leaves treated with gibberellin for 5 and 10 days and the CK group (**Figure 8**). The results showed that *DhFT1* and *DhFT3* were strongly induced by GA treatment at 10 days, and the expression levels were increased by 21.2 times and 6.2 times, respectively. The expression level of *DhFT2* remained stable after 5 days of treatment. The expression of *DhMFT* was relatively stable, with a slight increase at 5 days. Due to the inhibition of negative feedback regulation, the expression of *DhTFL1* decreased rapidly after GA treatment, especially *DhTFL1b* was hardly expressed at 10 days. The results indicated that *DhFT* and *DhTFL1* may have different regulatory roles in the flowering regulation of *D. huoshanense*.

CONCLUSION

Six PEBP genes were isolated from the *D. huoshanense* genome. The chromosome mapping suggested that the genes were distributed on chromosomes 6, 9, 15, and 17. The paralogs *DhTFL1b* and *DhTFL1a* that obtained by tandem duplication exhibited similar function. Gene structure and the conserved motif analysis further indicated 5 PEBP genes apart from *DhMFT* contained four exons and three introns entirely. The phylogeny showed that PEBP genes in *A. thaliana*, *Z. mays*, *S. lycopersicum*, *O. sativa*, and *D. huoshanense* can be classified into three subclades, FT, TFL, and MFT, which maintained a high homology with the same family in other species. The conserved domain of the amino acid demonstrated that two highly conserved short motifs DPDXP and GXHR embed in DhPEBPs. The 86th position of DhFTs was tyrosine (Y), while the 83th and 87th positions of DhTFL1s were both histidine (H), suggesting they had distinct functions in flowering regulation. The promoter analysis of six DhPEBPs revealed that several cis-elements related to hormone induction, light response and abiotic stress may be involved in the regulation of environmental stress and endogenous signaling pathways. The RT-PCR analysis in short-term days treated with GA3 indicated that the gene expressions of all *DhFTs* were gradually increased, whereas *DhTFL1* were decreased. Taken together, *DhPEBPs* have various regulatory functions in modulating flowering.

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

CS and HD designed the research. CS, GL, and JD conducted the experiments. CS and GL analyzed the data. CS wrote the manuscript. HD and CS revised the manuscript, improved the English, and acquired the funding. All authors have read, reviewed, and approved the submitted version.

FUNDING

This work was supported by the Natural Science Foundation of Anhui Province (1808085MH307), High-Level Talents Research Initiation Funding Project (WGKQ202001011), and Quality Engineering Project in Anhui Province (2020jyxm2137).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.687689/full#supplementary-material

Supplementary Figure 1 | The phylogenic tree of five species by NJ method.

Supplementary Figure 2 | The orthologous clusters of five species and the main clusters involved in DhPEBPs.

Supplementary Figure 3 | Orthologous genes related to DhPEBPs in several clusters.

REFERENCES

Adeyemo, O. S., Hyde, P. T., and Setter, T. L. (2019). Identification of FT family genes that respond to photoperiod, temperature and genotype in relation to flowering in cassava (Manihot esculenta, Crantz). *Plant Reprod.* 32, 181–191. doi: 10.1007/s00497-018-00354-5

Ahluwalia, J. S., and Hatsukami, D. (2015). FLOWERING LOCUS T4 (HvFT4) delays flowering and decreases floret fertility in barley. 4, 1–19. **as well as**J query.

Auge, G. A., Blair, L. K., Karediya, A., and Donohue, K. (2018). The autonomous flowering-time pathway pleiotropically regulates seed germination in Arabidopsis thaliana. *Ann. Bot.* 121, 183–191. doi: 10.1093/aob/mcx132

Bi, Z., Huang, H., and Hua, Y. (2019). Cloning and Characterization of Two FLOWERING LOCUS T-like Genes from Rubber Tree (Hevea brasiliensis). *J. Plant Growth Regul.* 38, 919–930. doi: 10.1007/s00344-018-9902-z

Carvalho, S. G., Guerra-Sá, R., de, C., and Merschmann, L. H. (2015). The impact of sequence length and number of sequences on promoter prediction performance. *BMC Bioinformatics* 16:S5. doi: 10.1186/1471-2105-16-S19-S5

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant*. 8, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, M., and Penfield, S. (2018). Feedback regulation of COOLAIR expression controls seed dormancy and flowering time. *Science* 12, 1014–1017. doi: 10.1126/science.aar7361

Ge, W., Zhang, Y., Cheng, Z., Hou, D., Li, X., and Gao, J. (2017). Main regulatory pathways, key genes and microRNAs involved in flower formation and development of moso bamboo (Phyllostachys edulis). *Plant Biotechnol. J.* 15, 82–96. doi: 10.1111/pbi.12593

González-Suárez, P., Walker, C. H., and Bennett, T. (2020). Bloom and bust: understanding the nature and regulation of the end of flowering. *Curr. Opin. Plant Biol.* 57, 24–30. doi: 10.1016/j.pbi.2020.05.009

Guo, D., Li, C., Dong, R., Li, X., Xiao, X., and Huang, X. (2015). Molecular cloning and functional analysis of the FLOWERING LOCUS T (FT) homolog GhFT1 from Gossypium hirsutum. *J. Integr. Plant Biol.* 57, 522–533. doi: 10.1111/jipb.12316

Han, B., Jing, Y., Dai, J., Zheng, T., Gu, F., Zhao, Q., et al. (2020). A Chromosome-Level Genome Assembly of Dendrobium Huoshanense Using Long Reads and Hi-C Data. *Genome Biol. Evol.* 12, 2486–2490. doi: 10.1093/gbe/evaa215

Jiang, Z., Sun, L., Wei, Q., Ju, Y., Zou, X., Wan, X., et al. (2020). A new insight into flowering regulation: Molecular basis of flowering initiation in magnolia × soulangeana 'changchun.' *Genes (Basel).* 11, 1–18. doi: 10.3390/genes11010015

Jin, H., Tang, X., Xing, M., Zhu, H., Sui, J., Cai, C., et al. (2019). Molecular and transcriptional characterization of phosphatidyl ethanolamine-binding proteins in wild peanuts Arachis duranensis and Arachis ipaensis. *BMC Plant Biol.* 19:3. doi: 10.1186/s12870-019-2113-3

Jin, S., Nasim, Z., Susila, H., and Ahn, J. H. (2021). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin. Cell Dev. Biol.* 109, 20–30. doi: 10.1016/j.semcdb.2020.05.007

Johnson, M. A., Harper, J. F., and Palanivelu, R. (2019). A Fruitful Journey: Pollen Tube Navigation from Germination to Fertilization. *Annu. Rev. Plant Biol.* 70, 809–837. doi: 10.1146/annurev-arplant-050718-100133

Li, C., Luo, L., Fu, Q., Niu, L., and Xu, Z. F. (2014). Isolation and functional characterization of JcFT, a FLOWERING LOCUS T (FT) homologous gene from the biofuel plant Jatropha curcas. *BMC Plant Biol.* 14:125. doi: 10.1186/1471-2229-14-125

Li, C., Luo, L., Fu, Q., Niu, L., and Xu, Z. F. (2015). Identification and Characterization of the FT/TFL1 Gene Family in the Biofuel Plant Jatropha curcas. *Plant Mol. Biol. Report.* 33, 326–333. doi: 10.1007/s11105-014-0747-8

Liu, H., Song, S., and Xing, Y. (2019). Beyond heading time: FT-like genes and spike development in cereals. *J. Exp. Bot.* 70, 1–3. doi: 10.1093/jxb/ery408

Luo, Y., Li, H., Xiang, Z., and He, N. (2018). Identification of Morus notabilis MADS-box genes and elucidation of the roles of MnMADS33 during endodormancy. *Sci. Rep.* 8, 1–16. doi: 10.1038/s41598-018-23985-0

Mackenzie, K. K., Coelho, L. L., Lütken, H., and Müller, R. (2019). Phylogenomic analysis of the PEBP gene family from Kalanchoë. *Agronomy* 9, 1–16. doi: 10.3390/agronomy9040171

Melzer, R. (2017). Regulation of flowering time: A splicy business. *J. Exp. Bot.* 68, 5017–5020. doi: 10.1093/jxb/erx353

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

Ospina-Zapata, D. A., Madrigal, Y., Alzate, J. F., and Pabón-Mora, N. (2020). Evolution and Expression of Reproductive Transition Regulatory Genes FT/TFL1 With Emphasis in Selected Neotropical Orchids. *Front. Plant Sci.* 11:469. doi: 10.3389/fpls.2020.00469

Plackett, A. R. G., Powers, S. J., Phillips, A. L., Wilson, Z. A., Hedden, P., and Thomas, S. G. (2018). The early inflorescence of Arabidopsis thaliana demonstrates positional effects in floral organ growth and meristem patterning. *Plant Reprod.* 31, 171–191. doi: 10.1007/s00497-017-0320-3

Schiessl, S. V., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. J. (2017). Flowering time gene variation in Brassica species shows evolutionary principles. *Front. Plant Sci.* 8:01742. doi: 10.3389/fpls.2017.01742

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 12, 2725–2729. doi: 10.1093/molbev/mst197

Wang, M., Tan, Y., Cai, C., and Zhang, B. (2019). Identification and expression analysis of phosphatidy ethanolamine-binding protein (PEBP) gene family in cotton. *Genomics* 111, 1373–1380. doi: 10.1016/j.ygeno.2018.09.009

Wang, Z., Yang, R., Devisetty, U. K., Maloof, J. N., Zuo, Y., Li, J., et al. (2017). The divergence of flowering time modulated by FT/TFL1 is independent to their interaction and binding activities. *Front. Plant Sci.* 8:697. doi: 10.3389/fpls.2017.00697

Wang, Z., Zhou, Z., Liu, Y., Liua, T., Li, Q., Ji, Y., et al. (2015). Functional evolution of phosphatidylethanolamine binding proteins in soybean and arabidopsis. *Plant Cell* 27, 323–336. doi: 10.1105/tpc.114.135103

Xiao, G., Li, B., Chen, H., Chen, W., Wang, Z., Mao, B., et al. (2018). Overexpression of PvCO1, a bamboo CONSTANS-LIKE gene, delays flowering by reducing expression of the FT gene in transgenic Arabidopsis. *BMC Plant Biol.* 18:1469. doi: 10.1186/s12870-018-1469-0

Yang, Z., Chen, L., Kohnen, M. V., Xiong, B., Zhen, X., Liao, J., et al. (2019). Identification and Characterization of the PEBP Family Genes in Moso Bamboo (Phyllostachys heterocycla). *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-51278-7

Yu, X., Liu, H., Sang, N., Li, Y., Zhang, T., Sun, J., et al. (2019). Identification of cotton mother of FT and TFL1 homologs, GhMFT1 and GhMFT2, involved in seed germination. *PLoS One* 14:e0215771. doi: 10.1371/journal.pone.0215771

Zhang, X., Wang, C., Pang, C., Wei, H., Wang, H., Song, M., et al. (2016). Characterization and functional analysis of PEBP Family genes in upland cotton (Gossypium hirsutum L.). *PLoS One* 11:e0161080. doi: 10.1371/journal.pone.0161080

Zhao, S., Wei, Y., Pang, H., Xu, J., Li, Y., Zhang, H., et al. (2020). Genome-wide identification of the PEBP genes in pears and the putative role of PbFT in flower bud differentiation. *PeerJ* 2020, 1–19. doi: 10.7717/peerj.8928

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 July 2021
doi: 10.3389/fgene.2021.678804

Investigation of Thermomorphogenesis-Related Genes for a Multi-Silique Trait in *Brassica napus* by Comparative Transcriptome Analysis

Liang Chai[1†], Jinfang Zhang[1†], Haojie Li[1], Cheng Cui[1], Jun Jiang[1], Benchuan Zheng[1], Lintao Wu[2] and Liangcai Jiang[1]*

[1] Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu, China, [2] School of Biological Sciences, Guizhou Education University, Guiyang, China

OPEN ACCESS

Edited by:
Lin Zhang,
Central South University of Forestry and Technology, China

Reviewed by:
Sachin Kumar,
Chaudhary Charan Singh University, India
Rakesh Srivastava,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

***Correspondence:**
Liangcai Jiang
jlcrape@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 10 March 2021
Accepted: 18 June 2021
Published: 23 July 2021

Citation:
Chai L, Zhang J, Li H, Cui C, Jiang J, Zheng B, Wu L and Jiang L (2021) Investigation of Thermomorphogenesis-Related Genes for a Multi-Silique Trait in Brassica napus by Comparative Transcriptome Analysis. Front. Genet. 12:678804. doi: 10.3389/fgene.2021.678804

In higher plants, the structure of a flower is precisely controlled by a series of genes. An aberrance flower results in abnormal fruit morphology. Previously, we reported multi-silique rapeseed (*Brassica napus*) line zws-ms. We identified two associated regions and investigated differentially expressed genes (DEGs); thus, some candidate genes underlying the multi-silique phenotype in warm area Xindu were selected. However, this phenotype was switched off by lower temperature, and the responsive genes, known as thermomorphogenesis-related genes, remained elusive. So, based on that, in this study, we further investigated the transcriptome data from buds of zws-ms and its near-isogenic line zws-217 grown in colder area Ma'erkang, where both lines showed normal siliques only, and the DEGs between them analyzed. We compared the 129 DEGs from Xindu to the 117 ones from Ma'erkang and found that 33 of them represented the same or similar expression trends, whereas the other 96 DEGs showed different expression trends, which were defined as environment-specific. Furthermore, we combined this with the gene annotations and ortholog information and then selected BnaA09g45320D (chaperonin gene *CPN10*-homologous) and BnaC08g41780D [Seryl-tRNA synthetase gene *OVULE ABORTION 7* (*OVA7*)-homologous] the possible thermomorphogenesis-related genes, which probably switched off the multi-silique under lower temperature. This study paves a way to a new perspective into flower/fruit development in *Brassica* plants.

Keywords: *Brassica napus*, differentially expressed gene, environmental effect, multi-silique, RNA-seq

INTRODUCTION

Flower development, as well as the subsequent fruit formation, is vital to the crop life cycle. Siliques are important to rapeseed (*Brassica napus*, AACC = 38), the leading source of plant oil worldwide, which offers more than 13% of the global vegetable oil (Geng et al., 2016). In rapeseed, siliques supply nutrients from photosynthesis, transport carbohydrates from the vegetative organs to the seeds, and ensure their development (Shen et al., 2019). In addition, silique number is a crucial component determining seed yield per plant (Yang et al., 2012; Wang et al., 2016; Shen et al., 2019).

We previously reported a multi-silique phenotype in rapeseed line zws-ms (Chai et al., 2019, 2020a,b), which presents three pistils instead of one typical pistil in flower, and then develops three siliques on a carpopodium. This trait is different from the multilocular phenotype in *Brassica* plants, such as tetra-locular *Brassica rapa* (Yadava et al., 2014; Lee et al., 2018), multilocular *Brassica juncea* (Xiao et al., 2013; Xu et al., 2014), etc., which increases the number of locules in a silique. However, similar phenomena described as "multi-pistil" have been reported in wheat (*Triticum aestivum*) (Yang et al., 2011, 2017; Duan et al., 2015; Wei, 2017; Guo et al., 2019; Zhu et al., 2019; Yu et al., 2020), rice (*Oryza sativa*) (Zheng et al., 2019), sweet cherry (*Prunus avium*) (Liu et al., 2019; Wang et al., 2019), and alfalfa (*Medicago sativa*) (Nair et al., 2008). In wheat, increasing the number of grains per spike is considered to be vital for maximizing its yield potential (Wei, 2017; Zhu et al., 2019; Yu et al., 2020). The multi-pistil traits in wheat were found to be controlled by a recessive gene, two recessive non-complementary genes, or a single dominant gene; F_2 populations, BC_6F_2 populations, or near-isogenic lines (NILs) were constructed to map the underlying locus, and they found to be located on 2DL, 5DS, 6BS, and 6B (Zhu et al., 2019). However, there has been no gene cloned so far.

The environment can influence various aspects of plants. Temperature, a key environmental factor, affects the growth, development, and geographical distribution of plants, as well as the quality and productivity of crops (Ding et al., 2020). Therein, "thermomorphogenesis" is defined as the effect of temperature on the morphogenesis of plants (Casal and Balasubramanian, 2019). Take soybean (*Glycine max*) for example; a 4°C rise in the temperature could increase its stem height more than 3-fold (Sionit et al., 1987); additionally, the temperature raised from 10 to 15°C could increase the leaf number in wheat (*T. aestivum*) (Friend, 1965), whereas, in model plant *Arabidopsis thaliana*, higher temperatures reduced both silique number per plant and seed number per silique (Ibañez et al., 2017).

Similarly, as described in earlier studies (Chai et al., 2019, 2020a), the multi-silique trait in zws-ms is significantly affected by the environment. Precisely, zws-ms showed stable multi-silique trait in Xindu (with an annual average temperature of 16.2°C) for successive years; however, when grown in a colder area such as Ma'erkang (with an annual average temperature of 8.6°C), zws-ms switched off the formation of multi-silique, and all siliques appeared normal. In previous studies, we identified two associated regions on chromosome A09 and C08 and screened out some potential candidate genes by the combination of bulked-segregant analysis and whole-genome re-sequencing; we also analyzed genes differentially expressed between NILs zws-ms (multi-silique) and zws-217 (normal silique) from Xindu and selected potential underlying genes based on annotations. However, the transcriptome from Ma'erkang, where the multi-silique morphology is switched off, as well as the comparison of differentially expressed genes (DEGs) between Xindu and Ma'erkang, remains unclear. In other words, genes involved in the thermomorphogenesis of this multi-silique trait require further investigation.

Thus, in this study, we identified DEGs based on transcriptome sequencing (RNA-seq) from Ma'erkang and then compared them with those from Xindu. The variations in DEGs, combined with their annotations and information of orthologs in *Arabidopsis*, were then analyzed, and then some potential underlying genes related to thermomorphogenesis were identified.

MATERIALS AND METHODS

Plant Materials, Growth Conditions, and Sample Collection

The rapeseed line zws-ms and its NIL, zws-217, were kept in the Crop Research Institute, Sichuan Academy of Agricultural Sciences, China. The multi-silique plant was originally discovered in progenies of *B. napus* × *B. rapa* and was successively self-pollinated for six successive generations until the homozygous multi-silique zws-ms line was obtained, whereas the single-silique offspring were continuously backcrossed with zws-ms (current parent) for six generations, followed by six continuous generations of self-pollination until zws-217 with the single-silique was obtained. The two lines were simultaneously grown in September in Xindu District of Chengdu in the Sichuan Basin under normal environmental conditions (with an annual average temperature of 16.2°C). Moreover, both lines were also grown concurrently in early June in Ma'erkang, a mountainous area of western Sichuan, with an annual average temperature of 8.6°C.

In each location, buds were sampled at their budding stage. In Xindu, three randomly selected individual plants of the zws-ms were assigned as T01, T02, and T03, and three plants in the zws-217 line were assigned as T04, T05, and T06, whereas in colder area Ma'erkang, plants in zws-ms were defined as T01′, T02′, and T03′, and plants in zws-217 were defined as T04′, T05′, and T06′. Eight to 10 buds were sampled from each plant and then quick-frozen and stored in liquid nitrogen.

RNA Isolation and the Library Preparation

The total RNA was isolated as Chai et al. (2019) described. The OD260/OD230 value and concentration were determined on NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, United States) for the quality control of RNA. The sequencing libraries were generated by RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, United States), according to its instructions.

Transcriptome Sequencing

The transcriptome sequencing (RNA-seq) was performed on the Illumina HiSeq X-ten platform. In-house Perl scripts were used to remove adapter sequences and read containing poly-*N*, or low-quality reads, to process the initially generated raw reads into clean reads. Clean reads were then aligned to the *B. napus* "Darmor-*bzh*" reference genome[1] by using Tophat2

[1] www.genoscope.cns.fr/brassicanapus/data/

tools (Kim et al., 2013) to screen out the reads with a perfect match or one mismatch for the next investigation.

Differentially Expressed Gene Analysis

DEGs were detected by the DESeq R package (Love et al., 2014). The *P*-value was adjusted by controlling the false discovery rate (FDR) (Benjamini and Hochberg, 1995), and genes with an adjusted fold change (FC) > 4 ($\log_2$FC > 2) and an FDR < 0.001 were then identified as DEGs.

Annotation of Genes

Gene Ontology (GO) database[2] and GOseq R packages (Young et al., 2010) were used to provide gene annotations and calculate GO enrichment of the DEGs. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database[3] and KOBAS software (Mao et al., 2005) were used to explore the high-level functions and utilities of the biological system (Kanehisa et al., 2007) and test the statistical enrichment of the DEGs in the various KEGG pathways. The TAIR database[4] was utilized to provide ortholog information from the model plant *Arabidopsis*. Sequences of genes from rapeseed were blasted on the website, and orthologs in *Arabidopsis* were then screened out. Orthologs in *Arabidopsis* were sufficiently annotated and able to give abundant information, as they have been well studied in the model plant and reported.

RESULTS

Multi-Silique Trait Under Normal Conditions and Its Absence in Colder Areas

Compared with its NIL zws-217 with normal siliques, the zws-ms line displayed stable multi-pistil phenotype in successive years in Xindu, where a zws-ms plant showed approximately 32–53% of flowers with the multi-pistil trait, and that further developed into a multi-silique trait, precisely, two or three siliques on a carpopodium (**Figure 1A**), like we previously described (Chai et al., 2019). When grown in Ma'erkang, a colder place in a mountain area of western Sichuan Province, zws-ms switched off this multi-silique phenotype; in other words, both lines produced normal siliques there (**Figure 1B**).

Transcriptome Sequencing (RNA-Seq) From Two Environments

As described previously (Chai et al., 2019, 2020a), the two lines were grown at Xindu and Ma'erkang. At each place, flower buds from three individual plants from each line were selected randomly for RNA isolation. Then, their total RNA was extracted, and sequencing libraries were generated, followed by being sequenced on a HiSeq X-Ten platform; the validation for the RNA-seq was also confirmed by real-time quantitative polymerase chain reaction (Chai et al., 2019, 2020a).

[2]http://geneontology.org/

[3]http://www.genome.jp/kegg/

[4]www.arabidopsis.org/

Comparison of Differentially Expressed Genes Between Two Environments

Genes in an environment with expression level fold change > 4 ($\log_2$FC > 2) between zws-ms and zws-217 and FDR < 0.001 were identified as DEGs. In our earlier report (Chai et al., 2019), 129 DEGs were found between zws-ms and zws-217 in Xindu, among which 67 genes were upregulated, whereas 62 were downregulated. In this study, both lines were grown in colder area Ma'erkang, where they did not show any phenotypic differences to each other and were further subjected to RNA-seq, and 117 genes were found expressed differentially in stamen and pistils between zws-ms and zws-217 (**Table 1**), including 63 upregulated and 54 downregulated in zws-ms (**Supplementary Table 1**). Samples from the two environments generated different DEGs between zws-ms and zws-217, but in either environment, chromosome A09 and C08, where two associated regions (Chai et al., 2019) underlying this trait were identified, provided most DEGs: in Xindu, 16 (12.4%) and 30 (23.26%) genes were located on chromosome A09 and C08, respectively, whereas 7 (5.98%) and 23 (19.66%) genes were found on chromosome A09 and C08, respectively, of samples from Ma'erkang.

Further analysis found that among these DEGs, there were some expressed only in zws-ms or zws-217, which were assigned "line-specific expressed genes" in this study. Herein, we discovered 21 genes that were line-specific expressed in zws-ms and 25 in zws-217 from Xindu (**Supplementary Table 2**), whereas from Ma'erkang, 18 and one line-specific expressed genes were detected in zws-ms and zws-217, respectively (**Supplementary Table 3**).

Moreover, the comparison of DEGs between two environments identified 33 DEGs that represented the same or similar expression trends (**Supplementary Table 4**). In other words, these genes were up- (or down-) regulated in both environments, and they were assigned as "stable DEGs," whereas the other genes showed different expression trends, which meant zws-ms and zws-217 displayed differential expression level of those genes when grown in Xindu, but no obvious difference while grown in Ma'erkang (**Supplementary Table 5**). These genes were defined as "environment-specific DEGs" in this study.

Annotations for the Differentially Expressed Genes

As mentioned earlier, we divided these DEGs into two classifications (stable and environment-specific DEGs) and then analyzed their GO annotations. GO terms were usually divided into three categories: biological processes, cellular components (CCs), and molecular functions (MFs). In classification (**Figure 2A**) for 33 stable DEGs, the biological process terms with the highest levels of enrichment included "metabolic process (GO:0008152)" and "cellular process (GO:0009987)," involving 17 and 16 DEGs; in the CC category, most enriched terms were "cell (GO: 0005623)" and "cell part (GO: 0044464)," whereas in the MF category, "catalytic activity (GO: 0003824)" and "binding (GO: 0005488)" accounted the top enriched terms, containing 13 and 10 DEGs, respectively. In the second classification (**Figure 2B**) for 96 environment-specific DEGs,

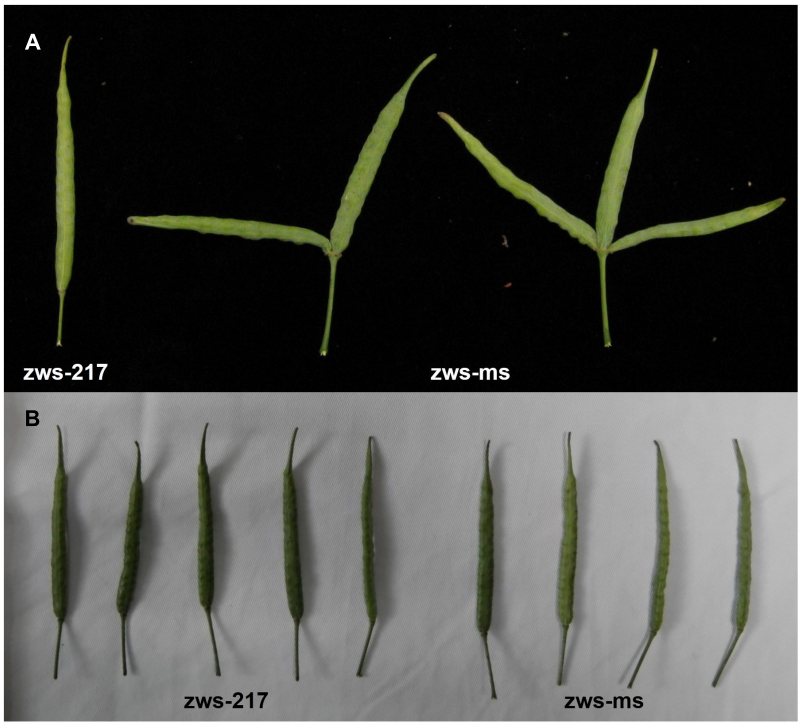


FIGURE 1 | Multi-silique phenotype in zws-ms. **(A)** Compared with normal silique in its near-isogenic line (NIL) zws-217, zws-ms had three individual siliques sharing one carpopodium in Xindu. Sometimes, one of silique degrade. **(B)** Both zws-ms and zws-217 showed normal siliques in colder Ma'erkang.

the terms "single-organism process (GO:0044699)" and "cellular process (GO:0009987)" got the highest numbers of a gene, at 38 and 37, respectively; the CC and MF categories showed similar most-enriched terms to those in the first classification.

Two stable DEGs and eight environment-specific DEGs got annotations related to flower development or environment response (**Tables 2**, **3**): on the one hand, (1) BnaC08g39130D was line-specific only in zws-ms in both environments, and it was annotated to "ovule development (GO:0048481)" and "response to cold (GO:0009409);" (2) BnaC08g42280D, associated with "vegetative to reproductive phase transition of the meristem (GO:0010228)" and "cellular response to cold (GO:0070417)," was expressed only in zws-217 in Xindu, and in Ma'erkang, it was significantly downregulated in zws-ms, which showed a similar tendency. On the other hand, as to the environment-specific DEGs, BnaA09g45320D, BnaA10g07970D, BnaAnng35580D, BnaC08g29060D, BnaC08g41780D, BnaC08g42450D, and BnaA09g44370D were downregulated in zws-ms or only expressed in zws-217 in Xindu but showed no significant difference between two lines in Ma'erkang, whereas BnaC08g40320D showed variation in Xindu only, with line-specific expression in zws-ms there. To be specific, (1) BnaA09g45320D was annotated to environment-responsive terms such as "response to heat (GO:0009408)," "response to cold (GO:0009409)," and "ovule development (GO:0048481)," its ortholog from *Arabidopsis* is *Chaperonin 10* (*CPN10*); (2) BnaA10g07970D was found a heat shock protein about "response to abiotic stimulus (GO:0009628)" and homologous to AT5G51440; (3) BnaAnng35580D was related to "response to cold (GO:0009409)" and "vegetative to the reproductive phase transition of meristem (GO:0010228)." The ortholog encoded a glycine-rich protein; (4) BnaC08g29060D was in connection with "stamen development (GO:0048443);" (5) BnaC08g41780D was relevant to "vegetative to reproductive phase transition of meristem (GO:0010228)" and "ovule development (GO:0048481)." Its ortholog, AT1G11870, encoded the seryl-tRNA synthetase and was identified as *OVA7* gene; (6) BnaC08g42450D was annotated as "stamen development (GO:0048443);" both (7) BnaA09g44370D and (8) BnaC08g40320D had a Myb-like DNA-binding domain.

TABLE 1 | Number of differentially expressed genes between zws-ms and zws-217 in two environments.

	Number of DEGs	Upregulated genes	Downregulated genes
Xindu	129	67	62
Ma'erkang	117	63	54

KEGG pathway enrichment revealed that the stable DEG classification, the two enriched pathways (**Figure 3A**) with most genes, were "RNA transport (ko03013)," with BnaC08g38300D and BnaC08g40410D involved. Pathways "Biotin metabolism (ko00780)" and "Fatty acid elongation (ko00062)" got the highest enrichment factor values, at 27.4 and 18.4, and associated with BnaC01g43270D and BnaC03g65980D, respectively. As

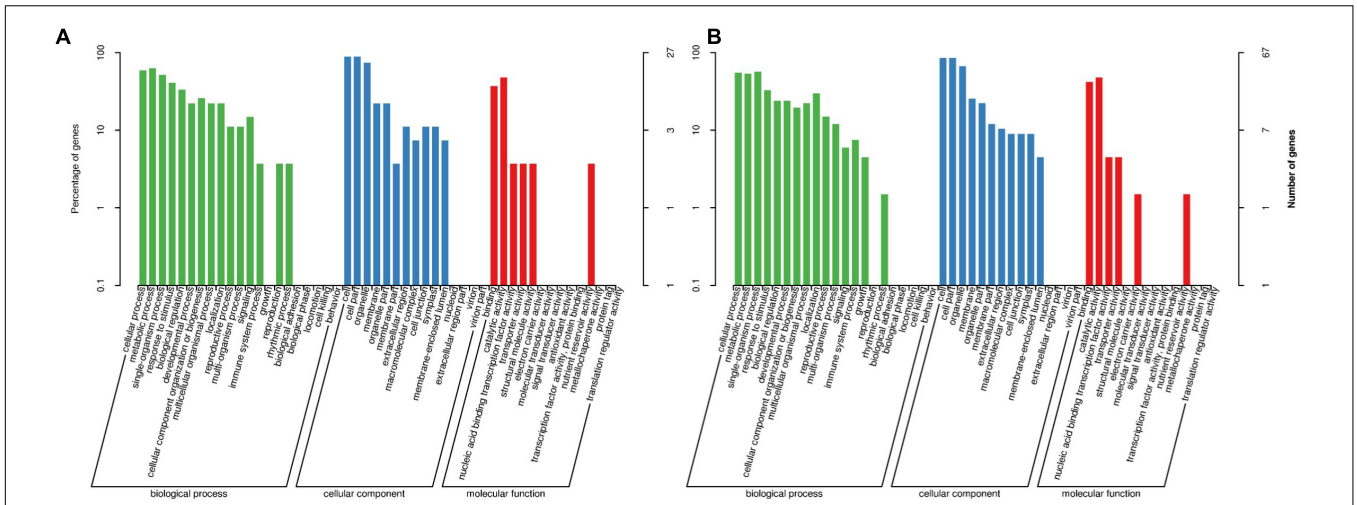


FIGURE 2 | Gene Ontology (GO) terms associated with two classifications of differentially expressed genes (DEGs). **(A)** Thirty-three stable DEGs representing same or similar expression trends in both environments. **(B)** Ninety-six environment-specific DEGs showing different expression trends from Xindu to Ma'erkang. *X*-axis shows GO categories and subclasses of DEGs; *y*-axis shows number or percentage of annotated DEGs.

to the environment-specific DEG group (**Figure 3B**), "Protein processing in endoplasmic reticulum (ko04141)" got the most genes, whereas "Lysine degradation (ko00310)" with BnaA07g09660D and "Aminoacyl-tRNA biosynthesis (ko00970)" with BnaC08g41780D and a new gene got the highest values of enrichment factors.

DISCUSSION

Morphology of higher plants results from the interaction of genotype and environment. The multi-silique trait in rapeseed line zws-ms was found stable in Xindu for generations but absent in Ma'erkang where the climate is colder (Chai et al., 2019). In each place, there are some genes expressed differentially between zws-ms and its NIL zws-217. In Xindu, these DEGs may be the causal factors that distinguish zws-ms from zws-217 in the multi-silique trait, whereas in Ma'erkang, some of those genes turn to no significant difference in expression level between the two lines, which may be regulated by environment and switch off the multi-silique trait in zws-ms. To investigate the potential environment-regulated genes, which switch on/off the multi-silique trait, we compared DEGs generated in Xindu with those in Ma'erkang.

In our earlier report (Chai et al., 2019), we investigated the potential causal variation between zws-ms and zws-217 by the whole-genome re-sequencing, which then led to the identification of two associated regions on chromosome A09 and C08, respectively, as well as some candidate genes within them. It was based on the genomic level conferring inherent and stable genetic variations between the multi-silique and normal-silique lines, which would not be affected by environmental factors. The success of NIL construction, which conferred high genetic similarity between zws-ms and zws-217 and made them different from each other only in this multi-silique trait, ensured the accuracy of these studies. However, the expression level of transcripts can obviously change with external factors. In Chai et al. (2019), we also performed an RNA-seq and found some DEGs between zws-ms and zws-217 in Xindu, where they distinguished from each other in this multi-/single-silique trait; thus, we found some potential causal genes. However, which gene(s) switched off the multi-silique in the colder area remained unclear. Thus, in this study, we added another RNA-seq based on two lines grown in Ma'erkang, where the multi-silique disappears in zws-217, and both lines display normal siliques. By comparing the DEGs from the two environments, we can find out the changed factors.

Thus, among the 129 DEGs from Xindu, we found 33 were stable DEGs. It is worth noting that genes such as BnaA01g10540D, which was downregulated in both Xindu ($\log_2$FC = −4.53) and Ma'erkang ($\log_2$FC = −2.33), were defined as "having the same expression tendency," whereas genes such as BnaA09g06740D, which was line-specifically expressed in zws-ms ($\log_2$FC = $+\infty$) from Xindu and strongly upregulated in it ($\log_2$FC = 7.57) from Ma'erkang, were defined as "having similar expression tendency." Two genes of them, BnaC08g39130D and BnaC08g42280D, got important annotations. BnaC08g39130D was line-specifically expressed in zws-ms and associated with "ovule development (GO:0048481)." Its ortholog in *Arabidopsis*, AT1G14980, encodes chaperonin 10, and this type of protein can be involved in physiological processes such as plant seed abortion (Hanania et al., 2007). BnaC08g42280D was strongly downregulated in zws-ms and annotated to "vegetative to reproductive phase transition of meristem (GO:0010228)." This implies that BnaC08g42280D probably serves as an inhibitor of the formation of multi-silique. Its ortholog, AT1G10930, is RECQ4A. Its mutant is hypersensitive to ultraviolet light and sensitive to methyl methanesulfonate. The two genes confer intrinsic and stable variations in zws-ms and are not influenced by the environment.

TABLE 2 | Annotations for important stable and environment-specific DEGs.

	Gene ID	GO annotation	KEGG pathway
Stable DEG	BnaC08g39130D	copper ion binding (GO:0005507); calmodulin binding (GO:0005516); ATP binding (GO:0005524); mitochondrion (GO:0005739); cytosol (GO:0005829); gluconeogenesis (GO:0006094); glycolytic process (GO:0006096); protein folding (GO:0006457); tryptophan catabolic process (GO:0006569); response to heat (GO:0009408); response to cold (GO:0009409); chloroplast thylakoid membrane (GO:0009535); chloroplast stroma (GO:0009570); response to high light intensity (GO:0009644); response to salt stress (GO:0009651); chloroplast organization (GO:0009658); indoleacetic acid biosynthetic process (GO:0009684); chloroplast envelope (GO:0009941); isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway (GO:0019288); cysteine biosynthetic process (GO:0019344); response to endoplasmic reticulum stress (GO:0034976); response to hydrogen peroxide (GO:0042542); response to cadmium ion (GO:0046686); apoplast (GO:0048046); ovule development (GO:0048481); chaperone binding (GO:0051087); positive regulation of superoxide dismutase activity (GO:1901671);	–
	BnaC08g42280D	telomere maintenance (GO:0000723); double-strand break repair via homologous recombination (GO:0000724); nucleic acid binding (GO:0003676); ATP binding (GO:0005524); nucleus (GO:0005634); DNA replication (GO:0006260); plasmodesma (GO:0009506); vegetative to reproductive phase transition of meristem (GO:0010228); ATP-dependent 3'-5' DNA helicase activity (GO:0043140); cellular response to cold (GO:0070417); cellular response to abscisic acid stimulus (GO:0071215);	Homologous recombination (ko03440)
Environment-specific DEG	BnaA09g44370D	DNA binding (GO:0003677); chromatin binding (GO:0003682); sequence-specific DNA binding transcription factor activity (GO:0003700); nucleus (GO:0005634); regulation of transcription, DNA-templated (GO:0006355); protein targeting to membrane (GO:0006612); response to salt stress (GO:0009651); response to ethylene (GO:0009723); response to auxin (GO:0009733); response to abscisic acid (GO:0009737); response to gibberellin (GO:0009739); response to salicylic acid (GO:0009751); response to jasmonic acid (GO:0009753); positive regulation of flavonoid biosynthetic process (GO:0009963); regulation of plant-type hypersensitive response (GO:0010363); response to cadmium ion (GO:0046686);	–
	BnaA09g45320D	copper ion binding (GO:0005507); calmodulin binding (GO:0005516); ATP binding (GO:0005524); mitochondrion (GO:0005739); cytosol (GO:0005829); gluconeogenesis (GO:0006094); glycolytic process (GO:0006096); protein folding (GO:0006457); tryptophan catabolic process (GO:0006569); response to heat (GO:0009408); response to cold (GO:0009409); chloroplast thylakoid membrane (GO:0009535); chloroplast stroma (GO:0009570); response to high light intensity (GO:0009644); response to salt stress (GO:0009651); chloroplast organization (GO:0009658); indoleacetic acid biosynthetic process (GO:0009684); chloroplast envelope (GO:0009941); isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway (GO:0019288); cysteine biosynthetic process (GO:0019344); response to endoplasmic reticulum stress (GO:0034976); response to hydrogen peroxide (GO:0042542); response to cadmium ion (GO:0046686); apoplast (GO:0048046); ovule development (GO:0048481); chaperone binding (GO:0051087); positive regulation of superoxide dismutase activity (GO:1901671);	–
	BnaA10g07970D	response to stress (GO:0006950); response to abiotic stimulus (GO:0009628); cellular process (GO:0009987);	Protein processing in endoplasmic reticulum (ko04141)
	BnaAnng35580D	nucleotide binding (GO:0000166); alternative mRNA splicing, via spliceosome (GO:0000380); double-stranded DNA binding (GO:0003690); single-stranded DNA binding (GO:0003697); mRNA binding (GO:0003729); protein kinase activity (GO:0004672); nucleus (GO:0005634); mitochondrion (GO:0005739); peroxisome (GO:0005777); cytosol (GO:0005829); gluconeogenesis (GO:0006094); glycolytic process (GO:0006096); mRNA export from nucleus (GO:0006406); water transport (GO:0006833); hyperosmotic response (GO:0006972); Golgi organization (GO:0007030); response to cold (GO:0009409); response to water deprivation (GO:0009414); plasmodesma (GO:0009506); chloroplast (GO:0009507); response to salt stress (GO:0009651); etioplast organization (GO:0009662); lignin biosynthetic process (GO:0009809); response to zinc ion (GO:0010043); regulation of stomatal movement (GO:0010119); response to chitin (GO:0010200); vegetative to reproductive phase transition of meristem (GO:0010228); RNA secondary structure unwinding (GO:0010501); carotenoid biosynthetic process (GO:0016117); brassinosteroid biosynthetic process (GO:0016132); cinnamoyl-CoA reductase activity (GO:0016621); DNA duplex unwinding (GO:0032508); negative regulation of circadian rhythm (GO:0042754); protein homodimerization activity (GO:0042803); innate immune response (GO:0045087); carotenoid isomerase activity (GO:0046608); response to cadmium ion (GO:0046686); apoplast (GO:0048046); defense response to fungus (GO:0050832);	–
	BnaC08g29060D	RNA splicing, via endonucleolytic cleavage and ligation (GO:0000394); inositol hexakisphosphate binding (GO:0000822); response to molecule of bacterial origin (GO:0002237); ubiquitin-protein transferase activity (GO:0004842); nucleus (GO:0005634); vacuolar membrane (GO:0005774); methionine biosynthetic process (GO:0009086); auxin-activated signaling pathway (GO:0009734);	

(Continued)

TABLE 2 | Continued

Gene ID	GO annotation	KEGG pathway
	auxin binding (GO:0010011); stomatal complex morphogenesis (GO:0010103); pollen maturation (GO:0010152); protein ubiquitination (GO:0016567); stamen development (GO:0048443); lateral root development (GO:0048527); photoperiodism, flowering (GO:0048573); cellular response to nitrate (GO:0071249); primary root development (GO:0080022);	–
BnaC08g40320D	chromatin binding (GO:0003682); sequence-specific DNA binding transcription factor activity (GO:0003700); nucleus (GO:0005634); regulation of transcription, DNA-templated (GO:0006355); membrane fusion (GO:0006944); identical protein binding (GO:0042802); sequence-specific DNA binding (GO:0043565); Golgi vesicle transport (GO:0048193);	–
BnaC08g41780D	sulfur amino acid metabolic process (GO:0000096); serine-tRNA ligase activity (GO:0004828); ATP binding (GO:0005524); mitochondrion (GO:0005739); rRNA processing (GO:0006364); seryl-tRNA aminoacylation (GO:0006434); mitochondrion organization (GO:0007005); cellular amino acid biosynthetic process (GO:0008652); serine family amino acid metabolic process (GO:0009069); chloroplast (GO:0009507); embryo development ending in seed dormancy (GO:0009793); chloroplast relocation (GO:0009902); leaf morphogenesis (GO:0009965); thylakoid membrane organization (GO:0010027); photosystem II assembly (GO:0010207); vegetative to reproductive phase transition of meristem (GO:0010228); iron-sulfur cluster assembly (GO:0016226); cell differentiation (GO:0030154); regulation of protein dephosphorylation (GO:0035304); cell wall modification (GO:0042545); transcription from plastid promoter (GO:0042793); positive regulation of transcription, DNA-templated (GO:0045893); ovule development (GO:0048481);	Aminoacyl-tRNA biosynthesis (ko00970)
BnaC08g42450D	response to molecule of bacterial origin (GO:0002237); protein serine/threonine kinase activity (GO:0004674); ATP binding (GO:0005524); plasma membrane (GO:0005886); N-terminal protein myristoylation (GO:0006499); protein targeting to membrane (GO:0006612); membrane fusion (GO:0006944); response to oxidative stress (GO:0006979); transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169); systemic acquired resistance (GO:0009627); seed germination (GO:0009845); stomatal complex morphogenesis (GO:0010103); regulation of plant-type hypersensitive response (GO:0010363); integral component of membrane (GO:0016021); negative regulation of programmed cell death (GO:0043069); protein autophosphorylation (GO:0046777); stamen development (GO:0048443); micropyle (GO:0070825);	–
Cole_newGene_2073	intracellular membrane-bounded organelle (GO:0043231); cytoplasmic part (GO:0044444);	Protein processing in endoplasmic reticulum (ko04141); Plant-pathogen interaction (ko04626)

Aiming to investigate the thermomorphogenesis-relative genes, we then turned to those environment-specific DEGs, which were defined as those genes such as BnaA05g21710D, which showed significant upregulation in Xindu ($\log_2$FC = 3.12) but no obvious change in Ma'erkang ($\log_2$FC = 0.93), that were considered as "environment-specific DEGs" herein. Among the 96 environment-specific DEGs, nine were noteworthy. BnaA09g45320D shared the same ortholog (AT1G14980) with BnaC08g39130D, but unlike the latter, it was only expressed in zws-217 in Xindu and showed no significant difference in Ma'erkang. Its annotation may explain this in response to heat (GO:0009408) and response to cold (GO:0009409). Moreover, it was also annotated to "ovule development (GO:0048481)." Take chaperonin 21 as an example: it was found differentially expressed in seedless and seeded grapes; its silencing resulted in seed abortion in transgenic tobacco (*Nicotiana benthamiana*) and seedless fruits in transgenic tomato (*Lycopersicum esculentum*) (Hanania et al., 2007). In Xindu, BnaA09g45320D was line-specific in the normal line, whereas in Ma'erkang, zws-ms and zws-217 showed no significant difference in it. This implied a potential relevance with the multi-silique. The ortholog of BnaC08g41780D, AT1G11870, encodes a seryl-tRNA synthetase, i.e., OVA7, of which disruption can result in ovule abortion in *Arabidopsis* (Berg et al., 2005). When zws-ms represents multi-silique in Xindu, the expression of BnaC08g41780D was not detected in it, which indicates a possibility of it modifying the trait in a warm area. Thus, these two genes, due to the consistency of their expression verities with the changes of environment, as well as their annotations, are considered the most important thermomorphogenesis-related genes of all the candidates.

Although the other environmental-specific DEGs also implicated some indirect clues: BnaA09g44370D and BnaC08g40320D were annotated to MYB-like or MYB/SANT-like DNA-binding domain; and their orthologs, AT1G19000 and AT1G13450, were both homeodomain-like superfamily proteins. Some MYB proteins were found regulated by phytochrome-interacting factor 4, an important thermomorphogenesis factor in *Arabidopsis* (Wang et al., 2018). In addition to flower color, MYB transcription factors are also involved in the regulation of flower development: transgenic tobacco overexpressing MdMYB3 from apple (Malus × domestica) got longer peduncles of flowers and styles of pistils (Vimolmangkang et al., 2013). BnaA10g07970D and Cole_newGene_2073 were identified as

TABLE 3 | Ortholog information for selected stable DEGs and nine environment-specific DEGs.

	Gene ID	Ortholog in Arabidopsis	Description
Stable DEG	BnaC08g39130D	AT1G14980	Chaperonin 10 (CPN10)
	BnaC08g42280D	AT1G10930	RECQ4A
Environment-specific DEG	BnaA09g44370D	AT1G19000	Homeodomain-like superfamily protein
	BnaA09g45320D	AT1G14980	Chaperonin 10 (CPN10)
	BnaA10g07970D	AT5G51440	HSP20-like chaperones superfamily protein
	BnaAnng35580D	AT2G21660	GLYCINE RICH PROTEIN 7 (ATGRP7)
	BnaC08g29060D	AT1G12820	Auxin signaling F-box 3 (AFB3)
	BnaC08g40320D	AT1G13450	Homeodomain-like superfamily protein
	BnaC08g41780D	AT1G11870	Seryl-tRNA synthetase (SRS), OVULE ABORTION 7 (OVA7)
	BnaC08g42450D	AT1G09970	LRR XI-23, RECEPTOR-LIKE KINASE 7 (RLK7)
	Cole_newGene_2073	AT3G07770	HEAT SHOCK PROTEIN 90.6, ATHSP90-6

heat shock proteins, which respond to temperature and are involved in signal transduction, protein trafficking, protein degradation, maintaining protein homeostasis, and so on (Raman and Suguna, 2015), but whether/how they are related to flower/fruit shape needs further research. AT2G21660 is homologous to BnaAnng35580D. It is *ATGRP7* and encodes a small glycine-rich RNA binding protein. Loss-of-function mutations are late flowering in a non-photoperiodic manner. BnaC08g29060D was annotated to "photoperiodism, flowering (GO: 0048573)" and identified as homologous to AT1G12820, which is auxin signaling F-box 3 and involved in primary and lateral root growth inhibition (Dong et al., 2006). AT1G09970 is an ortholog of BnaC08g42450D and identified as *RECEPTOR-LIKE KINASE 7*. According to the existing knowledge, it is involved in the control of germination speed and oxidant stress tolerance.

There is another sort of gene, which was not discussed. Although these genes got annotations regarding environment-response, they showed no difference resulting from environmental changes. Take BnaC08g39990D for example; it was annotated to "response to cold (GO:0009409)," which seemed potentially related to what we were seeking. However, further analysis indicated that it was always upregulated in zws-ms in both Xindu and Ma'erkang, showing no variety as the environment changed. Thus, its expression level had a low correlation with environmental factors acting on the switching on/off of the multi-silique formation. So, this kind of gene was not emphasized herein.

Rapeseed is not the first crop conferring the multi-pistil trait reported. In fact, wheat contributes relatively sufficient studies about it. Researchers reported their multi-pistil wheat materials, in which the trait can be controlled by a recessive gene, two recessive non-complementary genes, or a single dominant gene, and the loci were located in different chromosomes (Zhu et al., 2019). Even so, no underlying gene in wheat has not been cloned so far (Zhu et al., 2019).

Notably, the multi-pistil trait was also reported in sweet cherry (*P. avium* L.), co-regulated by genes *PaMADS3/4/5*, members

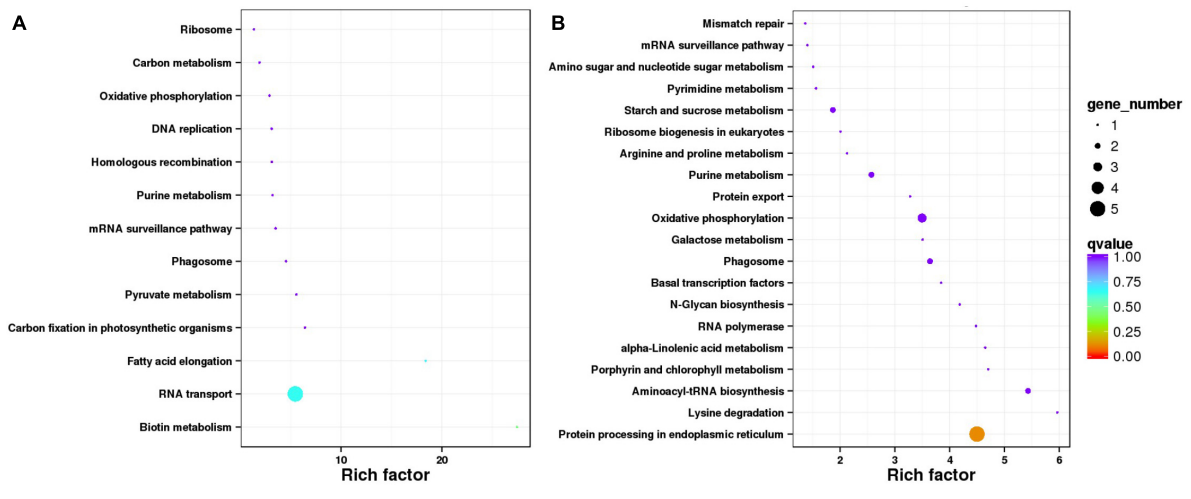


FIGURE 3 | Statistics of Kyoto Encyclopedia of Genes and Genomes pathway enrichment for two classifications of differentially expressed genes (DEGs). **(A)** Data for 33 stable DEGs representing same or similar expression trends in both environments. **(B)** Data for 96 environment-specific DEGs showing different expression trends from Xindu to Ma'erkang.

of the MADS-box family (Liu et al., 2019; Wang et al., 2019). Interestingly, that was also induced by temperature: it occurred more frequently in the warm region of Shanghai than in the cool region of Dalian. This coincides with our discovery from rapeseed. Notably, there are many environmental factors, but the temperature is the most significant one, varying from Xindu to Ma'erkang. Although some other factors, such as photoperiod, are also vital for plant development, there has been no report showing it can change pistil number in flower; it is known to regulate flowering time, as well as some physiological characters such as hypocotyl length, cotyledon angle, chlorophyll content, etc., so far. Instead, a change in temperature, as mentioned earlier (Liu et al., 2019; Wang et al., 2019), can promote the multi-pistil formation (in sweet cherry). Thus, we suppose it is the temperature that would switch on/off the multi-silique herein.

The multi-silique phenotype means different things to different plants. For some crops, the multi-silique phenotype was considered as an advantage: in wheat, for example, it is supposed to have the potential to increase yields (Wei, 2017; Zhu et al., 2019); on the other hand, for some other plants, such as sweet cherry, the multi-pistil is regarded as a disadvantage physiological disorder decreasing its commercial value (Liu et al., 2019). As to rapeseed, although how to use the multi-silique line to increase the seed yield, as well as the benefit of commercial use, still needs more investigations. We should explain that the mountain area (cold region such as Ma'erkang) is not the major rapeseed-producing area in Sichuan Province; we grow rapeseed there mainly for scientific purposes: the shuttle breeding makes it possible to produce two generations in 1 year, accelerating the population construction processes. Collectively, this study at least provides a new perspective into flower/fruit development in higher plants. In the next stage, the functional validation of these candidate genes will be carried out by investigating the phenotypic, biochemical, and molecular changes in the transgenic plants (both overexpression and knockdown).

CONCLUSION

By comparing DEGs between zws-ms and zws-217 in Xindu with those in Ma'erkang and referring to the gene annotations, we selected BnaA09g45320D (*chaperonin 10*-homologous) and BnaC08g41780D (*OVA7*-homologous) as the possible thermomorphogenesis-related genes, switching on/off the multi-silique under different environments.

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 and accession number PRJNA736189.

AUTHOR CONTRIBUTIONS

LC and LJ conceived the experiment. LC and JZ performed the research. HL, CC, JJ, and BZ contributed to data analysis. LC wrote the manuscript. LJ and LW reviewed and revised the manuscript. All authors reviewed and approved this submission.

FUNDING

This research was funded by the National Natural Science Foundation of China, grant number 31560402, Modern Agro-Industry Technology Research System of China, grant number CARS-12, National Key Research and Development Plan, grant number 2018YFE0108200, Sichuan Crop Breeding Community, grant numbers 2021YFYZ0018 and 2021YFYZ0020, Financial Innovation Ability Promotion Project of Sichuan Province, grant number 2021XKJS003, and Major Science and Technology Project of Sichuan Province, grant number 2018NZDZX0003.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.678804/full#supplementary-material

Supplementary Table 1 | The 117 DEGs between zws-ms and zws-217 from colder area Ma'erkang.

Supplementary Table 2 | Line-specific expressed genes from Xindu.

Supplementary Table 3 | Line-specific expressed genes from Ma'erkang.

Supplementary Table 4 | The 33 stable DEGs with the same or similar expression tendency in two environments.

Supplementary Table 5 | The 96 environment-specific DEGs with different expression tendency in two environments.

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berg, M., Rogers, R., Muralla, R., and Meinke, D. (2005). Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in Arabidopsis. *Plant J.* 44, 866–878. doi: 10.1111/j.1365-313X.2005.02580.x

Casal, J. J., and Balasubramanian, S. (2019). Thermomorphogenesis. *Annu. Rev. Plant Biol.* 70, 321–346. doi: 10.1146/annurev-arplant-050718-095919

Chai, L., Zhang, J., Li, H., Jiang, J., Cui, C., Zheng, B., et al. (2020b). Dynamic Transcriptome Analysis of A Multi-Silique Trait in Rapeseed (*Brassica napus* L.). *Int. J. Agric. Biol.* 24, 1625–1632.

Chai, L., Zhang, J., Li, H., Zheng, B., Jiang, J., Cui, C., et al. (2020a). Investigation for a multi-silique trait in *Brassica napus* by alternative splicing analysis. *PeerJ.* 8:e10135. doi: 10.7717/peerj.10135

Chai, L., Zhang, J., Lu, K., Li, H., Wu, L., Wan, H., et al. (2019). Identification of genomic regions associated with multi-silique trait in *Brassica napus*. *BMC Genomics* 20:304. doi: 10.1186/s12864-019-5675-4

Ding, Y., Shi, Y., and Yang, S. (2020). Molecular Regulation of Plant Responses to Environmental Temperatures. *Mol. Plant* 13, 544–564. doi: 10.1016/j.molp.2020.02.004

Dong, L., Wang, L., Zhang, Y. E., Zhang, Y., Deng, X., and Xue, Y. (2006). An Auxin-Inducible F-Box Protein CEGENDUO Negatively Regulates Auxin-Mediated Lateral Root Formation in *Arabidopsis*. *Plant Mol. Biol.* 60, 599–615. doi: 10.1007/s11103-005-5257-5

Duan, Z., Shen, C., Li, Q., Lü, G., Ni, Y., Yu, D., et al. (2015). Identification of a novel male sterile wheat mutant *dms* conferring dwarf status and multi-pistils. *J. Integr. Agr.* 14, 1706–1714. doi: 10.1016/S2095-3119(14)60936-9

Friend, D. J. C. (1965). TILLERING AND LEAF PRODUCTION IN WHEAT AS AFFECTED BY TEMPERATURE AND LIGHT INTENSITY. *Can. J. Bot.* 43, 1063–1076. doi: 10.1139/b65-123

Geng, X., Jiang, C., Yang, J., Wang, L., Wu, X., and Wei, W. (2016). Rapid Identification of Candidate Genes for Seed Weight Using the SLAF-Seq Method in *Brassica napus*. *PLoS One* 11:e147580. doi: 10.1371/journal.pone.0147580

Guo, J., Zhang, G., Song, Y., Li, Z., Ma, S., Niu, N., et al. (2019). Comparative proteomic analysis of multi-ovary wheat under heterogeneous cytoplasm suppression. *BMC Plant Biol.* 19:175. doi: 10.1186/s12870-019-1778-y

Hanania, U., Velcheva, M., Or, E., Flaishman, M., Sahar, N., and Perl, A. (2007). Silencing of chaperonin 21, that was differentially expressed in inflorescence of seedless and seeded grapes, promoted seed abortion in tobacco and tomato fruits. *Transgenic Res.* 16, 515–525. doi: 10.1007/s11248-006-9044-0

Ibañez, C., Poeschl, Y., Peterson, T., Bellstädt, J., Denk, K., Gogol-Döring, A., et al. (2017). Ambient temperature and genotype differentially affect developmental and phenotypic plasticity in Arabidopsis thaliana. *BMC Plant Biol.* 17:114. doi: 10.1186/s12870-017-1068-5

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Lee, Y., Kim, K., Lee, J., Cha, Y., Moon, Y., Song, Y., et al. (2018). Comprehensive Transcriptome Profiling in Relation to Seed Storage Compounds in Tetralocular *Brassica rapa*. *J. Plant Growth Regul.* 37, 867–882. doi: 10.1007/s00344-018-9784-0

Liu, J., Wang, J., She, W., Wang, L., Luo, M., Chen, Y., et al. (2019). MADS-Box Genes are Involved in Cultivar- and Temperature-Dependent Formation of Multi-pistil and Polycarpy in *Prunus avium* L. *J. Plant Growth Regul.* 38, 1017–1027. doi: 10.1007/s00344-018-09911-8

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Nair, R. M., Peck, D. M., Dundas, I. S., Samac, D. A., Moore, A., and Randles, J. W. (2008). Morphological characterisation and genetic analysis of a bi-pistil mutant (*bip*) in *Medicago truncatula* Gaertn. *Sex. Plant Reprod.* 21, 133–141. doi: 10.1007/s00497-008-0073-0

Raman, S., and Suguna, K. (2015). Functional characterization of heat−shock protein 90 from *Oryza sativa* and crystal structure of its N−terminal domain. *Acta Crystallogr. Sect. F: Struct. Biol. Commun.* 71, 688–696. doi: 10.1107/S2053230X15006639

Shen, W., Qin, P., Yan, M., Li, B., Wu, Z., Wen, J., et al. (2019). Fine mapping of a silique length- and seed weight-related gene in *Brassica napus*. *Theor. Appl. Genet.* 132, 2985–2996. doi: 10.1007/s00122-019-03400-6

Sionit, N., Strain, B. R., and Flint, E. P. (1987). Interaction of temperature and co_2 enrichment on soybean: growth and dry matter partitioning. *Can. J. Plant Sci.* 67, 59–67. doi: 10.4141/cjps87-007

Vimolmangkang, S., Han, Y., Wei, G., and Korban, S. S. (2013). An apple MYB transcription factor, *MdMYB3*, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 13:176. doi: 10.1186/1471-2229-13-176

Wang, J., Liu, J., Jiu, S., Li, Y., Whiting, M., She, W., et al. (2019). The MADS-box genes *PaMADS3/4/5* co-regulate multi-pistil formation induced by high temperature in *Prunus avium* L. *Sci. Hortic.-Amsterdam* 256:108593. doi: 10.1016/j.scienta.2019.108593

Wang, S., Ding, L., Liu, J. X., and Han, J. J. (2018). PIF4-Regulated Thermo-responsive Genes in *Arabidopsis*. *Biotechnol. Bull.* 34, 57–65.

Wang, X., Chen, L., Wang, A., Wang, H., Tian, J., Zhao, X., et al. (2016). Quantitative trait loci analysis and genome-wide comparison for silique related traits in *Brassica napus*. *BMC Plant Biol.* 16:71. doi: 10.1186/s12870-016-0759-7

Wei, S. H. (2017). Characterization and expression of *WAG-2* transcripts in a wheat three-pistil mutant line. *Russ. J. Plant Physl.* 64, 680–687. doi: 10.1134/S1021443717050156

Xiao, L., Zhao, H., Zhao, Z., Du, D., Xu, L., Yao, Y., et al. (2013). Genetic and physical fine mapping of a multilocular gene *Bjln1* in *Brassica juncea* to a 208-kb region. *Mol. Breeding* 32, 373–383. doi: 10.1007/s11032-013-9877-1

Xu, P., Lv, Z., Zhang, X., Wang, X., Pu, Y., Wang, H., et al. (2014). Identification of molecular markers linked to trilocular gene (*mc1*) in *Brassica juncea* L. *Mol. Breeding* 33, 425–434. doi: 10.1007/s11032-013-9960-7

Yadava, S. K., Paritosh, K., Panjabi-Massand, P., Gupta, V., Chandra, A., Sodhi, Y. S., et al. (2014). Tetralocular ovary and high silique width in yellow sarson lines of *Brassica rapa* (subspecies *trilocularis*) are due to a mutation in Bra034340 gene, a homologue of *CLAVATA3* in Arabidopsis. *Theor. Appl. Genet.* 127, 2359–2369. doi: 10.1007/s00122-014-2382-z

Yang, P., Shu, C., Chen, L., Xu, J., Wu, J., and Liu, K. (2012). Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 125, 285–296. doi: 10.1007/s00122-012-1833-7

Yang, Z., Chen, Z., Peng, Z., Yu, Y., Liao, M., and Wei, S. (2017). Development of a high-density linkage map and mapping of the three-pistil gene (*Pis1*) in wheat using GBS markers. *BMC Genomics* 18:567. doi: 10.1186/s12864-017-3960-7

Yang, Z., Peng, Z., Yang, H., Yang, J., Wei, S., and Cai, P. (2011). Suppression Subtractive Hybridization Identified Differentially Expressed Genes in Pistil Mutations in Wheat. *Plant Mol. Biol. Rep.* 29, 431–439. doi: 10.1007/s11105-010-0249-2

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

Yu, Z. Y., Luo, Q., Peng, Z. S., Wei, S. H., Yang, Z. J., and Yamamoto, N. (2020). Genetic mapping of the three-pistil gene *Pis1* in an F_2 population derived from a synthetic hexaploid wheat using multiple molecular marker systems. *Cereal Res. Commun.* 2020:078. doi: 10.1007/s42976-020-00078-1

Zheng, H., Zhang, J., Zhuang, H., Zeng, X., Tang, J., Wang, H., et al. (2019). Gene mapping and candidate gene analysis of multi-floret spikelet 3 (*mfs3*) in rice (*Oryza sativa* L.). *J. Integr. Agr.* 18, 2673–2681. doi: 10.1016/S2095-3119(19)62652-3

Zhu, X., Ni, Y., He, R., Jiang, Y., Li, Q., and Niu, J. (2019). Genetic mapping and expressivity of a wheat multi-pistil gene in mutant *12TP*. *J. Integr. Agr.* 18, 532–538. doi: 10.1016/S2095-3119(18)61935-5

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 September 2021
doi: 10.3389/fgene.2021.703688

Gibberellin Induced Transcriptome Profiles Reveal Gene Regulation of Loquat Flowering

Yuanyuan Jiang[1,2†], Yicun Liu[3†], Yongshun Gao[2,4,5], Jiangrong Peng[2], Wenbing Su[2,6], Yuan Yuan[1], Xianghui Yang[2], Chongbin Zhao[2], Man Wang[2], Shunquan Lin[2], Ze Peng[2*] and Fangfang Xie[2*]*

[1] Henry Fok College of Biology and Agriculture, Shaoguan University, Shaoguan, China, [2] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Horticulture, South China Agricultural University, Guangzhou, China, [3] College of Agriculture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou, China, [4] Beijing Academy of Forestry and Pomology Sciences, Beijing, China, [5] Beijing Engineering Research Center for Strawberry, Beijing, China, [6] Fruit Research Institute, Fujian Academy of Agricultural Science, Fuzhou, China

Flowering is an integral part of the life cycle of flowering plants, which is essential for plant survival and crop production. Most woody fruit trees such as apples and pears bloom in spring, but loquat blooms in autumn and winter. Gibberellin (GA) plays a key role in the regulation of plant flower formation. In this study, we sprayed loquat plants with exogenous GA_3, which resulted in vigorous vegetative growth rather than floral bud formation. We then performed a comprehensive RNA-seq analysis on GA_3-treated and control-treated leaves and buds over three time periods to observe the effects of exogenous GA_3 application on floral initiation and development. The results showed that 111 differentially expressed genes (DEGs) and 563 DEGs were down-regulated, and 151 DEGs and 506 DEGs were up-regulated in buds and leaves, respectively, upon treatment with GA_3. Among those that are homologs of the DELLA-mediated GA signal pathway genes, some may be involved in the positive regulation of flower development, including *EjWRKY75, EjFT, EjSOC1, EjAGL24, EjSPL, EjLFY, EjFUL,* and *EjAP1*; while some may be involved in the negative regulation of flower development, including *EjDELLA, EjMYC3, EjWRKY12,* and *EjWRKY13*. Finally, by analyzing the co-expression of DEGs and key floral genes *EjSOC1s, EjLFYs, EjFULs, EjAP1s*, 330 candidate genes that may be involved in the regulation of loquat flowering were screened. These genes belong to 74 gene families, including Cyclin_C, Histone, Kinesin, Lipase_GDSL, MYB, P450, Pkinase, Tubulin, and ZF-HD_dimer gene families. These findings provide new insights into the regulation mechanism of loquat flowering.

Keywords: GA_3, flowering, RNA-seq, co-expression, loquat

OPEN ACCESS

Edited by:
Yunpeng Cao,
Central South University Forestry and Technology, China

Reviewed by:
Zhongqi Fan,
Fujian Agriculture and Forestry University, China
Xu Liu,
University of Chinese Academy of Sciences, China
Rajesh Kumar Gazara,
Indian Institute of Technology Roorkee, India
Lester Young,
University of Saskatchewan, Canada

***Correspondence:**
Shunquan Lin
loquat@scau.edu.cn
Ze Peng
pengze1990121@gmail.com
Fangfang Xie
xiefangfang202012@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 April 2021*
Accepted: *24 August 2021*
Published: *10 September 2021*

Citation:
Jiang Y, Liu Y, Gao Y, Peng J, Su W, Yuan Y, Yang X, Zhao C, Wang M, Lin S, Peng Z and Xie F (2021) Gibberellin Induced Transcriptome Profiles Reveal Gene Regulation of Loquat Flowering. Front. Genet. 12:703688. doi: 10.3389/fgene.2021.703688

INTRODUCTION

The floral transformation of plants is affected by various endogenous and exogenous factors, forming a very sophisticated and complex regulatory network. It can accurately respond to internal and external signals and integrate them together to ensure that plants bloom at a favorable time and reproduce successfully. Plants can accurately sense photoperiod changes and adjust flowering time (Shim et al., 2017); in addition, gibberellin (GA), temperature, vernalization and age signals can also affect plant flowering (Moon et al., 2005; Amasino, 2010; Song et al., 2013; Teotia and Tang, 2015). These signals are not independent. In *Arabidopsis*, they integrate related signals to

regulate plant flower formation through integrators *FLOWERING LOCUS T (FT)*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*, *LEAFY (LFY)*, etc. (Blazquez et al., 1998; Kardailsky et al., 1999; Lee et al., 2000).

In *Arabidopsis*, the GA signal mainly regulates the flower formation of plants through the interaction with the photoperiod signal and regulates the expression of *FT* under the conditions of LD and SD (Osnato et al., 2012; Song et al., 2012). In leaves, MYB-type transcription factor ASYMMETRIC LEAVES positively regulates the expression of GA biosynthesis gene *GA20ox1* (Song et al., 2012). AS1 forms a complex with CO protein and regulates FT expression by directly binding to the FT promoter (Song et al., 2012). As a central inhibitor of the GA signaling pathway, DELLA has been proven to interact with many transcription factors in leaves and stem tips and regulate their activities, thereby regulating plant flowering (Bao et al., 2020). For example, under long-day conditions, DELLA directly binds to the CCT domain of the CO protein and sequentially separates CO from the binding of the FT promoter, thereby down-regulating the expression of *FT* (Wang et al., 2016; Xu et al., 2016). In addition, DELLA can inhibit the interaction between CO and Nuclear factor Y (NF-Y) subunit B (NF-YB), and DELLA can also regulate the expression of *FT* through the interaction between PHYTOCHROME INTERACTING FACTOR 4 (PIF4) and MYC3 (de Lucas et al., 2008; Xu et al., 2016; Bao et al., 2019, 2020).

In addition, GA can directly promote flowering by up-regulating flowering integrons *LFY*, *SOC1* and *AGAMOUS-LIKE 24 (AGL24)* independently of the photoperiod pathway (Blazquez and Weigel, 2000; Moon et al., 2003; Hisamatsu and King, 2008; Liu et al., 2008). Hou et al. (2014) found that GA signal can regulate the expression of *SOC1* through epigenetic modification mediated by NF-Y complex. Under short-day conditions, the promoter activity of *LFY* gradually increased during vegetative growth, and GA enhanced the promoter activity and accelerated plant flowering (Blazquez et al., 1998). GA mainly regulates miR159 by inhibiting the expression of DELLA protein, thereby regulating the transcription of downstream *MYB33*, ultimately regulating the expression of *LFY,* and regulating the floral transformation of plants (Blazquez and Weigel, 2000; Gocal et al., 2001; Achard et al., 2009; Davis, 2009). DELLA may recruit different SPLs to target various downstream target genes, so that GA can play a role in different development environments (Bao et al., 2020). In addition, GA signal can regulate the expression of *WRKY12, WRKY13* and *WRKY75* genes through DELLA protein to regulate plant flower formation (Li et al., 2016; Zhang et al., 2018). In summary, the role of GA signal in the regulation of plant flower formation is very important and complex, but there are relatively few studies on gibberellin-mediated flower formation in woody fruit trees.

Loquat (Eriobotrya japonica Lindl.) is an evergreen fruit tree, which belongs to the Maloideae subfamily of the Rosaceae family which is mainly planted in subtropical regions. In Rosaceae, the flowering transition time and flowering time usually occur in different years (flower buds differentiate in summer and autumn, and the flower buds bloom in the spring of the second year after dormancy), including apples, pears, and strawberries (Kurokura et al., 2013). However, the flowering transformation and flowering of loquat occur in the same year. Flower bud differentiation generally occurs from July to September, with flowering occurring between October and January of the same year (Lin, 2007). The phenomenon of autumn flowering and spring harvest of loquat is very unique among woody fruit trees. In spring, the selection of fresh fruit varieties is greatly reduced. Therefore, the market demand for fresh loquat fruits in spring is relatively high and the price is relatively high. However, in cold winters (especially extreme weather events), the newly opened loquat flowers or young fruits are very susceptible to freezing damage (Peng et al., 2021), resulting in a reduction in loquat production or even no harvest.

Recent research results show that although the start time of loquat flower bud differentiation is similar to that of apples and pears, it occurs from June to July; the difference is that the development of loquat flower buds is continuous and does not undergo dormancy, and it blooms in autumn and winter (Jiang et al., 2019c). In addition, after treating loquat plants with the exogenous hormone GA_3, the plants are in vigorous vegetative growth and cannot form flowers, and genes such as *EjSOC1s*, *EjAP1s* and *EjLFYs* are strongly inhibited (Jiang et al., 2019a,b,c). These studies show that gibberellin can regulate the flowering of loquat by regulating the expression of genes such as *EjSOC1s*, *EjAP1s* and *EjLFYs*. In Arabidopsis, *SOC1, AP1* and *LFY* genes are all downstream of the floral regulation network, and GA regulates their expression through the inhibition or activation of other transcription factors mediated by DELLAs (Bao et al., 2020). However, the mechanism of gibberellin regulating loquat flowering is not clear. Transcriptome sequencing technology is based on exogenous sequencing technology to quickly and comprehensively understand the difference level of transcripts. The application of transcriptome sequencing technology has accelerated the gene expression profile analysis and gene identification of many plant species. In this study, transcriptome analysis was performed on the materials of the GA_3 treatment and the control group to screen the key genes related to the regulation of loquat flowering, in order to analyze the regulation mechanism of loquat flowering.

MATERIALS AND METHODS

Plant Materials

The loquat tissue materials involved in the experiment were taken from 12-year-old "JieFangZhong" loquat plants in the loquat plant germplasm resource nursery of South China Agricultural University (Guangzhou, China, N23°09′N, 113°20′E). The experimental plants have entered the flowering and fruiting age for several years and have grown well. Loquat trees are planted in the loquat germplasm resource nursery and grow under natural conditions.

Exogenous GA_3 Treatment and Sample Collection

The trees were sprayed with an aqueous solution containing 0.1% (v/v) phosphoric acid and 0.025% (v/v) Triton X-100 as a

surfactant and 300 mg L^{-1} GA_3 (Dingguo Biotechnology Co., Ltd., Guangzhou, China). Spray the control plants with a solution containing only 0.1% (v/v) phosphoric acid and 0.025% (v/v) Triton X-100. The experimental treatment method was: spraying all leaves and top buds (soaked, the leaves began to drip) every 2 weeks, from May 18th to August 10th. The differentiation time of loquat flower buds is from the end of June to the beginning of July (Jiang et al., 2019c). Accordingly, the sampling time points for the GA_3 treatment group and the control group were set as: May 25, June 29, and August 17. Mature leaves and apical buds were used in this experiment (randomly mixed with tissue samples with the same maturity in different directions and different heights of the plant, as a biological repeat). Samples of the treatment group and the control group were taken at the same time. The sample was placed in a clean centrifuge tube that has been marked, immediately frozen and stored in liquid nitrogen, and then stored in an −80°C refrigerator until use. Three independent biological replicates (the biological replicates were from separate plants) were performed for each treatment.

RNA Extraction and Sequencing

Total RNA were extracted following the manufacturer of the RNA Prep Pure Plant Kit (TIANGEN, China). Their purity and integrity were checked and assessed using the NanoPhotometer® spectrophotometer (IMPLEN, CA, United States) and RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, United States), respectively. Subsequently, total 1 μg RNA of each sample was used as input material for the RNA sample preparations. mRNA was purified using poly-T oligo-attached magnetic beads. First-strand and second-strand cDNA was synthesized according to the manufacturer of M-MuLV Reverse Transcriptase (RNase H^-), and DNA Polymerase I and RNase H^-, respectively. cDNA fragments of 250∼300 bp in length were selected and purified with AMPure XP system (Beckman Coulter, Beverly, MA, United States). Besides, the library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Novaseq platform and 150 bp paired-end reads were generated.

Genes Annotation and Differentially Expressed Genes Analysis

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. Clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. Q20, Q30, GC content, and sequence duplication levels in the clean data were calculated (**Supplementary Table 1**). All the downstream analyses were based on the clean data with high quality. Raw reads of the RNA-seq data are uploaded to Sequence Read Archive (SRA) database of NCBI with Bioproject ID number PRJNA729650.

Reference genome and gene model annotation files were downloaded from genome website (Su et al., 2021). The building of index of the reference genome, and the alignment between clean reads and reference genome all using Hisat2 (version 2.0.5). FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) of each gene was calculated by featureCounts (version 1.5.0) (Florea et al., 2013).

DEGs (Differentially expressed genes) were defined by DESeq2 R package (version 1.16.1) with an adjusted P-value < 0.05. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (Benjamini and Hochberg, 1995). GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways enrichment analysis of DEGs was implemented by the clusterProfiler R package. Volcano plots, Venn diagrams and heatmaps were drawed by TBtools (Chen C. et al., 2020). WGCNA (weighted gene co-expression network analysis) was performed in R with the WGCNA package (Langfelder and Horvath, 2008) and visualized the networks by Cytoscape (version 3.8.2) (Shannon et al., 2003).

RESULTS

Loquat Cannot Bloom After GA_3 Treatment

At 10–20 days after treating loquat plants with 300 mg L^{-1} GA_3, vigorous vegetative growth was observed, and the stems grew rapidly. In the beginning of September when obvious inflorescence could be observed in the control group, plants in the GA_3 treatment group were still in vigorous vegetative growth (**Figures 1A,B**).

Summary Statistics of Transcriptome Sequencing

Transcriptome sequencing results showed that 41.68–56.71 Mb clean reads were obtained from the 18 samples in the control group; 41.39–61.024 Mb clean reads were obtained from the 18 samples in the GA_3 treatment group. The GC content of the GA_3 treatment group and the control group was similar, ranging from 45.59 to 47.08% (**Supplementary Table 1**). For 36 samples, 96.70% of the bases had a quality score greater than 20, and Q30 $\geq$ 91.39%. The sequencing data are of high quality and can meet the requirements of subsequent analysis.

Selection of DEGs in GA_3 Treatment Group and Control Group

Compared with the control group, there were 1,901 down-regulated DEGs and 1,268 up-regulated DEGs in the buds 7 days after GA_3 treatment on May 25th (**Figure 2** and **Supplementary Tables 2-1, 2**); 1,867 down-regulated DEGs and 2,003 up-regulated DEGs in the buds on June 29th (**Figure 2** and **Supplementary Tables 2-3, 4**); 5,172 down-regulated DEGs and 5,383 up-regulated DEGs in the buds on August 17th (**Figure 2** and **Supplementary Tables 2-5, 6**). In leaves, compared with the control group, there were 7,052 down-regulated DEGs and

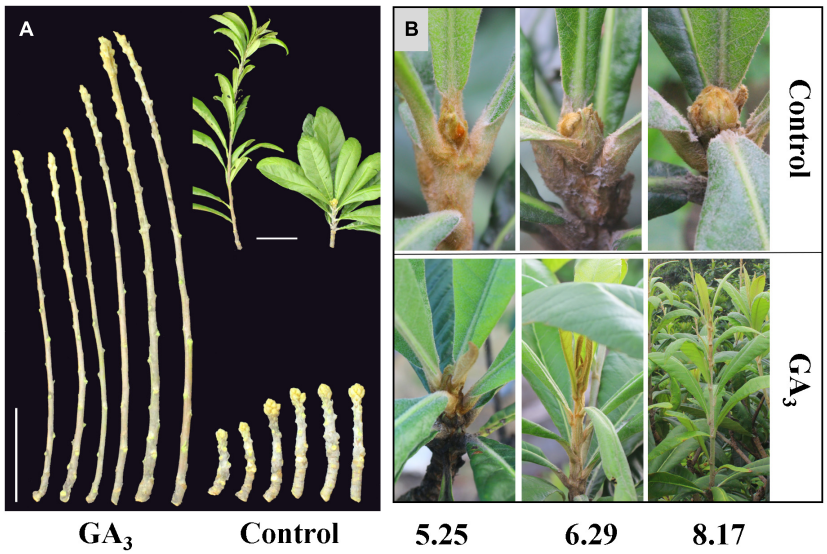


FIGURE 1 | The phenotype of loquat in the GA_3 treatment group and the control group. **(A)** GA_3 treatment promotes the growth of stems, bars = 10 cm; **(B)** Phenotypes of samples used for transcriptome sequencing at different stages.

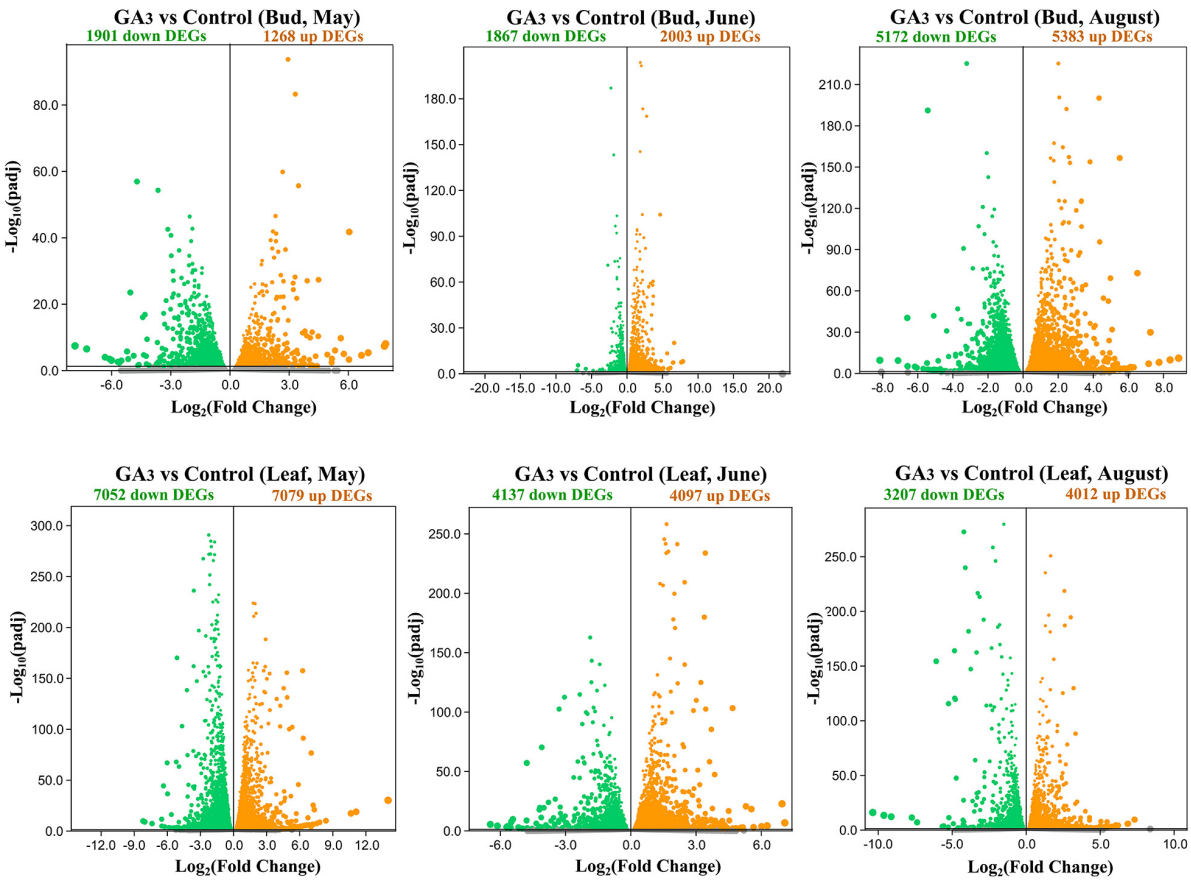


FIGURE 2 | Volcano map of DEGs in the buds and leaves between the GA_3 treatment group and the control group. Orange dots indicate up-regulated DEGs, and green dots indicate down-regulated DEGs, and gray dots indicate genes that were not differential expressed.

7,079 up-regulated DEGs in the leaves 7 days after GA_3 treatment on May 25th (**Figure 2** and **Supplementary Tables 3-1, 2**); 4,137 down-regulated DEGs and 4,097 up-regulated DEGs in the leaves on June 29th (**Figure 2** and **Supplementary Tables 3-3, 4**); 3,207 down-regulated DEGs and 4,012 up-regulated DEGs in the leaves on August 17th (**Figure 2** and **Supplementary Tables 3-5, 6**).

The results showed that compared with the control group, the number of DEGs in the apical buds of the GA_3 treatment group increased rapidly from June 29th to August 17th (**Figure 2**). It implies that during this period, GA_3 regulates the flower bud differentiation of loquat by up-regulating or down-regulating the expressions of a large number of flowering-related genes. In comparison, highly number of DEGs were expressed in May in leaves before the bud differentiation of loquat (**Figure 2**).

Functional Annotation and Expression Patterns of DEGs

In order to explore how loquat responds to GA_3 signals, GO and KEGG enrichment analysis were performed on DEGs the compare between GA_3 treatment and control of buds and leaves. GO enrichment showed that DEGs of buds mainly involved in biological signal binding and catalytic activity, such as heme binding, tetrapyrrole binding, hydrolase activity, acting on glycosyl bonds, etc. (**Figure 3** and **Supplementary Tables 4-1, 2, 3**). In leaves, DEGs mainly involved in metabolism, transcriptional activity, and biosynthesis, such as peptide metabolic process, amide biosynthetic process, peptide biosynthetic process, nucleic acid binding transcription factor activity, transcription factor activity (sequence-specific DNA binding), etc. (**Supplementary Figure 1** and **Supplementary Tables 4-4, 5, 6**).

The KEGG annotation shows that the 841 DEGs in the buds 7 days after GA_3 treatment were enriched in 110 KEGG pathways, among which plant hormone signal transduction (47 genes, 5.59%), phenylpropanoid biosynthesis (42 genes, 4.99%), and amino acids biosynthesis (35 genes, 4.16%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-1**). Total of 1,278 DEGs in the buds on June 29th were enriched in 115 KEGG pathways, among which plant hormone signal transduction (70 genes, 5.48%), protein processing in endoplasmic reticulum (61 genes, 4.77%), and plant-pathogen interaction (47 genes, 3.68%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-2**). Total of 3,327 DEGs in the buds on August 17th were enriched in 120 KEGG pathways, among which carbon metabolism (125 genes, 3.76%), ribosome (125 genes, 3.76%), and plant hormone signal transduction (124 genes, 3.73%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-3**).

In leaves, KEGG annotation shows that the 4,627 DEGs in the leaves on May 25th were enriched in 120 KEGG pathways, among which ribosome (302 genes, 6.53%), carbon metabolism (189 genes, 4.08%), biosynthesis of amino acids (153, 3.31%), and plant hormone signal transduction (137 genes, 2.96%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-4**). A total of 3,059 DEGs in the leaves on June 29th were enriched in 120 KEGG pathways, among which ribosome (171 genes, 5.59%), carbon metabolism (136 genes, 4.45%), biosynthesis of amino acids (135 genes, 4.41%), and plant hormone signal transduction (89 genes, 2.91%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-5**). 2,395 DEGs in the leaves on August 17th were enriched in 119 KEGG pathways, among which carbon metabolism (105 genes, 4.38%), plant hormone signal transduction (93 genes, 3.88%), biosynthesis of amino acids (93 genes, 3.88%), and ribosome (84 genes, 3.51%) were significantly enriched pathways (**Figure 3** and **Supplementary Table 5-6**).

The results showed that DEGs were mainly enriched in the plant hormone signal transduction pathway after GA_3 treatment, which also indicated that after GA_3 treatment, loquats mainly responded to GA_3 signals through these DEGs, and ultimately participated in the regulation of loquat flower bud differentiation.

In order to further understand the expression patterns of genes related to flower bud differentiation, we performed a cluster analysis on the selected DEGs. It was revealed that after GA_3 treatment, 111 DEGs were down-regulated in the three stages of buds (**Figures 4A,C** and **Supplementary Table 6-1**) and 151 DEGs were up-regulated (**Figures 4B,D** and **Supplementary Table 6-2**). In addition, we found in the leaves that 563 DEGs were down-regulated in the three stages after GA_3 treatment (**Figure 4A** and **Supplementary Table 6-3**), and 506 DEGs were up-regulated (**Figure 4B** and **Supplementary Table 6-4**).

Through the analysis of all down-regulated and up-regulated DEGs in buds with annotation results, it was found that 44 down-regulated DEGs belonged to 27 gene families (**Supplementary Figure 2A**); while the 58 up-regulated DEGs belonged to 28 gene families (**Supplementary Figure 2B**) including AP2, bZIP, F-box, MYB, WRKY and other gene family genes. The flower buds of loquat cannot differentiate after GA_3 treatment, which also implies that the down-regulated DEGs after treatment are possibly positive-regulatory genes involved in loquat flowering, and these up-regulated DEGs may be negative-regulatory genes for loquat flowering.

Expression Analysis of DELLA-Mediated GA Signal Regulatory Network in Loquat Flowering

In *Arabidopsis thaliana*, the GA pathway genes involved in the regulation of flower formation mainly include *DELLA, CO, MYC3, WRKY75, WRKY12, WRKY13, SOCl, SPL, FT, AGL24, LFY, FUL, AP1*, etc. (Bao et al., 2020). Thirty-three homologous genes in loquat were obtained through sequence alignment by blast the transcripts of loquat and coding sequence of *Arabidopsis thaliana* genome (**Figure 5** and **Supplementary Table 7**). Based on the expression patterns of these homologous genes in the GA_3 treatment group and the control group, thirteen candidate genes, including *EjWRKY75* (Eri011414), EjFT (Eri036481), EjSOC1 (Eri012338, Eri023104), EjAGL24 (Eri026753), EjSPL (Eri001949, Eri003494), EjLFY (Eri007397, Eri022269), EjFUL (Eri009416, Eri033768), EjAP1 (Eri000407, Eri030184), were highly expressed in control group than GA3 treatment group in buds, and may be involved in the positive regulation of flower development.

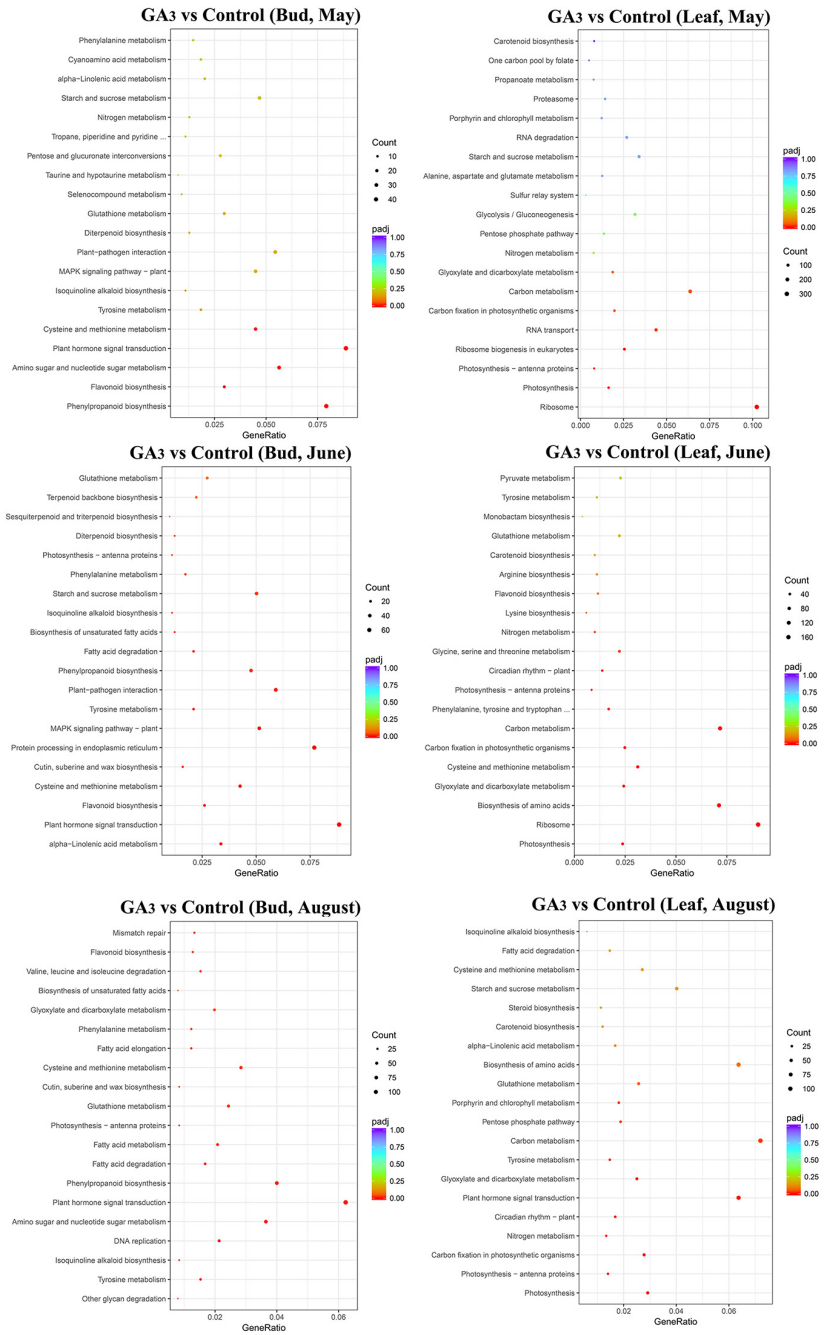


FIGURE 3 | KEGG analysis of DEGs in different tissues of GA_3 treatment group and control group at different periods.

Besides, eleven candidate genes, including EjDELLA (Eri016831, Eri029753, Eri030473, Eri031405, Eri038200, Eri038624), EjMYC3 (Eri010051, Eri034229), EjWRKY12 (Eri034804), EjWRKY13 (Eri012544, Eri035481), were highly expressed in GA_3 treatment group than control group in buds, and may be involved in the negative regulation of flower development. However, we found that the expressions of *EjCOs* in leaves did not decrease but increased after GA_3 treatment. It suggests that the *EjCOs* in loquat may be mainly regulated by photoperiod, rather than regulating loquat flower development by responding to GA_3 signals.

Co-expression Network Analysis of Genes Involved in the Regulation of Flower Formation

EjSOC1s play an active role in the flowering process of loquat, *EjAP1s* and *EjLFYs* can be used as marker genes for loquat

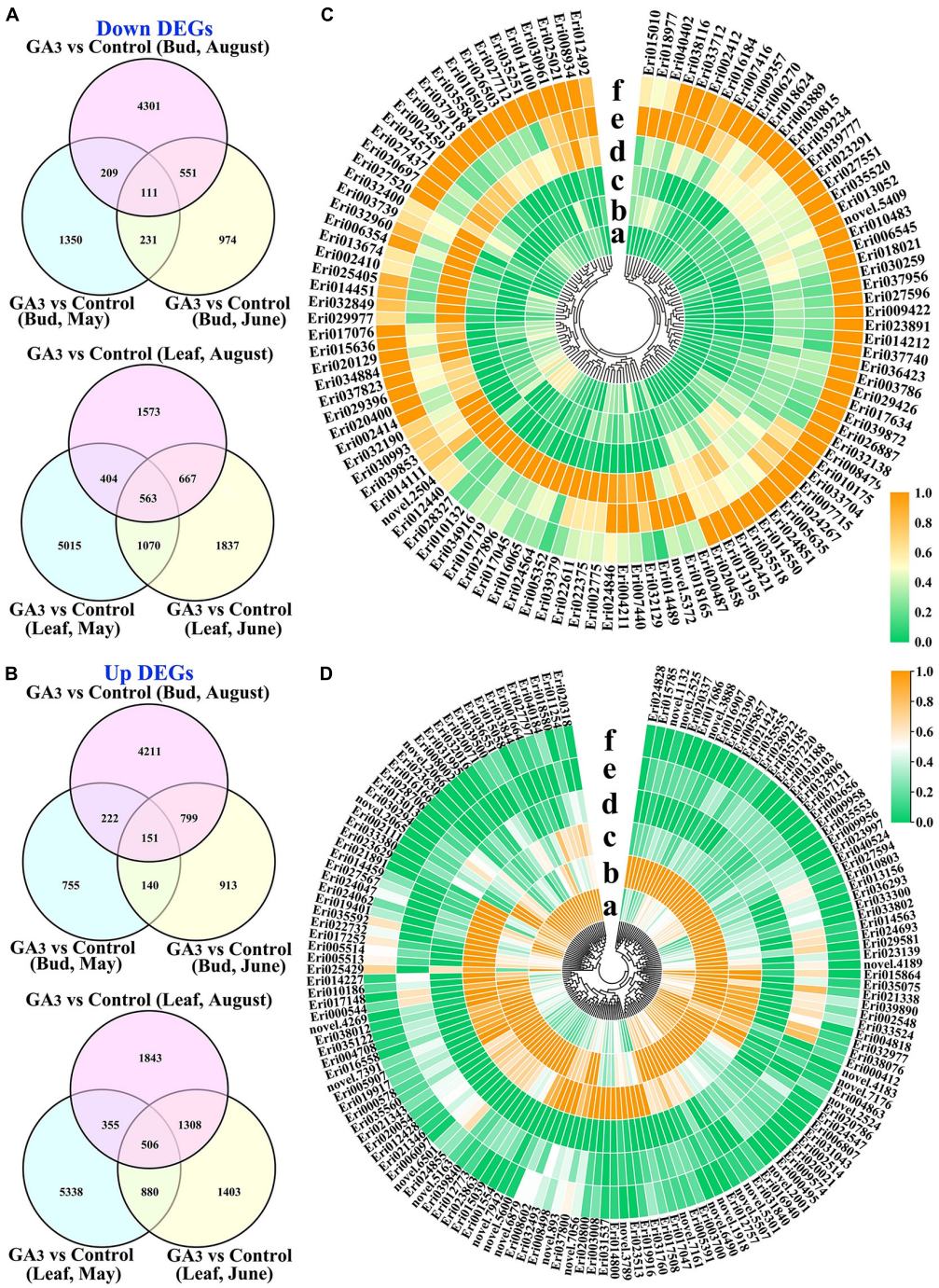


FIGURE 4 | DEGs analysis of loquat leaves and buds transcriptome datasets. **(A)** Venn diagrams of DEGs that are down-regulated in leaves and buds at different stages; **(B)** Venn diagrams of DEGs that are up-regulated in leaves and buds at different stages; **(C)** The expression patterns of 111 down-regulated DEGs in buds; **(D)** The expression patterns of 151 up-regulated DEGs in buds. a, b, and c respectively represent the material on May 25th, June 29th, and August 17th after GA3 treatment; d, e, and f represent the control group materials on May 25th, June 29th, and August 17th, respectively.

flowering regulation (Jiang et al., 2019c). Our experimental results also further confirmed this conclusion. In Arabidopsis, SOC1 or LFY activates the expressions of floral meristem identity genes *LFY, AP1,* and *FUL* to initiate flower bud differentiation (Blazquez et al., 1998; Blazquez and Weigel, 2000).

In order to further understand the regulatory network of loquat flowering genes and the regulatory relationship of flowering-related genes, WGCNA analysis was carried out to investigate the co-expression networks of DEGs in the transcriptome data, in which all the co-expressed genes were

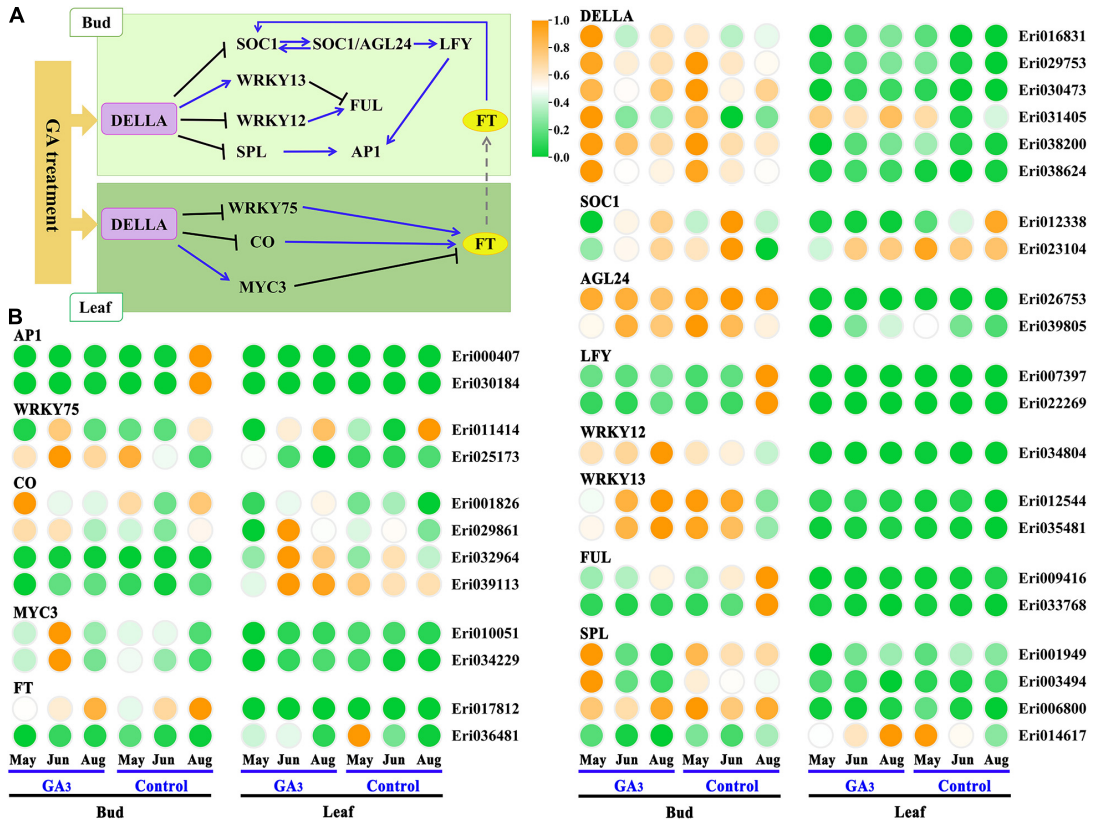


FIGURE 5 | Expression analysis of homologous genes from DELLA-mediated GA signal regulatory pathway in loquat. **(A)** DELLA-mediated GA signal regulation pathway in *Arabidopsis*. **(B)** The expression pattern of homologous genes from the GA pathway in loquat leaves and buds.

connected to each other with varying association strengths. DEGs are clustered into 27 modules according to the expression patterns (**Figure 6A**), interestingly, *EjSOC1-2* (Eri023104), EjLFY-1 (Eri007397), EjLFY-2 (Eri022269), EjFUL-1 (Eri009416), EjFUL-2 (Eri033768), EjAP1-1 (Eri000407) and EjAP1-2 (Eri030184) were assigned to the brown module, except for EjSOC1-1 (Eri012338) in the yellow module (weight < 0.1). Further analysis of the seven genes in the brown module revealed that 330 genes were co-expressed with them (**Figure 6B** and **Supplementary Table 8**), belonging to 74 gene families, including Cyclin_C, Histone, Kinesin, Lipase_GDSL, MYB, P450, Pkinase, Tubulin, ZF-HD_dimer gene family (**Figure 6C**). It showed that their expression patterns are similar to that of *EjSOC1-2, EjLFYs, EjFULs, EjAP1s*, and they are all inhibited by GA_3 treatment. It implied that these genes may play a positive regulatory role in the flowering of loquat. Therefore, these data provide new directions and useful candidate genes for research on loquat flowering regulation.

DISCUSSION

The Rosaceae family contains many fruit crop species, such as apple, pear, peach, strawberry and loquat. However, the unique flowering time of autumn flowering and spring fruit ripening of cultivated loquat suggests that it may have a flowering regulation system different from other woody plants in Rosaceae family during the evolutionary process. Although the flowering habit of loquat is very special, there are few reports about it. So far, several flowering-related genes have been cloned from cultivated loquat, including *EjAP1, EjAP3, EjFT, EjLFY, EjSOC1, EjSVP, EjSPL, EjFRI* and *EjTFL1* (Esumi et al., 2005; Liu et al., 2013, 2017; Reig et al., 2017; Jiang et al., 2019a,b,c; Chen W. et al., 2020). *EdCO, EdGI, EdFT* and *EdFD* have been cloned from wild loquat *Eriobotrya deflexa* Nakai forma *koshunensis* (Zhang L. et al., 2016; Zhang et al., 2019). However, the loquat floral regulation network is still unclear. This study enriched the loquat floral regulation network resources, which can provide an important reference for further analysis of the molecular mechanism of loquat flowering, and also provide a theoretical basis for the later research on the floral regulation of Rosaceae.

The two hormone systems "gibberellin" and "florigen" (FT) play a key role in crop yield and quality (Eshed and Lippman, 2019). Previous studies have shown that GA_3 treatment can inhibit the flower bud induction of nectarine, sweet cherry, mango, citrus and apple (García-Pallas et al., 2001; Lenahan et al., 2006; Nakagawa et al., 2012; Goldberg-Moeller et al., 2013; Zhang S. et al., 2016). The *miFT* in mangoes can regulate the flowering by responding to GA signals (Nakagawa et al., 2012); in citrus (*Citrus reticulata* Blanco × *Citrus temple* Hort.

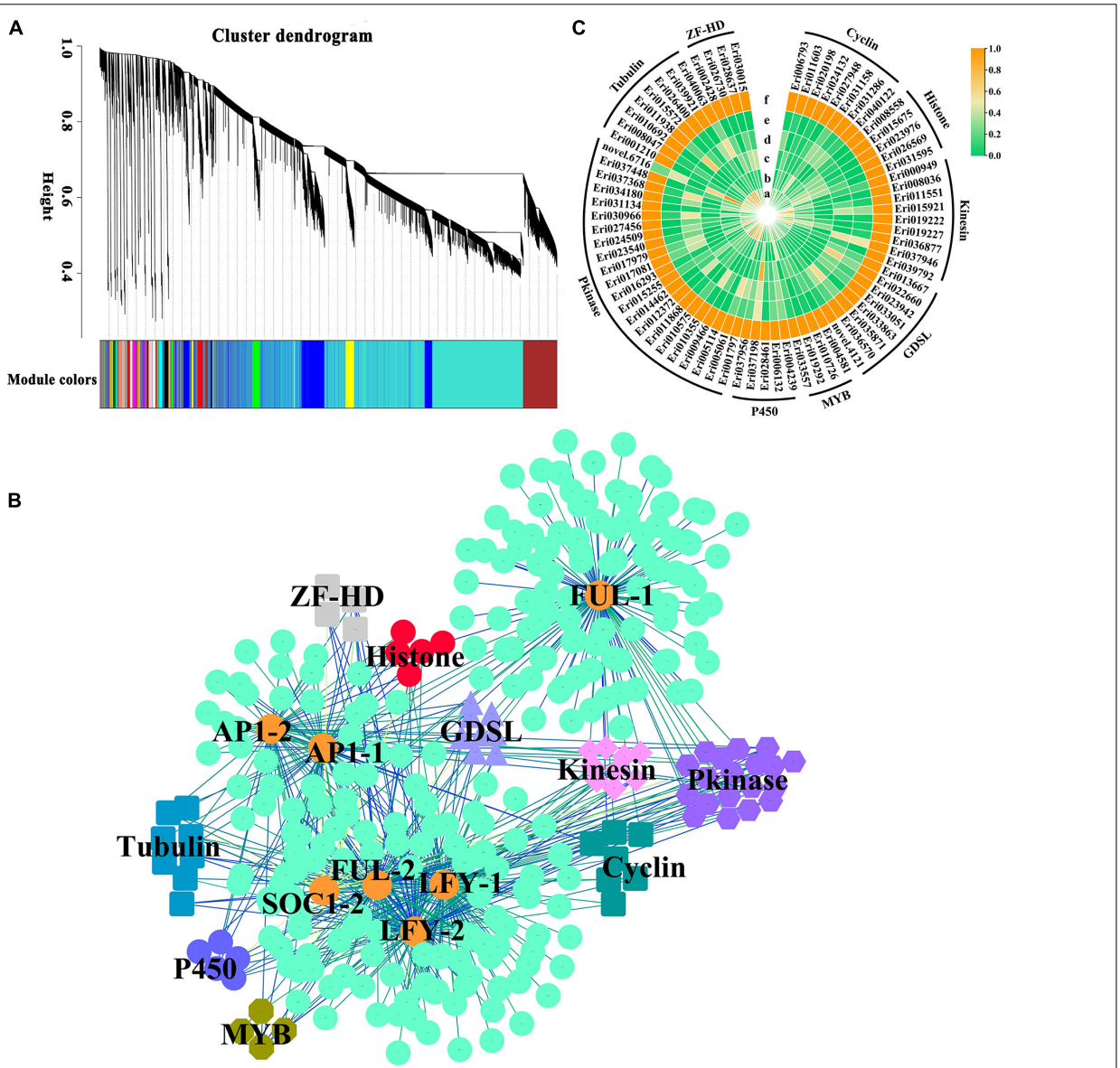


FIGURE 6 | WGCNA co-expression analysis of homolog genes from DELLA-mediated GA signal regulatory pathway in loquat. **(A)** Module trait relationships in the GA_3 treatment group and the control group. **(B)** The co-expression network connecting the key genes in DELLA-mediated GA signal regulatory pathway with candidate genes in brown module. Expression associations between homolog genes (orange circle) from GA pathway and candidate genes (other shapes except orange circle) are indicated with colored lines according to weight. **(C)** The expression pattern of candidate genes in loquat buds.

ex Y. Tanaka), *FT, AP1* and a few flower-organ-identity genes are inhibited by GA, but GA promotes the expression of *LFY* (Goldberg-Moeller et al., 2013); The expression levels of *MdSPLs, MdFT, MdSOC1* and *MdAP1* genes in apples are all inhibited by the GA_3 treatment (Zhang S. et al., 2016). In this study, we also found similar conclusions to that from these woody fruit trees. For example, GA_3 treatment inhibited the flowering of loquat, and the expressions of genes such as *EjFT, EjSOC1, EjSPL*, and *EjAP1* were inhibited. But different from the expression in citrus, the expression of *EjLFY* in loquat was inhibited by GA_3. Our results also showed that some genes are different from the expression patterns in model plants, such as *EjMYC3*, which mainly plays a regulatory role in Arabidopsis leaves, while the difference in expression of *EjMYC3* in loquat after GA_3 treatment mainly occurs in buds. There is little difference in expression in leaves, suggesting that it mainly plays a role in regulating flowering in buds. In addition, we found that the expressions of several *EjCO* genes did not decrease but increased after GA_3 treatment. We speculated that it may be mainly through response to photoperiod signals to regulate flower formation, rather than GA signals.

In this study, by analyzing all the down-regulated and up-regulated DEGs in the apical buds of different periods,

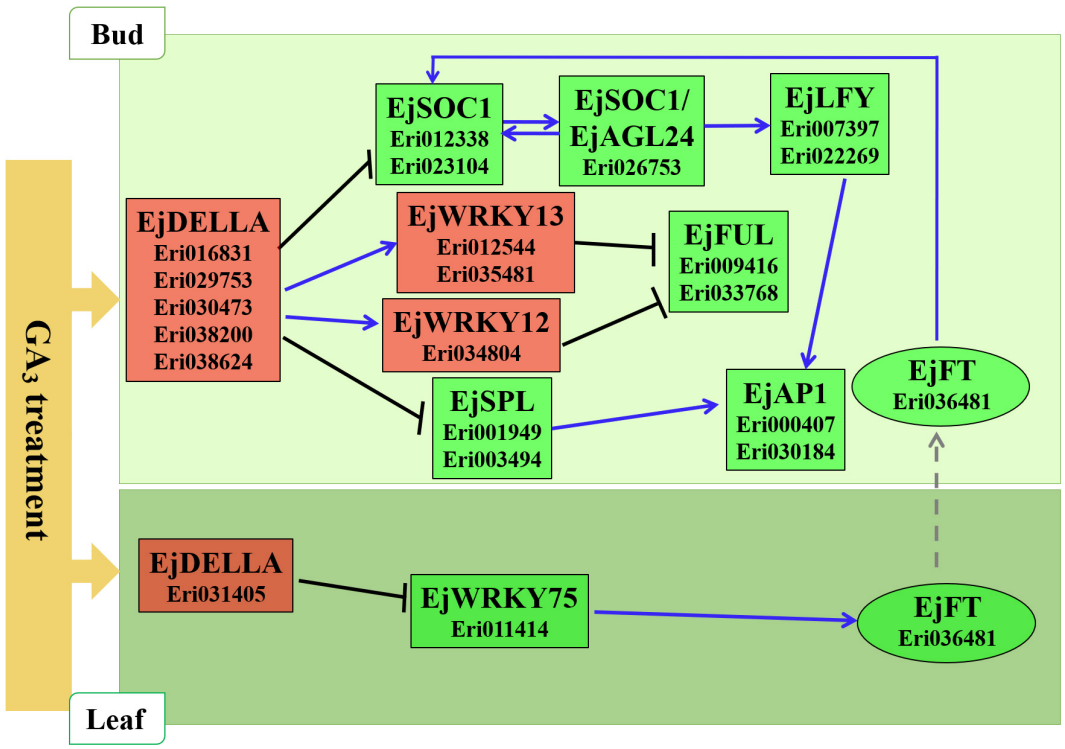


FIGURE 7 | Hypothetical model of loquat flowering regulation.

some candidate genes that may participate in loquat flower bud differentiation were screened, including AP2, bZIP, F-box, MYB, WRKY and other gene families (**Supplementary Figure 2**). In addition, through WGCNA analysis, candidate genes co-expressed with *EjSOC1-2, EjLFYs, EjFULs, EjAP1s*, and possibly involved in loquat flower bud differentiation, including Cyclin_C, Histone, Kinesin, Lipase_GDSL, MYB, P450, Pkinase, Tubulin, ZF- HD_dimer gene family. The discovery of these candidate genes has brought great convenience to the subsequent study of loquat flower formation. Future work will focus on verifying and analyzing the functions and mechanisms of these candidate genes in the formation of loquat flowers.

Referring to the DELLA-mediated GA signal regulation network diagram in Arabidopsis (**Figure 6A**), we constructed a hypothetical model for loquat flowering regulation network based on the expression patterns of related homologous genes in loquat (**Figure 7**). As follows: In leaves, GA_3 promotes the expression of DELLA family genes (Eri031405) to inhibit the expression of EjWRKY75 (Eri011414), and then EjWRKY75 down-regulates the expression of EjFT (Eri036481), and finally the EjFT protein is transported to SAMs to take effect. In SAMs, GA3 promotes the expression of DELLA family genes (Eri016831, Eri029753, Eri030473, Eri038200, and Eri038624). On the one hand, DELLA family genes inhibit the expression of EjFUL (Eri009416, Eri033768) by promoting the expression of EjWRKY12 (Eri034804) and EjWRKY13 (Eri012544, Eri035481). On the other hand, DELLA family genes inhibited the expression of flower-specific genes EjLFY (Eri007397, Eri022269) and EjAP1 (Eri000407, Eri030184) by inhibiting the expression of EjSOC1 (Eri012338, Eri023104), EjAGL24 (Eri026753) and EjSPL (Eri001949, Eri003494). Finally, the flower development of loquat was inhibited.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI using accession number PRJNA729650.

AUTHOR CONTRIBUTIONS

YJ, YL, and FX designed the research. YJ and FX mainly performed the research. YG, JP, WS, YY, XY, CZ, MW, and SL contributed reagents, materials, and analysis tools. YJ, FX, and ZP wrote the manuscript. ZP and FX revised and approved the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Natural Science Foundation of Guangdong Province (2021A1515011048), Young Innovative Talents Project in Ordinary Colleges and Universities in Guangdong Province (2020KQNCX075), University-level scientific research project of Shaoguan University (SZ2019ZK11),

Shaoguan City Science and Technology Planning Project (Social Development Direction-Supporting Scientific Research Workers Project) (200811094530739), and Collaborative innovation project from Fujian Provincial People's Government and Chinese Academy of Agricultural Sciences (KXXYJBG2021006).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.703688/full#supplementary-material

REFERENCES

Achard, P., Gusti, A., Cheminant, S., Alioua, M., Dhondt, S., Coppens, F., et al. (2009). Gibberellin signaling controls cell proliferation rate in Arabidopsis. *Curr. Biol.* 19, 1188–1193. doi: 10.1016/j.cub.2009.05.059

Amasino, R. (2010). Seasonal and developmental timing of flowering. *Plant J.* 61, 1001–1013. doi: 10.1111/j.1365-313X.2010.04148.x

Bao, S., Hua, C., Huang, G., Cheng, P., Gong, X., Shen, L., et al. (2019). Molecular basis of natural variation in photoperiodic flowering responses. *Dev Cell.* 50, 90–101.e3. doi: 10.1016/j.devcel.2019.05.018

Bao, S., Hua, C., Shen, L., and Yu, H. (2020). New insights into gibberellin signaling in regulating flowering in Arabidopsis. *J. Integr. Plant Biol.* 62, 118–131. doi: 10.1111/jipb.12892

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Blazquez, M. A., Green, R., Nilsson, O., Sussman, M. R., and Weigel, D. (1998). Gibberellins promote flowering of arabidopsis by activating the LEAFY promoter. *Plant Cell.* 10, 791–800.

Blazquez, M. A., and Weigel, D. (2000). Integration of floral inductive signals in Arabidopsis. *Nature* 404, 889–892. doi: 10.1038/35009125

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, W., Wang, P., Wang, D., Shi, M., Xia, Y., He, Q., et al. (2020). EjFRI, FRIGIDA (FRI) ortholog from *Eriobotrya japonica*, delays flowering in Arabidopsis. *Int J Mol Sci.* 21:1087. doi: 10.3390/ijms21031087

Davis, S. J. (2009). Integrating hormones into the floral-transition pathway of *Arabidopsis thaliana*. *Plant Cell Environ.* 32, 1201–1210. doi: 10.1111/j.1365-3040.2009.01968.x

de Lucas, M., Daviere, J. M., Rodriguez-Falcon, M., Pontin, M., Iglesias-Pedraz, J. M., Lorrain, S., et al. (2008). A molecular framework for light and gibberellin control of cell elongation. *Nature* 451, 480–484. doi: 10.1038/nature06520

Eshed, Y., and Lippman, Z. B. (2019). Revolutions in agriculture chart a coursresee for targeted breeding of old and new crops. *Science* 366:eaax0025. doi: 10.1126/science.aax0025

Esumi, T., Tao, R., and Yonemori, K. (2005). Isolation of LEAFY and TERMINAL FLOWER 1 homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae. *Sex. Plant Reprod.* 17, 277–287. doi: 10.1007/s00497-004-0239-3

Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res.* 2:188. doi: 10.12688/f1000research.2-188.v2

García-Pallas, I., Val, J., and Blanco, A. (2001). The inhibition of flower bud differentiation in 'Crimson Gold' nectarine with GA_3 as an alternative to hand thinning. *Sci. Hortic.* 90, 265–278. doi: 10.1016/S0304-4238(01)00229-1

Gocal, G. F., Sheldon, C. C., Gubler, F., Moritz, T., Bagnall, D. J., MacMillan, C. P., et al. (2001). GAMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. *Plant Physiol.* 127, 1682–1693.

Goldberg-Moeller, R., Shalom, L., Shlizerman, L., Samuels, S., Zur, N., Ophir, R., et al. (2013). Effects of gibberellin treatment during flowering induction period on global gene expression and the transcription of flowering-control genes in Citrus buds. *Plant Sci.* 198, 46–57. doi: 10.1016/j.plantsci.2012.09.012

Hisamatsu, T., and King, R. W. (2008). The nature of floral signals in Arabidopsis. II. Roles for FLOWERING LOCUS T (FT) and gibberellin. *J. Exp. Bot.* 59, 3821–3829. doi: 10.1093/jxb/ern232

Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L., and Yu, H. (2014). Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. *Nat. Commun.* 5:4601. doi: 10.1038/ncomms5601

Jiang, Y., Peng, J., Wang, M., Su, W., Gan, X., Jing, Y., et al. (2019a). The Role of EjSPL3, EjSPL4, EjSPL5, and EjSPL9 in regulating flowering in loquat (*Eriobotrya japonica* Lindl.). *Int. J. Mol. Sci.* 21:248. doi: 10.3390/ijms21010248

Jiang, Y., Peng, J., Zhang, Z., Lin, S., Lin, S., and Yang, X. (2019b). The role of EjSVPs in flower initiation in *Eriobotrya japonica*. *Int J Mol Sci.* 20:5933. doi: 10.3390/ijms20235933

Jiang, Y., Peng, J., Zhu, Y., Su, W., Zhang, L., Jing, Y., et al. (2019c). The role of EjSOC1s in flower initiation in *Eriobotrya japonica*. *Front. Plant Sci.* 10:253. doi: 10.3389/fpls.2019.00253

Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., et al. (1999). Activation tagging of the floral inducer FT. *Science* 286, 1962–1965.

Kurokura, T., Mimida, N., Battey, N. H., and Hytonen, T. (2013). The regulation of seasonal flowering in the Rosaceae. *J. Exp. Bot.* 64, 4131–4141. doi: 10.1093/jxb/ert233

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

Lee, H., Suh, S. S., Park, E., Cho, E., Ahn, J. H., Kim, S. G., et al. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev.* 14, 2366–2376. doi: 10.1101/gad.813600

Lenahan, O. M., Whiting, M. D., and Elfving, D. C. (2006). Gibberellic acid inhibits floral bud induction and improves 'Bing' sweet cherry fruit quality. *Hortscience* 41, 654–659.

Li, W., Wang, H., and Yu, D. (2016). Arabidopsis WRKY transcription factors WRKY12 and WRKY13 oppositely regulate flowering under short-day conditions. *Mol. Plant.* 9, 1492–1503. doi: 10.1016/j.molp.2016.08.003

Lin, S. (2007). World loquat production and research with special reference to China. *Proc. Sec. Int. Symp. Loquat*. 750, 37–43. doi: 10.17660/ActaHortic.2007.750.2

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, Y., Zhao, Q., Meng, N., Song, H., Li, C., Hu, G., et al. (2017). Over-expression of EjLFY-1 leads to an early flowering habit in strawberry (*Fragaria x ananassa*) and its asexual progeny. *Front. Plant Sci.* 8:496. doi: 10.3389/fpls.2017.00496

Liu, Y. X., Song, H. W., Liu, Z. L., Hu, G. B., and Lin, S. Q. (2013). Molecular characterization of loquat *EjAP1* gene in relation to flowering. *Plant Growth Regul.* 70, 287–296. doi: 10.1007/s10725-013-9800-0

Moon, J., Lee, H., Kim, M., and Lee, I. (2005). Analysis of flowering pathway integrators in Arabidopsis. *Plant Cell Physiol.* 46, 292–299. doi: 10.1093/pcp/pci024

Moon, J., Suh, S. S., Lee, H., Choi, K. R., Hong, C. B., Paek, N. C., et al. (2003). The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *Plant J.* 35, 613–623.

Nakagawa, M., Honsho, C., Kanzaki, S., Shimizu, K., and Utsunomiya, N. (2012). Isolation and expression analysis of FLOWERING LOCUS T-like and gibberellin metabolism genes in biennial-bearing mango trees. *Sci. Hortic.* 139, 108–117. doi: 10.1016/j.scienta.2012.03.005

Osnato, M., Castillejo, C., Matias-Hernandez, L., and Pelaz, S. (2012). TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in Arabidopsis. *Nat. Commun.* 3:808. doi: 10.1038/ncomms1810

Peng, J., Li, W., Yuan, Y., Han, Z., Cao, Y., Shahid, M. Q., et al. (2021). Removal of the Main Inflorescence to Induce Reflowering of Loquat. *Hortic. Plant J.* in press. doi: 10.1016/j.hpj.2021.03.009

Reig, C., Gil-Munoz, F., Vera-Sirera, F., Garcia-Lorca, A., Martinez-Fuentes, A., Mesejo, C., et al. (2017). Bud sprouting and floral induction and expression of FT in loquat [*Eriobotrya japonica* (Thunb.) Lindl.]. *Planta.* 246, 915–925. doi: 10.1007/s00425-017-2740-6

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

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. H., Lee, I., Lee, S. Y., Imaizumi, T., and Hong, J. C. (2012). CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of FLOWERING LOCUS T in photoperiodic flowering in Arabidopsis. *Plant J.* 69, 332–342. doi: 10.1111/j.1365-313X.2011.04793.x

Su, W., Jing, Y., Lin, S., Yue, Z., Yang, X., Xu, J., et al. (2021). Polyploidy underlies co-option and diversification of biosynthetic triterpene pathways in the apple tribe. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2101767118. doi: 10.1073/pnas.2101767118

Teotia, S., and Tang, G. (2015). To bloom or not to bloom: role of microRNAs in plant flowering. *Mol. Plant.* 8, 359–377. doi: 10.1016/j.molp.2014.12.018

Wang, H., Pan, J., Li, Y., Lou, D., Hu, Y., and Yu, D. (2016). The DELLA-CONSTANS transcription factor cascade integrates gibberellic acid and photoperiod signaling to regulate flowering. *Plant Physiol.* 172, 479–488. doi: 10.1104/pp.16.00891

Xu, F., Li, T., Xu, P. B., Li, L., Du, S. S., Lian, H. L., et al. (2016). DELLA proteins physically interact with CONSTANS to regulate flowering under long days in Arabidopsis. *FEBS Lett.* 590, 541–549. doi: 10.1002/1873-3468.12076

Zhang, L., Chen, L., and Yu, D. (2018). Transcription factor WRKY75 interacts with DELLA proteins to affect flowering. *Plant Physiol.* 176, 790–803. doi: 10.1104/pp.17.00657

Zhang, L., Jiang, Y., Zhu, Y., Su, W., Long, T., Huang, T., et al. (2019). Functional characterization of GI and CO homologs from *Eriobotrya deflexa* Nakai forma *koshunensis*. *Plant Cell Rep.* 38, 533–543. doi: 10.1007/s00299-019-02384-3

Zhang, L., Yu, H., Lin, S., and Gao, Y. (2016). Molecular characterization of FT and FD homologs from *Eriobotrya deflexa* Nakai forma *koshunensis*. *Front. Plant Sci.* 7:8. doi: 10.3389/fpls.2016.00008

Zhang, S., Zhang, D., Fan, S., Du, L., Shen, Y., Xing, L., et al. (2016). Effect of exogenous GA_3 and its inhibitor paclobutrazol on floral formation, endogenous hormones, and flowering-associated genes in 'Fuji' apple (*Malus domestica* Borkh.). *Plant Physiol. Biochem.* 107, 178–186. doi: 10.1016/j.plaphy.2016.06.005

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 handling editor declared a past collaboration with one of the author YJ.

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fgene.2021.727674

Identification, Molecular Characteristics, and Evolution of *GRF* Gene Family in Foxtail Millet (*Setaria italica* L.)

Huilong Chen[†] *and Weina Ge*[†]*

School of Life Science, North China University of Science and Technology, Tangsha, China

Growth-regulating factor (GRF) is a multigene family that plays a vital role in the growth and development of plants. In the past, the GRF family of many plants has been studied. However, there is not a report about identification and evolution of GRF in foxtail millet (*Setaria italia*). Here, we identified 10 *GRF* genes in foxtail millet. Seven (70.00%) were regulated by Sit-miR396, and there were 19 optimal codons in *GRF*s of foxtail millet. Additionally, we found that WGD or segmental duplication have affected *GRFs* in foxtail millet between 15.07 and 45.97 million years ago. Regarding the *GRF* gene family of land plants, we identified a total of 157 *GRF* genes in 15 representative land plants. We found that *GRF* gene family originated from Group E, and the *GRF* gene family in monocots was gradually shrinking. Also, more loss resulted from the small number of *GRF* genes in lower plants. Exploring the evolution of *GRF* and functional analysis in the foxtail millet help us to understand *GRF* better and make a further study about the mechanism of *GRF*. These results provide a basis for the genetic improvement of foxtail millet and indicate an improvement of the yield.

Keywords: growth-regulating factors, foxtail millet, structure, expression, loss

OPEN ACCESS

Edited by:
Million Tadege,
Oklahoma State University,
United States

Reviewed by:
Ghulam Qanmber,
Cotton Research Institute (CAAS),
China
Lifang Niu,
Biotechnology Research Institute
(CAAS), China

***Correspondence:**
Weina Ge
gwn-06@163.com

[†]*These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *19 June 2021*
Accepted: *20 December 2021*
Published: *03 February 2022*

Citation:
Chen H and Ge W (2022) Identification, Molecular Characteristics, and Evolution of GRF *Gene Family in Foxtail Millet (*Setaria italica *L.). Front. Genet. 12:727674. doi: 10.3389/fgene.2021.727674*

INTRODUCTION

Growth-regulating factor (GRF) is a plant-specific transcription factor that plays an important role in plant growth and development. The first member of the identified GRF family is OsGRF1, which plays a regulatory role in gibberellin (GA)-induced stem elongation (van der Knaap et al., 2000). GRF transcription factor has two conserved domains in its N-terminal region: QLQ (Gln, Leu, and Gln) and WRC (Trp, Arg, and Cys) (Rodriguez et al., 2016). The QLQ domain interacts with GRF interacting factor (GIF), and the resulting complex acts as a transcriptional co-activator (Wang et al., 2014). The WRC domain consists of a functional nuclear localization signal and a DNA binding motif (zinc finger structure), which is mainly involved in DNA binding. The C-terminal of some GRF proteins also consists of other domains, including TQL (Thr, Gln, and Leu), GGPL (Gly, Gly, Pro, and Leu), and FFD (Phe, Phe, and Asp) (Cao et al., 2016).

The *GRF* gene family is a small family; therefore, the functions of each member of the GRF family in the studied species can be studied more comprehensively. Studies have found that *GRF* genes are often expressed strongly in actively growing and developing tissues, such as germinating seeds, ears, shoots, flower buds, and young leaves (Kim et al., 2003; Choi et al., 2004; Zhang et al., 2008; Wang et al., 2014; Zhang et al., 2017). In addition, studies have shown that most *GRF*s are regulated by miRNA396. For example, in *Arabidopsis*, seven miRNA396 target genes were predicted, and

six *AtGRF*s were confirmed in the experiment (Jones-Rhoades and Bartel, 2004; Liu et al., 2009). With the completion of many plant genome sequences, *GRF* family members of some plants have been studied, such as *Arabidopsis* (Kim et al., 2003), rice (Choi et al., 2004), maize (Zhang et al., 2008), *Brachypodium distachyon* (Filiz et al., 2014), *Brassica rapa* (Wang et al., 2014), *Brassica napus* (Ma et al., 2017), *Solanu lycopersicum* (Khatun et al., 2017), *Nicotiana tabacum* (Zhang et al., 2017), Cucurbitaceae (Baloglu, 2014), *Manihot esculenta* (Shang et al., 2018), apple (Zheng et al., 2018), mulberry (Rukmangada et al., 2018), and so on. However, studies on *GRF* genes in foxtail millet and the evolutionary trajectory of *GRF* genes have not been available.

Foxtail millet is one of the oldest food crops in many regions of the world, especially in China and India, where it is still widely cultivated as a staple food. Although the genome of foxtail millet is small, it has a high inbreeding rate, strong C4 photosynthesis, and high nutritional value, which is usually higher than other grains, containing a large number of minerals, such as essential amino acids, carbohydrates, and vitamins (Li and Brutnell, 2011; Pandey et al., 2013; Jia et al., 2013; Ji et al., 2015; Li et al., 2018). With the sequencing and continuous updating of the foxtail millet genome, now the foxtail millet genome is about 515 Mb (Bennetzen et al., 2012; Zhang et al., 2012; Han et al., 2014; Yang et al., 2020). Together with other gramineous plants, foxtail millet was affected by a whole-genome duplication or tetraploidy approximately 100 million years ago (Wang et al., 2015). This event resulted in thousands of duplicated genes in the existing genome, providing evolutionary power for genetic and functional innovation. Studying *GRF*s in foxtail millet helps to improve crop genetics and contributes to in-depth study of *GRF* function and food production. In this study, we conducted a series of informatics analysis on the exploration and functional prediction of *GRF* using a more comprehensive bioinformatics method to lay the foundation for further study of *GRF* functions.

MATERIALS AND METHODS

Acquisition of Members of the *GRF* Gene Family

We selected 20 plants (5 dicots, 7 monocots, 1 basal angiosperm, 1 Pteridophyta, 1 Bryophyta, and 5 green algae) for *GRF* evolution analysis, in which the genome-wide of *Aegilops tauschii* was obtained from the literature (Luo et al., 2017). The remaining 19 species were obtained from the JGI database (http://genome.jgi.doe.gov/) [*Arabidopsis thaliana Araport11*, *Carica papaya ASGPBv0.4*, *Populus trichocarpa v3.1*, *Vitis vinifera v2.1*, *Solanum lycopersicum ITAG3.2*, *Zea mays Ensembl-18*, *Sorghum bicolor Rio v2.1*, *Setaria italica v2.2*, *Brachypodium distachyon Bd21-3 v1.1*, *Hordeum vulgare r1*, *Oryza sativa v7.0*, *Amborella trichopoda v1.0*, *Selaginella moellendorfii v1.0*, *Physcomitrella patens v3.3*, *Chlamydomonas reinhardtii v5.6*, *Volvox carteri v2.1*, *Coccomyxa subellipsoidea C-169 v2.0*, *Micromonas* sp. *RCC299 v3.0,* and *Ostreococcus lucimarinus v2.0*]. We downloaded the WRC (PF08879) and QLQ (PF08880) domains from the Pfam database (Bateman et al., 2013). The HMMER (version 3.2.1) software (Mistry et al., 2013) was used to identify GRF candidate members in 20 species. In addition, we also used local BLAST to screen GRF family members of all species again. Finally, we used Pfam (http://pfam.xfam.org), CDD (https://www.ncbi.nlm.nih.gov/cdd), and SMART databases (http://smart.embl-heidelberg.de/) to confirm GRF members that contain WRC and QLQ domains.

Phylogenetic Analysis of GRF Family

Full-length amino acid sequences of GRF in all species were aligned in MAFFT (version 7.037b) (Katoh and Standley, 2013) using auto strategy and were then manually adjusted in BioEdit (Hall, 1999). JTT + I + G + F model was determined to be the best model via ProtTest (version 3.4.2) (Darriba et al., 2011). PhyML 3.1 was used to construct ML trees with the above model and 1000 nonparametric bootstrap replicates (Guindon et al., 2010).

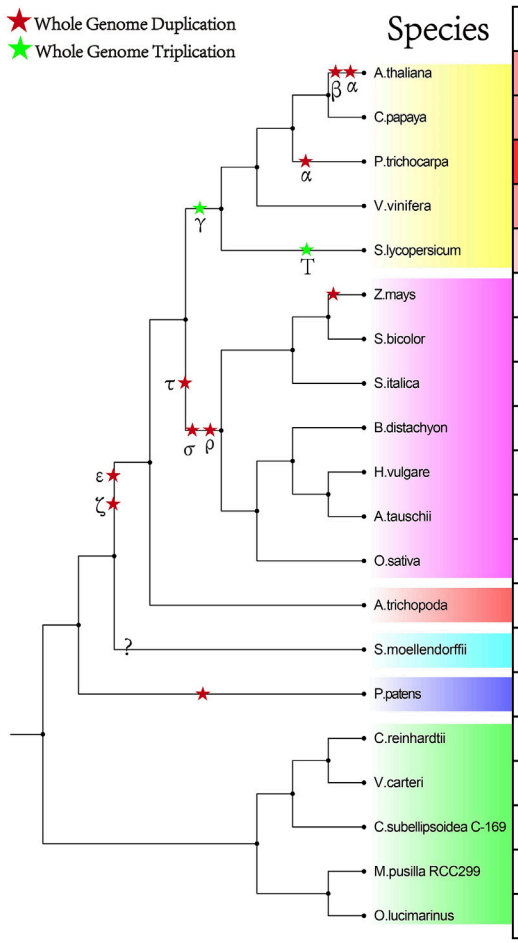
The amino acid sequences of 10 GRFs of foxtail millet were aligned by ClustalW (Thompson et al., 1994). We employed MEGA 7.0 to construct the phylogenetic trees of GRFs in foxtail millet by using the NJ method with the following parameters: Pairwise deletion and 1000 bootstrap replications (Kumar et al., 2016).

Characterization of GRFs in Foxtail Millet

The chromosome distribution of *GRFs* in foxtail millet was drawn by MapChart software (Voorrips, 2002). The online website MEME (http://meme-suite.org/) was employed to analyze GRF proteins in foxtail millet to identify as the conservative motifs (Bailey et al., 2009). The maximum number of motifs was set to be 5, and the remaining parameters were default. Isoelectric point value and theoretical molecular weight of GRF proteins in foxtail millet were calculated using the ProtParam tool of ExPaSy (https://web.expasy.org/protparam/) (Gasteiger et al., 2005). The subcellular localization of GRFs in foxtail millet was predicted by Plant-mPLoc database (Chou and Shen, 2010). Using the SOPMA website (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) to predict the secondary structure of GRF proteins in foxtail millet, the parameters were default. Using the Phyre2 website (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) to predict the three-dimensional structure of GRF proteins, the parameters were default. The gene structure of GRFs in foxtail millet was analyzed and drawn using GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2014) and CFVisual (version 2.1) (https://github.com/ChenHuilong1223/CFVisual) (Chen et al., 2021). Multiple sequence alignment of the amino acid sequences was performed by ClustalX (Wilm et al., 2007) and conserved regions were visualized using DNAMAN 8.0. In order to reduce errors, nine coding sequences that meet requirements were screened, based on literature criteria (Eyre-Walker, 1991). Afterward, the codon bias analysis of these sequences was performed via CodonW software (https://sourceforge.net/projects/codonw/).

Selection Pressure, Gene Duplication, and Collinearity Analysis

The amino acid sequences of GRFs in foxtail millet were aligned using MAFFT (version 7.037b), and the amino acid alignments



A1	A2	B1	B2	C1	C2	D1	D2	D3	E	Total GRF
2	0	0	0	2	0	0	1	4	0	9
2	0	0	1	2	0	0	0	2	0	7
4	0	0	4	4	0	0	3	4	0	19
2	0	0	1	2	0	0	2	3	0	10
1	0	0	3	2	0	0	2	5	0	13
0	4	4	0	3	1	2	2	0	0	16
0	2	3	0	2	1	1	1	0	0	10
0	2	3	0	2	1	1	1	0	0	10
0	2	3	0	2	1	2	2	0	0	12
0	2	4	0	2	2	3	1	0	0	14
0	3	3	0	1	2	3	1	0	0	13
0	2	3	0	3	1	2	1	0	0	12
0	0	0	1	1	1	1	1	1	0	6
0	0	0	0	0	0	0	0	0	4	4
0	0	0	0	0	0	0	0	0	2	2
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

FIGURE 1 | The number of *GRF* families in the collected species. The left of this figure shows the evolutionary relationships of the species; the right of this figure shows the number detail of the *GRF* family of each species.

were translated into coding sequence alignments via PAL2NAL (http://www.bork.embl.de/pal2nal/) (Suyama et al., 2006). After that, we employed the codeml program in PAML 4.9 h software (Yang, 1997) to calculate the selection pressure of each branch of the GRF phylogenetic tree. We chose the branch model to achieve this (Yang et al., 1998), which was based on the free ratio model and one ratio model (prob = 1.517e-04).

MCScanX software (Wang et al., 2012) was used to analyze the duplications of the GRF family of foxtail millet, and 34,584 protein sequences from foxtail millet (Bennetzen et al., 2012) were analyzed using all-vs-all BLAST search with e-value < 1e-05 (Camacho et al., 2009). The putative WGDs/segmental duplications of *GRF* genes located on chromosomes of foxtail millet are connected by red lines. *Ks* (synonymous substitution rate) and *Ka* (nonsynonymous substitution rate) values of WGDs/segmental duplications were calculated based on the coding sequence alignments using the method of Nei and Gojobori as implemented in KaKs_calculator 2.0 (Nei and Gojobori, 1986; Wang et al., 2010). The *Ks* value was translated into duplication time in millions of years based on the rate of λ substitutions per synonymous site per year. The duplication time of duplicated genes was calculated by $T = Ks/2\lambda \times 10^{-6}$ Mya ($\lambda = 6.5 \times 10^{-9}$ for grasses) (Lynch and Conery, 2000; Mehanathan et al., 2014; Wang et al., 2015; Chai et al., 2018). To reduce errors, we only analyzed the results for $Ks < 1$.

Orthologous pairs of *GRF* members among foxtail millet, *Arabidopsis*, and rice were identified using OrthoFinder software (version 2.2.6) (Emms and Kelly, 2015) and OrthoMCL (version 2.0.9) (Li et al., 2003). The results were visualized using Circos (version 0.69–6) (Krzywinski et al., 2009).

Expression and Regulation Analysis of *GRF*s in Foxtail Millet

We obtained the sequence of foxtail millet miRNA396 from the literature (Yadav et al., 2016), and then used psRNATargetsoftware (http://plantgrn.noble.org/psRNATarget/) to predict the binding site of miR396 in *GRF* genes of foxtail millet (Dai et al., 2018). PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to analyze the 1 Kb sequence

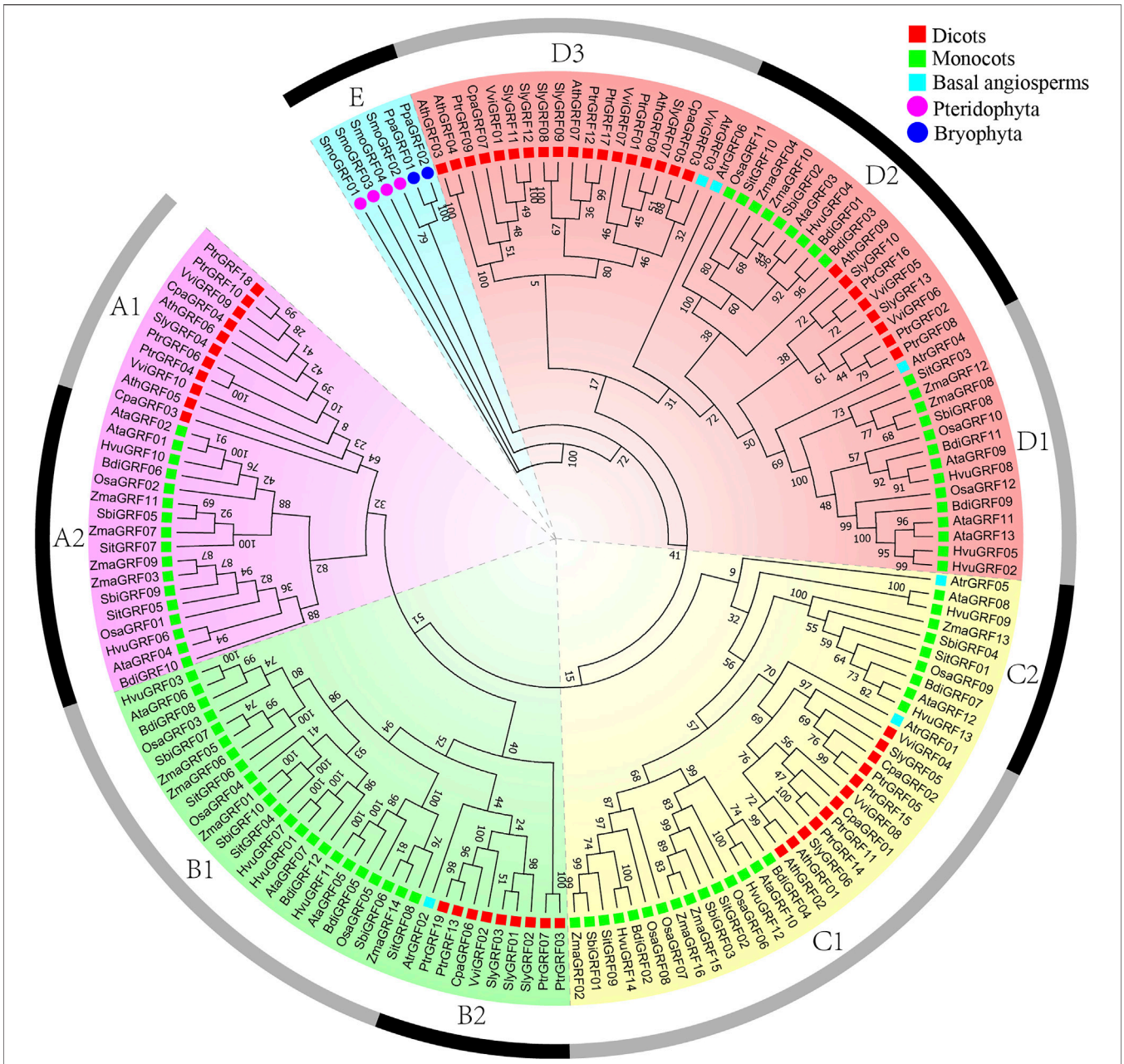


FIGURE 2 | Phylogenetic trees of *GRF* genes. Phylogenetic analysis of plant *GRF* proteins using PhyML 3.1 under JTT + I + G + F model with 1000 nonparametric bootstrap replicates.

upstream of *GRF* genes in foxtail millet (Lescot et al., 2002). We utilized an in-house Python script to extract *GRF* TPM values of foxtail millet from Illumina RNA-seq data reported previously (Yang et al., 2020). The heatmap was drawn via Morpheus software (https://software.broadinstitute.org/morpheus/) based on the transformed data of log2 (TPM+1) values. The String database (version 11.0) (https://string-db.org/) was used to predict interaction proteins of GRFs with the minimum required interaction score set to be high confidence (0.700) (Szklarczyk et al., 2015). The agriGO V2.0 was used for GO analysis of GRFs in foxtail millet (Tian et al., 2017).

Quantitative Real-Time PCR Analysis

The total RNA was extracted using RNAprep Pure Plant Plus Kit (TIANGEN) from three tissues: imbibed 3-day seed, 2-week-old seedling, and 2-week-old seedling root. First-strand cDNA was synthesized using Fastking RT Kit (with gDNase) (TIANGEN). The SuperReal PreMix Plus (SYBR Green)

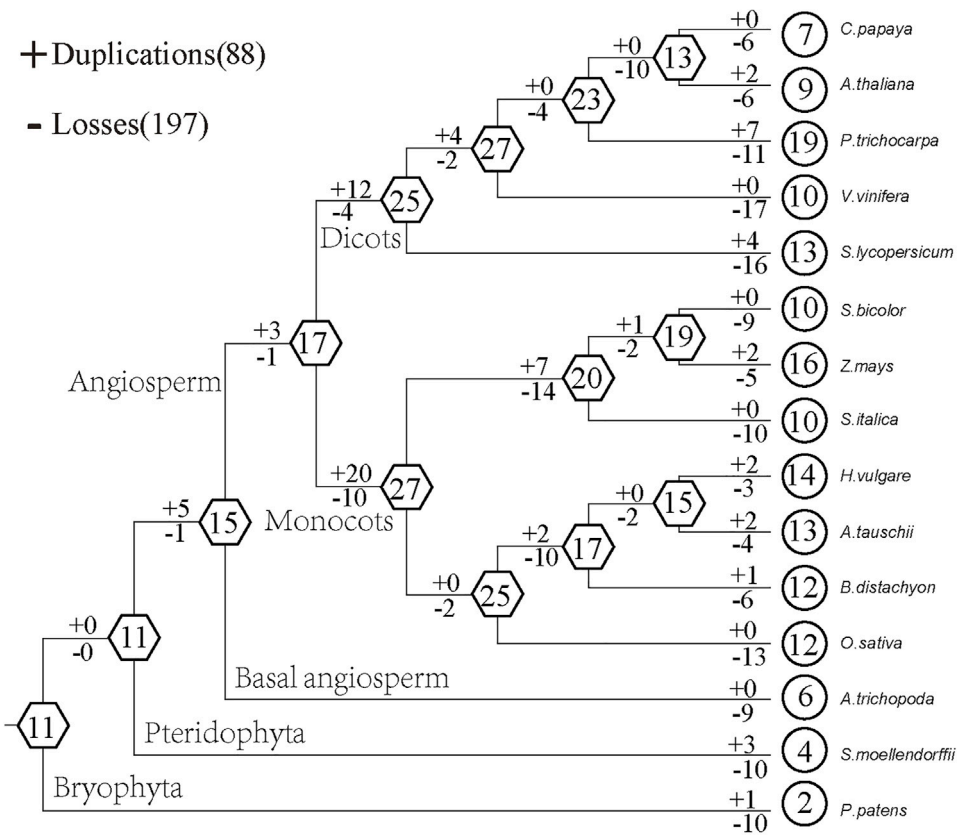


FIGURE 3 | Schematic diagram of gain and loss of *GRF* gene family in plants. The numbers in the hexagons and circles represent the number of *GRF* genes in ancestors and existing species, and the + and – signs represent the gain and loss of genes, respectively.

(TIANGEN) was used for real-time-qPCR analysis with 7900HT Fast Real-Time PCR System (American Applied Biosystems). The Primers were designed by Primer Premier6.0 and synthesized by GENEWIZ Biotechnology Co., Ltd. (**Supplementary Table S1**). EF-1α was the reference gene (Kumar et al., 2013).

RESULTS

Genome-wide Identification and Classification of *GRF* Genes in Plants

We identified a total of 157 *GRF* genes in 20 species (**Figure 1**, **Supplementary Table S2**). No *GRF* gene has been identified in green algae (*C.reinhardtii*, *V. carteri*, *C. subellipsoidea C-169*, *M.* sp. *RCC299,* and *O. lucimarinus*). In land plants, the least *GRF* genes (two) have been identified in *P. patens*, four *GRF* genes have been identified in *S. moellendorfii,* and six *GRF* genes have been identified in *A. trichopoda*. The number of *GRF*s in monocots (*Z.mays*, *S. bicolor*, *S. italica*, *B. distachyon*, *H. vulgare,* and *O. sativa*) ranges from 10 to 16, while the number of *GRF*s in dicots (*A.thaliana*, *C. papaya*, *P. trichocarpa*, *V. vinifera,* and *S. lycopersicum*) ranges from 7 to 19.

According to previous research and the phylogenetic tree topology (Song et al., 2018), phylogenetic analysis showed that 157 *GRF* genes could be clustered into five categories: A, B, C, D, and E (**Figure 2**). A class is subdivided into A1 and A2 subclass, and B class is subdivided into B1 and B2 subclass. C class is subdivided into C1 and C2 subclass, and D class is subdivided into D1, D2, and D3 subclass. According to statistics, the genes of the A1 subclass are dicots *GRF*s. The genes of the A2 subclass and B1 subclass are monocots *GRF*s. The genes of the B2 subclass and D3 subclass contain some dicots *GRF*s and one basal angiosperm *GRF*. The genes of the C2 subclass and D1 subclass contain some monocots *GRF*s and one basal angiosperm *GRF*. In addition, the genes of ancient E class are *GRF*s of all ancient species (*P. patens* and *S. moellendorfii*) (**Figure 1**).

Gain and Loss of *GRF* Genes in Plants

Based on the comparison between the species tree and the plant *GRF* gene tree, we used Notung software to analyze the gain and loss of *GRF* genes. The results show that the ancestors of land plants contained 11 *GRF* genes (**Figure 3**). The loss is more serious in lower plants. Among them, 10 *GRF* genes of *P. patens* and *S. moellendorfii* have been lost, and 1 *GRF* gene and 3 *GRF* genes have been obtained, respectively. The ancestor of angiosperms contains 15 *GRF* genes with 9 *GRF* genes being lost and 0 *GRF* genes being gained. This results in 6 existing *GRF* genes in *A. trichopoda*. There are 17 *GRF* genes in common

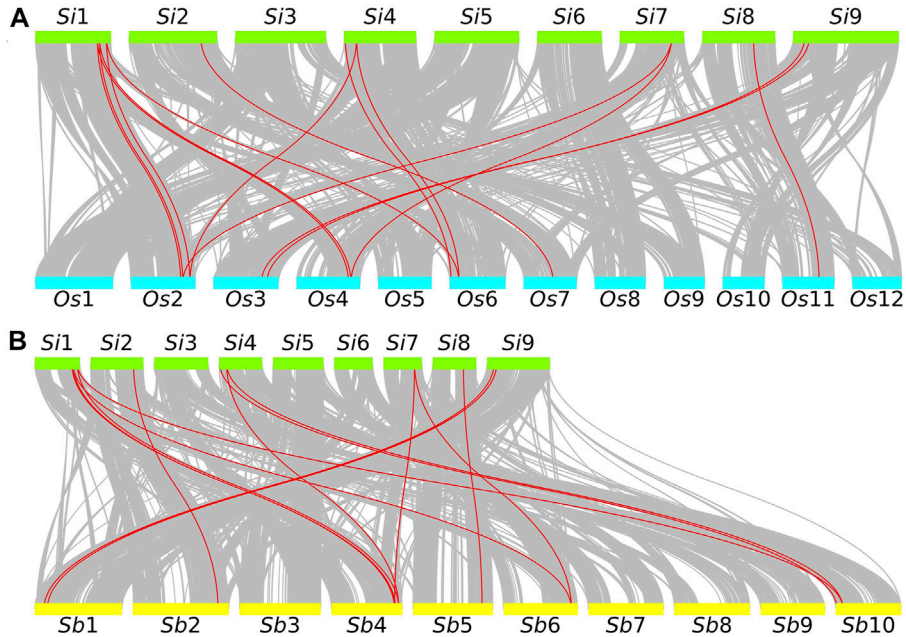


FIGURE 4 | Collinearity analysis of foxtail millet *GRF* and related species. **(A)** The green rectangular color block represents the foxtail millet chromosome. *Si* represents foxtail millet. The number represents the chromosome number. The blue rectangular color block represents the rice chromosome. *Os* represents rice. The number represents the chromosome number. **(B)** The green rectangular color block represents the foxtail millet chromosome. *Si* represents foxtail millet. The number represents the chromosome number. The yellow rectangle represents the sorghum chromosome. *Sb* represents sorghum. The number represents the chromosome number.

ancestors of dicots and monocots. After 12 *GRF* genes were gained and 4 *GRF* genes were lost, 25 *GRF* genes exist in dicots ancestor species, after 20 *GRF* genes were gained and 10 *GRF* genes were lost, 25 *GRF* genes exist in monocots ancestor species. This indicates that the *GRF* gene family in the ancestor species of angiosperms has expanded after the divergence of monocots and dicots. After comparing the gain and loss of dicots with that of monocots, we found that the *GRF* gene family in monocots was gradually shrinking. For example, the ancestor of foxtail millet, sorghum, and maize has undergone 14 losses and 7 gains, which result in the reduction from 27 *GRF* genes to 20 *GRF* genes. Although the number of existing species of dicots is less than that of ancestors, it does not show a gradual shrinking phenomenon. For example, the ancestor of *V. vinifera*, *P. pilosa*, *A. thaliana,* and *C. papaya* has undergone 2 losses and 4 gains, but the number of *GRF* genes of the ancestor increased from 25 to 27.

Strong Collinearity Between Foxtail Millet *GRFs* and Related Species and Weaker Positive Selection

Combining with the results of internal collinearity in foxtail millet, we found that 6 pairs of *GRF* genes (seven genes, accounting for 70.00%) are in the collinearity block. Estimates of divergence time indicate that the divergence time of fragment duplication ranges from 15.07 Mya to 45.97 Mya, and *GRF*s are subject to purification options (**Supplementary Table S3**). Unfortunately, we did not find a tandem repeat gene pair.

In addition, we used a self-made Python script (https://github.com/ChenHuilong1223) to draw the *GRF* collinearity relationship between foxtail millet and other closely related species. We identified 15 pairs of collinearity genes in the collinearity region of the genome of foxtail millet and rice. Chromosome 1 of foxtail millet has the most collinearity *GRF* gene pairs (40.00%) with rice. Among them, there are collinearity *GRF* genes with chromosome 2, 4, and 6 of rice, respectively (**Figure 4A**). Similarly, 15 collinearity gene pairs (33.33%) were identified in foxtail millet and sorghum. Chromosome 1 of foxtail millet has the most collinearity *GRF* gene pairs with sorghum. Among them, there are collinearity *GRF* genes with chromosome 4, 6, and 10 of sorghum (**Figure 4B**). An abundance of collinearity gene pairs indicate that *GRF* of foxtail millet has strong collinearity with closely related species.

The *GRF* phylogeny tree of foxtail millet shows that these 10 *GRF* genes can be assigned to A, B, C, and D class in the phylogenetic tree (**Figure 5A**). After selection pressure analysis, two (11.76%) of the 17 branches in the *GRF* gene tree of foxtail millet were detected to be positively selected. Therefore, this indicates that the *GRF* of foxtail millet has received weaker positive selection during the evolution process.

Characterization and Structure of *GRF*s in Foxtail Millet

We have identified 10 *GRF* genes in the foxtail millet genome. The amino acid lengths of GRFs in foxtail millet are between 232 and 590 amino acids. The molecular weights (MW) are between

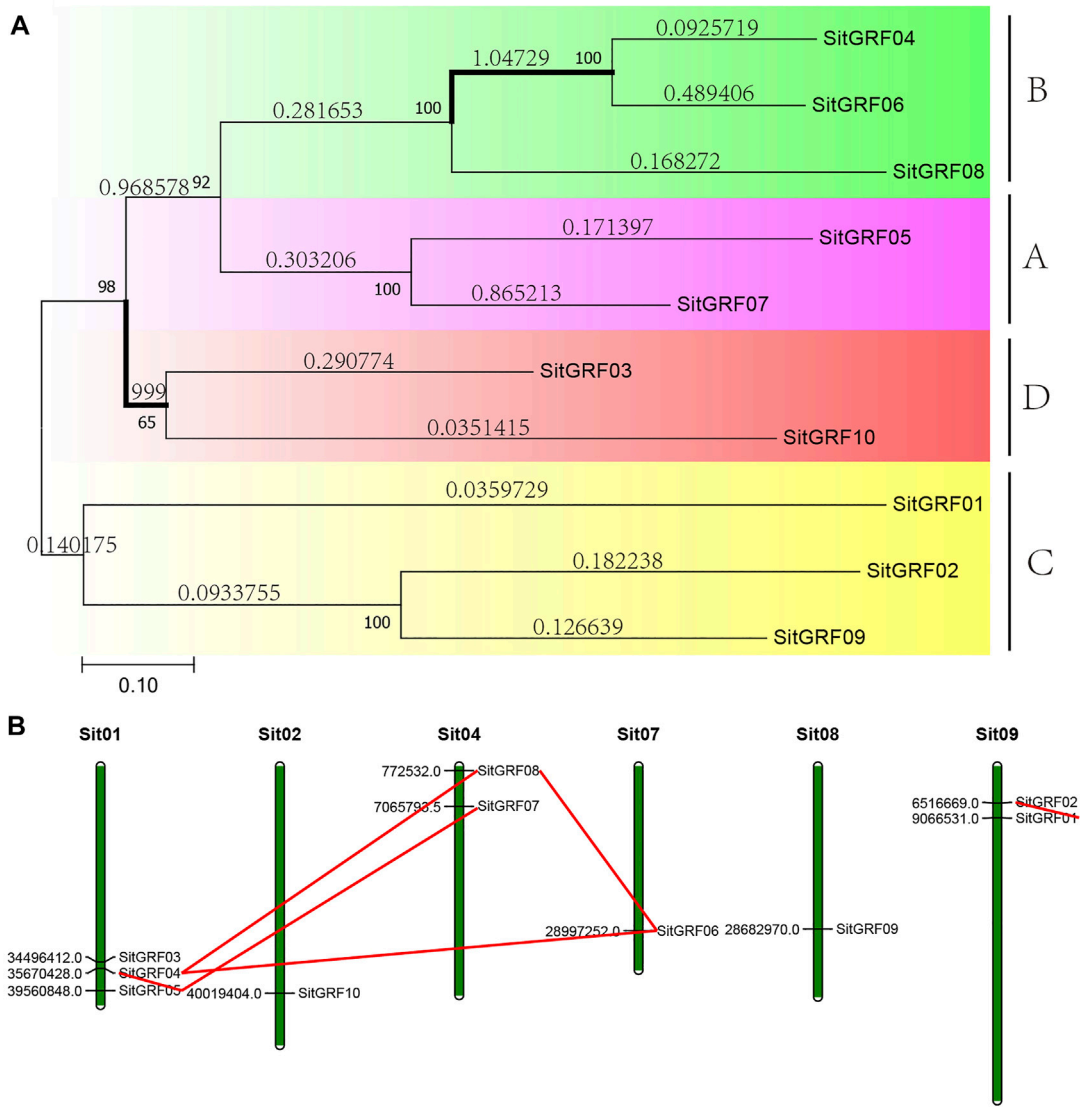


FIGURE 5 | Phylogenetic tree and chromosome location of the GRF family of foxtail millet. **(A)** Phylogenetic analysis of foxtail millet GRF proteins using MEGA 7.0 via the neighbor-joining (NJ) method with 1000 bootstrap replicates. Bold branches suggest that they may be positively selected. **(B)** Chromosomal distribution and gene duplications of foxtail millet *GRF* genes. Green bars represent chromosomes of foxtail millet. The putative whole genome duplication (WGD) or segmental duplication genes are linked by a red line.

34,981.10 and 61,607.68 Da. The predicted isoelectric point (PI) values are between 4.95 and 9.54. Prediction of subcellular location indicates that all 10 GRFs may be located in the nucleus (**Supplementary Table S4**).

The chromosome location shows that there is no *GRF* gene on chromosomes 3, 5, and 6 of the foxtail millet, no clustering phenomenon and scattered distribution on those chromosomes (**Figure 5B**). The remaining chromosomes have one to three *GRF* genes with the most being on chromosome 1.

Regarding the composition of the secondary structure of GRF in foxtail millet, a random coil occupies the largest proportion (49.63–67.83%). Alpha helix occupies the second largest proportion (18.26–36.57%), and extended strand occupies the third largest proportion (4.56–12.03%). Beta turn makes up the smallest proportion (3.04–6.08%) (**Supplementary Table S5**). In addition, we predicted the three-dimensional structure of the GRFs in foxtail millet. The results showed that the three-dimensional structure of the GRFs in foxtail millet is simple with no complicated spiral folding structures. The three-dimensional structure of the 10 GRFs is very similar (**Figure 6**).

The results of the amino acid sequence alignment of GRFs in foxtail millet indicate that all GRFs in foxtail millet contain QLQ and WRC (**Figure 7C**). The WRC domain contains a C3H motif spanning three cysteines and one histidine. The motif is CX9CX10CX2H, and the motif of the QLQ domain is QX3LX2Q. We also found that five GRF proteins contain FFD and TQL domains and are highly conserved (**Figure 7D**).

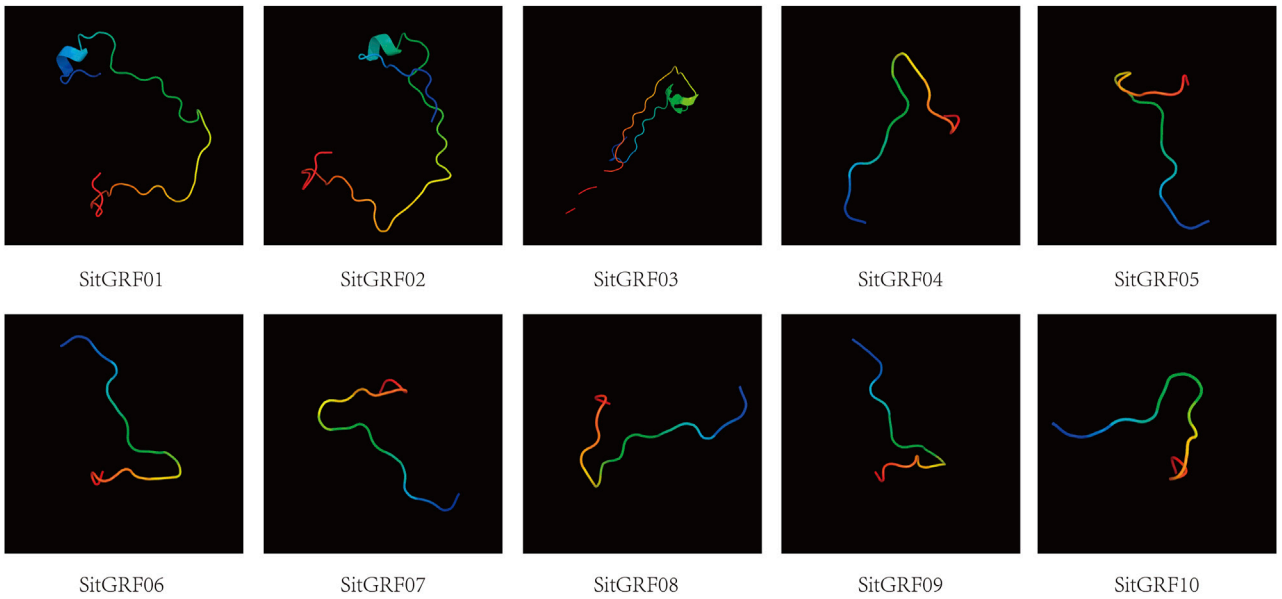


FIGURE 6 | Three-dimensional structures of GRF proteins in foxtail millet.

We used MEME software to predict the GRF motifs in foxtail millet (**Figure 7B**). The results indicated that the conservative motifs 1 through 4 correspond to the corresponding domains. For example, motif 1 corresponds to QLQ, and motif 2 corresponds to WRC. In addition, we predicted that motif 5 exists in most members (seven) of the GRF family in foxtail millet.

In the structure of *GRF*s in foxtail millet, each *GRF* has two to four coding sequences (**Figure 7A**). The number of introns varies from one to four. It is noteworthy that *SitGRF08* has no UTR, and the 3′ UTR of *SitGRF01* is particularly long. However, *SitGRF03* has the shortest length of the gene structure.

We also conducted the analysis of codon preference, and the results showed that the *GRF* gene family of foxtail millet is biased toward the use of G or C nucleotides. The ENC value is between 32.68 and 61.00, and 35.00 is the strength of codon preference. The closer to 61.00, the weaker the codon preference is, and vice versa. It shows that the overall preference of *GRF*s in foxtail millet is relatively weak (**Supplementary Table S6**). We also identified 19 optimal codons: UUU, UUG, AUU, GUA, UCU, CCG, ACG, GCA, UAA, CAU, CAG, AAU, AAG, GAU, GAA, UGC, CGC, AGU, and GGG (**Supplementary Table S7**). These results are helpful in the application of transgenic technology on foxtail millet.

cis-acting Elements and miRNA

We predicted *cis*-acting elements of the putative promoter region (upstream 1000 bp) of each *GRF* in foxtail millet (**Figures 8A,B**, **Supplementary Table S8**). We found that all gene promoters contained hormone-related *cis*-acting elements, and only four gene promoters were predicted to contain stress-related *cis*-acting elements. Seven hormone-related *cis*-acting elements were identified in the promoter region of *GRF*s in foxtail millet. These seven hormone-related *cis*-acting elements are ABRE (*cis*-acting element involved in abscisic acid responsiveness) (Hobo et al., 1999), CGTCA/TGACG-motif (*cis*-acting element involved in MeJA-responsiveness) (Rouster et al., 1997), GARE-motif/P-box (gibberellin-responsive *cis*-acting element) (Gubler and Jacobsen, 1992), TCA-element (*cis*-acting element involved in salicylic acid responsiveness) (Shah and Klessig, 1996), and TGA-element (auxin-responsive *cis*-acting element) (Khan et al., 2012). Three stress-related *cis*-acting elements were identified in the promoter region of *GRF*s in foxtail millet. The three stress-related *cis*-acting elements are ARE (*cis*-acting element essential for anaerobic induction), LTR (*cis*-acting element involved in low-temperature responsiveness), and MBS (MYB binding site involved in drought-inducibility) (Yoshida et al., 1998).

According to previous studies, most *GRF*s are regulated by miR396 (Jones-Rhoades and Bartel, 2004; Liu et al., 2009; Zheng et al., 2018). Therefore, we predicted the *GRF* target gene of Sit-miR396 (**Figure 7A**, **Supplementary Table S9**). The results show that seven *GRF* target genes are predicted, and they all have a cleavage effect on each *GRF* gene.

Tissue Expression Analysis of *GRF* Genes in Foxtail Millet at Different Periods

By analyzing the expression pattern of *GRF* genes in foxtail millet in different tissues (**Figure 9**, **Supplementary Table S10**), the results showed that the expression of *GRF*s in foxtail millet was the strongest in seed, panicle, and stem tissues. The weakest expression patterns of *GRF* genes in foxtail millet were in different periods and were found during the third day of imbibed seeds. Some genes displayed higher expression, such as *SitGRF06* (68.12 TPM) and *SitGRF10* (29.86 TPM), followed by *SitGRF08* (16.66 TPM). In the seedling stage on the 14th day, the expression of seven *GRF* genes (70.00%) all increased, while

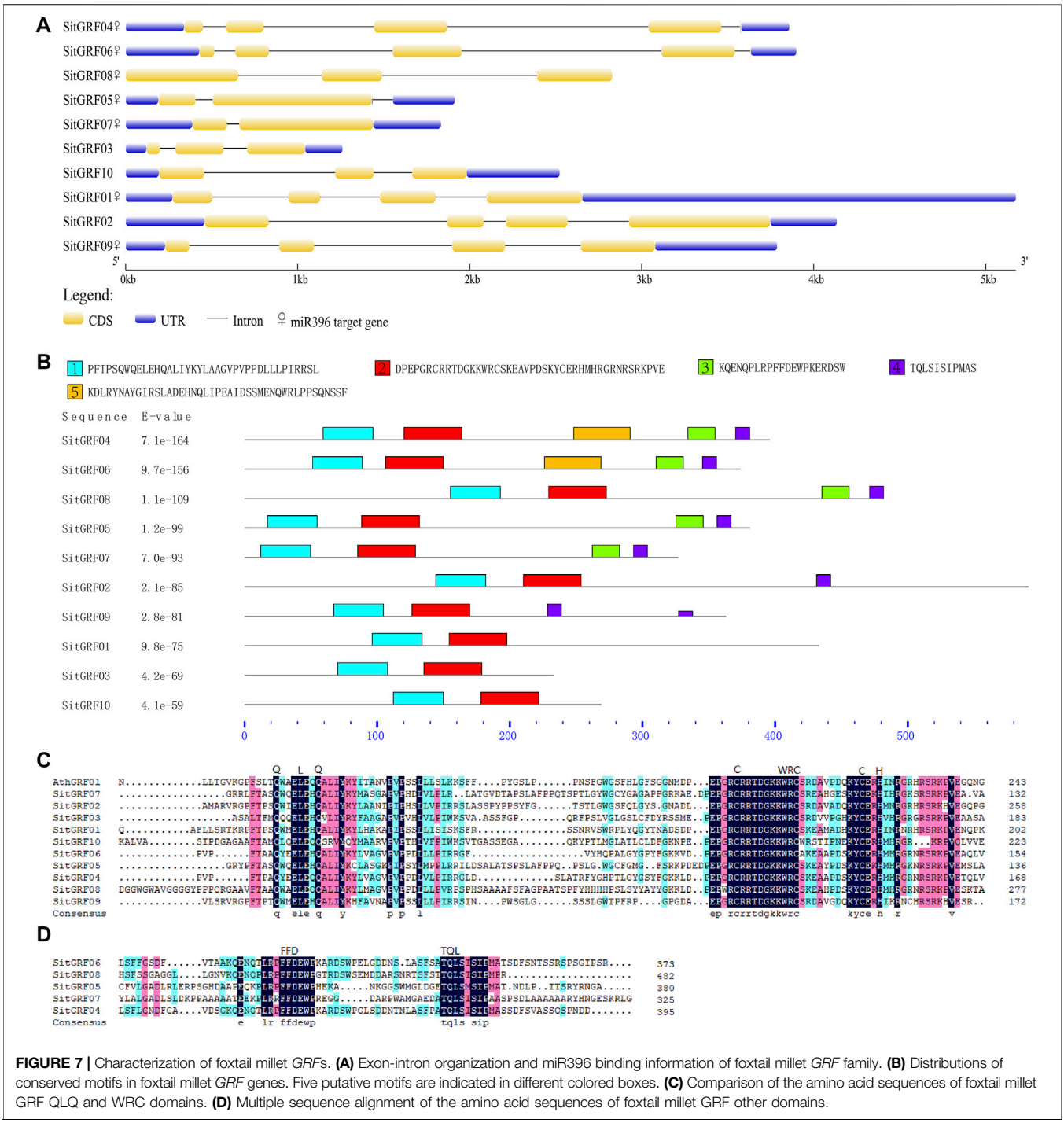


FIGURE 7 | Characterization of foxtail millet *GRF*s. **(A)** Exon-intron organization and miR396 binding information of foxtail millet *GRF* family. **(B)** Distributions of conserved motifs in foxtail millet *GRF* genes. Five putative motifs are indicated in different colored boxes. **(C)** Comparison of the amino acid sequences of foxtail millet GRF QLQ and WRC domains. **(D)** Multiple sequence alignment of the amino acid sequences of foxtail millet GRF other domains.

SitGRF06, *SitGRF10,* and *SitGRF07* decreased. When the top first leaf of a 2-week-old seedling is fully extended, the gene expression changed little while the expression of *SitGRF01*, *SitGRF09*, *SitGRF08,* and *SitGRF10* increased. In immature panicles, *SitGRF02*, *SitGRF03*, *SitGRF04*, *SitGRF05,* and *SitGRF06* were significantly increased. The overall performance of the *GRF* family decreased gradually in the panicle at the pollination stage and at the grain-filling stage. However, *SitGRF03* and *SitGRF10* increased in the panicle at the grain-filling stage. In the flag leaf and the fourth leaf, the expression of seven *GRF* genes was extremely low while *SitGRF07*, *SitGRF08,* and *SitGRF10* had higher expression. Additionally, the overall expression of the *GRF* family was extremely low and nine *GRF* genes (except *SitGRF07*) had high expression values in the stem tissue. Through observation, we found that *SitGRF08* and *SitGRF10*

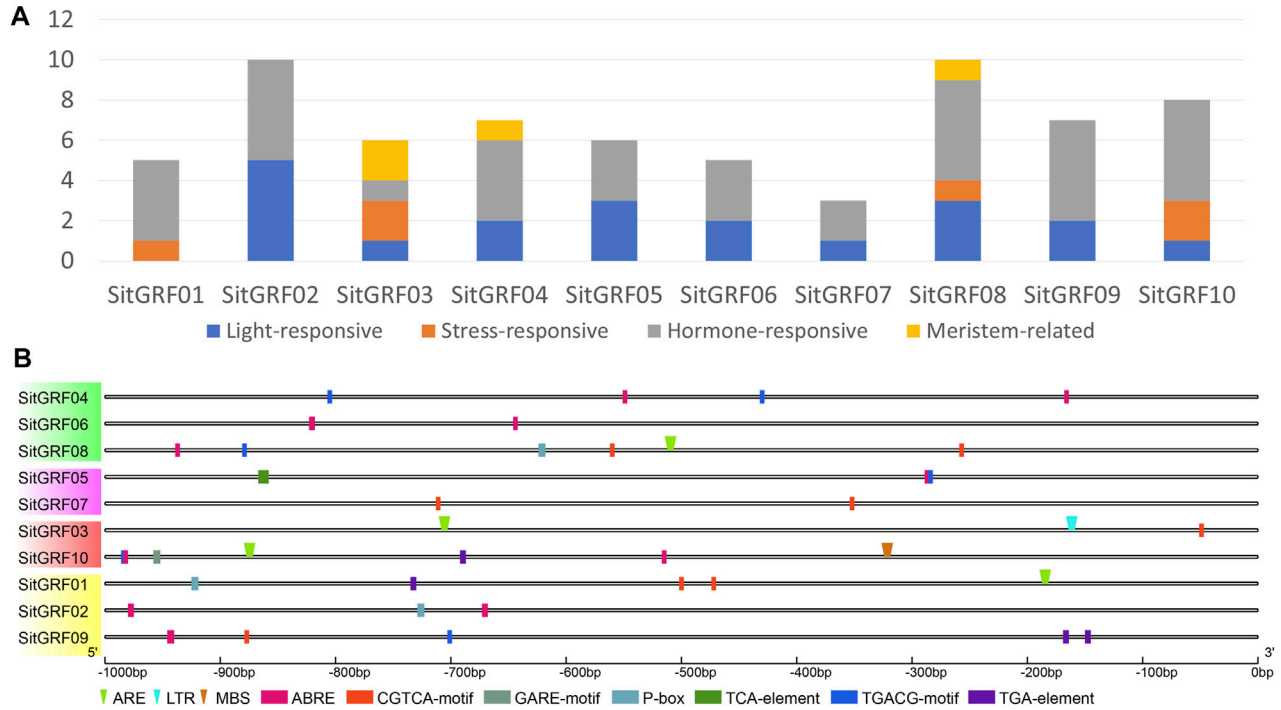


FIGURE 8 | Promoter analysis of foxtail millet *GRF* genes. **(A)** Distribution of four kinds of *cis*-regulatory elements in the promoter of foxtail millet *GRF* genes. **(B)** The 1 Kb promoter sequences of corresponding *GRF* genes were used to analyze hormone-related *cis*-elements and stress-related elements. Rectangle indicates stress-related elements, and the inverted triangle indicates hormone-related elements. Different *cis*-elements were indicated by different colored symbols and placed in their relative position on the promoter.

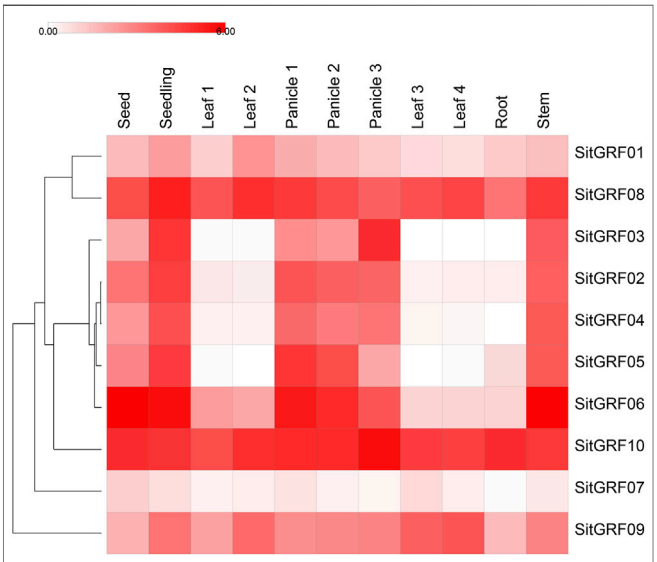


FIGURE 9 | Expression of foxtail millet *GRF* genes in the different organs and periods. The shade of the color corresponds to the expression value (log2 (TPM+1)). Seed represents 3 days imbibed seeds; seedling represents 2-week-old whole seedling. Leaf 1 represents the top first fully extended leaf of a 2-week-old seedling. Leaf 2 represents the top second leaf of 30-day-old plants. Panicle 1 represents an immature panicle. Panicle 2 represents a panicle at the pollination stage. Panicle 3 represents a panicleat at the grain-filling stage. Leaf 3 represents a flag leaf. Leaf 4 represents the fourth leaf. Root represents the root. Stem represents the stem.

maintained high expression in the organization of each period, and *SitGRF07* expressed lower in the organization of each period.

In order to definitely test the tissue expression patterns of *SitGRF* genes, qRT-PCR was used to detect the relative expression level of five *SitGRF* genes (*SitGRF02*, *SitGRF06*, *SitGRF08*, *SitGRF09,* and *SitGRF10*) in imbibed 3-day seed, 2-week-old seedling, and the root of 2-week-old seedling. The other five *SitGRF* genes were not detected due to the lack of screening specific primers across the introns. The results of qRT-PCR showed that among 15 pairs of comparison (five genes and three tissues), the expression trend of 13 pairs (86.67%) was consistent with the transcriptome data (**Figure 10**). For example, among the three tissues, *SitGRF08* and *SitGRF09* expressed the highest in the seedling and the lowest in the root, *SitGRF10* expressed the highest in the germinated seed and the lowest in the seedling, In the meanwhile, both *SitGRF02* and *SitGRF06* expressed higher in the seedling than in the germinated seed. Only the expression pattern in the root of *SitGRF02* and *SitGRF06* was inconsistent. In qRT-PCR, *SitGRF02* and *SitGRF06* expressed the highest in the root, while in transcriptome, they were the lowest in the root.

Protein Interaction Analysis

In order to further explore the mechanism of action of proteins expressed by the *GRF* genes of foxtail millet, we looked for an interaction protein for each GRF in the String database (**Figure 11**). Under high confidence (0.7) conditions, GRFs in foxtail millet have

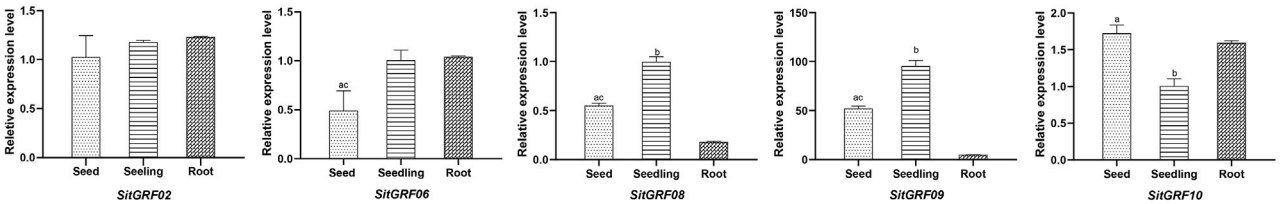


FIGURE 10 | Tissue specific expression analysis of *SitGRF* genes in germinated seed, 2-week-old seedling, and root. The bars represent the mean values of three replicates ± s.d. Significant differences in means are indicated by a, b, c, $p < 0.01$, according to one-way ANOVA test. a represents the comparation between germinated seed and seedling. b represents the comparation between seedling and root. c represents the comparation between germinated seed and root.

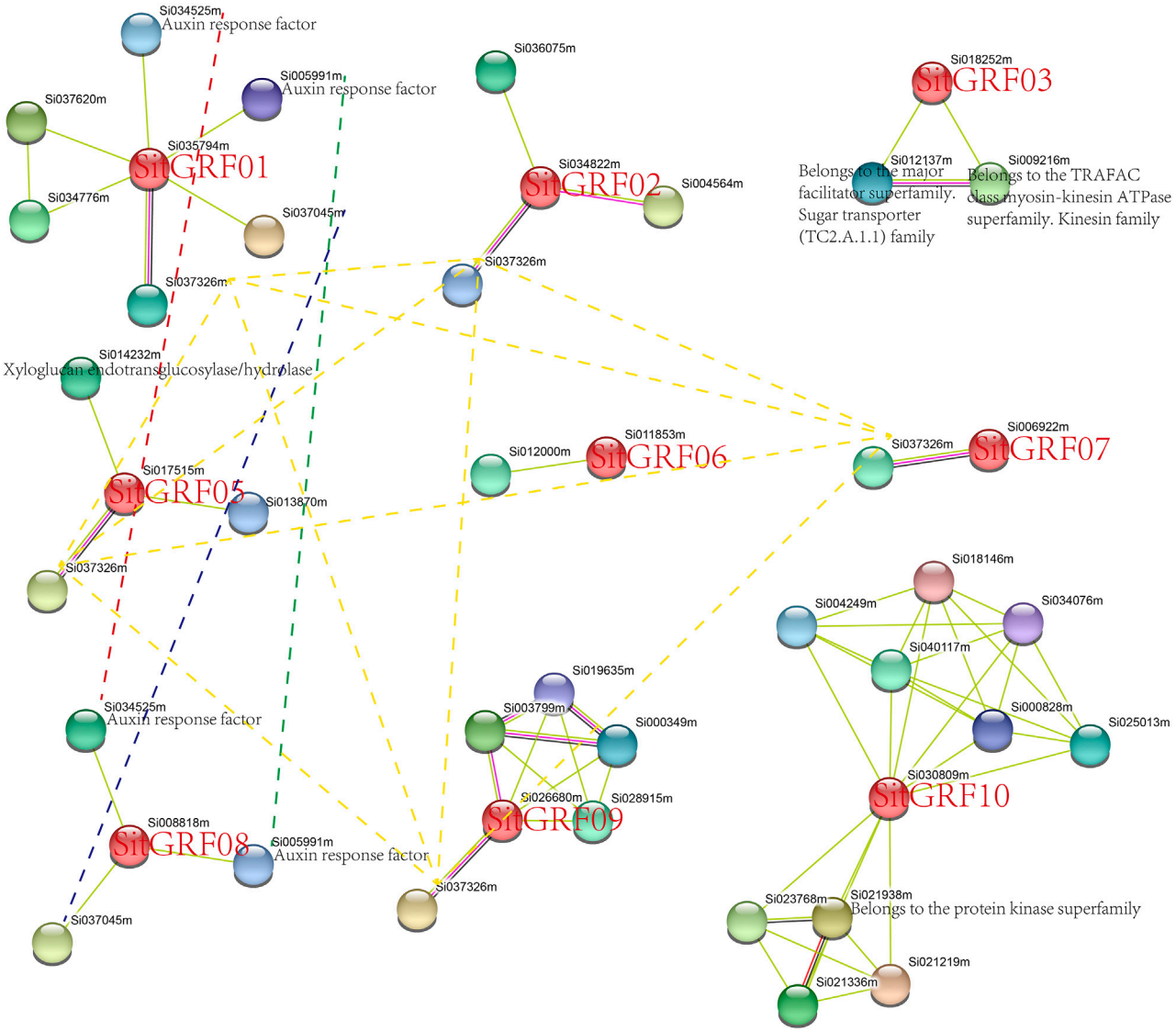


FIGURE 11 | Protein interaction analysis of foxtail millet GRFs. The same protein is connected by a dotted line. The unannotated protein indicates uncharacterized protein in the String database.

interaction proteins, except SitGRF04. Among them, SitGRF10 had the most interaction proteins. Each of SitGRF06 and SitGRF07 has only one interaction protein. These interaction proteins provide clues to the function and mechanism of each *GRF*. For example, both SitGRF01 and SitGRF08 could interact with auxin response factors, implying that they may participate in the network

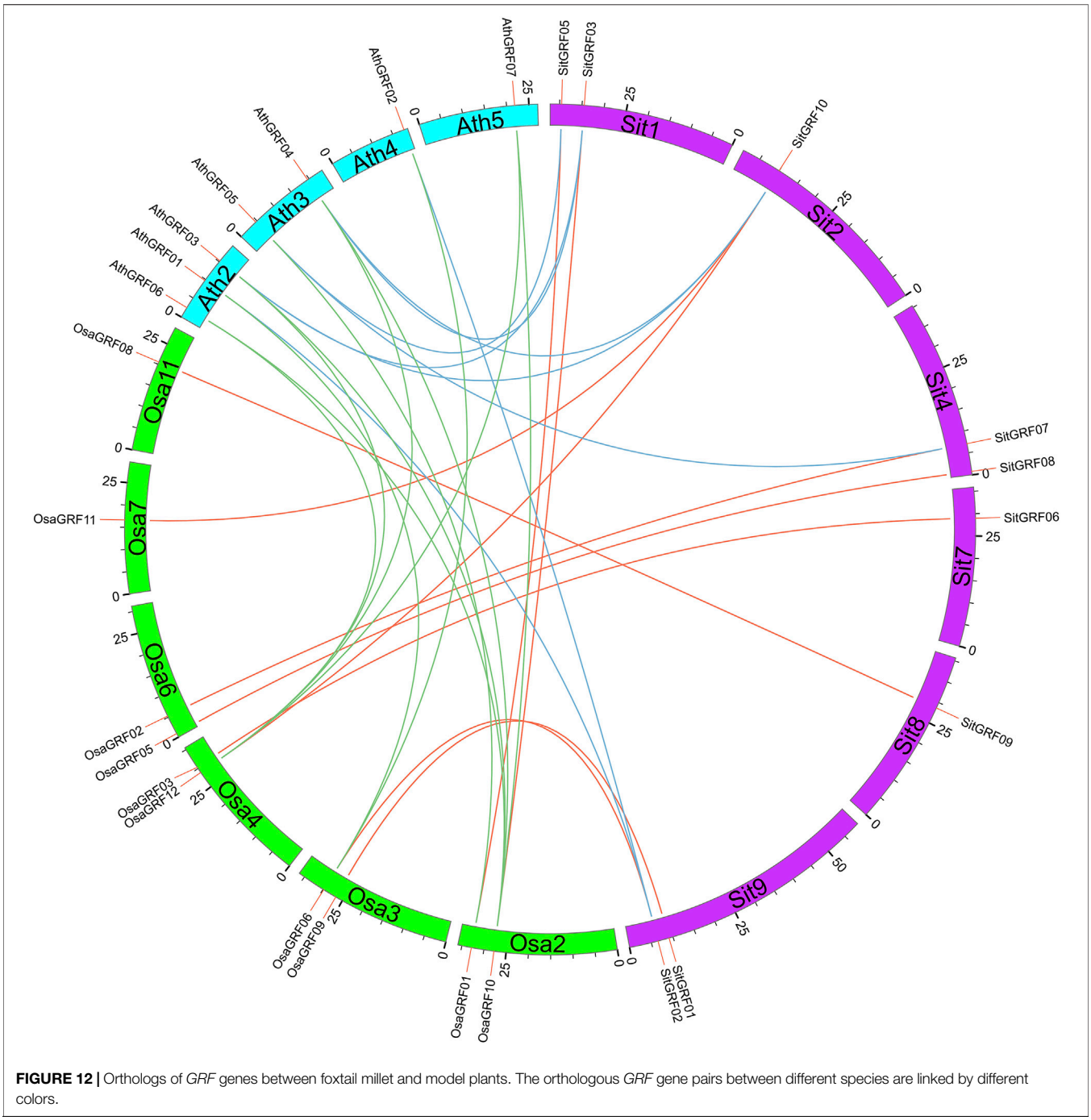


FIGURE 12 | Orthologs of *GRF* genes between foxtail millet and model plants. The orthologous *GRF* gene pairs between different species are linked by different colors.

regulation mechanism of auxin response factors. We also found that SitGRF01, SitGRF02, SitGRF05, SitGRF07, SitGRF08, and SitGRF09 can interact with Si037326m.

Identification of Orthologs with Model Plants and Functional Annotation of GO

The identification of orthologous genes with *GRF* genes in model plants is helpful for the function prediction of *GRF* genes in foxtail millet. Through the identification of orthologous gene pairs of *GRF* with rice and *Arabidopsis*, it is found that the *GRF* gene family of foxtail millet and rice can form 10 pairs of orthologous genes. the *GRF* gene family of foxtail millet and rice have eight orthologous gene pairs. In addition, there are 10 orthologous gene pairs of the *GRF* gene family between *Arabidopsis* and rice (**Figure 12**).

We performed GO function annotations and the results reported that all 10 GRFs of foxtail millet can participate in biological pathways, molecular functions, and cellular components (**Supplementary Table S11**). A total of 75 GO numbers could be annotated. The most annotated entries are in biological pathways

(64.00%), such as developmental process, regulation of cellular biosynthetic process, regulation of cellular metabolic process, etc. Molecular functions (ATP binding, purine ribonucleoside triphosphate binding, purine nucleoside binding, etc.) accounted for 24.00%. The cellular component (nucleus, intracellular membrane-bounded organelle, membrane-bounded organelle, etc.) was 12.00%. Overall, GO function annotations found that *GRF* mainly functions in molecular pathways.

DISCUSSION

In our research, the ancestors of land plants contain at least 11 *GRF* genes, which are not much different from the number of existing plants. This indicates that *GRF* genes have not expanded on a large scale. The *GRF* content of ancient species is very small, and it can always cluster in the ancient E branch in the phylogenetic tree. Thus, the *GRF* gene family of land plants originated from the E class. The number of *GRF* in higher plants is significantly elevated than that of ancient species (lower plants), meanwhile, the number of monocots is relatively stable and the number of dicots is more divergent. We found that the ancestors of angiosperms have experienced more gains and the family has expanded. However, the *GRF* gene family of monocots has been shrinking in the course of evolution, while the *GRF* gene family of dicots has not shown a gradual shrinking phenomenon. This points to the different evolutionary processes of monocots and dicots. Combining the results of identification and classification, we found that the low number of *GRF* in lower plants should be caused by excessive loss, rather than a small amount at the beginning.

Our research also showed that *GRF* in foxtail millet is affected by WGD or segmental duplication between 15.07 Mya and 45.97 Mya. The *GRF*s of foxtail millet and other closely related species have strong collinearity and homology. They were mainly subjected to purification selection in the past. These evolutionary phenomena indicate that the evolution of *GRF* is a conservative evolutionary model.

Comparing the qRT-PCR results with transcriptome data showed that 86.67% of the expression trend was consistent. Only 13.33% was inconsistent. This inconsistency has been reported in many literatures (Everaert et al., 2017). Celine Everaert et al. reported that about 85% of the genes showed consistent results between RNA-sequencing and qRT-PCR data. Our result was consistent with that reported. These results indicated that the transcriptome data was reliable.

DATA AVAILABILITY STATEMENT

The RNA-seq data presented in the study are deposited in the Beijing Institute of Genomics Data Center (https://bigd.big.ac.cn/) repository, accession number CRA001953.

AUTHOR CONTRIBUTIONS

W.G. conceived the study; H.C. and W.G. conducted experiments; H.C. and W.G. analyzed data; H.C. and W.G. wrote the manuscript. All authors revised and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.727674/full#supplementary-material

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* 37 (Suppl. l_2), W202–W208. doi:10.1093/nar/gkp335

Baloglu, M. C. (2014). Genome-wide In Silico Identification and Comparison of Growth Regulating Factor (GRF) Genes in Cucurbitaceae Family. *Plant Omics* 7 (4), 260–270. doi:10.1097/01.tp.0000332556.64365.e7

Bateman, A., Heger, A., Sonnhammer, E. L. L., Mistry, J., Clements, J., Tate, J., et al. (2013). Pfam: the Protein Families Database. *Nucleic Acids Res.* 42 (D1), D222–D230. doi:10.1093/nar/gkt1223

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference Genome Sequence of the Model Plant Setaria. *Nat. Biotechnol.* 30 (6), 555–561. doi:10.1038/nbt.2196

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics* 10 (1), 421. doi:10.1186/1471-2105-10-421

Cao, Y., Han, Y., Jin, Q., Lin, Y., and Cai, Y. (2016). Comparative Genomic Analysis of the GRF Genes in Chinese Pear (Pyrus Bretschneideri Rehd), Poplar (Populous), Grape (Vitis vinifera), Arabidopsis and Rice (Oryza Sativa). *Front. Plant Sci.* 7 (577), 1750. doi:10.3389/fpls.2016.01750

Chai, W., Si, W., Ji, W., Qin, Q., Zhao, M., and Jiang, H. (20182018). Genome-Wide Investigation and Expression Profiling of HD-Zip Transcription Factors in Foxtail Millet (Setaria Italica L.). *Biomed. Res. Int.* 2018, 8457614. doi:10.1155/2018/8457614

Chen, H. L., Wang, X. Y., and Ge, W. N. (2021). Comparative Genomics of Three-Domain Multi-Copper Oxidase Gene Family in Foxtail Millet (Setaria Italica L.). *Comput. Mol. Biol.* 11 (4), 1–13. doi:10.5376/cmb.2021.11.0004

Choi, D., Kim, J. H., and Kende, H. (2004). Whole Genome Analysis of the OsGRF Gene Family Encoding Plant-specific Putative Transcription Activators in Rice (Oryza Sativa L.). *Plant Cell Physiol.* 45 (7), 897–904. doi:10.1093/pcp/pch098

Chou, K.-C., and Shen, H.-B. (2010). Plant-mPLoc: a Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. *PloS one* 5 (6), e11335. doi:10.1371/journal.pone.0011335

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a Plant Small RNA Target Analysis Server (2017 Release). *Nucleic Acids Res.* 46 (W1), W49–W54. doi:10.1093/nar/gky316

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: Fast Selection of Best-Fit Models of Protein Evolution. *Bioinformatics* 27 (8), 1164–1165. doi:10.1093/bioinformatics/btr088

Emms, D. M., and Kelly, S. (2015). OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.* 16 (1), 157. doi:10.1186/s13059-015-0721-2

Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., Dinger, M. E., Hellemans, J., et al. (2017). Benchmarking of RNA-Sequencing Analysis Workflows Using Whole-Transcriptome RT-qPCR Expression Data. *Sci. Rep.* 7 (1), 1–11. doi:10.1038/s41598-017-01617-3

Eyre-Walker, A. C. (1991). An Analysis of Codon Usage in Mammals: Selection or Mutation Bias? *J. Mol. Evol.* 33 (5), 442–449. doi:10.1007/bf02103136

Filiz, E., Koç, İ., and TombuloğLu, H. (2014). Genome-wide Identification and Analysis of Growth Regulating Factor Genes in Brachypodium Distachyon: In Silico Approaches. *Turkish J. Biol.* 38, 296–306. doi:10.3906/biy-1308-57

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein Identification and Analysis Tools on the ExPASy Server," in *The Proteomics Protocols Handbook*. Editor J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi:10.1385/1-59259-890-0:571

Gubler, F., and Jacobsen, J. V. (1992). Gibberellin-responsive Elements in the Promoter of a Barley High-pI Alpha-Amylase Gene. *Plant Cell* 4 (11), 1435–1441. doi:10.1105/tpc.4.11.1435

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* 59 (3), 307–321. doi:10.1093/sysbio/syq010

Hall, T. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT[C]. *Nucleic Acids Symp. Ser.* 41, 95–98. doi:10.1021/bk-1999-0734.ch008

Han, J., Xie, H., Sun, Q., Wang, J., Lu, M., Wang, W., et al. (2014). Bioinformatic Identification and Experimental Validation of miRNAs from Foxtail Millet (Setaria Italica). *Gene* 546 (2), 367–377. doi:10.1016/j.gene.2014.05.050

Hobo, T., Asada, M., Kowyama, Y., and Hattori, T. (1999). ACGT-containing Abscisic Acid Response Element (ABRE) and Coupling Element 3 (CE3) Are Functionally Equivalent. *Plant J.* 19 (6), 679–689. doi:10.1046/j.1365-313x.1999.00565.x

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2014). GSDS 2.0: an Upgraded Gene Feature Visualization Server. *Bioinformatics* 31 (8), 1296–1297. doi:10.1093/bioinformatics/btu817

Ji, J. H., Zhou, Y. J., Wu, H. H., and Yang, L. M. (2015). Genome-wide Analysis and Functional Prediction of the Trihelix Transcription Factor Family in rice. *Yi Chuan* 37, 1228–1241. doi:10.16288/j.yczz.15-196

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., et al. (2013). A Haplotype Map of Genomic Variations and Genome-wide Association Studies of Agronomic Traits in Foxtail Millet (Setaria Italica). *Nat. Genet.* 45, 957–961. doi:10.1038/ng.2673

Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Mol. Cell* 14 (6), 787–799. doi:10.1016/j.molcel.2004.05.027

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi:10.1093/molbev/mst010

Khan, M. R., Hu, J., and Ali, G. M. (2012). Reciprocal Loss of CArG-Boxes and Auxin Response Elements Drives Expression Divergence of MPF2-like MADS-Box Genes Controlling Calyx Inflation. *PLoS One* 7, e42781. doi:10.1371/journal.pone.0042781

Khatun, K., Robin, A. H. K., Park, J.-I., Nath, U. K., Kim, C. K., Lim, K.-B., et al. (2017). Molecular Characterization and Expression Profiling of Tomato GRF Transcription Factor Family Genes in Response to Abiotic Stresses and Phytohormones. *Ijms* 18 (5), 1056. doi:10.3390/ijms18051056

Kim, J. H., Choi, D., and Kende, H. (2003). The AtGRF Family of Putative Transcription Factors Is Involved in Leaf and Cotyledon Growth inArabidopsis. *Plant J.* 36 (1), 94–104. doi:10.1046/j.1365-313x.2003.01862.x

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an Information Aesthetic for Comparative Genomics. *Genome Res.* 19 (9), 1639–1645. doi:10.1101/gr.092759.109

Kumar, K., Muthamilarasan, M., and Prasad, M. (2013). Reference Genes for Quantitative Real-Time PCR Analysis in the Model Plant Foxtail Millet (Setaria Italica L.) Subjected to Abiotic Stress Conditions. *Plant Cell Tiss Organ. Cult* 115 (1), 13–22. doi:10.1007/s11240-013-0335-x

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi:10.1093/nar/30.1.325

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13 (9), 2178–2189. doi:10.1101/gr.1224503

Li, P., and Brutnell, T. P. (2011). Setaria Viridis and Setaria Italica, Model Genetic Systems for the Panicoid Grasses. *J. Exp. Bot.* 62 (9), 3031–3037. doi:10.1093/jxb/err096

Li, S., Dong, X., Fan, G., Yang, Q., Shi, J., Wei, W., et al. (2018). Comprehensive Profiling and Inheritance Patterns of Metabolites in Foxtail Millet. *Front. Plant Sci.* 9, 1716. doi:10.3389/fpls.2018.01716

Liu, D., Song, Y., Chen, Z., and Yu, D. (2009). Ectopic Expression of miR396 suppressesGRFtarget Gene Expression and Alters Leaf Growth inArabidopsis. *Physiologia Plantarum* 136 (2), 223–236. doi:10.1111/j.1399-3054.2009.01229.x

Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome Sequence of the Progenitor of the Wheat D Genome Aegilops Tauschii. *Nature* 551 (7681), 498–502. doi:10.1038/nature24486

Lynch, M., and Conery, J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science* 290 (5494), 1151–1155. doi:10.1126/science.290.5494.1151

Ma, J.-Q., Jian, H.-J., Yang, B., Lu, K., Zhang, A.-X., Liu, P., et al. (2017). Genome-wide Analysis and Expression Profiling of the GRF Gene Family in Oilseed Rape (Brassica Napus L.). *Gene* 620, 36–45. doi:10.1016/j.gene.2017.03.030

Mehanathan, M., Rohit, K., Chandra Bhan, Y., Venkata Suresh, B., Yusuf, K., and Manoj, P. (2014). Identification and Molecular Characterization of MYB Transcription Factor Superfamily in C4 Model Plant Foxtail Millet (Setaria Italica L.). *Plos One* 9 (10), e109920. doi:10.1371/journal.pone.0109920s

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 (12), e121. doi:10.1093/nar/gkt263

Nei, M., and Gojobori, T. (1986). Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol. Biol. Evol.* 3 (5), 418–426. doi:10.1093/oxfordjournals.molbev.a040410

Pandey, G., Misra, G., Kumari, K., Gupta, S., Parida, S. K., Chattopadhyay, D., et al. (2013). Genome-Wide Development and Use of Microsatellite Markers for Large-Scale Genotyping Applications in Foxtail Millet [Setaria Italica (L.)]. *DNA Res.* 20 (2), 197–207. doi:10.1093/dnares/dst002

Rodriguez, R. E., Ercoli, M. F., Debernardi, J. M., and Palatnik, J. F. (2016). "Growth-Regulating Factors, A Transcription Factor Family Regulating More Than Just Plant Growth," in *Plant Transcription Factors*. Editor D. H. Gonzalez (Boston: Academic Press), 269–280. doi:10.1016/b978-0-12-800854-6.00017-8

Rouster, J., Leah, R., Mundy, J., and Cameron-Mills, V. (1997). Identification of a Methyl Jasmonate-Responsive Region in the Promoter of a Lipoxygenase 1 Gene Expressed in Barley Grain. *Plant J.* 11 (3), 513–523. doi:10.1046/j.1365-313x.1997.11030513.x

Rukmangada, M. S., Sumathy, R., Sivaprasad, V., and Naik, V. G. (2018). Genome-wide Identification and Characterization of Growth-Regulating Factors in mulberry (Morus spp.). *Trees* 32 (6), 1695–1705. doi:10.1007/s00468-018-1744-6

Shah, J., and Klessig, D. F. (1996). Identification of a Salicylic Acid-Responsive Element in the Promoter of the Tobacco Pathogenesis-Related Beta-1,3-Glucanase Gene, PR-2d. *Plant J.* 10 (6), 1089–1101. doi:10.1046/j.1365-313x.1996.10061089.x

Shang, S., Wu, C., Huang, C., Tie, W., Yan, Y., Ding, Z., et al. (2018). Genome-Wide Analysis of the GRF Family Reveals Their Involvement in Abiotic Stress Response in Cassava. *Genes* 9 (2), 110. doi:10.3390/genes9020110

Song, X., Ma, X., Li, C., Hu, J., Yang, Q., Wang, T., et al. (2018). Comprehensive Analyses of the BES1 Gene Family in Brassica Napus and Examination of Their Evolutionary Pattern in Representative Species. *BMC genomics* 19 (1), 346–415. doi:10.1186/s12864-018-4744-4

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Res.* 34 (Suppl. l_2), W609–W612. doi:10.1093/nar/gkl315

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* 43 (Database issue), D447–D452. doi:10.1093/nar/gku1003

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi:10.1093/nar/22.22.4673

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO Analysis Toolkit for the Agricultural Community, 2017 Update. *Nucleic Acids Res.* 45 (W1), W122–W129. doi:10.1093/nar/gkx382

van der Knaap, E., Kim, J. H., and Kende, H. (2000). A Novel Gibberellin-Induced Gene from Rice and its Potential Regulatory Role in Stem Growth. *Plant Physiol.* 122 (3), 695–704. doi:10.1104/pp.122.3.695

Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J. Hered.* 93 (1), 77–78. doi:10.1093/jhered/93.1.77

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics* 8 (1), 77–80. doi:10.1016/s1672-0229(10)60008-3

Wang, F., Qiu, N., Ding, Q., Li, J., Zhang, Y., Li, H., et al. (2014). Genome-wide Identification and Analysis of the Growth-Regulating Factor Family in Chinese Cabbage (Brassica Rapa L. Ssp. Pekinensis). *BMC Genomics* 15, 807. doi:10.1186/1471-2164-15-807

Wang, X., Wang, J., Jin, D., Guo, H., Lee, T.-H., Liu, T., et al. (2015). Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. *Mol. Plant* 8 (6), 885–898. doi:10.1016/j.molp.2015.04.004

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40 (7), e49. doi:10.1093/nar/gkr1293

Wilm, A., Higgins, D. G., Valentin, F., Blackshields, G., McWilliam, H., Wallace, I. M., et al. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics* 23 (21), 2947–2948. doi:10.1093/bioinformatics/btm404

Yadav, A., Khan, Y., and Prasad, M. (2016). Dehydration-responsive miRNAs in Foxtail Millet: Genome-wide Identification, Characterization and Expression Profiling. *Planta* 243 (3), 749–766. doi:10.1007/s00425-015-2437-7

Yang, Z., Nielsen, R., and Hasegawa, M. (1998). Models of Amino Acid Substitution and Applications to Mitochondrial Protein Evolution. *Mol. Biol. Evol.* 15 (12), 1600–1611. doi:10.1093/oxfordjournals.molbev.a025888

Yang, Z. (1997). PAML: a Program Package for Phylogenetic Analysis by Maximum Likelihood. *Bioinformatics* 13 (5), 555–556. doi:10.1093/bioinformatics/13.5.555

Yang, Z., Zhang, H., Li, X., Shen, H., Gao, J., Hou, S., et al. (2020). A Mini Foxtail Millet with an Arabidopsis-like Life Cycle as a C4 Model System. *Nat. Plants* 6 (9), 1167–1178. doi:10.1038/s41477-020-0747-7

Yoshida, H., Haze, K., Yanagi, H., Yura, T., and Mori, K. (1998). Identification of the Cis-Acting Endoplasmic Reticulum Stress Response Element Responsible for Transcriptional Induction of Mammalian Glucose-Regulated Proteins. *J. Biol. Chem.* 273 (50), 33741–33749. doi:10.1074/jbc.273.50.33741

Zhang, D.-F., Li, B., Jia, G.-Q., Zhang, T.-F., Dai, J.-R., Li, J.-S., et al. (2008). Isolation and Characterization of Genes Encoding GRF Transcription Factors and GIF Transcriptional Coactivators in Maize (Zea mays L.). *Plant Sci.* 175 (6), 809–817. doi:10.1016/j.plantsci.2008.08.002

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., et al. (2012). Genome Sequence of Foxtail Millet (Setaria Italica) Provides Insights into Grass Evolution and Biofuel Potential. *Nat. Biotechnol.* 30 (6), 549–554. doi:10.1038/nbt.2195

Zhang, J., Li, Z., Jin, J., Xie, X., Zhang, H., Chen, Q., et al. (2017). Genome-wide Identification and Analysis of the Growth-Regulating Factor Family in Tobacco (Nicotiana Tabacum). *Gene* 639, 117–127. doi:10.1016/j.gene.2017.09.070

Zheng, L., Ma, J., Song, C., Zhang, L., Gao, C., Zhang, D., et al. (2018). Genome-wide Identification and Expression Analysis of GRF Genes Regulating Apple Tree Architecture. *Tree Genet. Genomes* 14 (4), 54. doi:10.1007/s11295-018-1267-8

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership